BREEDING FOR SUSTAINABILITY IN COMMON BEAN (PHASEOLUS VULGARIS L.)

By

Madison Clare Whyte

A THESIS

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

Plant Breeding, Genetics, and Biotechnology - Master of Science

ABSTRACT

Common bean (*Phaseolus vulgaris* L.) is an important legume for human consumption and has an important role in cropping systems as a rotational crop. Improving the sustainability in agriculture is necessary for meeting the food demands of a growing global population while lessening the environmental impact of cropping systems. Developing efficient methods of improving host-plant resistance to dry bean anthracnose (Colletotrichum lindemuthianum) and the symbiotic nitrogen fixation ability (SNF) can enhance the sustainability of common bean as a food crop. A QTL study with the black bean cultivar 'TU', known to possess the C. *lindemuthianum* race 109 resistance gene Co-5, was conducted to develop molecular markers to deploy in the MSU Dry Bean Breeding Program. Resistance to anthracnose was investigated in an F₂ population developed from a cross between 'B19504' (a susceptible breeding line) and TU. 25 SNPs were identified between 6.84 and 24.62 Mb on linkage group 07. Improving SNF in common bean requires a method of efficiently evaluating breeding lines for the trait. Predictive models were developed from remote sensing-derived vegetation indices and machine learning algorithms to assess their ability to accurately and reliably estimate percent nitrogen derived from the atmosphere. A Random Forest model developed to predict nitrogen derived from the atmosphere (Ndfa) using yield and remote sensing (RS) data resulted in an average accuracy of r = 0.54. This model is promising in low nitrogen trials as an early selection tool to identify lines with higher SNF ability. Two prediction models for yield as an indirect indicator of SNF were developed using stepwise general linear modeling (StepwiseGLM) and Bayesian regularized artificial neural network (BRNeural Network) were determined to be accurate and reliable (StepwiseGLM r = 0.64; BRNeural Network r = 0.65). These models are promising in low nitrogen trials as an early selection tool to identify lines with higher SNF ability.

I would like to dedicate this work to my family: My fiancé, Isaac, who has given me his unyielding support and has always been a willing listener; My mother, Terri, father, Jim, and brother, Alex, who have always encouraged me to pursue my passions and interests; and my uncles Will and James for their counsel and guidance.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to those who helped me achieve my master's degree. I would like to thank my major advisor, Dr. Francisco Gomez for giving me the opportunity to pursue this degree, and for his guidance and support throughout its completion. I would also like to thank everyone in the bean lab, Dr. Leonardo Volpato, Evan Wright, Halima Awale, and Molly Irvin, for their support. Without them, the work and the experience would be far lesser.

TABLE OF CONTENTS

CHAPTER ONE: GENERAL INTRODUCTION	
INTRODUCTION	
ANTHRACNOSE	
ANTHRACNOSE VARIABILITY	
ANTHRACNOSE RESISTANCE IN HOST	
BREEDING FOR ANTHRACNOSE RESISTANCE	
MOLECULAR MARKERS	6
ANTHRACNOSE CONTROL METHODS	7
NITROGEN (N)	9
SYMBIOTIC NITROGEN FIXATION IN LEGUMES	
METHODOLOGIES FOR MEASURING SNF	
BREEDING FOR SNF IN COMMON BEAN	
REMOTE SENSING	
MACHINE LEARNING AS A BREEDING TOOL	
REFERENCES	
APPENDIX	
CHAPTER TWO: MAPPING THE ANTHRACNOSE RESISTANCE GENE CO-5	5
ABSTRACT	
INTRODUCTION	
METHODS AND MATERIALS	
RESULTS	
DISCUSSION	
CONCLUSION	
TABLES	
FIGURES	
REFERENCES	
CHAPTER THREE: UTILIZING MACHINE LEARNING TO PREDICT	SYMBIOTIC
NITROGEN FIXATION IN COMMON BEAN	
ABSTRACT	
INTRODUCTION	
MATERIALS AND METHODS	
RESULTS	
DISCUSSION	
CONCLUSION	
TABLES	
FIGURES	
REFERENCES	

CHAPTER ONE: GENERAL INTRODUCTION INTRODUCTION

The need for developing sustainable agricultural systems with low environmental impacts has become more apparent as the global population grows. Another driving factor is the reduction of available land for agriculture alongside diminishing natural resources. Developing agricultural technologies and implementing the use of crops with enhanced sustainability is one solution to mitigate these issues. Common bean (*Phaseolus vulgaris*) is one such crop.

Common bean is the most important legume used for human consumption due to its high protein content, dietary fiber, minerals, and carbohydrates. Additionally, common bean has also had a positive impact on cropping systems as a rotational crop. Common bean is a host plant for nitrogen fixing rhizobia bacteria, able to improve soil fertility by enriching the soil with nitrogen fixed from the atmosphere while also requiring less synthetic nitrogen application to produce yield (Reinprecht et al., 2020; Uebersax et al., 2022). Harnessing symbiotic nitrogen fixation carries the additional advantage of reducing fossil fuel usage to produce and apply fertilizer. Common bean is perceived to be a poor fixer compared to other legumes, however findings from numerous studies have noted a wide variance between genotypes with some varieties fixing as much as 70% of their nitrogen from fixation (Heilig et al., 2017; Kamfwa et al., 2015; Wilker et al., 2019). With the specificity of the relationship between hostplant and rhizobia genotypes beginning to be explored, the trait stands to be exploited through breeding (Gunnabo et al., 2019).

Integrated pest management is integral to agricultural sustainability wherein preventative crop protection practices are the foundations of the approach. Host plant resistance is an important aspect of preventative management. In common bean, *Colletotrichum lindemuthianum*

is one of the most economically important pathogens affecting the sustainability of production (Ferreira et al., 2013; Oblessuc et al., 2012). *C. lindemuthianum* (Sacc. and Magn.), or common bean anthracnose, is a seed borne pathogen that is readily spread between local fields and across longer distances between and within countries (Tu, 1994). Anthracnose outbreaks are managed using clean disease-free seed, crop rotation, and preventative foliar fungicidal treatments; however, these methods tend to be effective only in the short-term as well as dependent on environmental conditions. Genetic resistance employing the use of genes conferring broad resistance can be a long-term and economical solution to decreasing the use of fungicides and fossil fuels consumed during application.

ANTHRACNOSE

Colletotrichum lindemuthianum, commonly known as anthracnose, is a seed-borne hemibiotrophic fungal pathogen that is one of the most devastating diseases affecting bean production (Costa et al., 2021; Ferreira et al., 2013). Its importance among the biotic factors affecting commercial production is due to its lethality in susceptible cultivars and its high virulence diversity between races (L. C. Costa et al. 2021; Padder et al. 2017). Infected fields of susceptible cultivars in favorable, humid, and cool environments can have reduced yields up to 100% (Nunes et al. 2021).

Anthracnose has been reported in many African, Latin American, and European countries, and has also been found in numerous fields in both the United States and Canada (Mohammed 2013). The fungus commonly creates reddish-brown to black lesions on the leaf petiole and underside in addition to along the leaf veins, causing vein necrosis (Boersma et al. 2014). As the disease progresses through the stem it creates round, brown eyespots and lesions will also appear on the pods and deterioration will occur in the seeds. Infected seeds are both

unmarketable to consumers and sources of further infection and spreading of the disease. Ultimately, anthracnose can lead to premature defoliation, early flower and pod drop, and plant death (S. J. Boersma et al. 2020; Campa et al. 2014; Mohammed 2013; Tu 1983).

Anthracnose is easily spread to new regions through infected seed and between plants by irrigation water, rain drops, or mechanical movement such as farm equipment or wildlife moving through a field. Additionally, it can also affect future common bean crops. Anthracnose spores can survive winter conditions residing in both seed and plant residues left in the field after harvest and can remain virulent in seeds for years (Schwartz and Corrales 1989; Tu 1983). A multi-year study performed by Conner et al. found that the survivability of anthracnose spores in a field was influenced by environment, the type of infected tissue, and whether the samples were buried or not. When studied under a three-year crop rotation, *C. lindemuthanium* spores were still viable and able to infect bean crops in the third year under no-tillage conditions (Conner et al., 2019).

ANTHRACNOSE VARIABILITY

Pathogenic variability in *C. lindemuthianum* was first noted by Barrus (1911) wherein two races of anthracnose displayed differing levels of virulence against 139 bean cultivars. These two races were the first to be classified as distinct races with the denotation α and β . With this first step, further understanding of the hyper-variability of the pathogen would follow. Numerous races were characterized in the following years resulting in a need for a standardized method of determination as some localities used different codes for identifying anthracnose races rather than the Greek letters, which impeded attempts to understand the global variability (Melotto, Balardin, and Kelly 2000). A method of isolate race standardization was proposed by Pastor-Corrales (1991) that utilized 12 differential bean cultivars from both Andean and Mesoamerican

gene pools. The cultivars were assigned binary numbers that are used to identify a specific anthracnose race (Table 1). The values of the cultivars showing susceptibility when inoculated with an unknown race are summed to give the binary value of that race. For example: an unknown race is tested against the differential cultivars and susceptibility is shown in Michelite (1), Perry Marrow (4), Cornell 49242 (8), Kaboon (32), and Mexico 222 (64). The race would be characterized as 109. Using this method, 182 races have been identified world-wide (Padder et al. 2017).

ANTHRACNOSE RESISTANCE IN HOST

Coevolution has been observed between common bean and *C. lindemuthianum* following the gene-for-gene theory established by Flor (1955). The theory states that for every gene that conditions resistance in the host there is a complementary gene in the parasite that conditions avirulence. In common bean, resistance is conferred by individual, independently segregating loci in a family called *Co*. These genes are grouped in clusters across 7 chromosomes, Pv01, Pv02, Pv03, Pv04, Pv07, Pv08, and Pv11. Like other resistance genes, these clusters are often encoded by nucleotide-binding leucine-rich repeat (NB-LRR) proteins (Richard W. Michelmore, Christopoulou, and Caldwell 2013).

An incompatible interaction between a host cultivar carrying a *Co* gene and an avirulent race of anthracnose leads to a hypersensitive reaction (HR) characterized by localized host cell death and the formation of necrotic spots. This is to prevent the spread of the fungus to further cells. However, like many other *Colletotrichum* species, C. lindemuthianum is hemibiotrophic. When the fungus first penetrates the cell wall, the infection hyphae grows between the cell wall and membrane and does not trigger defense responses either by masking the presence of the hyphae or actively suppressing defense responses (Münch et al. 2008). In *C. lindemuthianum*, the

⁴

hyphae is masked by the release of glycoproteins that separate the structure from the plant membrane, protecting the fungus from recognition (Perfect et al. 1998). The biotrophic phase can last up to three days before the necrotrophic phase is induced and HR occurs within the plant.

BREEDING FOR ANTHRACNOSE RESISTANCE

There are many methods that can be used to manage plant diseases like anthracnose, the most efficient being sowing anthracnose-resistant cultivars to stave off locally known races (Balardin, Jarosz, and Kelly 1997; Strange and Scott 2005). Integrating gene-specific disease resistance into cultivars requires the identification of resistant plants to cross with agronomically favorable, but susceptible, plants (Strange and Scott 2005). Thus far, 20 *Co* genes have been mapped in the common bean genome with the Mesoamerican gene pool containing most of the anthracnose resistance genes in common bean (Nunes et al. 2021). New sources of resistance have been identified in recent years, particularly *Co-Bf* (Marcon et al. 2020) and *Co-Pa* (de Lima Castro et al. 2017) and numerous *Co* genes have been recharacterized as allelic forms rather than unique loci. Efforts to investigate new sources of resistance are integral to adapt to the continuously developing variability in *C. lindemuthianum*.

Resistance to anthracnose can be evaluated using molecular markers or through greenhouse inoculation. Inoculations are commonly performed using a spray suspension of 1.0 x 10^6 conidia ml⁻¹ of selected races of *C. lindemuthianum* onto seedlings. The plants are kept at high humidity (>80%) for 72 hours with disease symptoms appearing 7-10 days after inoculation. Response rating is often performed using the broad characterization of 'susceptible' and 'resistant', assuming a qualitative trait. A 1-9 or 0-5 scale describing the range of reactions are also used to identify quantitative genes conferring partial resistance to anthracnose (Drijfhout & Davis, 1989; Pastor-Corrales et al., 1998).

MOLECULAR MARKERS

Molecular markers have been used to efficiently identify resistance in breeding lines or segregating populations. Markers tightly linked to a region controlling a trait of interest, called quantitative trait loci (QTL), can be used to integrate that trait into elite cultivars. QTL-seq is a method of identifying DNA markers tightly linked to the causal gene for a given phenotype using the combined methodologies of bulked segregant analysis (BSA) and whole genome sequencing (WGS) to hasten the identification of QTLs (Takagi et al. 2013). BSA uses the progeny of parents with contrasting phenotypes that are scored and bulked for segregation of that phenotype (Giovannoni et al. 1991; R W Michelmore, Paran, and Kesseli 1991). Comparative genetic analysis is performed on the bulks to identify markers linked to the traits of interest. Like BSA, QTL-seq uses the progeny of parents with contrasting phenotypes of the trait of interest in a mapping population. The segregation of the phenotype is scored and DNA from the extremes are bulked to generate 'high' and 'low' groups. The proportion of short reads that correspond to the parental genomes that can be identified by a SNP is evaluated. The sequencing data is aligned to the reference sequence and the number of differing SNPs is counted. The SNP-index is defined as the percentage of differing SNPs in the total short reads within a genomic region (Takagi et al. 2013). If the index value is close to 1 if all the short reads are representative of the genome of the non-reference parent and 0 if representative of the parent used as the reference. The methodology of QTL-seq has been altered and improved since its development. Recent techniques utilize only the bulked samples with a focus on improving SNP filtering through improving alignment and allele accuracy (Korani et al. 2021).

In dry bean, random amplified polymorphic DNA (RAPD), and restriction fragment length polymorphism, sequenced characterized amplified region (SCAR), simple-sequence

repeats (SSR), single nucleotide polymorphism (SNP), and most recently kompetitive allelespecific polymerase chain reaction (KASP) markers have been used to map and identify *Co* genes in resistant lines (Burt et al. 2015; Gilio et al. 2020; de Lima Castro et al. 2017; R. Young et al. 1998; R. A. Young and Kelly 1997). KASP is a low-cost, accurate and reliable SNP genotyping platform that has gained prominence in trait-specific marker development (Cortés, Chavarro, and Blair 2011; Gilio et al. 2020; Valentini et al. 2017). KASP markers have been used in common bean breeding for anthracnose resistance, drought tolerance, rust resistance, and color retention after canning (Bornowski et al., 2020; de Lima Castro et al., 2017; Diaz et al., 2018; Gilio et al., 2020; Hurtado-Gonzales et al., 2017; Villordo-Pineda et al., 2015).

The publication of the common bean reference genome by Schmutz et al. (2014) has allowed for comparing and mapping of the positions of molecular markers. The progression of genotyping technology and the availability of sequencing information has made SNP markers incredibly useful in MAS. Numerous SNP genotyping platforms have been developed that utilize a variety of allele detection and discrimination techniques, as well as reaction formats (Chen and Sullivan 2003; Sobrino, Brión, and Carracedo 2005). Genotyping by next-generation sequencing (GBS) is a recent application of SNP genotyping that has become popularized in QTL discovery in several crops including common bean, soybean, and wheat (Ariani, Berny Mier y Teran, and Gepts 2016; Hart and Griffiths 2015; Iquira, Humira, and Francois 2015; Li et al. 2015). GBS is a platform that simultaneously discovers and genotypes many SNPs in multiplexed libraries with or without the use of reference genomes (Elshire et al. 2011).

ANTHRACNOSE CONTROL METHODS

General preventative management includes removing or burying plant debris from the fields after harvest, rotating crops using non-hosts for two years at a minimum, fungicide

treatments on the seeds, and cleaning seed storage facilities (Conner et al., 2019). To prevent spores from transferring from one plant to another via water movement, overhead irrigation should be avoided and physical movement through the field should be delayed until the leaves are dry. Contact and systemic fungicides are the only methods of treating anthracnose present in the field, and function only as preventative treatments to limit disease spread, typically by inhibiting spore germination. To date, no curative fungicides are known to treat diseased plants once infection has occurred.

Anthracnose control begins with planting disease-free seed as seed transportation is the most common method of introducing the disease (S. J. Boersma et al. 2020). Before planting, most dry bean seed is treated with a fungicide as a preventative measure. Commonly used fungicides for seed treatment are azoxystrobin, fludioxonil, thiamethoxam, and metalaxyl-m. Various foliar fungicides are used to protect yield and increase seed quality in dry bean, but their efficacy is dependent on the severity of the infection and timing of application (S. J. Boersma et al. 2020; Conner et al. 2009; Negera and Dejene 2018). Likewise, the success of seed treatments is dependent on the severity of infection, thus dry bean yields are best protected when seed treatments are utilized with foliar fungicides (Gillard, Ranatunga, and Conner 2012).

Bioagents have been deployed to manage anthracnose as cost-effective and less ecologically harmful alternatives. The results of previous studies evaluated the efficacy of alternative fungicides and antagonistic bioagents suggest that seed treatments with bioagents like *Pseudomonas fluorescens* and *Trichoderma viride* are effective for inhibiting the pathogen (Padder et al. 2010, 2010). Inorganic salts of phosphorous acid (H₃PO₃), phosphites, and plant extracts have been investigated as alternative anthracnose treatment products due to their fungistatic and fungicidal effects, respectively. Phosphites have been determined to increase

peroxidase and phenylalanine ammonia lyase activity, enzymes that are positively correlated with anthracnose resistance (Jhonata et al. 2015). In addition to increasing defensive enzyme activity, phosphites can reduce anthracnose severity by directly inhibiting mycelial growth (B. H. G. Costa et al. 2018; Gadaga et al. 2017). Previous success has been noted in controlling anthracnose of soybean, *Colletotrichum dematium*, and other plant diseases with plant extracts (Shovan et al. 2008). Dry bean and cowpea seed treatments with acetone- and water-based extracts of *Allium sativum*, *Agapanthus caulescens*, *Carica papaya*, and *Syzygium cordatum* expressed inhibitory activity against anthracnose (Masangwa, Aveling, and Kritzinger 2013).

NITROGEN (N)

Proteins are composed of 16% nitrogen (N), making the element necessary for survival and growth in plants and animals (Frink, Waggoner, and Ausubel 1999). In crop production, yields are often determined by N. As plants cannot directly assimilate molecular N, it is often supplied to crops either as NH₄⁺ and NO₃⁻ (Franche, Lindström, and Elmerich 2009; Ladha et al. 2005; Lam et al. 1996). The synthesis of ammonia through the Haber-Bosch process allowed for the mass production of usable N. The usage of synthetic fertilizers was a major contributing factor to the success of the "green revolution" and they are currently an integral part of modern agriculture. An estimated 40% of the world's population relied on fertilizer inputs for food at the end of the twentieth century and currently it is estimated roughly 50% of the population is supported by synthetic nitrogen (Erisman et al. 2008).

Global N inputs have greatly increased in crop production from 37 Tg N per year in 1961 to 163 Tg N per year in 2009 (Lassaletta et al. 2016). These increases do not come without drawbacks, however, as N has become a major pollutant (Ladha et al. 2005; C. Wang et al. 2017; Xu et al. 2019; X. Zhang et al. 2015). A majority of the N applied to crops is not absorbed by the

plant and instead lost to the environment through ammonia volatilization, denitrification, soil leaching, and eutrophication (Akter, Lupwayi, and Balasubramanian 2017; Asghari and Cavagnaro 2011; Turner and Rabalais 1994; Turner, Rabalais, and Justic 2008). Wang et al 2017 determined that there has been an approximate tenfold increase in N exported from croplands to the hydrosphere and a fivefold increase to the atmosphere has occurred since 1860. In addition, the Haber-Bosch process requires large quantities of fossil fuels to produce ammonia and consequently it releases large amounts of greenhouse gasses (Razon 2014). Projected population growth and the rising demand for food necessitates an increase in agricultural production (M. Han et al. 2015; Mulvaney, Khan, and Ellsworth 2009; X. Wang et al. 2019). Meeting this demand poses long-term ramifications to the ecosystem and human health.

SYMBIOTIC NITROGEN FIXATION IN LEGUMES

Biological nitrogen fixation (BNF) is an alternative approach for inputting N in cropping systems and reducing synthetic N usage. BNF is a natural process of converting atmospheric N, N₂, into the usable form NH₄⁺. This process is performed exclusively by archaea and bacteria. These organisms can be classified into three categories: associative, free-living, and symbiotic fixers. Associative and symbiotic N fixers are found in the rhizosphere of legume and nonlegumes (Santi, Bogusz, and Franche 2013). Free-living N fixers encompass microbes that fix N independent of other organisms. Photosynthetic diazotrophs provide their own energy required to chemically convert N₂ while nonphotosynthetic diazotrophs rely on a chemical energy source (Saikia and Jain 2007). Associative N fixing diazotrophs reside in near proximity to plant roots, relying on root exudates to fund their fixation (Mus et al. 2016). Symbiotic N fixation (SNF) describes the specific interaction between legumes and rhizobia and is the most effective in N fixing (Soumare et al. 2020). Rhizobia is the collective term for the genera *Rhizobium*, *Azorhizobium*, *and Bradyrhizobium*.

The symbiotic relationship begins with the host and rhizobia exchanging chemical signals. Flavonoids are secreted by the host plant's roots that activate specific signaling compounds, called Nod factors, in compatible rhizobia (Fisher and Long 1992; Jimenez-Jimenez et al. 2019; C.-W. Liu and Murray 2016). A signaling cascade is triggered in the host plant upon its recognition of the Nod factors that lead to the formation of intracellular structures called infection threads. Infection threads allow the rhizobia to access root hair tissues where they undergo endocytosis, forming nodule cells (Cissoko et al. 2018; Suzaki et al. 2019). Once inside the nodule, the bacteria begin to fix nitrogen.

The fixation ability of rhizobia-host symbiosis is dependent on many factors. Soil, genetic (host-rhizobia) interactions, and competition with other microorganisms can affect N-fixation (Soumare et al. 2020). Soil conditions can alter rhizobia signaling secretions. In general, rhizobia are pH-sensitive with their optimal range for growth between 6.0 and 7.0 (Hungria and Vargas 2000). Acidic soil affects the host-rhizobia relationship by causing a reduction in flavonoid and Nod factor secretion in the rhizobia, by impeding attachment to the root hairs, and by impeding nodule formation (Ferguson et al. 2019; Lira Junior 2015; McKay and Djordjevic 1993). Soil lacking moisture availability over long periods of time decreases nodule formation and nodule-specific activity (Serraj, Sinclair, and Purcell 1999). Under drought conditions, the rhizobia lack mobility for root infection and growth. Within the nodule, nitrogenase activity decreases as oxygen is unable to enter the nodule due to a decrease in permeability and respiration is restricted (Durand, Sheehy, and Minchin 1987; Walsh 1995). Drought conditions

Igor et al. 2011; Kunert et al. 2016). Nitrogen availability in the soil causes a downregulation in both nodule formation and activity (J. Streeter and Wong 1988). The process of fixation requires energy and resources from the hostplant and thus if there is sufficient N, the host-plant reduces its expenditure. In the rhizobia, N in the form of nitrate inhibits the synthesis of flavonoids and the expression of the nodulation transcription factor NODULE INCEPTION (NIN) (Barbulova et al. 2007; Van Noorden et al. 2016). In contrast, N in the form of ammonium has been observed to stimulate nodulation formation in low concentrations (Bollman and Vessey 2006).

The level of compatibility between host and rhizobia strain can determine fixation ability. In legumes the symbiosis is highly specific, controlled at multiple levels within the host and rhizobia (D. Wang et al. 2012). These controls allow for the differentiation of rhizobia from pathogens in the host and genotype or species discrimination in the rhizobia (Sadowsky et al. 1991; S. Yang et al. 2010). Incompatibility between host genotype and strain can occur during the process of symbiosis where either nodules fail to form or formed nodules fail to fix nitrogen (J. Liu et al. 2014; Q. Wang et al. 2017, 2018; S. Yang et al. 2010). *Rhizobium* and *Bradyrhizobium* species are considered to be promiscuous, having a broad host-range for infection (Perret, Staehelin, and Broughton 2000). While able to infect many host-genotypes, their fixation efficiency can vary greatly with different combinations of host and strain (Schumpp and Deakin 2010; Simsek et al. 2007; Tirichine, de Billy, and Huguet 2000). Unfortunately, despite the observed variation in strain-specific N fixation, the mechanisms behind the behavior are unknown (D. Wang et al. 2012).

Rhizobia strains with superior N-fixing ability have been identified over the decades and have been used in agricultural practices (Brockwell and Bottomley 1995; J. Brockwell, Bottomley, and Thies 1995; J. G. Streeter 1994). However, the use of inoculants does not

guarantee the successful infection of that strain. Resident rhizobial strains are just as, if not more likely to inhabit nodules on a host-plant while also being less efficient in N-fixation (J. G. Streeter 1994; Triplett and Sadowsky 1992). The success of native rhizobial strains in occupying nodules is often thought to be due to 'out-competing' inoculated strains, being better adapted to the environment and in greater numbers. However, there is evidence that host-plant preference may play a role in determining nodule occupancy and symbiotic efficiency (Gunnabo et al. 2019; Simms and Taylor 2002; Yates et al. 2011).

METHODOLOGIES FOR MEASURING SNF

Quantifying N-fixation often requires the use of an N isotope other than ¹⁴N. The stable isotope ¹⁵N has been proven useful in fixation experiments and has been widely adopted (Giller 2001). The isotope is utilized across both direct and indirect methods of measuring N-fixation.

The direct method of quantifying fixation in plant tissue involves incubating the sample in an enclosed atmosphere enriched with ¹⁵N₂. After incubation, the sample is purified and the proportion of ¹⁵N present is determined with mass spectrometry. While precise, this method requires knowledge of the ¹⁵N enrichment of the experimental atmosphere and the incubation time is dependent on the rate of fixation relative to the amount of N present in the tissue (Bergersen 1980).

A simple method of quantifying fixed N over a growing season is to measure the concentration of N in the tissue and multiply it by the weight of the plant material produced (Giller 2001). This method is sensitive to N contamination and is performed using N-free media, however it can also be conducted in soil if a non-fixing reference plant is used to estimate the amount of N in the soil or if the gains and losses of N are accounted for. The first extension of

this method is referred to as the 'N-difference' method and the second is the 'N-balance' method. The equation for the 'N-balance' method is as follows:

$$\% Ndfa = \frac{N \text{ yield}_{fixing \text{ plant}} - N \text{ yield}_{non-fixing \text{ plant}}}{N \text{ yield}_{fixing \text{ plant}}} x 100$$
Eq. 1

Where %Ndfa is the percentage of N in the seed at harvest that is derived from the atmosphere through fixation, N yield is the product of the plant's seed yield in kg/ha and the percent of N present in the seed from the respective fixing plant and the non-fixing plant.

Measuring the acetylene reduction activity (ARA) has been used as an indirect method of estimating N fixation (Stewart, Fitzgerald, and Burris 1967). Rhizobia nitrogenase, which reduces N₂ to NH₃, can also reduce acetylene to ethylene. The gaseous ethylene can then be measured with gas chromatography. N fixation can be estimated from this using the theoretical conversion ratio of 4 mol acetylene reduced to 1 mol N fixed (Boddey and Knowles 1987; Giller 2001). ARA is conducted under controlled environments and on root nodules from plants grown in the field (Riar et al. 2018; Zablotowicz and Reddy 2007).

Ureide assays can be performed to estimate fixed N as N fixed by nodule-inhabiting rhizobia is assimilated into glutamine and then metabolized into ureide compounds like allantoin and allantoic acid. These compounds can be sampled from the xylem and analyzed with colorimetric assays with the results expressed as the relative % ureide content (Herridge 1982; Tegeder 2014). Ureide assays estimate only the proportion of N in the plant derived from N fixation and sequential estimates are necessary for evaluating the total amount of fixed N.

A common method of measuring N fixation is the ¹⁵N isotope enrichment method. Fertilizer composed entirely of ¹⁵N is used as the plant's sole source of N which results in all the plant's N consisting of ¹⁵N. A plant that is able to fix atmospheric N, ¹⁴N₂, will have a proportion of ¹⁵N less than 100 atom % ¹⁵N, with the difference calculated as the amount of N derived from the atmosphere (Chalk 1985). A reference plant that does not fix N is often used to measure the amount of ¹⁵N enrichment in the soil. The equation for isotope enrichment is as follows:

$$N from fixation = \left[1 - \left(\frac{atom \% 15N \, excess_{fixing \, plant}}{atom \% 15N \, excess_{fertilizer \, or \, reference \, plant}}\right)\right]$$
Eq. 2

Where N from fixation is the amount of N fixed by the plant from the atmosphere, atom % ¹⁵N excess represents the percentage of N in the reference or test plant from the ¹⁵N fertilizer.

The isotope dilution method of ¹⁵N evaluation is similar to the dilution method, with the exclusion of applying the ¹⁵N fertilizer. The differences in ¹⁵N enrichment in the test plants and the reference plants indicate the plants' dependence on atmospheric N and would be used to calculate N fixation (Schwenke et al. 1998; Shearer and Kohl 1986). This method additionally requires the natural abundance of the N fixing plant grown in N-free media to account for isotopic fractionation of atmospheric N during fixation lest the calculation of N fixed would be overestimated. The amount of N derived from fixation is calculated using the following equation:

$$N from fixation = \left(\frac{\delta^{15} N_{reference \ plant} - \delta^{15} N_{fixing \ plant}}{\delta^{15} N_{reference \ plant} - B}\right)$$
Eq. 3

Where $\delta^{15}N$ is the ¹⁵N content of the reference or the test plant and B refers to a measure of the ¹⁵N content of the test legumes grown under N-free conditions where their only source of N is from fixation.

BREEDING FOR SNF IN COMMON BEAN

The genetic variability found between common bean genotypes for fixation ability has presented the possibility of breeding for improvement (Buttery, Park, and Berkum 1997; Elizondo Barron et al. 1999; Kamfwa, Cichy, and Kelly 2015). Additionally, Graham & Rosas (1977) determined via acetylene reduction that fixation ability varies significantly between common bean growth types with climbing, indeterminate cultivars showing consistently greater fixation. Unfortunately, the complexity of the trait has been an obstacle in improving the fixation ability in common bean through breeding. Numerous genes control the process of fixation from nodule formation to photosynthesis and their expression is sensitive to environmental changes and stresses. The latter can obfuscate evaluations for SNF, leading to limited improvement. Identifying QTLs and molecular markers for SNF have been proposed to overcome these challenges.

Previous QTL mapping studies utilizing the ¹⁵N natural abundance method and nodule counting to evaluate SNF identified QTLs on nearly every chromosome. The nodulation count studies found QTLs on Pv01, Pv02, Pv03, Pv05, Pv06, Pv07, Pv09, Pv10, and Pv11 (Nodari et al. 1993; Souza et al. 2000; Tsai et al. 1998). Studies using the natural abundance method to directly assess SNF also found QTLs on Pv01, Pv03, Pv06, Pv07, and Pv09, Pv10, and Pv11 with additional QTLs found on Pv04 and Pv08 (Lucy M. Diaz et al. 2017; Heilig et al. 2016; Kamfwa, Cichy, and Kelly 2015, 2019; Ramaekers et al. 2013). A genome-wide association study identified three genes as candidate genes for Ndfa: *Phvul.009G136200* and *Phvul.007G050500* on Pv07 (Kamfwa, Cichy, and Kelly 2015). *Phvul.009G136200* and *Phvul.007G050500* encodes leucine-rich repeat receptor-like protein kinase, which has been reported to be critical in the signal transduction necessary for nodule formation (Stracke et al. 2002). The third gene is involved in the signaling cascade that induces nodule formation (Lévy et al. 2004).

The wide breadth of genomic regions found for SNF could be due in part to differences between the gene pools of the genotypes used in these studies. Middle American and Andean bean genotypes have been found to have differing SNF abilities and SNF-related traits (Farid and

Navabi 2015). Middle American genotypes tend to be better fixers under optimal conditions while many Andean genotypes had the advantage in nodule development.

SNF has been associated with seed yield in numerous studies with the implication that SNF can be indirectly selected for under low N conditions via seed yield (Barbosa et al. 2018; Buttery, Park, and Berkum 1997; Farid and Navabi 2015; Fenta, Beebe, and Kunert 2020). Selecting lines with high yields and SNF ability for variety development is ideal for breeders, but it is not uncommon that lines with high SNF are not among the highest grain yielders (Farid et al. 2017; Reinprecht et al. 2020). Leaf chlorophyll content (SPAD) has often been found to have a strong, positive correlation with SNF, suggesting that this trait is an indicator of SNF (Farid et al. 2017; Jiang et al. 2020; Kamfwa, Cichy, and Kelly 2015). Significant, positive correlations with nitrogen fixation have also been found in soybean and peanut (Dinh et al. 2013; Gwata et al. 2004). Its high heritability in bean suggests the use of the trait as a means of indirectly selecting for fixation ability and high yield as long as measurements are taken under low N conditions (Farid and Navabi 2015).

REMOTE SENSING

Traditional methods of assessing biochemical and physiological traits in a breeding program are time consuming and can be costly. Spectral reflectance has been proposed as a timeefficient and nondestructive method of measuring plant characteristics (Babar et al. 2006). Reflectance in the visible/near infrared region of the electromagnetic spectrum and indices based off of emitted wavelengths in those regions have been related to chlorophyll and nitrogen status, water status, disease response, and biomass across numerous crops (Boshkovski et al. 2021; Grüner, Astor, and Wachendorf 2021; Hansen and Schjoerring 2003; Hunt et al. 2013; Sandino Mora et al. 2018; Zhou et al. 2018).

Spectral remote sensing collects data by measuring the radiance emitted, transmitted, or reflected from objects or surfaces. In the field, these measurements are impacted by environmental conditions, specifically conditions affecting light availability such as cloud cover and the time of day. The sensors used in spectral imagery are often either red-green-blue (RGB), multispectral, or hyperspectral imaging sensors. These sensors have varying levels of sensitivity; RGB sensors only capture visible light wavebands (400 - 700 nm), multispectral sensors capture visible light in addition to red edge and near infrared (700 - 1300 nm) bands, and hyperspectral sensors can capture hundreds of wavebands (500 - 2500 nm). The measured reflectance is dependent on the chemical and morphological characteristics of the surface imaged and changes with plant type, developmental stage, vigor, water content, biomass, and pigments present (Babar et al. 2006). Spectral reflectance indices, hereafter referred to as vegetation indices or VIs, were developed from formulas usually in the form of ratios or differences between reflectances at given wavelengths. The information gathered from the captured images is interpreted by differences and changes of the canopies' spectral characteristics and changes in the leaves of plants. VIs are validated through correlations between the indices and the traits of interest.

Many VIs have been developed over the years as a means of quick, cost-efficient highthroughput phenotyping and many studies have addressed comparisons of performance between sensor types. Hyperspectral sensors capture the most bands, but like multispectral sensors, they are often expensive and heavy in addition to being more sensitive to ambient light conditions (Herzig et al. 2021). Multispectral sensors are commonly used in agriculture for their application-specific band selection and the information collected is generally more reliable and repeatable due to their radiometric calibration (Haghighattalab et al. 2016; Nebiker et al. 2016). RGB cameras offer the lowest cost and the narrowest spectral range, but RGB-derived VIs have

been reported to perform equally or better than multispectral VIs for specific plant traits and measurements (Di Gennaro et al. 2018; Gracia-Romero et al. 2017; Marcial-Pablo et al. 2019; Travlos et al. 2017). VIs across all imaging sensor types have been shown to be incredibly useful in crop management and prediction models for yield and nitrogen status (Bascon et al. 2022; Candiago et al. 2015; N. Han et al. 2022; Lu et al. 2019).

Spectral data has been captured using a wide variety of vehicles and instruments from mobile ground-based vehicles to unmanned aerial systems (UAS) and satellites. Each method of data collection comes with their own benefits and limitations. Satellites capture data from large areas, but often have limited spatial resolution and are inflexible in regard to scheduling and sensor selection (C. Zhang, Marzougui, and Sankaran 2020). Images captured by ground-based vehicles can be particularly susceptible to wind and have a limited field of view, but the sensors they carry can be customized and the close proximity to field plots allows for high resolution images (Virlet et al. 2017). UAS can obtain low-cost, high-resolution images from plots and fields (Aasen et al. 2018). UAS are limited to what sensors they can carry due to weight restrictions, however they are able to carry RGB, multispectral, and hyperspectral cameras (Fu et al. 2021; Gracia-Romero et al. 2017; Herzig et al. 2021).

MACHINE LEARNING AS A BREEDING TOOL

With the emergence of big data technologies and high-performance computing, machine learning (ML) has become a method to investigate and navigate collected information. ML, as first defined by Arthur Samuel in 1959, is the study of computational techniques that gives machines the ability to learn and identify patterns without explicit programming (Samuel 2000). Many ML models have been developed and deployed across numerous scientific fields including agriculture, bioinformatics (Rahman et al. 2021), and medicine (Petersen and Aung 2022).

ML methodologies set out to perform a task using patterns 'learned' from training data. Model performance for a specific task is measured using a metric that improves over time, the metrics being any of the various statistical and mathematical models used to evaluate regression or classification models. After the pattern is developed using the training data, the trained model can be applied to perform the task on a new data set, usually called the testing data. ML tasks are divided into categories depending on the learning type (supervised or unsupervised) or learning models (regression, classification, clustering, etc.). The goal of supervised learning is to determine a rule or equation that relates the variables to the response. Unsupervised learning is meant to find hidden patterns from the input data with no distinction of an output or response variable.

ML methods have been applied in agriculture to assist in data-driven decision making and predictive modeling. In recent years, different ML algorithms have been used to accurately predict yield for different crops. Artificial neural networks have been demonstrated to be accurate (Ashapure et al. 2019; Fortin et al. 2011; Q. Yang et al. 2019), as has K-nearest neighbor (L. Zhang et al. 2010), gradient boosting (Shendryk, Davy, and Thorburn 2021; Stas et al. 2016), and random forest (Kim and Lee 2016).

A recent study by J. Han et al. (2020) presented a comparative study of random forest, Knearest neighbor, back-propagation neural network, decision tree, support vector machines, Gaussian process regression, boosting and bagging trees for winter wheat yield using remote sensing, climate, and soil data. They used three accuracy metrics to validate the models: Root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R²). Three models proved to be the best performing methods: support vector machines, random forest, and Gaussian process regression. Another comparative study by Gonzalez-Sanchez et al.

(2014) investigated the accuracies of multiple linear regression, M5-Prime regression trees, perception multilayer neural networks, support vector regression, and K-nearest neighbor in predicting yields for ten crops. In contrast to the prior study, the variables used were growth and climate measurements. Additionally, K-nearest neighbors was found to be one of the most accurate models alongside M5-Prime regression trees. The type of predictor variables included in the models may play a strong role in determining model accuracy.

ML methods have also been employed in estimating plant N status as a quick, nondestructive method. Shi et al. (2021) utilized RGB images in evaluating three regression methods, simple nonlinear regression, back-propagation neural networks, and random forest, for estimating rice shoot dry matter, N accumulation, and leaf area index. Here, random forest was found to be the most accurate model for each response variable. Random forest was also determined to be accurate in estimating N in barley and grass silage using RGB, multispectral, and hyperspectral bands and indices (Näsi et al. 2018). A comparative study performed by (Yao et al. 2015) evaluated six algorithms for estimating wheat leaf N in eight experiments over nine years with hyperspectral bands. Two nonlinear ML methods, artificial neural networks and support vector machines were compared against linear regression models stepwise multiple linear regression, partial least squares regression, and models built from VIs and continuum removal. This study found greater accuracy with support vector machines and an overall trend of increasing accuracy with a greater inclusion of wavelengths in the models.

REFERENCES

- Aasen, Helge, Eija Honkavaara, Arko Lucieer, and Pablo J. Zarco-Tejada. 2018. "Quantitative Remote Sensing at Ultra-High Resolution with UAV Spectroscopy: A Review of Sensor Technology, Measurement Procedures, and Data Correction Workflows." *Remote Sensing* 10(7): 1091.
- Akter, Zafrin, Newton Z Lupwayi, and Parthiba Balasubramanian. 2017. "Nitrogen Use Efficiency of Irrigated Dry Bean (Phaseolus Vulgaris L.) Genotypes in Southern Alberta." *Canadian Journal of Plant Science*: CJPS-2016-0254.
- Ariani, Andrea, Jorge Carlos Berny Mier y Teran, and Paul Gepts. 2016. "Genome-Wide Identification of SNPs and Copy Number Variation in Common Bean (Phaseolus Vulgaris L.) Using Genotyping-by-Sequencing (GBS)." *Molecular Breeding* 36(7): 87.
- Arrese-Igor, Cesar et al. 2011. "Physiological Responses of Legume Nodules to Drought." *Plant Stress* 5(1): 24–31.
- Asghari, H., and Timothy Cavagnaro. 2011. "Arbuscular Mycorrhizas Enhance Plant Interception of Leached Nutrients." *Functional Plant Biology* 38: 219–26.

Ashapure, Akash et al. 2019. "Unmanned Aerial System Based Tomato Yield Estimation Using Machine Learning." In Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV, SPIE, 171–80. https://www.spiedigitallibrary.org/conference-proceedings-ofspie/11008/110080O/Unmanned-aerial-system-based-tomato-yield-estimation-usingmachine-learning/10.1117/12.2519129.full (August 22, 2022).

- Babar, M. A. et al. 2006. "Spectral Reflectance to Estimate Genetic Variation for In-Season Biomass, Leaf Chlorophyll, and Canopy Temperature in Wheat." *Crop Science* 46(3): 1046–57.
- Balardin, R. S., A. M. Jarosz, and J. D. Kelly. 1997. "Virulence and Molecular Diversity in *Colletotrichum Lindemuthianum* from South, Central, and North America." *Phytopathology*® 87(12): 1184–91.
- Barbosa, Norma et al. 2018. "Genotypic Differences in Symbiotic Nitrogen Fixation Ability and Seed Yield of Climbing Bean." *Plant and Soil* 428(1): 223–39.
- Barbulova, Ani et al. 2007. "Differential Effects of Combined N Sources on Early Steps of the Nod Factor–Dependent Transduction Pathway in Lotus Japonicus." Molecular Plant-Microbe Interactions 20(8): 994–1003.
- Barrus, M. F. 1911. "Variation of Varieties of Beans in Their Susceptibility to Anthracnose." *Phytopathology* 1(6): 190–95.
- Bascon, Maria Victoria et al. 2022. "Estimating Yield-Related Traits Using UAV-Derived Multispectral Images to Improve Rice Grain Yield Prediction." *Agriculture* 12(8): 1141.

- Bergersen, F. J. 1980. "Measurement of Nitrogen Fixation by Direct Means." *Measurement of nitrogen fixation by direct means*.: 65–110.
- Boddey, Robert M., and Roger Knowles. 1987. "Methods for Quantification of Nitrogen Fixation Associated with Gramineae." *Critical Reviews in Plant Sciences* 6(3): 209–66.
- Boersma, J. G. et al. 2014. "Combining Resistance to Common Bacterial Blight, Anthracnose, and Bean Common Mosaic Virus into Manitoba-Adapted Dry Bean (Phaseolus Vulgaris L.) Cultivars." *Canadian Journal of Plant Science* 94(2): 405–15.
- Boersma, Stephen J., Don J. Depuydt, Richard J. Vyn, and Chris L. Gillard. 2020. "Fungicide Efficacy for Control of Anthracnose of Dry Bean in Ontario." *Crop Protection* 127: 104979.
- Bollman, Mavis I., and J. Kevin Vessey. 2006. "Differential Effects of Nitrate and Ammonium Supply on Nodule Initiation, Development, and Distribution on Roots of Pea (*Pisum Sativum*)." *Canadian Journal of Botany* 84(6): 893–903.
- Boshkovski, Blagoja et al. 2021. "Relationship between Physiological and Biochemical Measurements with Spectral Reflectance for Two *Phaseolus Vulgaris* L. Genotypes under Multiple Stress." *International Journal of Remote Sensing* 42(4): 1230–49.
- Brockwell, J., and P. J. Bottomley. 1995. "Recent Advances in Inoculant Technology and Prospects for the Future." *Soil Biology and Biochemistry* 27(4): 683–97.
- Burt, Andrew J. et al. 2015. "Candidate Gene Identification with SNP Marker-Based Fine Mapping of Anthracnose Resistance Gene Co-4 in Common Bean." *PLOS ONE* 10(10): e0139450.
- Buttery, B. R., S-J. Park, and P. van Berkum. 1997. "Effects of Common Bean (*Phaseolus Vulgaris* L.) Cultivar and Rhizobium Strain on Plant Growth, Seed Yield and Nitrogen Content." *Canadian Journal of Plant Science* 77(3): 347–51.
- Campa, Ana, Cristina Rodríguez-Suárez, Ramón Giraldez, and Juan José Ferreira. 2014. "Genetic Analysis of the Response to Eleven Colletotrichum Lindemuthianum Races in a RIL Population of Common Bean (Phaseolus Vulgaris L.)." *BMC Plant Biology* 14(1): 115.
- Candiago, Sebastian et al. 2015. "Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images." *Remote Sensing* 7(4): 4026–47.
- Chalk, Phillip M. 1985. "Estimation of N2 Fixation by Isotope Dilution: An Appraisal of Techniques Involving 15N Enrichment and Their Application." Soil Biology and Biochemistry 17(4): 389–410.
- Chen, X, and P F Sullivan. 2003. "Single Nucleotide Polymorphism Genotyping: Biochemistry, Protocol, Cost and Throughput." *Pharmacogenomics Journal* 3(2): 77.

- Cissoko, Maimouna et al. 2018. "Actinorhizal Signaling Molecules: Frankia Root Hair Deforming Factor Shares Properties With NIN Inducing Factor." *Frontiers in Plant Science* 9. https://www.frontiersin.org/articles/10.3389/fpls.2018.01494 (November 23, 2022).
- Conner, R. L. et al. 2009. "Seedborne Infection Affects Anthracnose Development in Two Dry Bean Cultivars." *Canadian Journal of Plant Pathology* 31(4): 449–55.
- Cortés, Andrés J., Martha C. Chavarro, and Matthew W. Blair. 2011. "SNP Marker Diversity in Common Bean (Phaseolus Vulgaris L.)." *Theoretical and Applied Genetics* 123(5): 827– 45.
- Costa, Bruno Henrique Garcia et al. 2018. "Potassium Phosphites in the Protection of Common Bean Plants against Anthracnose and Biochemical Defence Responses." *Journal of Phytopathology* 166(2): 95–102.
- Costa, Larissa Carvalho et al. 2021. "Different Loci Control Resistance to Different Isolates of the Same Race of Colletotrichum Lindemuthianum in Common Bean." *Theoretical and Applied Genetics* 134(2): 543–56.
- Di Gennaro, Salvatore Filippo et al. 2018. "UAV-Based High-Throughput Phenotyping to Discriminate Barley Vigour with Visible and near-Infrared Vegetation Indices." *International Journal of Remote Sensing* 39(15–16): 5330–44.
- Diaz, Lucy M. et al. 2017. "Phenotypic Evaluation and QTL Analysis of Yield and Symbiotic Nitrogen Fixation in a Common Bean Population Grown with Two Levels of Phosphorus Supply." *Molecular Breeding* 37(6): 76.
- Diaz, Lucy Milena et al. 2018. "QTL Analyses for Tolerance to Abiotic Stresses in a Common Bean (Phaseolus Vulgaris L.) Population." *PLOS ONE* 13(8): e0202342.
- Dinh, H. T. et al. 2013. "Biological Nitrogen Fixation of Peanut Genotypes with Different Levels of Drought Tolerance under Mid-Season Drought." *SABRAO Journal of Breeding and Genetics* 45(3): 491–503.
- Durand, J. L., J. E. Sheehy, and F. R. Minchin. 1987. "Nitrogenase Activity, Photosynthesis and Nodule Water Potential in Soyabean Plants Experiencing Water Deprivation." *Journal of Experimental Botany* 38(2): 311–21.
- Elizondo Barron, J et al. 1999. "Response to Selection for Seed Yield and Nitrogen (N2) Fixation in Common Bean (Phaseolus Vulgaris L.)." *Field Crops Research* 62(2): 119–28.
- Elshire, Robert J. et al. 2011. "A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species." *PLOS ONE* 6(5): e19379.
- Erisman, Jan Willem et al. 2008. "How a Century of Ammonia Synthesis Changed the World." *Nature Geoscience* 1(10): 636–39.

- Farid, Mehdi, Hugh J. Earl, K. Peter Pauls, and Alireza Navabi. 2017. "Response to Selection for Improved Nitrogen Fixation in Common Bean (Phaseolus Vulgaris L.)." *Euphytica* 213(4): 99.
- Farid, Mehdi, and Alireza Navabi. 2015. "N2 Fixation Ability of Different Dry Bean Genotypes." *Canadian Journal of Plant Science* 95: 15.
- Fenta, Berhanu Amsalu, Stephen E. Beebe, and Karl J. Kunert. 2020. "Role of Fixing Nitrogen in Common Bean Growth under Water Deficit Conditions." *Food and Energy Security* 9(1): e183.
- Ferguson, Brett J. et al. 2019. "Legume Nodulation: The Host Controls the Party." *Plant, Cell & Environment* 42(1): 41–51.
- Ferreira, Juan José, Ana Campa, and James D. Kelly. 2013. "Organization of Genes Conferring Resistance to Anthracnose in Common Bean." In *Translational Genomics for Crop Breeding*, John Wiley & Sons, Ltd, 151–81. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118728475.ch9 (September 27, 2022).
- Fisher, Robert F., and Sharon R. Long. 1992. "Rhizobium–Plant Signal Exchange." *Nature* 357(6380): 655–60.
- Flor, H. H. 1955. "Host-Parasite Interaction in Flaxrust-Its Genetics and Other Implications." *Phytopathology* 45: 680–85.
- Fortin, Jérôme G., François Anctil, Léon-Étienne Parent, and Martin A. Bolinder. 2011. "Site-Specific Early Season Potato Yield Forecast by Neural Network in Eastern Canada." *Precision Agriculture* 12(6): 905–23.
- Franche, Claudine, Kristina Lindström, and Claudine Elmerich. 2009. "Nitrogen-Fixing Bacteria Associated with Leguminous and Non-Leguminous Plants." *Plant and Soil* 321(1): 35– 59.
- Frink, Charles R., Paul E. Waggoner, and Jesse H. Ausubel. 1999. "Nitrogen Fertilizer: Retrospect and Prospect." *Proceedings of the National Academy of Sciences* 96(4): 1175– 80.
- Fu, Yuanyuan et al. 2021. "An Overview of Crop Nitrogen Status Assessment Using Hyperspectral Remote Sensing: Current Status and Perspectives." *European Journal of* Agronomy 124: 126241.
- Gadaga, Stélio Jorge Castro, Mário Sobral de Abreu, Mário Lúcio Vilela de Resende, and Pedro Martins Ribeiro. 2017. "Phosphites for the Control of Anthracnose in Common Bean." *Pesquisa Agropecuária Brasileira* 52: 36–44.
- Gilio, Thiago Alexandre Santana et al. 2020. "Fine Mapping of an Anthracnose-Resistance Locus in Andean Common Bean Cultivar Amendoim Cavalo." *PLOS ONE* 15(10):

e0239763.

Gillard, C. L., N. K. Ranatunga, and R. L. Conner. 2012. "The Control of Dry Bean Anthracnose through Seed Treatment and the Correct Application Timing of Foliar Fungicides." Crop Protection 37: 81–90.

Giller, Ken E. 2001. Nitrogen Fixation in Tropical Cropping Systems. CABI.

- Giovannoni, James J., Rod A. Wing, Martin W. Ganal, and Steven D. Tanksley. 1991. "Isolation of Molecular Markers from Specific Chromosomal Intervals Using DNA Pools from Existing Mapping Populations." *Nucleic Acids Research* 19(23): 6553–68.
- Gonzalez-Sanchez, Alberto, Juan Frausto-Solis, and Waldo Ojeda-Bustamante. 2014. "Predictive Ability of Machine Learning Methods for Massive Crop Yield Prediction." *Spanish Journal of Agricultural Research* 12(2): 313–28.
- Gracia-Romero, Adrian et al. 2017. "Comparative Performance of Ground vs. Aerially Assessed RGB and Multispectral Indices for Early-Growth Evaluation of Maize Performance under Phosphorus Fertilization." *Frontiers in Plant Science* 8. https://www.frontiersin.org/articles/10.3389/fpls.2017.02004 (October 25, 2022).
- Grüner, Esther, Thomas Astor, and Michael Wachendorf. 2021. "Prediction of Biomass and N Fixation of Legume–Grass Mixtures Using Sensor Fusion." *Frontiers in Plant Science* 11. https://www.frontiersin.org/articles/10.3389/fpls.2020.603921 (August 4, 2022).
- Gunnabo, A. H. et al. 2019. "Genetic Interaction Studies Reveal Superior Performance of Rhizobium Tropici CIAT899 on a Range of Diverse East African Common Bean (Phaseolus Vulgaris L.) Genotypes." Applied and Environmental Microbiology 85(24): e01763-19.
- Gwata, E. T., D. S. Wofford, P. L. Pfahler, and K. J. Boote. 2004. "Genetics of Promiscuous Nodulation in Soybean: Nodule Dry Weight and Leaf Color Score." *Journal of Heredity* 95(2): 154–57.
- Haghighattalab, Atena et al. 2016. "Application of Unmanned Aerial Systems for High Throughput Phenotyping of Large Wheat Breeding Nurseries." *Plant Methods* 12(1): 35.
- Han, Jichong et al. 2020. "Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China." *Remote Sensing* 12(2): 236.
- Han, Mei et al. 2015. "The Genetics of Nitrogen Use Efficiency in Crop Plants." *Annual Review* of Genetics 49(1): 269–89.
- Han, Nana et al. 2022. "Rapid Diagnosis of Nitrogen Nutrition Status in Summer Maize over Its Life Cycle by a Multi-Index Synergy Model Using Ground Hyperspectral and UAV Multispectral Sensor Data." Atmosphere 13(1): 122.

Hansen, P. M., and J. K. Schjoerring. 2003. "Reflectance Measurement of Canopy Biomass and

Nitrogen Status in Wheat Crops Using Normalized Difference Vegetation Indices and Partial Least Squares Regression." *Remote Sensing of Environment* 86(4): 542–53.

- Hart, John P., and Phillip D. Griffiths. 2015. "Genotyping-by-Sequencing Enabled Mapping and Marker Development for the By-2 Potyvirus Resistance Allele in Common Bean." *The Plant Genome* 8(1): plantgenome2014.09.0058.
- Heilig, James A. et al. 2016. "QTL Analysis of Symbiotic Nitrogen Fixation in a Black Bean Population." *Crop Science* 57. https://acsess.onlinelibrary.wiley.com/doi/full/10.2135/cropsci2016.05.0348 (May 16, 2022).
- Herridge, David F. 1982. "Relative Abundance of Ureides and Nitrate in Plant Tissues of Soybean as a Quantitative Assay of Nitrogen Fixation." *Plant Physiology* 70(1): 1–6.
- Herzig, Paul et al. 2021. "Evaluation of RGB and Multispectral Unmanned Aerial Vehicle (UAV) Imagery for High-Throughput Phenotyping and Yield Prediction in Barley Breeding." *Remote Sensing* 13(14): 2670.
- Hungria, Mariangela, and Milton A. T. Vargas. 2000. "Environmental Factors Affecting N2 Fixation in Grain Legumes in the Tropics, with an Emphasis on Brazil." *Field Crops Research* 65(2): 151–64.
- Hunt, E. Raymond et al. 2013. "A Visible Band Index for Remote Sensing Leaf Chlorophyll Content at the Canopy Scale." *International Journal of Applied Earth Observation and Geoinformation* 21: 103–12.
- Hurtado-Gonzales, Oscar P et al. 2017. "Fine Mapping of Ur-3, a Historically Important Rust Resistance Locus in Common Bean." *G3 Genes/Genetics* 7(2): 557–69.
- Iquira, Elmer, Sonah Humira, and Belzile Francois. 2015. "Association Mapping of QTLs for Sclerotinia Stem Rot Resistance in a Collection of Soybean Plant Introductions Using a Genotyping by Sequencing (GBS) Approach | BMC Plant Biology | Full Text." https://bmcplantbiol.biomedcentral.com/articles/10.1186/s12870-014-0408-y (November 17, 2022).
- J. Brockwell, Peter J. Bottomley, and Janice E. Thies. 1995. "Manipulation of Rhizobia Microflora for Improving Legume Productivity and Soil Fertility: A Critical Assessment." *Plant and Soil* 174: 143–80.
- Jhonata, Lemos da Silva et al. 2015. "Essential Oil of Cymbopogon Flexuosus, Vernonia Polyanthes and Potassium Phosphite in Control of Bean Anthracnose." *Journal of Medicinal Plants Research* 9(8): 243–53.
- Jiang, Yunfei et al. 2020. "Evaluation of Beneficial and Inhibitory Effects of Nitrate on Nodulation and Nitrogen Fixation in Common Bean (Phaseolus Vulgaris)." *Legume Science* 2(3): e45.

- Jimenez-Jimenez, Saul et al. 2019. "Differential Tetraspanin Genes Expression and Subcellular Localization during Mutualistic Interactions in Phaseolus Vulgaris." *PLOS ONE* 14(8): e0219765.
- Kamfwa, Kelvin, Karen A. Cichy, and James D. Kelly. 2015. "Genome-Wide Association Analysis of Symbiotic Nitrogen Fixation in Common Bean." *Theoretical and Applied Genetics* 128(10): 1999–2017.
- Kamfwa, Kelvin, Karen A. Cichy, and James D. Kelly. 2019. "Identification of Quantitative Trait Loci for Symbiotic Nitrogen Fixation in Common Bean." *Theoretical and Applied Genetics* 132(5): 1375–87.
- Kim, Nari, and Yang-Won Lee. 2016. "Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State." *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography* 34(4): 383–90.
- Korani, Walid et al. 2021. Accurate Analysis of Short Read Sequencing in Complex Genomes: A Case Study Using QTL-Seq to Target Blanchability in Peanut (Arachis Hypogaea). Genomics. preprint. http://biorxiv.org/lookup/doi/10.1101/2021.03.13.435236 (May 14, 2022).
- Kunert, Karl J. et al. 2016. "Drought Stress Responses in Soybean Roots and Nodules." *Frontiers in Plant Science* 7. https://www.frontiersin.org/articles/10.3389/fpls.2016.01015 (November 28, 2022).
- Ladha, Jagdish K. et al. 2005. "Efficiency of Fertilizer Nitrogen in Cereal Production: Retrospects and Prospects." In *Advances in Agronomy*, Academic Press, 85–156. https://www.sciencedirect.com/science/article/pii/S0065211305870038 (November 22, 2022).
- Lam, Hon-Ming et al. 1996. "The Molecular-Genetics of Nitrogen Assimilation into Amino Acids in Higher Plants." Annual review of plant physiology and plant molecular biology 47: 569–93.
- Lassaletta, Luis et al. 2016. "Nitrogen Use in the Global Food System: Past Trends and Future Trajectories of Agronomic Performance, Pollution, Trade, and Dietary Demand." *Environmental Research Letters* 11(9): 095007.
- Lévy, Julien et al. 2004. "A Putative Ca²⁺ and Calmodulin-Dependent Protein Kinase Required for Bacterial and Fungal Symbioses." *Science* 303(5662): 1361–64.
- Li, Huihui et al. 2015. "A High Density GBS Map of Bread Wheat and Its Application for Dissecting Complex Disease Resistance Traits." *BMC Genomics* 16(1): 216.
- de Lima Castro, Sandra Aparecida et al. 2017. "Genetics and Mapping of a New Anthracnose Resistance Locus in Andean Common Bean Paloma." *BMC Genomics* 18(1): 306.

Lira Junior, Mario. 2015. "Legume-Rhizobia Signal Exchange: Promiscuity and Environmental

Effects." *Frontiers in Microbiology* 6. https://www.frontiersin.org/articles/10.3389/fmicb.2015.00945 (November 28, 2022).

- Liu, Cheng-Wu, and Jeremy D. Murray. 2016. "The Role of Flavonoids in Nodulation Host-Range Specificity: An Update." *Plants* 5(3): 33.
- Liu, Jinge, Shengming Yang, Qiaolin Zheng, and Hongyan Zhu. 2014. "Identification of a Dominant Gene in Medicago Truncatula That Restricts Nodulation by Sinorhizobium Meliloti Strain Rm41." *BMC Plant Biology* 14(1): 167.
- Lu, Ning et al. 2019. "Improved Estimation of Aboveground Biomass in Wheat from RGB Imagery and Point Cloud Data Acquired with a Low-Cost Unmanned Aerial Vehicle System." *Plant Methods* 15(1): 17.
- Marcial-Pablo, Mariana de Jesús et al. 2019. "Estimation of Vegetation Fraction Using RGB and Multispectral Images from UAV." *International Journal of Remote Sensing* 40(2): 420– 38.
- Marcon, João Ricardo Silva et al. 2020. "Genetic Resistance of Common Bean Cultivar Beija Flor to *Colletotrichum Lindemuthianum*." *Acta Scientiarum*. *Agronomy* 43. http://www.scielo.br/j/asagr/a/MbhdFC45gvM5j59Dfjm89Wj/ (November 17, 2022).
- Masangwa, J. I. G., T. a. S. Aveling, and Q. Kritzinger. 2013. "Screening of Plant Extracts for Antifungal Activities against Collectorichum Species of Common Bean (Phaseolus Vulgaris L.) and Cowpea (Vigna Unguiculata (L.) Walp)." *The Journal of Agricultural Science* 151(4): 482–91.
- McKay, Ian A., and Michael A. Djordjevic. 1993. "Production and Excretion of Nod Metabolites by Rhizobium Leguminosarum Bv. Trifolii Are Disrupted by the Same Environmental Factors That Reduce Nodulation in the Field." *Applied and Environmental Microbiology* 59(10): 3385–92.
- Melotto, Maeli, R. S. Balardin, and J. D. Kelly. 2000. "Host-Pathogen Interaction and Variability of Colletotrichum Lindemuthianum." In *Colletotrichum Host Specifity, Pathology, Host-Pathogen Interaction*, APS press, St. Paul, MN, 346–61.
- Michelmore, R W, I Paran, and R V Kesseli. 1991. "Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations." *Proceedings of the National Academy of Sciences* 88(21): 9828–32.
- Michelmore, Richard W., Marilena Christopoulou, and Katherine S. Caldwell. 2013. "Impacts of Resistance Gene Genetics, Function, and Evolution on a Durable Future." *Annual Review of Phytopathology* 51(1): 291–319.
- Mohammed, Amin. 2013. "An Overview of Distribution, Biology and the Management of Common Bean Anthracnose." *Journal of Plant Pathology and Microbiology* 04.

- Mulvaney, R. L., S. A. Khan, and T. R. Ellsworth. 2009. "Synthetic Nitrogen Fertilizers Deplete Soil Nitrogen: A Global Dilemma for Sustainable Cereal Production." *Journal of Environmental Quality* 38(6): 2295–2314.
- Münch, Steffen et al. 2008. "The Hemibiotrophic Lifestyle of Colletotrichum Species." *Journal* of Plant Physiology 165(1): 41–51.
- Mus, Florence et al. 2016. "Symbiotic Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes" ed. R. M. Kelly. *Applied and Environmental Microbiology* 82(13): 3698– 3710.
- Näsi, Roope et al. 2018. "Estimating Biomass and Nitrogen Amount of Barley and Grass Using UAV and Aircraft Based Spectral and Photogrammetric 3D Features." *Remote Sensing* 10(7): 1082.
- Nebiker, Stephan, Natalie Lack, Martin Abächerli, and Sonja Läderach. 2016. "LIGHT-WEIGHT MULTISPECTRAL UAV SENSORS AND THEIR CAPABILITIES FOR PREDICTING GRAIN YIELD AND DETECTING PLANT DISEASES." *ISPRS -International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLI-B1: 963–70.
- Negera, Abraham, and Mashilla Dejene. 2018. "Effect of Integrating Variety, Seed Treatment, and Foliar Fungicide Spray Timing on Managing Common Bean Anthracnose at Bako, Western Ethiopia." *East African Journal of Sciences* 12(2): 111–26.
- Nodari, R. O. et al. 1993. "Toward an Integrated Linkage Map of Common Bean. III. Mapping Genetic Factors Controlling Host-Bacteria Interactions." *Genetics* 134(1): 341–50.
- Nunes, Maria Paula Barion A. et al. 2021. "Relationship of Colletotrichum Lindemuthianum Races and Resistance Loci in the Phaseolus Vulgaris L. Genome." *Crop Science* 61(6): 3877–93.
- Padder, B.A. et al. 2010. "Evaluation of Bioagents and Biopesticides against Collectrichum Lindemuthianum and Its Integrated Management in Common Bean | Notulae Scientia Biologicae." https://notulaebiologicae.ro/index.php/nsb/article/view/4772 (September 27, 2022).
- Padder, B.A., P.N. Sharma, H.E. Awale, and J.D. Kelly. 2017. "COLLETOTRICHUM LINDEMUTHIANUM, THE CAUSAL AGENT OF BEAN ANTHRACNOSE." *Journal* of Plant Pathology 99(2): 317–30.
- Pastor-Corrales, Marcial A. 1991. "Standardization of Differential Varieties and Race Designation of Colletotrichum Lindemuthianum." *Phytopathology* 81: 694.
- Perfect, Sarah E. et al. 1998. "Expression Cloning of a Fungal Proline-Rich Glycoprotein Specific to the Biotrophic Interface Formed in the Colletotrichum–Bean Interaction." *The Plant Journal* 15(2): 273–79.

- Perret, Xavier, Christian Staehelin, and William J. Broughton. 2000. "Molecular Basis of Symbiotic Promiscuity." *Microbiology and Molecular Biology Reviews* 64(1): 180–201.
- Petersen, Steffen E., and Nay Aung. 2022. "Benefits of Machine Learning to Predict Survival Using Stress Perfusion CMR and Basic Clinical Information*." *JACC: Cardiovascular Imaging* 15(11): 1914–15.
- Rahman, Md Habibur et al. 2021. "Bioinformatics and Machine Learning Methodologies to Identify the Effects of Central Nervous System Disorders on Glioblastoma Progression." *Briefings in Bioinformatics* 22(5): bbaa365.
- Ramaekers, Lara et al. 2013. "Identifying Quantitative Trait Loci for Symbiotic Nitrogen Fixation Capacity and Related Traits in Common Bean." *Molecular Breeding* 31(1): 163– 80.
- Razon, Luis F. 2014. "Life Cycle Analysis of an Alternative to the Haber-Bosch Process: Non-Renewable Energy Usage and Global Warming Potential of Liquid Ammonia from Cyanobacteria." *Environmental Progress & Sustainable Energy* 33(2): 618–24.
- Reinprecht, Yarmilla et al. 2020. "Effects of Nitrogen Application on Nitrogen Fixation in Common Bean Production." *Frontiers in Plant Science* 11. https://www.frontiersin.org/article/10.3389/fpls.2020.01172 (April 20, 2022).
- Riar, Mandeep K. et al. 2018. "Expression of Drought-Tolerant N2 Fixation in Heterogeneous Inbred Families Derived from PI471938 and Hutcheson Soybean." *Crop Science* 58(1): 364–69.
- Sadowsky, Michael J., Perry B. Cregan, Francisco Rodriguez-Quinones, and Harold H. Keyser. 1991. "Microbial Influence on Gene-for-Gene Interactions in Legume-Rhizobium Symbioses." In *The Rhizosphere and Plant Growth: Papers Presented at a Symposium Held May 8–11, 1989, at the Beltsville Agricultural Research Center (BARC), Beltsville, Maryland*, Beltsville Symposia in Agricultural Research, eds. Donald L. Keister and Perry B. Cregan. Dordrecht: Springer Netherlands, 173–80. https://doi.org/10.1007/978-94-011-3336-4_33 (November 28, 2022).
- Saikia, S P, and Vanita Jain. 2007. "Biological Nitrogen Fixation with Non-Legumes: An Achievable Target or a Dogma?" *CURRENT SCIENCE* 92(3): 6.
- Samuel, A. L. 2000. "Some Studies in Machine Learning Using the Game of Checkers." *IBM Journal of Research and Development* 44(1.2): 206–26.
- Sandino Mora, Juan, Geoff Pegg, Felipe Gonzalez, and Grant Smith. 2018. "Aerial Mapping of Forests Affected by Pathogens Using UAVs, Hyperspectral Sensors, and Artificial Intelligence." Sensors 18(4): Article number: 944 1-17.
- Santi, Carole, Didier Bogusz, and Claudine Franche. 2013. "Biological Nitrogen Fixation in Non-Legume Plants." *Annals of Botany* 111(5): 743–67.
- Schmutz, Jeremy et al. 2014. "A Reference Genome for Common Bean and Genome-Wide Analysis of Dual Domestications." *Nature Genetics* 46(7): 707–13.
- Schumpp, Olivier, and William J. Deakin. 2010. "How Inefficient Rhizobia Prolong Their Existence within Nodules." *Trends in Plant Science* 15(4): 189–95.
- Schwartz, Howard F., and Marcial A. Pastor Corrales. 1989. In *Bean Production Problems in the Tropics*, CIAT, 77–104.
- Schwenke, G. D., M. B. Peoples, G. L. Turner, and D. F. Herridge. 1998. "Does Nitrogen Fixation of Commercial, Dryland Chickpea and Faba Bean Crops in North-West New South Wales Maintain or Enhance Soil Nitrogen?" *Australian Journal of Experimental Agriculture* 38(1): 61.
- Serraj, Rachid, Thomas R Sinclair, and Larry C Purcell. 1999. "Symbiotic N2 Fixation Response to Drought." *Journal of Experimental Botany* 50(331): 143–55.
- Shearer, G., and D. H. Kohl. 1986. "N2-Fixation in Field Settings: Estimations Based on Natural 15N Abundance." *Functional Plant Biology* 13(6): 699–756.
- Shendryk, Yuri, Robert Davy, and Peter Thorburn. 2021. "Integrating Satellite Imagery and Environmental Data to Predict Field-Level Cane and Sugar Yields in Australia Using Machine Learning." *Field Crops Research* 260: 107984.
- Shi, Peihua et al. 2021. "Rice Nitrogen Nutrition Estimation with RGB Images and Machine Learning Methods." *Computers and Electronics in Agriculture* 180: 105860.
- Shovan, L.R., J.A. Bhuiyan, B. Pervez, and Z. Pervez. 2008. "IN VITRO CONTROL OF COLLETOTRICHUM DEMATIUM CAUSING ANTHRACNOSE OF SOYBEAN BY FUNGICIDES, PLANT EXTRACTS AND TRICHODERMA HARZ." International Journal of Sustainable Crop Production 3(3): 10–17.
- Simms, Ellen L., and D. Lee Taylor. 2002. "Partner Choice in Nitrogen-Fixation Mutualisms of Legumes and Rhizobia1." *Integrative and Comparative Biology* 42(2): 369–80.
- Simsek, Senay, Tuula Ojanen-Reuhs, Samuel B. Stephens, and Bradley L. Reuhs. 2007. "Strain-Ecotype Specificity in Sinorhizobium Meliloti-Medicago Truncatula Symbiosis Is Correlated to Succinoglycan Oligosaccharide Structure." *Journal of Bacteriology* 189(21): 7733–40.
- Sobrino, Beatriz, María Brión, and Angel Carracedo. 2005. "SNPs in Forensic Genetics: A Review on SNP Typing Methodologies." *Forensic Science International* 154(2): 181–94.
- Soumare, Abdoulaye et al. 2020. "Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable Agriculture." *Plants* 9(8): 1011.
- Souza, Alessandra A. et al. 2000. "Effects of Phaseolus Vulgaris QTL in Controlling Host-Bacteria Interactions under Two Levels of Nitrogen Fertilization." *Genetics and*

Molecular Biology 23(1): 155–61.

- Stas, Michiel et al. 2016. "A Comparison of Machine Learning Algorithms for Regional Wheat Yield Prediction Using NDVI Time Series of SPOT-VGT." In 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), , 1–5.
- Stewart, W D, G P Fitzgerald, and R H Burris. 1967. "In Situ Studies on N2 Fixation Using the Acetylene Reduction Technique." *Proceedings of the National Academy of Sciences* 58(5): 2071–78.
- Stracke, Silke et al. 2002. "A Plant Receptor-like Kinase Required for Both Bacterial and Fungal Symbiosis." *Nature* 417(6892): 959–62.
- Strange, Richard, and Peter Scott. 2005. "Plant Disease: A Threat to Global Food Security." *Annual review of phytopathology* 43: 83–116.
- Streeter, John G. 1994. "Failure of Inoculant Rhizobia to Overcome the Dominance of Indigenous Strains for Nodule Formation." *Canadian Journal of Microbiology* 40(7): 513–22.
- Streeter, John, and Peter P. Wong. 1988. "Inhibition of Legume Nodule Formation and N ₂ Fixation by Nitrate." *Critical Reviews in Plant Sciences* 7(1): 1–23.
- Suzaki, Takuya et al. 2019. "LACK OF SYMBIONT ACCOMMODATION Controls Intracellular Symbiont Accommodation in Root Nodule and Arbuscular Mycorrhizal Symbiosis in Lotus Japonicus." *PLOS Genetics* 15(1): e1007865.
- Takagi, Hiroki et al. 2013. "QTL-Seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole Genome Resequencing of DNA from Two Bulked Populations." *The Plant Journal* 74(1): 174–83.
- Tegeder, Mechthild. 2014. "Transporters Involved in Source to Sink Partitioning of Amino Acids and Ureides: Opportunities for Crop Improvement." *Journal of Experimental Botany* 65(7): 1865–78.
- Tirichine, Leïla, Françoise de Billy, and Thierry Huguet. 2000. "Mtsym6, a Gene ConditioningSinorhizobium Strain-Specific Nitrogen Fixation InMedicago Truncatula 1." *Plant Physiology* 123(3): 845–52.
- Travlos, I. et al. 2017. "The Use of RGB Cameras in Defining Crop Development in Legumes." *Advances in Animal Biosciences* 8(2): 224–28.
- Triplett, Eric W., and Michael J. Sadowsky. 1992. "Genetics of Competition for Nodulation of Legumes." *Annual Review of Microbiology* 46(1): 399–422.
- Tsai, Siu et al. 1998. "QTL Mapping for Nodule Number and Common Bacterial Blight in Phaseolus Vulgaris L." *Plant and Soil* 204: 135–45.

- Tu, J. C. 1983. "Epidemiology of Anthracnose Caused by Collectorichum Lindemuthianum on White Bean (Phaseolus Vulgaris) in Southern Ontario: Survival of the Pathogen." *Plant Disease* 67(4): 402–4.
- Turner, R. Eugene, and Nancy N. Rabalais. 1994. "Coastal Eutrophication near the Mississippi River Delta." *Nature* 368(6472): 619–21.
- Turner, R. Eugene, Nancy N. Rabalais, and Dubravko Justic. 2008. "Gulf of Mexico Hypoxia: Alternate States and a Legacy." *Environmental Science & Technology* 42(7): 2323–27.
- Valentini, Giseli et al. 2017. "High-Resolution Mapping Reveals Linkage between Genes in Common Bean Cultivar Ouro Negro Conferring Resistance to the Rust, Anthracnose, and Angular Leaf Spot Diseases." *Theoretical and Applied Genetics* 130(8): 1705–22.
- Van Noorden, Giel E. et al. 2016. "Molecular Signals Controlling the Inhibition of Nodulation by Nitrate in Medicago Truncatula." *International Journal of Molecular Sciences* 17(7): 1060.
- Villordo-Pineda, Emiliano et al. 2015. "Identification of Novel Drought-Tolerant-Associated SNPs in Common Bean (Phaseolus Vulgaris)." *Frontiers in Plant Science* 6. https://www.frontiersin.org/articles/10.3389/fpls.2015.00546 (November 18, 2022).
- Virlet, Nicolas, Kasra Sabermanesh, Pouria Sadeghi-Tehran, and Malcolm J. Hawkesford. 2017. "Field Scanalyzer: An Automated Robotic Field Phenotyping Platform for Detailed Crop Monitoring." *Functional Plant Biology* 44(1): 143.
- Walsh, K. B. 1995. "Physiology of the Legume Nodule and Its Response to Stress." *Soil Biology and Biochemistry* 27(4): 637–55.
- Wang, Chao, Benjamin Z. Houlton, Weiwei Dai, and Edith Bai. 2017. "Growth in the Global N2 Sink Attributed to N Fertilizer Inputs over 1860 to 2000." Science of The Total Environment 574: 1044–53.
- Wang, Dong, Shengming Yang, Fang Tang, and Hongyan Zhu. 2012. "Symbiosis Specificity in the Legume Rhizobial Mutualism." *Cellular Microbiology* 14(3): 334–42.
- Wang, Qi et al. 2017. "Host-Secreted Antimicrobial Peptide Enforces Symbiotic Selectivity in Medicago Truncatula." *Proceedings of the National Academy of Sciences* 114(26): 6854– 59.
 - ———. 2018. "Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in *Medicago Truncatula*." *Molecular Plant-Microbe Interactions*® 31(2): 240–48.
- Wang, Xiukang et al. 2019. "Chapter Three The Effects of Mulch and Nitrogen Fertilizer on the Soil Environment of Crop Plants." In *Advances in Agronomy*, ed. Donald L. Sparks. Academic Press, 121–73. https://www.sciencedirect.com/science/article/pii/S0065211318300786 (November 22,

2022).

- Xu, Rongting et al. 2019. "Global Ammonia Emissions from Synthetic Nitrogen Fertilizer Applications in Agricultural Systems: Empirical and Process-Based Estimates and Uncertainty." *Global Change Biology* 25(1): 314–26.
- Yang, Qi et al. 2019. "Deep Convolutional Neural Networks for Rice Grain Yield Estimation at the Ripening Stage Using UAV-Based Remotely Sensed Images." *Field Crops Research* 235: 142–53.
- Yang, Shengming et al. 2010. "R Gene-Controlled Host Specificity in the Legume–Rhizobia Symbiosis." *Proceedings of the National Academy of Sciences* 107(43): 18735–40.
- Yao, Xia et al. 2015. "Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration." *Remote Sensing* 7(11): 14939–66.
- Yates, Ron John, John Gregory Howieson, Wayne Gerald Reeve, and Graham William O'Hara.
 2011. "A Re-Appraisal of the Biology and Terminology Describing Rhizobial Strain Success in Nodule Occupancy of Legumes in Agriculture." *Plant and Soil* 348(1): 255.
- Young, RA, Maeli Melotto, Rubens Nodari, and James Kelly. 1998. "Marker-Assisted Dissection of the Oligogenic Anthracnose Resistance in the Common Bean Cultivar, 'G2333." *TAG Theoretical and Applied Genetics* 96: 87–94.
- Young, Roberto A., and James D. Kelly. 1997. "RAPD Markers Linked to Three Major Anthracnose Resistance Genes in Common Bean." *Crop Science* 37(3): cropsci1997.0011183X003700030039x.
- Zablotowicz, Robert M., and Krishna N. Reddy. 2007. "Nitrogenase Activity, Nitrogen Content, and Yield Responses to Glyphosate in Glyphosate-Resistant Soybean." *Crop Protection* 26(3): 370–76.
- Zhang, Chongyuan, Afef Marzougui, and Sindhuja Sankaran. 2020. "High-Resolution Satellite Imagery Applications in Crop Phenotyping: An Overview." Computers and Electronics in Agriculture 175: 105584.
- Zhang, Lingxiao, Joe Zhang, Stephen Kyei-Boahen, and Minghua Zhang. 2010. "Simulation and Prediction of Soybean Growth and Development under Field Conditions." & Environ. Sci 7: 374–85.
- Zhang, Xin et al. 2015. "Managing Nitrogen for Sustainable Development." *Nature* 528(7580): 51–59.
- Zhou, Jianfeng et al. 2018. "Low Altitude Remote Sensing Technologies for Crop Stress Monitoring: A Case Study on Spatial and Temporal Monitoring of Irrigated Pinto Bean." *Precision Agriculture* 19(3): 555–69.

APPENDIX

Differential Cultivar	Gene Pool	Host Genes	Binary Number
Michelite	Middle American	<i>Co-11</i>	1
MDRK	Andean	Co-1	2
Perry Marrow	Andean	<i>Co-1</i> ³	4
Cornell 49242	Middle American	Co-2	8
Widusa	Andean	<i>Co-1</i> ⁵	16
Kaboon	Andean	<i>Co-1</i> ²	32
Mexico 222	Middle American	Со-3	64
PI 207262	Middle American	<i>Co-3</i> ³ ; <i>Co-4</i> ³	128
ТО	Middle American	Co-4	256
TU	Middle American	<i>Co-5</i>	512
AB 136	Middle American	Со-б	1024
G 2333	Middle American	<i>Co-3⁵</i> ; <i>Co-4</i> ² ; <i>Co-5</i> ²	2048

Table 1.1. Anthracnose differential series, host gene pool, resistance genes, and the binary number of each cultivar used to characterize the races of anthracnose in common bean.

Binary numbers are calculated from 2ⁿ where n is equivalent to the position of the cultivar in the series starting with 0. The sum of the cultivars with susceptible reactions gives the binary number of a specific race (Pastor-Corrales, 1991). For example, race 109 is virulent on Mexico 222 [64], Kaboon [32], Cornell 4942 [8], Perry Marrow [4], and Michelite [1].

CHAPTER TWO: MAPPING THE ANTHRACNOSE RESISTANCE GENE CO-5

ABSTRACT

Anthracnose is a highly variable and destructive seed-borne disease caused by the fungus *Colletotrichum lindemuthianum*. The most efficient method of managing the disease is to use resistant cultivars, but emergent races threaten their durability. The black bean cultivar 'TU' possesses the *Colletotrichum* resistance gene *Co-5* and is resistant to race 109, an emergent race in Michigan (2017). The objectives of this study were to i) map the resistance gene *Co-5* and ii) develop KASP markers to facilitate future marker assisted selection (MAS) for resistance. An F2 population developed from a cross between the susceptible Michigan State University (MSU) Dry Bean Breeding Program breeding line B19504 and TU was utilized for mapping. QTL-seq was performed to identify significant SNP markers associated with race 109 resistance and twenty-five SNPs were identified on the *Phaseolus vulgaris* linkage group Pv07 between 6.838810 Mb and 24.62480 Mb. Validating KASP markers associated with *Co-5* will allow breeding programs to efficiently integrate resistance to race 109 into their breeding lines, improving the durability of their cultivars.

INTRODUCTION

Bean anthracnose, caused by *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cav., is a seed-borne fungal pathogen and causal agent of one of the most devastating diseases affecting bean production worldwide (Costa et al., 2021). The pathogen is devastating in susceptible cultivars and displays high virulence diversity between races (Padder et al., 2017). Infected fields of susceptible cultivars in favorable, humid, and cool environments can have up to 100% yield reduction (Nunes et al., 2021). The pathogen is endemic in numerous African, Latin

American, and European countries in addition to many fields in both the United States and Canada (Mohammed, 2013). *Colletotrichum lindemuthianum* is easily spread to new regions by infected seed and between plants by rain drops, irrigation water, or mechanical movement such as wildlife or farm equipment through the field. As a hemibiotrophic pathogen, the hypersensitive response (cell death) in plants is suppressed until the infection has advanced. The fungus creates rust red to black lesions on the leaf petiole and along the leaf veins, causing vein necrosis (Boersma et al., 2014). Brown eyespots and lesions will also appear on the stem and pods, followed by deterioration and infection in the seeds. Ultimately, anthracnose can lead to premature defoliation, early flower and pod drop, and plant death (Boersma et al., 2020; Tu, 1983). Infected seeds left in the soil avter harvest can also infect future bean crops for an average of two to three years (Conner et al., 2019). Anthracnose spores can survive winter conditions inside seed and plant debris left in the field and become a source of inoculum for the next crop (Schwartz & Corrales, 1989).

Management strategies such as fungicide application, using certified disease-free seed, fungicide seed treatments, and crop rotation can be beneficial, but the most effective and sustainable approach is the use of resistant cultivars (Balardin, Jarosz, and Kelly 1997). In common bean, resistance follows the gene-for-gene model and is conferred by individual, independently segregating loci in *Colletotrichum* resistance gene family denoted *Co* followed by a number. Currently, 20 resistance genes from the *Co* family have been identified and mapped in the common bean genome, with resistance gene clusters present on Pv01, Pv02, Pv03, Pv04, Pv07, Pv08, and Pv11 (Nunes et al., 2021). Integrating gene-specific disease resistance in improved cultivars relies on the identification of resistant plants to cross with agronomically favorable plants (Strange & Scott, 2005). Resistance to anthracnose can be evaluated directly through greenhouse inoculation or indirectly using molecular markers. Inoculations are commonly performed by spraying a conidia

spore suspension of a selected race onto common bean seedlings. The plants are kept at 80% humidity for 72 hours and disease symptoms become apparent between 7 and 10 days after inoculation. The responses are rated using either a broad categorization of 'susceptible' and 'resistant' or rated on a scale of 1-9 or 0-5 (Drijfhout & Davis, 1989; Pastor-Corrales et al., 1998).

In previous years, two races of anthracnose were present in Michigan: races 7 and 73 (Kelly et al., 1994). Due to their prevalence, the resistance gene $Co-I^2$ on Pv01 was commonly deployed in navy and black bean cultivars to avoid infection. In 2017, a new isolate of *C. lindemuthianum* was characterized in Michigan that was able to overcome $Co-I^2$ (Awale et al. 2018). This race was first detected on the cultivar 'Zenith' that possesses the Co-1² gene and characterized as 109. Severe infections were observed in previously durable varieties across all market classes. In the proceeding years the resistance genes Co-I, $Co-4^2$, Co-5, and Co-6 were found to confer resistance to race 109. $Co-4^2$ was deployed in the MSU Dry Bean Breeding Program. Two commercial varieties 'Adams' and 'Eiger' resistant to Race 109 were released using marker assisted selection (MAS) with a single nucleotide polymorphism (SNP) marker assay and greenhouse inoculation (Kelly et al., 2021a, 2021b).

MAS is the most efficient means of introgressing resistance (Choudhary et al., 2018). Molecular markers tightly linked to resistance quantitative trait loci (QTLs), or genes allow for the efficient deployment of anthracnose resistance into dry bean germplasm and cultivars threatened by emergent races. One resistance gene can confer resistance to a number of anthracnose races, thus combining complementary genes can strengthen the durability of resistance in cultivars (Balardin, Jarosz, and Kelly 1997). Before the advent of whole genome sequencing, random amplified polymorphic DNA (RAPD) and sequence-characterized amplified region (SCAR) markers were used to map and identify *Co* genes in common bean (Vallejo & Kelly 2001; Young & Kelly 1997; Young & Kelly 1998). Traditionally, bulked segregant analysis was used to identify DNA markers linked to a gene associated with disease resistance (Michelmore, Paran, and Kesseli 1991). The progression of sequencing technology has led to the combination of bulked segregant analysis with sequencing data in a method referred to as QTL-seq which uses the difference in SNP expression between the bulks to rapidly identify QTLs (Takagi et al. 2013). Kompetitive allele-specific polymerase chain reaction (KASP) markers have now gained prominence in traitspecific marker development (Gilio et al., 2020; Valentini et al., 2017). KASP identifies SNP variations in a population by utilizing PCR-based amplifications and has been used widely to develop trait-specific markers and genetic mapping due the ability to combine several markers in a single assay (Cao et al. 2021; Hurtado-Gonzales et al. 2017; Semagn et al. 2014). KASP markers were developed for $Co-4^2$ (Kelly et al., 2021a), however these new generation markers have not been developed for Co-5 and breeders currently rely on the classical SCAR marker developed by Vallejo & Kelly (2001). Co-5 confers broad resistance to 31 races (Balardin et al 1997) including races 3, 6, 7, 31, 38, 39, 102, 109, 449, 3481, and 3545 (Mahuku & Riascos, 2004). Developing a Co-5 KASP molecular marker for modern MAS would allow for efficient deployment in breeding programs, in addition to capitalizing on the gene's broad resistance to reinforce the durability of developed cultivars.

The differential cultivar TU contains *Co-5* as its single known resistance gene (Fouilloux, 1976). This cultivar is therefore an acceptable source of resistance; however, its agronomic traits are undesirable for commercial black bean production. The MSU breeding line B19504 conversely has preferred agronomic traits, including the *I* gene which confers resistance to *bean common mosaic necrosis virus* (BCMNV), but lacks anthracnose resistance. BCMNV is one of the most common and destructive viruses that affect common bean and, like anthracnose, can

40

cause up to 100% yield loss (Singh & Schwartz, 2010). The *I* gene is another important resistance gene to consider when breeding for disease resistance. This study aimed to identify and map SNPs associated with the *Co-5* locus on Pv07 associated with resistance to race 109 in an F₂ population developed from a cross between B19504 and TU. The goal of this study was to identify molecular markers tightly linked to *Co-5* for future development into high throughput KASP markers for rapid deployment in breeding programs to facilitate more effective marker-assisted selection. A secondary goal of this study was to identify lines from the F₂ population that contain *Co-5* and the *I* gene to utilize in back-crossing resistance into black bean breeding lines in the MSU Dry Bean Breeding Program.

METHODS AND MATERIALS

Plant materials

A mapping population consisting of 446 F_2 individuals were derived from a cross between the anthracnose race 109-susceptible black bean B19504 and the resistant cultivar TU (Figure 2.1). F1 seeds were selfled and grown in the MSU greenhouses for increase, and then the F2 seed were bulked. B19504 is a near isogenic black bean line to 'Adams', but lacks the resistance conferred by *Co-4*². TU is a differential cultivar that carries the Middle American *Co-5* gene, first characterized by Fouilloux (1976).

Inoculation procedure and disease scoring

In 2022, 446 F_2 individuals and 12 anthracnose differential cultivars (Pastor-Corrales, 1991) were evaluated in the MSU greenhouses, East Lansing, MI. All F_2 individuals including the parents were grown to the first trifoliate stage in 72-cell trays containing SureMix potting soil. Inoculation of anthracnose race 109 was performed by spraying a suspension of 1.2 x 10⁶ C.

lindemuthianum conidia ml⁻¹ onto the leaves and stems of the seedlings. The plants were maintained under 80% humidity in a mist chamber for three days following inoculation. Anthracnose symptoms were observed and evaluated in the population 7 days after inoculation. A scale of 1-5 was used to differentiate resistance and susceptibility (Drijfhout & Davis, 1989). Plants were assessed using the following classification: 1, no symptoms; 2, minor hypersensitive response; 3, pinpoint lesions or small lesions, not sunken; 4, large, sunken lesions; 5, plant death by pathogen (Figure 2.2).

DNA extraction

DNA was collected from young leaf tissues of the F₂ and parental genotypes prior to inoculation using a modified CTAB (Hexadecyltrimethyl ammonium bromide) extraction protocol (Doyle 1987). The DNA concentrations were measured using a Qubit Flex dsDNA Broad Range Assay kit (Thermo Fisher Scientific, Waltham, MA) and quality was checked on an agarose gel.

Genetic analyses

Simple inheritance of resistance to anthracnose race 109 was confirmed using a chi-square test. DNA from the leaves of the resistant and susceptible F₂ individuals were bulked for whole genome sequencing (WGS). The bulks consisted of 158 individuals given a rating of 1 and 25 individuals rated 5 during evaluations. The entire population was sent to the HudsonAlpha Genome Sequencing Center (Huntsville, AL) for PCR-free cDNA library construction and Illumina paired-end sequencing. Two Illumina Trueseq DNA libraries were prepared from the bulks with an insert size of 550 bp. WGS was performed with a NovaSeq using the NovaSeq6000 S4 Reagent Kit at 3x coverage for the entire population. The resulting data was used for QTL-seq analysis of the race 109 resistance trait using a QTL-seq pipeline developed at HudsonAlpha

Institute for Biotechnology (Korani et al. 2021). Short reads were aligned with the common bean genomic DNA sequence (Phaseolus vulgaris v.2.1) at Phytozome, DOE, and JGI (https://phytozome-next.jgi.doe.gov). The Khufu var program was utilized to filter and improve the data for high-quality reads (Korani et al. 2021). SNPs missing more than 75% of calls, monomorphic between parents, or any that were heterozygous in the parents were removed. To validate the Khufu variant calling and analysis, a secondary procedure was performed. The raw paired-end sequencing reads were processed to discard low quality reads that were shorter than 80 bp in length and did not meet the default quality threshold of 20 using SICKLE software (Joshi & Fass, 2021). The Burrows-Wheeler Alignment Tool (BWA-mem v0.7.17) (Li & Durbin, 2009) software was used to map the filtered reads against the common bean reference genome using the default parameters. The mapped results were then sorted, indexed, and pileup with SAMtools v1.15.1 (Li et al., 2009). The mplieup2snp module of VarScan v2.3.9 (Koboldt et al, 2012) was used to call the SNPs using a minimum coverage of 3 and minimum read quality of 22. Multiallelic SNPs were discarded alongside markers missing more than 5% of their sites, more than 2.5% heterozygous, and with a minor allele frequency higher than 30%.

The SNP genotyping data from the secondary filtering method was inspected using the *QTL* package in R (Broman et al., 2003) and identical SNPs and individuals were removed. The independence logarithm of the odds (LOD) was used to develop the linkage groups beginning at a minimum of 8 and a maximum of 20 to extract groups. Linkage groups were assigned to a chromosome based on the chromosome assignment of the markers clustered within the linkage groups. The SNP-based genetic map was developed using Kosambi's mapping function (Kosambi, 1944) in JoinMap 4 (Van Ooijen, 2006) with the remaining JoinMap default parameters used for linkage analysis.

The anthracnose response scoring was analyzed in a single-trait QTL analysis using the entire population. QTL were identified using the phenotypic data collected from the individuals of the population. The *QTL* package in R was utilized to conduct the QTL analysis with Haley-Knott regression (Haley & Knott, 1992). The significance of each QTL was tested using the permutation test (Churchill & Doerge, 1994) using 1000 permutations and expressed as a p-value on the -log₁₀ scale. The QTL and linked SNPs were selected using a significance threshold of 0.01.

SCAR marker screening

Resistant individuals from the F₂ population were screened for the *I* gene conferring resistance to *bean common mosaic necrosis virus* (BCMNV) using the SW13 marker developed by Melotto, Afanador, and Kelly (1996). This screening was performed to select individuals for back crossing that were resistant to both anthracnose race 109 and BCMNV. PCR amplifications were conducted using Programmable Thermal Controller (MJ Research, Inc, St. Bruno, QC) in 25 µl solutions containing 1 µl of the forward and reverse primer, 13.6 µl PCR water, 4.8 µl of 5 mM dNTP, 3.0 µl 1x buffer, 2.5 µl MgCl₂, and 0.3 µl Taq polymerase (Invitrogen citation). The PCR procedure consisted of 33 cycles of 10 seconds at 94°C, 40 seconds at 67°C, 2 minutes at 72°C, followed by one cycle at 5 minutes at 72°C. Amplification products were detected on a 1.2% agarose gel prepared with 1x TAE (Tris-Borate EDTA) buffer and 0.1µg/100ml ethidium bromide. PCR products were loaded into the gel for electrophoretic separation with an electric potential maintained at 80 V for 1 hour and visualized under UV light using a Bio-Rad Gel Doc EZ Imager (Bio-Rad laboratories, Inc., Hercules, CA).

RESULTS

Segregation of resistance to race 109

The observed segregation of resistance within the F₂ population deviated from the anticipated 3:1 ratio for a major dominant qualitative trait. There were significantly more individuals displaying resistance to race 109 within the population than the expected 75% ($X^2 = 14.23$, p < 0.0001) (Table 2.1) (Figure 2.3). The observed ratio also deviated from the expected ratio under a two dominant gene segregating population ($X^2 = 92.35$, p < 0.0001). However, the segregation of resistance in the population did meet the expectations of a dominant suppression epistasis ($X^2 = 0.646$, p = 0.42) (Table 2.1).

I gene

The *I* gene was observed in nearly 75% of the resistant genotypes and the chi-square analysis confirmed adherence to the expected phenotypic ratio of a trait controlled by a dominant major gene ($X^2 = 0.002$, p = 0.9675) (Table 2.1). The presence of the *I* gene in B19504 and absence in TU is confirmed in Table 2.2.

QTL-seq

QTL-sequencing and subsequent filtering resulted in 48,110 polymorphic SNPs. QTL-seq analysis revealed two possible QTL for anthracnose race 109 resistance on chromosome Pv07. The first region was found at an interval between 6.6 Mb and 13.7 Mb, and the second was found between 20.6 Mb and 24.5 Mb (Figure 2.4). These regions had a deltaVAR of 1.0. Two QTL were identified on chromosome Pv03 between 34.8 Mb and 51.4 Mb (Figure 2.5). The calculated deltaVAR for the QTL on this chromosome was also approximately 1.0.

QTL mapping

A genetic map for the entire B19504 x TU population was constructed with 617 SNPs after inspection with the *QTL* package in R. The map consisted of 16 linkage groups with each linkage group representing most chromosomes. Chromosome 3 separated into three groups and chromosome 7 separated into two linkage groups. Anthracnose inoculation response in F₂ plants identified two potential QTL for race 109 resistance (Figure 2.6). The strongest QTL identified is approximately located between the SNPs S07_ 9010419 and S07_9484401 (LOD = 31.84). Both markers show strong additive effects with S07_ 9010419 having a value of 0.75 and S07_9484401 having a value of 0.66. These flanking SNPs are located on Pv07 at 9.010419 and 9.484401 Mb (Table 2.3). The second QTL on Pv07 was located between SNPs S07_24624785 and S07_24624800 at 24.624785 and 24.6248 Mb (LOD = 12.54). The third QTL on Pv07 between SNPs S07_156765 at 0.156765 Mb and S07_156864 at 0.156864 Mb (LOD = 5.27) (Table 2.3). A fourth, weak QTL was also detected on Pv03 located between S03_51830541 and S03_52047723 at 51.830541 Mb and 52.047723 Mb, respectively (LOD = 4.07) (Table 2.3).

DISCUSSION

Genetic analyses

This study utilized a WGS approach via QTL-seq methods for genetic analysis paired with genetic analysis through genotyping. The resistant parent, TU, is included in the differential panel used for race characterization and thus was identified as resistant to race 109 when it was first identified in Manitoba and Ontario and later, in Michigan (Awale et al. 2018; Conner et al. 2020). TU has been utilized in previous anthracnose studies investigating systems of resistance (Campa, Giraldez, and Ferreira 2009; Campa, Trabanco, and Ferreira 2017). *Co-5* and the *I* gene are major dominant genes conferring resistance to their respective diseases, anthracnose and BCMNV. The

deviation from the expected 3:1 ratio of a major gene observed in the segregation analysis of the former could be due to early losses in the population from a lack of germination and because the population was a randomly selected subset from a larger population. However, the possibility of dominance suppression could indicate the possibility of a *Co* gene in B19504. The susceptible parent is nearly isogenic to the commercial line 'Adams' but lacks the resistance conferred *Co-4*². *Co-4*² also confers resistance to races 7 and 73 which are commonly used to evaluate resistance in the MSU Dry Bean Breeding Program, thus it is possible that the additional gene segregating in the population also confers resistance to at least one of 7 and 73, but not 109.

Fouilloux (1976) first characterized the mesoamerican resistance gene Co-5 in the differential cultivar TU that confers resistance to races 3, 6, 7, 31, 38, 39, 102, 109, 449, 3481, and 3545. In comparison to the other differential cultivars in a study by Paulino et al. (2022), Co-5 was found to have the third highest resistance index, showing resistance to 81 out of 89 races evaluated. Co-5 was later reported to be found in the differential cultivar G2333 and RAPD markers were then developed from a F₂ population using the G2333-derived line SEL1360 (Young and Kelly 1997). The marker, OAB3₄₅₀, was mapped in coupling phase $(5.9 \pm 1.7 \text{ cM})$ to Co-5 (Young et al. 1998; and Young and Kelly 1997). SCAR markers were developed from the RAPD marker by Vallejo and Kelly (2001) mapped in coupling-phase (12.98 cM), the inconsistency between the two markers thought to be from the use of different mapping populations as the original population was no longer available. Both populations utilized resistant parents derived from G2333, the carrier of Co-5. In a study by Mahuku and Riascos (2004) reported the gene present in G2333 displayed differing susceptibility reactions to race 3481 compared to TU. This led to the conclusion in Vallejo and Kelly (2009) that the gene carried by G2333 was an allelic form of Co-5, named Co- 5^2 . Sousa et al. (2014) mapped the gene to a region stretching from 6.98 Mb to 7.02 Mb on Pv07.

Co-5 would also be found in the differential cultivar AB136 alongside another resistance gene on Pv07, *Co-6* (Campa, Trabanco, and Ferreira 2017). Allelism tests would confirm that the two genes are independent of each other with *Co-6* located at 9.62 Mb (Campa, Trabanco, and Ferreira 2017).

The locations of the SNPs identified in this study coincide with the findings of Nunes et al (2020). Interestingly, only one of the QTL identified via QTL-seq agrees with the previously established location of Co-5. It cannot be stated that the range captured by this study is accurately representative of the gene size as the QTL-seq captured many significant polymorphic SNPs across Pv07 and the whole genome. B19504 and TU are contrasting in phenotypic traits that go beyond anthracnose resistance and it is possible that the bulks included these additional traits. B19504 is an upright, indeterminate bush type, matte black bean line and TU is an indeterminate bush type with glossy black seeds and high expression of anthocyanin in the seed pods and stem. In particular, the QTL identified on linkage group 7.1 could be associated with the Asp gene, a major dominant gene that controls the thickness of the seed epicuticular wax layer (Cichy et al. 2014). This trait can be phenotyped by the 'glossiness' of the common bean seed. The 'shiny' trait in the black bean market class is undesirable in black bean production as it is a major determinant in water uptake during cooking (S. Diaz et al., 2021). The SNPs identified on Pv03 are of note as no genes were known to be present on that chromosome in either of the parents. Previous studies have found genes on Pv03 near the SNPs identified in this study relating to angular leaf spot resistance (Vidigal Filho et al. 2020), anthracnose resistance gene Co-17 (Trabanco et al., 2015), and genes encoding nucleotide binding sites with leucine rich repeats (NBS-LRR) (Vaz Bisneta and Gonçalves-Vidigal 2020). The second QTL found on linkage group 7.2 is also unknown; no anthracnose genes have been mapped to that region nor is it located within the range of the P locus which controls seed coat color (McClean et al., 2018).

Implications in plant breeding

Molecular markers provide breeding programs with the means of efficiently deploying effective combinations of complementary resistance genes. Utilizing multiple tightly linked markers to genes that confer broad resistance to many isolates increases the durability of cultivars as they lower the threat of emergent races. In previous years, the MSU dry bean breeding program utilized the same source of resistance for anthracnose races 7 and 73, $Co-l^2$ (Kelly et al., 2001; Zuiderveen et al., 2016). More recently, KASP markers have been developed for the anthracnose resistance gene $Co-4^2$, which has been mapped to Pv08, is reported to confer resistance to a broad range of C. lindemuthianum races, demonstrating resistance to 33 out of 34 races of anthracnose from 9 countries in a study by Balardin et al (1997). Developing KASP markers from the validated SNPs identified in this study would result in efficient development of lines resistant to anthracnose race 109. Like $Co-4^2$, Co-5 confers resistance to a broad range of races. In the same Balardin et al (1997) study, Co-5 conferred resistance to 31 out of 34 races. Identifying breeding lines with Co-5 through MAS rather than inoculation testing would increase efficiency of resistance evaluations. This study is important for breeders utilizing Co-5 as their source of anthracnose resistance to pyramid the gene into their germplasm. For broad, durable resistance in Michigan to races 7, 73, and 109, Co-5 should be used in tandem with $Co-4^2$ in breeding programs.

CONCLUSION

Integrating the *Co-5* gene into cultivars has become a present concern due to the emergence of a new *C. lindemuthianum* race in Michigan. A major QTL for anthracnose resistance to race 109 was identified on Pv07 within a region located between 6.838810 Mb and 24.62480 Mb between SNPs S07_6838810 and S07_24624800. The strongest peak of this QTL

lies between SNPs S07_9010419 and S07_9484401 at 9.010419 and 9.484401 Mb, respectively. Additionally, a second, weaker QTL was identified on Pv07 between SNPs S07_156765 at 0.156765 Mb and S07_156864 at 0.156864 Mb. The location does not coincide with previous literature regarding the location of *Co-5*; however, this weaker QTL may be linked with the *Asp* gene on Pv07 which controls the thickness of the epicuticular wax layer in common bean seeds. This QTL may be useful for developing markers for MAS to select against the *Asp* gene and undesirable shiny seed coat phenotype. The genetic information provided by the SNP markers identified in this study flanking the *Co-5* locus will be useful for the future development of molecular markers necessary for MAS. KASP marker development and validation will confirm these markers are linked and co-segregate with the *Co-5* gene. The availability of high throughput markers will expedite efforts to pyramid multiple anthracnose resistance genes in improved germplasm for increased cultivar durability to rapidly evolving pathogen populations in Michigan and other bean production regions worldwide.

TABLES

Population	Locus Tested	Generation	Expected Ratio	Observed Ratio	X^2	Probability
B19504/TU			3:1		14.233	<i>p</i> < 0.0001
	Co-5	Γ.	15:1	369:77	92.35	<i>p</i> < 0.0001
		Γ2	13:3		0.646	0.42
	Ι		3:1	151:50	0.002	0.97

Table 2.1. Chi square (X^2) for the *Co-5* resistance gene and *I* gene in a segregating F₂ population.

Plant ID #	SW13 marker I gene
B19504	+
TU	-
1	+
2	+
3	+
6	+
7	+
8	-
9	-
10	+
11	+
12	-
13	-
14	+
15	-
16	+
17	+
18	-
19	+
20	+
21	+
22	+
23	+
24	+
25	+
26	+
27	+
28	+
29	+
30	+
31	+
32	+
33	+
34	+
35	+
36	-
37	+
38	+
39	+
40	+
41	+
42	+
43	+
44	+

Table 2.2. Subset of F₂ population with amplification products of gene-based markers.

(/		
	45	+
	46	+
	47	+
	48	-
	49	+
	50	+
	51	+
	52	+
	53	-
	54	-
	55	-
	56	-
	57	-
	58	-
	60	-
	61	+
	62	+
	63	-
	64	-
	65	-
	66	-
	67	-
	68	-
	69	-
	70	-
	71	-
	72	+
	73	+
	74	+
	75	+
	76	+
	77	+
	79	+
	80	+
	82	+
	83	+
	84	+
	85	+
	86	+
	87	+
	88	• +
	89	-
	90	+
	91	· +
	02	Т
		ļ.

Table 2.2 ((cont'd)
-------------	----------

93	+
94	+
95	+
96	+
97	+
98	+
99	+
100	+
101	+
103	+
104	+
105	+
106	+
107	+
108	+
109	+
110	+
111	-
112	+
113	-
115	+
116	+
117	+
118	+
119	+
120	+
121	+
122	+
124	-
125	+
126	+
127	-
128	<u>-</u>
129	+
130	-
131	+
132	+
132	
133	
135	' +
135	т _
130	⊤ ⊥
137	т _
130	T
137	Ť
140	-

Table 2.2 (cont'd)	
141	-
142	+
143	+
144	+
145	+
146	+
147	+
148	+
149	+
150	+
151	+
152	+
153	-
154	+
155	+
156	+
157	-
158	+
159	+
160	+
161	+
162	-
163	+
164	-
165	+
166	+
167	+
168	+
169	+
170	+
171	+
172	+
173	+
174	+
175	+
176	-
177	-
178	+
179	+
180	+
181	+
182	+

183

184

185

+

+

-

Table 2.2 (cont'd)	
186	+
187	+
188	-
189	+
190	+
191	-
193	+
194	+
195	-
196	+
197	+
198	-
199	+
200	-
201	+
202	-
203	+
204	+
205	+
206	-
207	-
208	+

208 + Band present, individual contains the marker +; Band absent, individual lacks the marker -; individual is heterozygous for the marker +/-.

SNP marker	Chromosome	Physical Position (Mb)	SNP position (cM)	LOD
S07_157260	7.1	0.157260	13.085124	4.14
S07_158956	7.1	0.158956	39.051732	5.16
S07_158957	7.1	0.158957	38.548093	5.16
S07_6838810	7.2	6.838810	69.4459	17.86
S07_8535219	7.2	8.535219	81.8984	19.67
S07_9010419	7.2	9.010419	92.7622	30.93
S07_9484401	7.2	9.484401	101.6100	26.20
S07_11886839	7.2	11.886839	131.7475	19.36
S07_11886840	7.2	11.886840	137.8678	18.27
S07_12209488	7.2	12.209488	150.4089	13.57
S03_51830541	3.3	51.830541	8.784066	4.05
S03_52047723	3.3	52.047723	17.134810	3.75

Table 2.3. SNP markers, location, and position on chromosome Pv07 used to identify major QTL for resistance to anthracnose race 109.

FIGURES



Figure 2.1. Phenotype of the population parents, B19504 and TU, in response to anthracnose race 109 inoculation. B19504, the maternal parent, is susceptible to race 109. TU is resistant.



Figure 2.2. Phenotypic rating scale of the F2 population response to anthracnose race 109 inoculation. 1, no symptoms; 2, minor hypersensitive response; 3, pinpoint lesions or small lesions, not sunken; 4, large, sunken lesions; 5, plant death by pathogen.



Figure 2.3. Distribution of anthracnose disease severity across a F_2 population segregating for disease resistance to race 109. The parental genotypes, B19504 (susceptible) and TU (resistant), are counted among the rating 5 and rating 1 categories, respectively.



Figure 2.4. Significant QTL for anthracnose resistance identified on chromosome Pv07 in dry bean. deltaVAR values represent SNP expression across the F_2 population: values of 0 signify equivalent expression across all individuals; values of 1 signify highly segregating SNPs in the population.



Figure 2.5. Significant QTL for anthracnose resistance identified on chromosome Pv03 in dry bean. deltaVAR values represent SNP expression across the F_2 population: values of 0 signify equivalent expression across all individuals; values of 1 signify highly segregating SNPs in the population.



Figure 2.6. QTL for the anthracnose race 109 resistance gene *Co-5*. The dashed green line represents the significance threshold of 3.94 (p = 0.01).]



Figure 2.7. Additive and dominance effects for all SNP markers carrying alleles from 'TU' on Pv07.2.

REFERENCES

- Awale, H.E. et al. 2018. "Characterization and Distribution of a New Emerging Race of Anthracnose in Michigan." *Annual Reporter Bean Improvement Cooperative* 61: 113–14.
- Balardin, R. S., A. M. Jarosz, and J. D. Kelly. 1997. "Virulence and Molecular Diversity in *Colletotrichum Lindemuthianum* from South, Central, and North America." *Phytopathology*® 87(12): 1184–91.
- Balardin, R. S., and J. D. Kelly. 1998. "Interaction between Collectorichum Lindemuthianum Races and Gene Pool Diversity in Phaseolus Vulgaris." *Journal of the American Society for Horticultural Science* 123(6): 1038–47.
- Bello, Marco H. et al. 2014. "Application of in Silico Bulked Segregant Analysis for Rapid Development of Markers Linked to Bean Common Mosaic Virusresistance in Common Bean | BMC Genomics | Full Text." https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-15-903 (May 16, 2022).
- Cadle-Davidson, M. M., and M. M. Jahn. 2005. "Resistance Conferred against Bean Common Mosaic Virus by the Incompletely Dominant I Locus of Phaseolus Vulgaris Is Active at the Single Cell Level." Archives of Virology 150(12): 2601–8.
- Campa, Ana, Ramón Giraldez, and Juan José Ferreira. 2009. "Genetic Dissection of the Resistance to Nine Anthracnose Races in the Common Bean Differential Cultivars MDRK and TU." *Theoretical and Applied Genetics* 119(1): 1–11.
- Campa, Ana, Noemí Trabanco, and Juan José Ferreira. 2017. "Identification of Clusters That Condition Resistance to Anthracnose in the Common Bean Differential Cultivars AB136 and MDRK." *Phytopathology*® 107(12): 1515–21.
- Cao, Yanyan et al. 2021. "Development of KASP Markers and Identification of a QTL Underlying Powdery Mildew Resistance in Melon (Cucumis Melo L.) by Bulked Segregant Analysis and RNA-Seq." *Frontiers in Plant Science* 11: 593207.
- Choudhary, Neeraj et al. 2018. "Gene/QTL Discovery for Anthracnose in Common Bean (Phaseolus Vulgaris L.) from North-Western Himalayas." *PLOS ONE* 13(2): e0191700.
- Cichy, Karen A. et al. 2014. "QTL Analysis of Canning Quality and Color Retention in Black Beans (Phaseolus Vulgaris L.)." *Molecular Breeding* 33(1): 139–54.
- Conner, Robert L. et al. 2020. "Identification of Anthracnose Races in Manitoba and Ontario from 2005 to 2015 and Their Reactions on Ontario Dry Bean Cultivars" ed. Brian Beres. *Canadian Journal of Plant Science* 100(1): 40–55.
- Doyle, J. J. 1987. "A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue." *Phytochemical Bulletin* 19: 11–15.
- Feng, Xue, Gardenia E. Orellana, James R. Myers, and Alexander V. Karasev. 2018. "Recessive

Resistance to *Bean Common Mosaic Virus* Conferred by the *Bc-1* and *Bc-2* Genes in Common Bean (*Phaseolus Vulgaris*) Affects Long-Distance Movement of the Virus." *Phytopathology*® 108(8): 1011–18.

- Fouilloux, G. 1976. "L'anthracnose Du Haricot (Colletotrichum Lindemuthianum, Sacc et Magn): Nouvelles Sources de Résistance et Nouvelles Races Physiologiques." Ann Amélior Plantes 26: 443–53.
- Gonçalves-Vidigal, M. Celeste, and James D. Kelly. 2006. "Inheritance of Anthracnose Resistance in the Common Bean Cultivar Widusa." *Euphytica* 151(3): 411–19.
- Hurtado-Gonzales, Oscar P et al. 2017. "Fine Mapping of Ur-3, a Historically Important Rust Resistance Locus in Common Bean." *G3 Genes/Genomes/Genetics* 7(2): 557–69.
- Joshi, N.A., & Fass, J. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]
- Kelly, J. D., and Veronica Vallejo. 2004. "A Comprehensive Review of the Major Genes Conditioning Resistance to Anthracnose in Common Bean." *HortScience* 39(6): 1196– 1207.
- Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., & Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22, 568–576. https://doi.org/10.1101/gr.129684.111
- Korani, Walid et al. 2021. Accurate Analysis of Short Read Sequencing in Complex Genomes: A Case Study Using QTL-Seq to Target Blanchability in Peanut (Arachis Hypogaea). Genomics. preprint. http://biorxiv.org/lookup/doi/10.1101/2021.03.13.435236 (May 14, 2022).
- Lacanallo, Giselly Figueiredo, and Maria Celeste Gonçalves-Vidigal. 2015. "Mapping of an Andean Gene for Anthracnose Resistance (Co-13) in Common Bean (Phaseolus Vulgaris L.) Jalo Listras Pretas Landrace." : 7.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25, 1754–1760. <u>https://doi.org/10.1093/bioinformatics/btp324</u>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, 1000 Genome Project Data Processing. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352
- Mahuku, George S., and Jhon Jaime Riascos. 2004. "Virulence and Molecular Diversity within Colletotrichum Lindemuthianum Isolates from Andean and Mesoamerican Bean Varieties and Regions." *European Journal of Plant Pathology* 110(3): 253–63.

McClean, P. E., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., Geffroy,

V., Miklas, P., & Mamidi, S. 2018. White seed color in common bean (Phaseolus vulgaris) results from convergent evolution in the P (pigment) gene. *New Phytologist*, *219*(3), 1112–1123.

- Melotto, M. et al. 2004. "The Anthracnose Resistance Locus Co-4 of Common Bean Is Located on Chromosome 3 and Contains Putative Disease Resistance-Related Genes." *Theoretical and Applied Genetics* 109(4): 690–99.
- Melotto, M., L. Afanador, and J. D. Kelly. 1996. "Development of a SCAR Marker Linked to the *I* Gene in Common Bean." *Genome* 39(6): 1216–19.
- Michelmore, R W, I Paran, and R V Kesseli. 1991. "Identification of Markers Linked to Disease-Resistance Genes by Bulked Segregant Analysis: A Rapid Method to Detect Markers in Specific Genomic Regions by Using Segregating Populations." *Proceedings of the National Academy of Sciences* 88(21): 9828–32.
- Nunes, Maria Paula Barion A. et al. 2021. "Relationship of Colletotrichum Lindemuthianum Races and Resistance Loci in the Phaseolus Vulgaris L. Genome." *Crop Science* 61(6): 3877–93.
- Padder, B.A., P.N. Sharma, H.E. Awale, and J.D. Kelly. 2017. "COLLETOTRICHUM LINDEMUTHIANUM, THE CAUSAL AGENT OF BEAN ANTHRACNOSE." *Journal* of Plant Pathology 99(2): 317–30.
- Paulino, Pollyana Priscila Schuertz et al. 2022. "Occurrence of Anthracnose Pathogen Races and Resistance Genes in Common Bean across 30 Years in Brazil." Agronomy Science and Biotechnology 8: 1–21.
- Semagn, Kassa, Raman Babu, Sarah Hearne, and Michael Olsen. 2014. "Single Nucleotide Polymorphism Genotyping Using Kompetitive Allele Specific PCR (KASP): Overview of the Technology and Its Application in Crop Improvement." *Molecular Breeding* 33(1): 1–14.
- Singh, Shree P., and Howard F. Schwartz. 2010. "Breeding Common Bean for Resistance to Diseases: A Review." *Crop Science* 50(6): 2199–2223.
- Schmutz, J., Mcclean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., TorresTorres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., . . . Jackson, S. A. (2014). A reference genome for common bean and genomewide analysis of dual domestications. Nature Genetics, 46, 707—713. https://doi.org/10.1038/ng.3008
- Sousa, Lorenna L. et al. 2014. "Genetic Mapping of the Resistance Allele 'Co-52' to 'Colletotrichum Lindemuthianum' in the Common Bean MSU 7-1 Line." *Australian Journal of Crop Science* 8(2): 317–23.
- Takagi, Hiroki et al. 2013. "QTL-Seq: Rapid Mapping of Quantitative Trait Loci in Rice by Whole Genome Resequencing of DNA from Two Bulked Populations." *The Plant*

Journal 74(1): 174-83.

- Vallejo, Veronica, and J. D. Kelly. 2001. "Development of a SCAR Maker Linked to Co-5 Locus in Common Bean." *Annual Reporter Bean Improvement Cooperative* 44: 121–22.
- Vallejo, Veronica, and James Kelly. 2009. "New Insights into the Anthracnose Resistance of Common Bean Landrace G 2333." *The Open Horticulture Journal* 2: 29–33.
- Van Ooijen, J.W. 2006. "'JoinMap 4'. Software for the Calculation of Genetic Linkage Maps in Experimental Populations." *Kyazma BV, Wageningen, Netherlands* 33.
- Vaz Bisneta, Mariana, and Maria Celeste Gonçalves-Vidigal. 2020. "Integration of Anthracnose Resistance Loci and RLK and NBS-LRR-Encoding Genes in the Phaseolus Vulgaris L. Genome." Crop Science 60(6): 2901–18.
- Vidigal Filho, Pedro S. et al. 2020. "Genome-Wide Association Study of Resistance to Anthracnose and Angular Leaf Spot in Brazilian Mesoamerican and Andean Common Bean Cultivars." *Crop Science* 60(6): 2931–50.
- Young, RA, Maeli Melotto, Rubens Nodari, and James Kelly. 1998. "Marker-Assisted Dissection of the Oligogenic Anthracnose Resistance in the Common Bean Cultivar, 'G2333." *TAG Theoretical and Applied Genetics* 96: 87–94.
- Young, Roberto A., and James D. Kelly. 1997. "RAPD Markers Linked to Three Major Anthracnose Resistance Genes in Common Bean." *Crop Science* 37(3): cropsci1997.0011183X003700030039x.

CHAPTER THREE: UTILIZING MACHINE LEARNING TO PREDICT SYMBIOTIC NITROGEN FIXATION IN COMMON BEAN

ABSTRACT

Nitrogen is a major yield limiting factor in common bean (*Phaseolus vulgaris* L.). Common bean like other legumes can fix atmospheric nitrogen (N) through a symbiotic relationship formed with specific Rhizobia species that develop complex nodules on the roots of bean plants. However, this trait is rarely utilized by growers or selected for in breeding programs due to the costly and time-consuming evaluation techniques. Remote sensing techniques utilizing unmanned aerial systems (UAS) is a potential solution to this bottleneck while also providing a high-throughput phenotyping method for trait evaluation. In this study, we investigated the use of vegetation indices (VIs) and machine learning methods in estimating symbiotic nitrogen fixation (SNF). Nitrogen derived from the atmosphere (Ndfa) and yield were used as direct and indirect measurements of SNF, respectively. Forty-two black bean breeding lines from the Michigan State University Dry Bean Breeding Program were grown and compared under both high and low N conditions. A Random Forest model developed to predict Ndfa using yield and remote sensing (RS) data resulted in an average accuracy of r = 0.54. A three-year evaluation of these trials in Michigan demonstrated how seed yield under unfertilized conditions could be used as an indirect indicator of SNF ability. Seven black bean breeding lines and cultivars showed consistent yields (Zenith, B16504, B18204, Adams, B19309, B19330 and Black Bear) under low N conditions. Zenith, B16504, Adams, B19330, B19309 and Black Bear were additionally found to have high Ndfa when evaluated in 2021. Two prediction models for yield developed using stepwise general linear modeling (StepwiseGLM) and Bayesian regularized artificial neural network (BRNeural Network) were determined to be accurate and reliable (StepwiseGLM r =
0.64, range of accuracy = 0.42-0.80; BRNeural Network r = 0.65, range of accuracy = 0.42-0.80). This model is promising in low nitrogen trials as an early selection tool to identify lines with higher SNF ability.

INTRODUCTION

The use of synthetic N in agriculture has soared over the last four decades, significantly improving yields to meet growing global population and food demands (Mulvaney et al., 2009; Han et al., 2015). To maximize yield in common bean, it is recommended that the applied synthetic N should be tailored to the amount of N measured in the soil, such that the cumulative total lies between 18-27 kg N/a (Warncke et al., 2009). However, this rate is often exceeded by growers regardless of soil testing, thus a majority of the added N is unabsorbed by the plants and lost into the environment (Turner and Rabalais, 1994; Turner et al., 2008; Asghari and Cavagnaro, 2011; Akter et al., 2017). This overuse is an economic issue for growers as the cost of fertilizer has risen 133% in the last year according to the Texas A&M Agricultural and Food Policy Center (Outlaw et al., 2022). Geopolitical events impact fertilizer and natural gas prices which directly impact agricultural production costs (Broom, 2023). Increased demands coinciding with limited supply have also driven manure prices higher (Rembert, 2022). Legumes like common bean offer the possibility of reducing the reliance on N fertilizers by exploiting the plant's ability to convert atmospheric nitrogen (N₂) to NH₃ through symbiotic nitrogen fixation (SNF).

However, the nitrogen fixing ability of common bean is low compared to other legumes (Hardarson et al., 1993). Yet, several studies have found evidence for genetic variation across bean genotypes for nitrogen derived from the atmosphere (Ndfa) (Douxchamps et al., 2010; Farid et al., 2016; Polania et al., 2016; Akter et al., 2017; Barbosa et al., 2018) suggesting the possibility of improving the SNF ability of common bean through selection. Furthermore, quantitative trail loci

(QTLs) and associated markers linked to SNF traits (nodule number, biomass, N in seed and shoot, and root biomass) have been performed in numerous studies for use in marker assisted breeding (L. M. Diaz et al., 2017; Heilig et al., 2016, 2017; Kamfwa et al., 2019; Muñoz-Azcarate et al., 2017; Ramaekers et al., 2013). For example, a study performed by Heilig et al. (2017) identified several QTL clusters associated with SNF traits in the black bean market class on chromosomes Pv01, Pv06, and Pv08 which may be potential targets for SNF improvement. Ramaekers et al. (2013) found several QTLs linked to various SNF traits using recombinant inbred lines generated from the landrace 'G2333' as did Kamfwa et al. (2015) with the landrace 'Solwezi'. Similarly, single nucleotide polymorphisms (SNPs) for SNF and its related traits were found in a genomewide association study performed on the Andean diversity panel by Kamfwa et al. (2017). Other works have also identified significant SNPs for Ndfa and Farid et al. (2017) demonstrated a 13% genetic gain in response to selection for SNF in recombinant inbred lines under optimal moisture conditions (Kamfwa et al., 2015).

Identifying parents with high SNF and yield potential has also been the focus of studies aiming to improve the trait through breeding. For example, Kamfwa et al., (2015) identified seven climbing bean lines from the Andean bean breeding program at CIAT with high amounts of atmospheric N fixation. Additionally, Wilker et al. (2019) found five heirloom genotypes that fixed more than 60% of their N from the atmosphere in their study in addition to finding that modern breeding had not reduced SNF capacity. Commercial lines like Zorro, OAC Inferno, and Red Rider were amongst the top N fixers in this study overall. While it is possible commercial lines are able to fix a moderate or relatively high amount of N, SNF ability of common bean has been disregarded as rhizobia activity downregulates with fertilizer use common to conventional production practices (Farid and Navabi, 2015a; Heilig et al., 2017; Wilker et al., 2019; Jiang et al., 2020). SNF

evaluation methods also require specialized equipment and expensive analyses, which can be a barrier to breeding programs (Lee & Lee, 2013; Li et al., 2019).

Each SNF evaluation method varies in their means of quantifying the amount of N fixed by rhizobia activity. Methods for evaluating N fixation in field experiments commonly relies on the stable isotope ¹⁵N and its differentiation from fertilizer-sourced ¹⁴N (Giller, 2001). The Ndifference evaluation method utilizes the total amount of N2 fixed over the experiment and can be extended to soil-based field experiments with the use of a reference non-fixing plant to account for soil N. This method, alongside other soil-based evaluations, heavily relies on the assumption that the non-fixing reference plant takes up the same amount of soil N as the test plants. In common bean, this assumption is met with the use of the non-nodulating navy bean mutant, R99 (Farid & Navabi, 2015; Heilig et al., 2016, 2017; Kamfwa et al., 2017; Wilker et al., 2019). All evaluation methods rely on assumptions regarding either N uptake in the reference plants or rhizobia activity, which can effect estimates if not met (Giller, 2001). Estimations of SNF can be further compounded by environmental factors, as rhizobia activity is dependent on many environmental factors outside of genetic, and host-rhizobia interactions (Soumare et al., 2020). Due to these responses, Ndfa is best measured and rated as a categorical quantitative trait as Ndfa values can greatly vary between environments and conditions (van Kessel & Hartley, 2000). To evaluate a breeding line's SNF ability would thus be reliant on repeated measures across multiple locations and years. The current evaluation methods are not conducive for integrating SNF improvement into most breeding programs with regularity due to the necessity of high-volume sampling. A highthroughput phenotyping method could improve the efficiency of trait evaluation and thus allow more consideration towards adding SNF in breeding schemes.

Unmanned aerial systems (UAS) when equipped with optical sensors can be efficient tools for rapid phenotyping in field conditions. The sensors are able to capture high-throughput, highresolution data from multiple plots (Shi et al., 2016). Light reflectance ratios, commonly known as vegetation indices (VIs), captured by these remote sensing (RS) methods have been used to estimate nutrient status and plant parameters of various crops (Song et al., 2018; Sankaran et al., 2018; Liang et al., 2018; Lu et al., 2019). Furthermore, RS can be cost-effective alternative for crop N monitoring using the crop's leaf chlorophyll content (Li et al., 2010; Saberioon et al., 2014). Previous studies have used leaf chlorophyll content to indicate N status in many crops including cereals, common bean, and *Solanum tuberosum* (potato) (Hansen and Schjoerring, 2003; Zheng et al., 2018; Boshkovski et al., 2021). Yang et al (2019) was able to accurately estimate aboveground N content in wheat using a combination of VIs and wavelet features ($R^2 = 0.90$, RMSE = 0.33) and Li et al (2010) demonstrated that one VI was able to accurately estimate leaf N content ($R^2 = 0.73$, RMSE = 0.38). Soil Plant Analysis Development, or SPAD meters were used to measure leaf chlorophyll content via handheld equipment and by extension N status before the further development of RS methods. Monitoring N applications with SPAD meters have been demonstrated to be successful in improving N use efficiency and yield in rice and wheat (Zhang et al., 2020). Additionally, SPAD meters have been used in SNF experiments as indicators of Ndfa (Farid et al., 2016; Kamfwa et al., 2015). Numerous VIs have been found to be correlated to leaf chlorophyll content and may prove to be more advantageous due UAS being able to capture more data on a whole field scale (Babar et al., 2006).

Machine learning (ML) methods have also been utilized in agriculture to assist in datadriven decision making and in predictive modeling. Many of the ML algorithms have been deployed to accurately predict yield in different crops including artificial neural networks (Ashapure et al., 2019), K-nearest neighbor (Zhang et al., 2010), gradient boosting (Shendryk et al., 2021), and random forest (Kim and Lee 2016). Outside of yield predictions, ML methods have found success in estimating plant N status in tandem with RS due to their ability to handle the analysis of the large amount of information captured by the sensors and in improving modeling prediction power (Chlingaryan et al., 2018; Moghimi et al., 2020; Qin et al., 2018; P. Shi et al., 2021). Yao et al (2015) found high prediction accuracies for leaf N content in winter wheat when evaluating ML methods against multivariate models as did Zha et al., (2020) for predicting rice N nutrition indices (Yao et al., 2015; Zha et al., 2020).

Considering the complexities of the phenotyping methods currently used to evaluate SNF and the lack of trait integration into breeding programs despite its potential for genetic improvement, lowering the barrier of evaluation with a simpler, high efficiency tool would allow for greater consideration toward adding selection for SNF into breeding programs. The current study was designed to evaluate the performance and SNF ability of advanced dry bean breeding lines from the Michigan State University Dry Bean Breeding Program and evaluate RS tools and ML techniques to develop a prediction tool to quantify SNF. Therefore, the objectives of this study were to evaluate the potential of UAS-based multispectral and RGB imaging for phenotyping dry bean breeding trials for SNF under low N conditions and to develop an earlyseason predictive model for SNF estimation. Additionally, as a secondary objective, the use of yield production under low N conditions was evaluated as an indirect indicator of SNF to develop a post-harvest predictive model for SNF estimation.

72

MATERIALS AND METHODS

Field trials

A set of 6 independent advanced yield trials (AYT) from the Michigan State University (MSU) dry bean breeding program were evaluated under two nitrogen conditions across three growing seasons (2019-2021) at the Saginaw Valley Research and Extension Center (SVREC) near Frankenmuth, Michigan. Each year the AYT was replicated under no added (-N) and added (+N) fertilizer conditions. The +N trials grown in 2019, 2020, and 2021, received routine N applications of 21, 24, and 29 kg/ha at sowing, respectively, whereas the -N trial received no N fertilizer application. Thus, the only source of N available for the -N trial was the residual soil N. The experiment design was a 6x7 or 6x6 lattice with four replicates in all years. A total of 42, 36, and 42 entries were grown in 2019, 2020, and 2021, respectively. Individual trials consisted of MSU black bean breeding lines, check cultivars, and the non-nodulating mutant navy bean R99 (Park and Buttery, 2006). Field plots (experimental units) consisted of four 6 m rows in length and spaced 50 cm apart. All trials were carried out under rainfed conditions utilizing industry standard agronomic practices with exception of the differential fertilizer application. At maturity, plots were trimmed to 4.6 m and the center two rows of each plot were harvested with a Wintersteiger Classic plot combine. Seed yield was recorded and standardized to 18% moisture measured in kilograms per acre (kg/ha). In 2019 and 2021 only, four soil samples were taken from randomly selected locations within each replication of each trial. The samples were analyzed for nitrate content at the Michigan State University Soil and Plant Nutrient Laboratory.

Remote sensing data collection and preprocessing for 2021

SPAD measurements were taken using the MultispeqQ v1.0 device produced by PhotosynQ (PhotosynQ inc., East Lansing, MI) for the 2021 trials only. Measurements were

collected at the R1 stage on the topmost open trifoliate leaf from four randomly selected plants in the center two rows of each plot and averaged.

The UAS RGB (red-green-blue) digital images were collected using a high-resolution digital RGB camera mounted onto a DJI Phantom 4 Pro v2 (DJI Technology Co., Ltd.). Images were taken twice weekly from initial growth stages (June 16, 2021) to full maturity (August 30, 2021). The number of days between flights varied if weather conditions prevented flying (rain, wind, and cloudiness) and flights altitude using the RGB camera were conducted at an altitude of 20 m and a speed of 5 m/s. Multispectral flights (MS) were performed with a DJI Matrice 210 v2 (DJI Technology Co., Ltd.) drone with a Micasense RedEdge-MX camera attached. The multispectral camera collects images in the R (centered at 668 nm, bandwidth of 14 nm), G (centered at 560 nm, bandwidth of 27 nm), B (centered at 495 nm, 32 nm bandwidth), red edge (centered at 717 nm, 12 nm bandwidth), and near infra-red (centered at 842 nm, bandwidth 57 nm) bands. The camera also contains a DSL-2 module that measures irradiance and sun ray angle. The Micasense RedEdge-MX camera settings were self-regulated according to the ambient light determined by the DLS-2 model. Two MS flights were performed over the season on 19 July and 2 August at an altitude of 25 m and a speed of 6 m/s. RGB and MS images were captured with 75% end, and side lap with the area covered greater than the trial area.

White circular lids were placed randomly throughout the field as ground control points (GCPs) and remained for the duration of the growing season to use as reference markers for accurate georeferencing of the images. GCPs were surveyed with a Global Navigation Satellite System (GNSS) receiver using a real-time kinematic (RTK) correction (Trimble R4 GNSS system, Trimble, Sunnyvale, CA, United States). Othomosaic images and reflectance maps were generated using the Pix4D Mapper software (Pix4D SA, 2022). Individual experimental plots were identified

74

using polygons shapefiles and soil masking was applied to extract the wavelength channels solely from the vegetation using a hue based VI (threshold = 0.7) for both the RGB and MS data. The average digital number values of canopy red, blue, and green channels were extracted using the R package FIELDimageR (Matias et al., 2020). A total of 80 VIs were calculated for each plot using the captured red, green, blue, near infrared, and far-red wavelengths (Henrich et al., 2012). The mean, median, and standard deviation were derived from each VI measurement to be used as variables for the prediction models.

Seed Isotope Analysis

Seed from the breeding lines grown in 2021 under -N were evaluated for SNF as percent nitrogen derived from atmosphere (%Ndfa). A 30g seed subsample from each plot was placed in an envelope and dried at 60° C for 72h. The seeds were then ground using a Christy-Turner Willey Mill to pass through a 1-mm mesh screen and ground samples were then stored at room temperature. Approximately 6mg of ground seed were prepared and shipped to the Stable Isotope Facility at University of California, Davis, California, USA, for measurements of total N and ¹⁵N natural abundance. Ndfa in the seed was measured using the N difference method (Boddey & Knowles, 1987; Heilig et al., 2017). The following two equations were used for estimating %Ndfa using the ¹⁵N balance method and N yield respectively.

%Ndfa= (N yield-fixer - N yield-non-fixer)/N yield-fixer x 100 Eq. 1

N yield= Seed yield (kg/ha) x %N Eq. 2

Where %Ndfa is the percentage of N in the seed at harvest that is derived from the atmosphere through N fixation and %N is the percent of nitrogen in the seed from the respective N-fixer and non-fixer (R99).

Phenotypic data analysis

A linear mixed model was fit for individual and combined environment data using the *lmer* function of the *lme4* package in the R coding language (Bates et al., 2015). Best linear unbiased estimators (BLUEs) of the genotypes grown in the -N trial in 2021 were estimated for %Ndfa and yield. BLUEs were obtained for the -N trial in 2021 using the following statistical model:

$$Y_{ijl} = \mu + G_i + R_j + B_{l(j)} + \varepsilon_{ijl}$$
 Eq. 3

where Y_{ijl} is the phenotypic observation of the *i*th genotype on the *j*th replicate in the *l*th incomplete block. The effects in the models are as follows: μ is the grand mean; G_i is the fixed effect of the genotype *I*; *R* is the random effect of replicate *j*; B is the random effect of the incomplete block within the replication *l*(*j*); and ε_{ijl} is the random residual. Variance components were estimated via the restricted maximum likelihood (REML) method (Patterson and Thompson, 1971) using the *lmer* function of the *lme4* R package (Bates et al., 2015), and its significance assessed by the likelihood ratio test (LRT) using the *ranova* function of the *lmeTest* R package (Kuznetsova et al., 2017). Repeatability (H²) on an entry-mean basis was estimated for Ndfa and yield for each flight date using the following equation:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_e^2}{r}}$$
 Eq. 4

where σ_G^2 is the genotypic variance, σ_e^2 is the error variance, and *r* is the number of replicate plots. Repeatability was used to select flight dates that captured the most genetic variation between breeding lines. The accuracies of prediction models utilizing the full season of flight dates were compared against prediction models built only from the selected dates.

A second model was developed to incorporate flight date in estimating BLUEs of both responses to account for variation from flight date timing.

$$Y_{ijlkmn} = \mu + [G(flight)]_{ik} + [R(flight)]_{jk} + [B(flight)]_{lk(j)} + [P(flight)]_{mk} +$$
Eq. 5

$[A(flight)]_{nk} + \varepsilon_{ijlkmn}$

where Y_{ijlkmn} is the phenotypic observation, %Ndfa and yield, of the *i*th genotype on the *j*th replicate in the *l*th incomplete block in the *m*th pass and *n*th range on the *k*th flight date. The effects in the models are as follows: μ is the grand mean; G_{ik} is the fixed effect of the genotype *i* on flight date *k*; *R* is the random effect of replicate *j* on flight date *k*; B is the random effect of the incomplete block within the replication *l*(*j*) on flight date *k*; P is the random effect of pass *m* on flight date *k*; A is the random effect of range *n* on flight date *k* and e_{ijlkmn} is the random residual.

Machine Learning Regression

The VIs from the RGB and MS images estimates from Eq. [3 & 5] were used to assess the predictive ability of %Ndfa and yield. Pearson's correlations for the VI's were estimated and any correlation coefficient above 0.95 were removed to avoid overfitting and multicollinearity issues, which can lead to variance inflation factors among predictors (James et al., 2021). This filtering in addition to filtering out VIs missing more than 10% of data left 76 VIs for analysis. Several prediction models were fit using multiple machine learning methods including random forest (RForest) (Breiman, 2001), extreme gradient boosting (XGBoosting) (Natekin and Knoll, 2013; Chen and Guestrin, 2016), K-nearest neighbor (KNNeighbors) (Altman, 1992), Bayesian regularized artificial neural network (BRNeural Network) (Burden and Winkler, 2009), partial least squares (PLSR) (Tobias, 1995), and general linear model stepwise regression (StepwiseGLM). All modeling methods were implemented using the caret package in R (Kuhn, 2008). The hyperparameter in the RForest model mtry (the number of variables randomly selected to be sampled at each tree split) was set as mtry = $n_{predictors}/3$. Parameters for XGBoosting were run using 100 iterations, a learning rate of 0.3, a subsampling rate of 0.75, gamma = 6, and package

defaults for the remaining parameters. Default parameters were used for KNNeighbors to determine the optimal k-value. The required specification for neuron number in BRNeural Network was set to a constant range of 1 through 5. Default parameters were utilized to find the optimal number of components for the PLSR prediction models.

Variable selection was performed to control for models that did not include variable selection in their protocols and to identify informative VIs. Selection was performed using Akaike information criterion (AIC) as selection criteria. Two models were developed for predicting %Ndfa: (1) %Ndfa = VIs + SPAD; (2) %Ndfa = VIs + SPAD + Yield (Figure 2.1). Model 1 is applicable during the season, before harvest. Model 2 is only applicable post-harvest. One model was developed for predicting yield: Yield = VIs + SPAD (Figure 2.2).

The training and testing populations were created using an 80/20 split of the dataset with the split repeated 500 times to assess model reliability. A k-fold (k = 10) cross-validation was performed for each model using 10 repeats. Models were evaluated for selection using root mean square error (RMSE), the Person's correlation between the predicted and actual values (*r*), and mean absolute error (MAE):

$$RMSE = \sqrt{\frac{1}{n}\Sigma(y-\hat{y})^2}$$
Eq. 6

$$MAE = \frac{\Sigma |y - \hat{y}|}{n}$$
 Eq. 7

where *n* is the sample size, and *y* and \hat{y} are the observed and predicted values in the model, respectively.

RESULTS

Multi-Year Yield Analysis

Yield response varied by year and by treatment across three years (Figure 3.3). Only the 2019 and 2021 trials for both treatments were significantly different (P < 0.05), with the fertilized plots producing higher yields. Overall, the highest yield was recorded in the 2021 added N trials with a mean of 1667.4 kg/ha (CV = 9.2%) followed by 2020 with an average of 1555.3 kg/ha (CV = 11.0%), and 2019 with 1102.1 kg/ha (CV = 9.0%), respectively (Table 3.1). The highest yield in the unfertilized trials was observed in 2020 with a mean of 1534.8 kg/ha (CV = 11.2%) followed by 2021 and 2019 with 1457.1 kg/ha (CV = 10.3%) and 979.4 kg/ha (CV = 13.2%), respectively. Yields in 2020 did not significantly differ between -N and N trials (p = 0.62).

Repeatability of Spectral Traits Across Plant Development

Repeatability for spectral trials across all flight dates ranged from 0.47-0.86 (Figure 3.4). The three dates selected by having the highest mean repeatability, Flight 3, Flight 8, and Flight 9, represent the V8 through R3 stages of common bean development, from first flower development to early pod development. Repeatability showed a consistent pattern across the season, matching the plant's developmental stages. From germination, repeatability increased significantly from 0.71 to 0.86 at R2 before falling at the start of the R3 stage (H² = 0.76). Repeatability increased again through R3 and decreased at R4 (H² = 0.63) and R5 (H² = 0.51).

Vegetation Index Correlations

Overall, the correlations were moderate to weak (r = 0.01 - 0.60) between the variables and Ndfa (Table 3.2). The variable with the strongest Pearson's correlations to Ndfa was yield (r = 0.57). Moderate correlations were also present between the calculated BLUEs for both Ndfa and

yield, with the median of the green soil adjusted vegetation index (GSAVI) (r = -0.12 - 0.38) having the strongest correlation out of all the VIs. Many MS and RGB VI bands did show significant strong correlations with each other in the positive and negative direction, and MS VIs tended to have stronger correlations to the response variables overall. SPAD had weak correlations with yield and Ndfa BLUEs (r = 0.04 - 0.19) but showed moderate correlations with a number of VIs (r = -0.47 - 0.57).

Nitrogen Derived from Atmosphere Prediction under -N

Black bean genotypes evaluated in 2021 varied for %Ndfa ($\alpha = 0.05$) (Table 3.3). The breeding line B20602 had the highest %Ndfa and B20632 had the lowest average %Ndfa. Figure 3.5 shows the correlation between the predicted and actual %Ndfa for each model using BLUEs from both equations (Field Design and Flight Date Variation) and predictive models (Model 1 and Model 2). In general, the prediction accuracies of all models ranged between -0.20 - 0.71. Both BRNeural Network and StepwiseGLM models had the highest average prediction accuracies of 0.25 and 0.24, respectively for the Field Design BLUEs, Model 1. The RMSE for all prediction models ranged between 0.57 and 1.17 and MAE for the prediction models ranged between 0.45 and 0.88 (Figure 3.6 and 3.7). Between BRNeural Network and StepwiseGLM, BRNeural Network had the lowest average RMSE of 0.83. Flight date selection did not improve the prediction accuracies of the VIs + SPAD model. Overall prediction accuracies ranged between -0.18 – 0.40 for Flight Date Variation BLUES, Model 1 (Figure 3.5). StepwiseGLM and BRNeural Network had the highest average prediction accuracies of 0.15 and 0.14, respectively. The RMSE of all models ranged between 0.57 - 1.02 and MAE of all models ranged between 0.44 - 0.78. BRNeural Network had the lowest average RMSE and MAE of 0.77 and 0.60, respectively.

The inclusion of yield greatly increased the ML models' reliability and accuracy (Figure 3.5). The average accuracy of the RForest model when utilizing all flights increased to 0.43 and saw a decrease in RMSE and MAE from 0.87 and 0.68 to 0.79 and 0.62, respectively (Figure 3.6, 3.7). Yield also improved the accuracies of all models utilizing flight date selection by between 5 - 41.1%. The average correlation between the predicted and actual values of the RForest model increased to 0.54. The RMSE and MAE of all models remained low, with an average RMSE of 0.67 and average MAE of 0.53 for the RForest model.

The prediction accuracy of models using BLUEs from the flight date dependent Flight Date Variation BLUEs in the VIs + SPAD model was greater in StepwiseGLM compared to the other models. This model was also greater than the accuracies of Field Design BLUEs VIs + SPAD models with an average accuracy of 0.28. RMSE and MAE remained low across all models, however StepwiseGLM had neither the lowest RMSE or MAE. Using only the flight dates with the highest heritability in the Flight Date Variation BLUEs VIs + SPAD model lowered the prediction accuracies overall. Under these conditions, PSLR had the highest average accuracy of 0.15 with an RMSE of 0.63 and MAE of 0.47.

The inclusion of yield did not significantly affect the prediction accuracies in Flight Date Variation BLUEs compared to the results of Field Design BLUEs. StepwiseGLM and BRNeural Network were the most accurate models with an average correlation of 0.30. Utilizing date selection did result in greater prediction accuracies, especially in the RForest model. RForest had the greatest average accuracy of 0.50 and an RMSE of 0.57 and MAE of 0.43.

Yield Prediction

Significant differences (Tukey HSD, $\alpha = 0.050$) were present between the yield means of the genotypes (Table 3.4). Genotypic variation was a significant source of variation in yield, with a difference of 733.9 kg/ha between the highest and lowest yield producers.

All Field Design BLUEs ML models showed moderate accuracies (Figure 3.8). BRNeural Network and StepwiseGLM had the highest average accuracies of 0.65 and 0.64, respectively. Between the two, BRNeural Network had the lowest RMSE and MAE (RMSE = 114.8; MAE = 91.4) (Figure 3.9, 3.10). These models also had a narrow range of variation (r = 0.42 - 0.80). Implementing flight date selection in the VI + SPAD model did not improve model accuracies, RMSE, and MAE. BRNeural Network remained the model with the highest average prediction accuracy of 0.47 with the second lowest RMSE of 131.9 and lowest MAE of 105.6.

The average prediction accuracies for the BLUEs from the Flight Date Variation model were much lower compared to the previous models. StepwiseGLM had the highest average accuracy of 0.39 while BRNeural Network had the lowest with 0.11. Additionally, the variation between accuracies across the resampling was wider with a range of 0.08 - 0.59. The RMSE of StepwiseGLM remained the lowest out of all six models (RMSE = 140.3), but RForest obtained the lowest MAE of 111.5. As with the previous implementation of flight date selection, the prediction accuracies and model metrics were not improved. StepwiseGLM and PLSR had the highest average accuracies of 0.28 with StepwiseGLM having the lowest RMSE and MAE of 136.1 and 114.1, respectively.

DISCUSSION

SNF Prediction

SNF is an important, but underutilized trait in dry bean breeding. Improving the fixation ability in dry bean would produce cultivars less reliant on N fertilizer for yield production. Evaluating SNF relies on measuring Ndfa, however it is a difficult trait to evaluate not only due to the cost and equipment specialization required, but also due to its dependance on environmental conditions (Beebe, 2012; Fageria et al., 2013; Farid and Navabi, 2015; Wilker et al., 2020). To breed for greater SNF ability by selecting for high Ndfa, repeated evaluations of Ndfa are required not only to identify parents, but to evaluate progeny across multiple environments. Remote sensing in combination with machine learning methods have the potential for use in agriculture and breeding programs through fast, reliable crop assessment.

Strong correlations between VIs and nitrogen accumulation and associated traits are important for accurate modeling. Yield and chlorophyll content measured by PhotosynQ meters were found in this study to be the most important indirect variables for estimating seed Ndfa, however correlations between the VIs used in this study and Ndfa were, in some cases, moderate overall. The correlation analysis between Ndfa and the VIs were much lower as compared to other studies utilizing spectral imagery to evaluate nitrogen content in crops. (Lee and Lee, 2013; Li et al., 2019; Ge et al., 2021; Shi et al., 2021; Han et al., 2022; Fu et al., 2022). This suggests that while VIs can evaluate nitrogen content, they are unable to specifically distinguish N¹⁵. Thus, models relying solely on VIs were neither accurate nor reliable. Despite the lack of relationship with Ndfa, the VIs captured changes in reflectance in the canopy over the growing season. This supports previous studies utilizing spectral imaging captured from wheat and maize in a single season (Benincasa et al., 2018; Wittwer and van der Heijden, 2020; Han et al., 2022; Fu et al., 2023; Fu et al., 2023; Fu et al., 2024; Fu et al.,

2022). These studies associated this variation with changing leaf nitrogen content as nitrogen from the plant is sequestered into the seed as it develops. This study used seed as opposed to shoot or leaf tissue as a previous study by Polania et al 2016 suggested that seed would be a valid substitute for measuring Ndfa as the researchers found a strong correlation (r = 0.83) between the Ndfa measured from shoot tissue and Ndfa measured from dry seed. Using seed Ndfa is valuable from a breeding perspective as it is a measure indicating genotypes that are better at partitioning and remobilizing N¹⁵ to the seed (Kamfwa et al., 2015).

RForest regression predicting BLUEs from the Field Design under the VI + SPAD + Yield model had the highest average accuracy (r = 0.54). RForest models have also been proven to be accurate in previous studies estimating leaf N content (Näsi et al., 2018; Ge et al., 2021; Shi et al., 2021). RForest models utilize bootstrap aggregation and random variable selection to reduce overfitting, which can lead to great accuracy regardless of data splitting (Breiman, 2001). In this model, yield proved to be an important variable for Ndfa prediction, having the greatest correlation out of all the variables. This is in line with previous findings where seed yield was significantly and positively associated with SNF ability (Farid and Navabi, 2015b; Barbosa et al., 2018). Its inclusion improved the models' accuracy and RMSE. This suggests yield production under unfertilized field conditions is an indicator of the plant's ability to fix atmospheric nitrogen. SPAD measurements were an important predictor for Ndfa, especially when yield was not involved with model development. Due to its strong relationship with yield and Ndfa, leaf chlorophyll content may be useable as an indirect indicator of those traits (Ramaekers et al., 2013; Jaramillo et al., 2013; Kamfwa et al., 2015; Reinprecht et al., 2020; Vollmann et al., 2022). These results suggest this proposed model has applicable use for estimating Ndfa in dry beans grown without fertilizer.

In addition to yield, the Field Design BLUEs, Model 2 RForest model relied on six VIs: the quotient of the green band divided by the red band, the difference between the red and blue bands, the combination 2 index (COM2) (Guerrero et al., 2012), the triangular green index (TGI), and the difference between the red green ratio (Gamon & Surfus, 1999) and the green blue ratio indices (Sellaro et al., 2010). The TGI has previously been used to estimate chlorophyll content (Hunt Jr. et al., 2011). Interestingly, this VI was not strongly correlated with SPAD (r = -0.31). This is in contrast to previous comparisons between TGI and chlorophyll meter measurements (Hunt et al., 2013). This discrepancy could be due to the development period chosen for taking the SPAD measurement not being equivalent to the three dates TGI was evaluated from. COM2, green blue ratio, and red green ratio indices have previously been used to evaluate desertification and the classification of vegetation coverage (Xu et al., 2022). While chlorophyll content has clear indications towards N content present, canopy coverage, and by extension, plant biomass, is dependent on the availability of N have been previously used as SNF-related traits (Heilig et al., 2016, 2017).

The VIs taken from the flight dates with the highest repeatability only improved accuracy when yield was included as a variable. The timing of image and sensory data capture has been reported to influence prediction accuracy in previous studies. Nevavuori et al. (2020) noted the accuracy in predicting yields of wheat, barley, and oats was marginally higher when using a combination of data collected from 4 weeks of flights as opposed to the complete 5 weeks. Crops like maize have periods of exponential growth wherein the timing of the flight dates is crucial as noted in Anderson II et al. (2019). They also found that correlations between plant height and yield increased over time, suggesting the use of late season flights to predict yield. Late season measurements for predicting vineyard yields were also found to be more accurate in Ballesteros et

al., (2020). While this study supports the importance of flight timing, we found that flight dates between flowering and pod filling captured the greatest variation between genotypes. This could be due to a possible relationship between N in the canopy during these periods and Ndfa in the seed. As ¹⁵N is prioritized for the seed during pod development, the Ndfa in the canopy during prepod development stages may be informative for seed Ndfa.

Yield Prediction

UAS imagery has demonstrated that VIs can accurately predict yield across a wide range of crops for the purposes of efficient and nondestructive management and monitoring (Báez-González et al., 2002; Lobell et al., 2007). We investigated whether VIs could accurately predict yield under low N conditions in this study as an efficient means of indirectly selecting for SNF. Valid estimations of SNF ability rely on a lack of N added to the system due to the sensitivity of rhizobia activity to N fertilizer (Heilig et al., 2017; Wilker et al., 2019). Grain yield as an indirect indicator of SNF was established in previous research and is an important and strongly correlated variable in this study. SPAD proved to be important in the early prediction models, suggesting that it may be useful as a measure in the simultaneous selection for high fixation ability and yield. In contrast to the Ndfa models, relying on only VIs and SPAD for predicting yield provided both accurate and reliable predictions. The models in this study utilized seventeen VIs in addition to SPAD. Six of which were linear equations using the red, green, blue, and near-infrared bands. Like the combination 2 index, the hue index has previously been used to distinguish vegetation from soil (Meyer & Neto, 2008). The brightness index and the shape index is also used for soil measurements (Bakacsy et al., 2023; Schmidt & Karnieli, 2001). The normalized difference rededge (NDRE), difference (DVI), chlorophyll green (CIgreen) and chlorophyll red-edge (CIred-edge) vegetation indices have all been used to quantify chlorophyll content, and vegetation density in previous studies (Ahamed et al., 2011; Gitelson et al., 2003, 2005; Thompson et al., 2019; Tucker, 1979). Lastly, the blue band and two modified VIs were also utilized in model development: a modified brightness index and a modified red chromatic coordinate. The latter of which is used to distinguish vegetation from non-vegetative material (Woebbecke et al., 1995). These VIs indicate the relevance for chlorophyll content, vegetation density, and canopy coverage towards predicting yield.

VIs using only RGB bands have been shown to be useful in accurately predicting yields of numerous crops. For instance, RGB bands were able to accurately predict wheat yield as in Zeng et al (2021) ($R^2 = 0.55 - 0.68$) and maize yield in Gracia-Romero et al. (2017) ($R^2 = 0.66$) and Buchaillot et al. (2019) ($R^2 = 0.60$). There is also a precedent for the use of only multispectral image-derived VIs and in combination with RGB VIs to accurately estimate yield. Ballester et al (2017) utilized multispectral VIs associated with biomass and plant to accurately predict cotton yield ($R^2 = 0.64$) and Zhou et al (2017) ($R^2 = 0.71 - 0.75$) found comparable results in predicting rice grain yield using MS and digital imaging. This study utilized a combination of RGB and multispectral VIs. Combining the sensor data can lead to greater accuracies as near-infrared and red edge bands have been noted to be efficient in status monitoring (Zheng et al., 2018; Herzig et al., 2021; Bascon et al., 2022).

Yield, like SNF, is a complex trait dependent on both genetic and environmental factors. Combining VIs with other metrics, including physical plant measurements and climate data, may also improve predictive modeling as shown in Lu et al. (2019), Zeng et al. (2021) and Han et al. (2020). It is notable that the stepwise general linear model performed just as well as the Bayesian regularized artificial neural network model. BRNeural Network are robust models that are designed to be difficult to overtrain and overfit through the combination of conditional probability and weight regularization, but like other artificial neural networks they can be difficult to interpret how the variables affect model activity (Burden & Winkler, 2009). Unlike with predicting Ndfa, flight date selection negatively impacted the performance of the machine learning models under both the alpha-lattice and nested flight date designs. This may suggest that while specific periods during plant development are more important in evaluating Ndfa, yield in dry bean, a relatively short season crop, may not have developmental stages that are more informative in making predictions. Thus, accurately predicting yield may be more reliant on season-long measurements. Further repetitions of the experiment are required to parse the applicability of both models under varying environmental conditions.

Yield as an Indicator of SNF Under Low N

The multi-year study shows the significance of genotype and environment regarding yield production under low N conditions (Table 2.1). The trials that received N fertilizer showed an advantage in yield overall except in 2020 where added N had no significant effect. These findings agree with previous studies conducted by Farid and Navabi (2015), Farid et al. (2016), Heilig et al. (2016), and Heilig et al. (2017) which investigated genetic variation and environmental effects on N fixation and subsequent seed yield.

Improving SNF ability by selecting genotypes that perform well under low N, or SNF dependent, and N-dependent conditions has been demonstrated by Farid et al. (2016) and Fageria et al. (2013). Using yield as an indirect indicator under low N for SNF would allow for improvement of fixation ability without disregarding yield as a target trait. The equivalent yield performance in 2020 between the two trials is possibly indicative of environmental conditions where the SNF ability of bean genotypes was equal to genotypes grown with standard fertilizer application. 29 of the 36 lines planted in 2020 are repeated lines from either 2019 or 2021, most

of which responded to fertilizer application in those environments. Seven genotypes, Zenith, B16504, B18204, Adams, B19309, B19330, and Black Bear show consistent, high yield production under no N and added N across repeated years. The four lines grown in the 2021 trial from this selection, Zenith, B16504, B19309, and B19330, showed high %Ndfa when evaluated as well. Adams and the maternal parent of Black Bear, Jaguar, were developed in the MSU Dry Bean Breeding program under conventional breeding trial conditions. Zenith was selected under organic conditions during variety development. Zenith, Adams, and Black Bear share lineage with Jaguar, which may have influenced their SNF abilities. Additionally, Zenith's paternal parent, Zorro was tested both in this study and in Wilker et al (2019). While not among the top fixers in this study, Zorro was found to be one of the overall top fixers of the conventionally bred lines in the Wilker (2019) study. Ndfa is a moderately heritable trait that has been demonstrated to be conserved without purposeful selection (Farid et al., 2017). The genotypes indicated in this study may be good parental candidates for improving SNF ability indirectly by breeding for improved yield under low N conditions.

Limitations

VIs proved useful in leaf nitrogen estimation in previous studies. However, leaf canopy N content may be a poor direct indicator of Ndfa because the N captured in the spectral data is neither the same as nor proportional to the ¹⁵N analyzed from the seed. As opposed to N from the soil or fertilizer, atmospheric nitrogen is generally prioritized for grain storage rather than for depositing in the soil or remaining in the plant (Westermann et al., 1985; Fustec et al., 2010). The models relying on VIs alone were neither accurate nor reliable and this could be due to the weak relationship between leaf nitrogen content and seed Ndfa. Prediction accuracy of Ndfa using VIs could be strengthened by exploring the relationship between leaf nitrogen content and seed Ndfa.

in future studies. The models predicting yield did fare better in prediction accuracy and reliability despite the relationship between the VIs and yield being as weak as with seed Ndfa.

The best performing Ndfa model relies on the inclusion of yield for predicting Ndfa. Utilizing this model requires RS imagery taken from plants grown under low-N conditions where no fertilizer is added. Adding fertilizer decreases nitrogen fixation and any predictions made with fertilized plants may not accurately reflect their fixation abilities (Reinprecht et al., 2020). While the model can be utilized as a suggestive indication of fixation ability, the model should be used in multi-year, multi-location trials to evaluate its efficacy under varying environments before fully deploying it in a breeding program. The same can be said for the yield models. This study has demonstrated potential application of indirectly selecting for SNF using yield and has trained an accurate and reliable yield prediction model. However, the model was also developed from data collected under low-N conditions. Additionally, both models' accuracy may vary between years and locations because soil characteristics and environmental conditions influence SNF, and it may be beneficial to use control plots to calibrate the model's precision (Thilakarathna and Raizada, 2018).

Applications in Plant Breeding

Rising fertilizer costs and global food demand have driven increased efforts in developing crops with high yield production and lower input requirements. The ability in common bean to fix enhanced levels of atmospheric nitrogen and decrease reliance on applied fertilizer presents the opportunity to meet this aim. Dry bean has been characterized as a poor nitrogen fixer relative to other legumes (Hardarson et al., 1993), however previous studies have found moderate to high fixation ability in commercial cultivars, elite breeding lines in the MSU Dry Bean Breeding program, and among the Andean Diversity panel (avg %Ndfa = 12.4 - 63.3) (Kamfwa et al., 2015;

Heilig et al., 2017). This genetic variability can be leveraged for improving fixation ability through breeding, however, the time and cost of measuring this trait is a barrier to improved variety development due to the frequency of measurements required to evaluate potential parents and progeny. UAS imagery offers a cost-effective, quick, and efficient method of estimating crop metrics using VIs and has been shown to adequately capture nitrogen content in crops.

Using prediction models as a selection tool in a breeding program requires both accuracy and reliability. Additionally, an ideal model would be applied early in the season to identify parents that could be rapidly advanced into a breeding pipeline. The BRNeural Network and StepwiseGLM yield models described in this study can be applied pre-harvest. Using either of these models would negate the time and cost of preparing and submitting plant material samples for N^{15} analysis, with results from the prediction model available before harvest.

CONCLUSION

This study explored the potential of UAS imaging in estimating Ndfa and seed yield. The goal was to develop an accurate and reliable predictive model to deploy in breeding programs to develop dry bean lines that require less fertilizer to produce yields comparable to conventional commercial practices. We found that Ndfa measured in the seed could not be estimated reliably or accurately. This is inferred to be due to the lack of distinction between soil-derived N and SNF derived N in the canopy captured by the imaging. The random forest model developed with the addition of yield to RS data showed improvement in accuracy and reliability, providing the groundwork for future research in the application of VIs as indirect measurements of Ndfa. Further validation is required for this model.

We also investigated the use of yield as an indirect measure of SNF. The three-year N study showed the positive effect of increasing N applications on yield and revealed genotypes that

produce comparable yields under -N fertilization over repeated years. Seven breeding lines and cultivars (Zenith, B18204, Adams, B16504, B19309, B19330 and Black Bear) were identified as having consistent yields under low N conditions. These lines show promise as potential parents in breeding programs that aim to increase yield potential and SNF ability. In addition to this, we were able to develop two accurate and reliable yield prediction models using StepwiseGLM and BRNeural Network and only using RS data. Overall, using remote sensing and machine learning show promise to develop prediction tools to improve SNF in common bean.

TABLES

Year	Number of entries	Trial	Mean (kg/acre)	CV	P-value
2010	40	N+	1102.1	0.09	< 0.0001
2019	42	N-	979.4	0.13	< 0.0001
2020	26	N+	1555.3	0.11	0.62
2020	30	N-	1534.8	0.11	0.02
2021	40	N+	1667.4	0.09	< 0.0001
2021	42	N-	1457.14	0.1	< 0.0001

Table 3.1. Seed-yield production for the multi-year nitrogen trial. N- no nitrogen trial; N+ added nitrogen trial; CV coefficient of variation.

			Yield	Yield	Ndfa	Ndfa
Vegetation Indices	SPAD	Yield	BLUEs	BLUEs	BLUEs	BLUEs
-			Field	Flight	Field	Flight
B std RGB	-0.38**	-0.21	-0.76***	-0.89***	-0.71***	-0.30*
CI std RGB	-0.44***	-0.01	-0.56***	-0.80***	-0.59***	-0.36**
GdivB median RGB	-0.73***	0.01	-0.33*	-0.66***	-0.44***	-0.32*
GdivR mean RGB	-0.61***	0.39**	-0.08	-0.48***	-0.25	-0.29*
GmnR std RGB	0.71***	-0.07	0.08	0.33*	0.09	-0.06
HUE mean RGB	0.76***	-0.39**	0.05	0.43***	0.19	0.09
MRCCbyAlper mean RGB	-0.39**	-0.46***	-0.86***	-0.88***	-0.75***	-0.18
MRCCbyAlper std RGB	-0.27*	-0.56***	-0.91***	-0.88***	-0.79***	-0.25
MyIndexi std RGB	-0.25	-0.40**	-0.82***	-0.86***	-0.75***	-0.37**
RCC mean RGB	0.14	-0.85***	-0.70***	-0.38**	-0.46***	0.01
RmnB std RGB	-0.49***	-0.40**	-0.78***	-0.88***	-0.74***	-0.29*
TGI mean RGB	-0.61***	-0.1	-0.65***	-0.84***	-0.64***	-0.14
B median MS	0.55***	-0.38**	-0.38**	0.05	-0.08	0.38**
B std MS	0.63***	-0.14	0.31*	0.65***	0.40**	0.30*
RE median MS	-0.67***	-0.09	-0.52***	-0.78***	-0.57***	-0.27*
RE std MS	-0.61***	-0.57***	-0.42**	-0.57***	-0.63***	-0.49***
NIR std MS	0.53***	0.03	0.58***	0.77***	0.51***	0.13
CCCI std MS	-0.68***	-0.23	-0.54***	-0.77***	-0.65***	-0.35**
CVI std MS	0.53***	0.15	0.65***	0.82***	0.59***	0.19
redEdgeNDVI median MS	-0.72***	0.19	-0.17	-0.55***	-0.34**	-0.32*
GRNDVI std MS	0 62***	-0 36**	0.17	0 53***	0.26*	0.17
CIrededge mean MS	0.54***	0.34**	0.72***	0.85***	0.67***	0.25
CIrededge std MS	0.41**	0.27*	0.72	0.83***	0.58***	0.17
GSAVI median MS	0.26*	0 59***	0.86***	0.81***	0.50	0.16
NormG std MS	-0.55***	-0.54***	-0.76***	-0.84***	-0.76***	-0.30*
DVI std MS	0.13	0.26*	0 79***	0.73***	0 53***	-0.05
NDRE std MS	-0.62***	-0.55***	-0.64***	-0.74***	-0.77***	-0.37**
BCC std MS	0.41**	0.04	0.60***	0.75***	0.48***	0.1
BdivG std MS	0.63***	-0.03	0.47***	0.74***	0.50***	0.21
BI std MS	0.02	-0.91***	-0.72***	-0.42**	-0.58***	-0.03
BIM std MS	0.55***	-0.42***	0.12	0.49***	0.19	0.19
CI mean MS	0.35**	-0.84***	-0.48***	-0.11	-0.25	0.01
CI std MS	-0.75***	-0.11	-0.24	-0.56***	-0.48***	-0.42***
EXR std MS	0.35**	-0.80***	-0.25	0.09	-0.17	-0.06
GCC std MS	0.56***	-0.26*	0.36**	0.63***	0.36**	0.08
GmnR std MS	-0.65***	-0.30*	-0.60***	-0.81***	-0.66***	-0.35**
I median MS	-0.51***	-0.29*	-0.74***	-0.86***	-0.67***	-0.18
IF std MS	0.65***	-0.05	0.44***	0.72***	0.48***	0.21

Table 3.2. Pearson's correlation plot of vegetation indices, SPAD, yield, and BLUEs for %Ndfa and yield calculated from the field design and flight date variation model.

Tabl	le 3	.2 C	ont'd	

MSRGR std MS	-0.19	-0.74***	-0.18	-0.17	-0.35**	-0.48***
RdiB std MS	-0.55***	-0.65***	-0.60***	-0.70***	-0.68***	-0.42***
RmnB std MS	0.62***	-0.40**	0.11	0.48***	0.23	0.2
TGI std MS	-0.64***	-0.30*	-0.65***	-0.83***	-0.68***	-0.29*
VEG std MS	-0.68***	-0.34**	-0.48***	-0.73***	-0.61***	-0.46***
SPAD	1	0.11	0.02	0.35**	0.24	0.16
Yield	0.11	1	0.50***	0.28*	0.47***	0.15
Yield BLUEs Field	0.02	0.50***	1	0.81***	0.88^{***}	0.11
Yield BLUEs Flight	0.35**	0.28*	0.81***	1	0.71***	0.52***
Ndfa BLUEs Field	0.24	0.47***	0.88^{***}	0.71***	1	0.08
Ndfa BLUEs Flight	0.16	0.15	0.11	0.52***	0.08	1

<u> </u>	I ottor	%Ndfa BLUFs
B20602	Δ	07 32
B20501		97.52
D20371 D10345		06.80
D19343 D10344	ABC	90.89
D19344 D20527	ADCD	90.74
D20527		90.70
D20579		90.33
B20549	ABCDEF	96.50
B20617	ABCDEFG	96.48
B20532	BCDEFGH	96.43
Zenith	BCDEFGHI	96.36
B20597	BCDEFGHI	96.35
B20642	CDEFGHI	96.31
B19330	CDEFGHIJ	96.30
B19309	CDEFGHIJ	96.26
B20620	CDEFGHIJ	96.25
B16504	CDEFGHIJ	96.21
B20536	CDEFGHIJK	96.16
B20627	CDEFGHIJKL	96.12
B20629	DEFGHIJKLM	95.95
Adams	DEFGHIJKLMN	95.94
Black Bear	DEFGHIJKLMN	95.92
B19339	DEFGHIJKLMN	95.90
B20542	EFGHIJKLMN	95.83
B18236	FGHIJKLMN	95.79
B19341	FGHIJKLMNO	95.74
B20590	FGHIJKLMNO	95.71
Nimbus	FGHIJKLMNO	95.71
B20623	FGHIJKLMNO	95.69
B20599	FGHIJKLMNO	95.67
Zorro	FGHIJKLMNO	95.67
B20639	GHIJKLMNO	95.62
B19332	HIJKLMNO	95.59
B20538	IJKLMNO	95.53
B19340	GHIJKLMNO	95.53
B20616	JKLMNO	95.44
B20621	KLMNO	95.31
B20582	LMNO	95.12
Black Beard	NO	95.07
B20547	0	94.88
B20632	P	93.28
R99	Q	0

Table 3.3. Tukey's honest significant difference of percent nitrogen derived from the atmosphere (%Ndfa) of 42 black bean genotypes grown in 2021. R99 is the reference plant.

Line	Letter	Yield BLUEs
B20536	А	1827.3
B19344	AB	1774.8
B20542	ABC	1760.3
B16504	ABCD	1730.8
B20597	ABCD	1725.3
Zenith	ABCDE	1707.5
B20599	ABCDE	1703.2
B20602	ABCDE	1701.5
B20591	ABCDE	1695.2
B20590	ABCDEF	1694.8
Adams	ABCDEF	1690.7
B20617	ABCDEF	1688.8
B19309	ABCDEF	1684.7
B20616	ABCDEF	1654.9
B19330	ABCDEF	1641.5
B20532	ABCDEF	1635.6
B20549	ABCDEFG	1613.0
B20538	ABCDEFG	1606.7
B20527	ABCDEFG	1604.4
B20579	ABCDEFG	1597.9
B20547	ABCDEFG	1590.9
B20642	ABCDEFG	1569.3
B19332	ABCDEFG	1546.9
B19339	ABCDEFG	1527.2
B19345	ABCDEFG	1522.2
B19341	ABCDEFG	1521.3
B20639	ABCDEFG	1488.2
B20623	ABCDEFG	1485.0
B19340	ABCDEFG	1484.9
B20621	ABCDEFG	1468.2
B20620	BCDEFG	1439.0
B20629	ABCDEFG	1438.5
Zorro	BCDEFG	1422.9
B20627	BCDEFG	1418.3
B18236	BCDEFG	1411.4
Eclipse	BCDEFG	1403.7
Nimbus	CDEFG	1358.2
B20582	DEFG	1349.6

Table 3.4. Tukey's honest significant difference of seed yield (kg/a) of 42 black bean genotypes grown in 2021.

Table 3.4 (cont'd)		
Black Bear	EFG	1316.3
B20632	FG	1308.0
Black Beard	G	1240.4
R99	Н	51.8

FIGURES



Figure 3.1. Prediction model design and development methodology for estimating %Ndfa.



Figure 3.2. Prediction model design and development methodology for estimating yield.



Figure 3.3. Distribution of yield (kg/ha) over 2019, 2020, and 2021 between the two N treatments.



Figure 3.4. Average repeatability for 12 RGB imaging dates taken over the 2021 season. Dates not connected by the same letter are significantly different (P < 0.05) according to Tukey's honest significant difference.



Figure 3.5. Percent nitrogen derived from the atmosphere (%Ndfa) predictions accuracies for the ML models grouped by predictive model and BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.



Figure 3.6. MAE for the ML models predicting %Ndfa grouped by predictive model and BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.



Figure 3.7. RMSE for the ML models predicting %Ndfa grouped by predictive model and BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.



Figure 3.8. Prediction accuracies for the ML models predicting Yield grouped by BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.



Figure 3.9. MAE for the ML models predicting Yield grouped by BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.



Figure 3.10. RMSE for the ML models predicting Yield grouped by BLUE model. KNNeighbors, K-nearest neighbors; StepwiseGLM, stepwise general linear model; BRNeural Network, Bayesian regularized neural network; XGBoosting, extreme gradient boosting; PLSR, partial least squares; RForest, random forest.
REFERENCES

- Akter, Z., Lupwayi, N. Z., & Balasubramanian, P. (2017). Nitrogen use efficiency of irrigated dry bean (Phaseolus vulgaris L.) genotypes in southern Alberta. *Canadian Journal of Plant Science*, CJPS-2016-0254. https://doi.org/10.1139/CJPS-2016-0254
- Asghari, H., & Cavagnaro, T. (2011). Arbuscular mycorrhizas enhance plant interception of leached nutrients. *Functional Plant Biology*, 38, 219–226. https://doi.org/10.1071/FP10180
- Barbosa, N., Portilla, E., Buendia, H. F., Raatz, B., Beebe, S., & Rao, I. (2018). Genotypic differences in symbiotic nitrogen fixation ability and seed yield of climbing bean. *Plant and Soil*, 428(1), 223–239. https://doi.org/10.1007/s11104-018-3665-y
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, 25(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7
- Boddey, R. M., & Knowles, R. (1987). Methods for quantification of nitrogen fixation associated with gramineae. *Critical Reviews in Plant Sciences*, 6(3), 209–266. https://doi.org/10.1080/07352688709382251
- Boshkovski, B., Tzerakis, C., Doupis, G., Zapolska, A., Kalaitzidis, C., & Koubouris, G. (2021). Relationship between physiological and biochemical measurements with spectral reflectance for two *Phaseolus vulgaris* L. genotypes under multiple stress. *International Journal of Remote Sensing*, 42(4), 1230–1249. https://doi.org/10.1080/01431161.2020.1826061
- Burden, F., & Winkler, D. (2009). Bayesian Regularization of Neural Networks. In D. J. Livingstone (Ed.), *Artificial Neural Networks: Methods and Applications* (pp. 23–42). Humana Press. https://doi.org/10.1007/978-1-60327-101-1_3
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785
- Chlingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151, 61–69. https://doi.org/10.1016/j.compag.2018.05.012
- Douxchamps, S., Humbert, F.-L., van der Hoek, R., Mena, M., Bernasconi, S. M., Schmidt, A., Rao, I., Frossard, E., & Oberson, A. (2010). Nitrogen balances in farmers fields under alternative uses of a cover crop legume: A case study from Nicaragua. *Nutrient Cycling in Agroecosystems*, 88(3), 447–462. https://doi.org/10.1007/s10705-010-9368-2
- Farid, M., Earl, H. J., & Navabi, A. (2016). Yield Stability of Dry Bean Genotypes across Nitrogen-Fixation-Dependent and Fertilizer-Dependent Management Systems. Crop Science, 56(1), 173–182. https://doi.org/10.2135/cropsci2015.06.0343

- Farid, M., Earl, H. J., Pauls, K. P., & Navabi, A. (2017). Response to selection for improved nitrogen fixation in common bean (Phaseolus vulgaris L.). *Euphytica*, 213(4), 99. https://doi.org/10.1007/s10681-017-1885-5
- Giller, K. E. (2001). Nitrogen Fixation in Tropical Cropping Systems. CABI.
- Good, K. (2021, December 12). Fertilizer Prices- Still No Reprieve Farm Policy News. *Farm Policy News*. https://farmpolicynews.illinois.edu/2021/12/fertilizer-prices-still-no-reprieve/
- Han, M., Okamoto, M., Beatty, P. H., Rothstein, S. J., & Good, A. G. (2015). The Genetics of Nitrogen Use Efficiency in Crop Plants. *Annual Review of Genetics*, 49(1), 269–289. https://doi.org/10.1146/annurev-genet-112414-055037
- Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4), 542–553. https://doi.org/10.1016/S0034-4257(03)00131-7
- Irrigated Dry Bean Nutrient Requirements for Southern Alberta. (2013). Alberta Agriculture, Food and Rural Development.
- Kamfwa, K., Cichy, K. A., & Kelly, J. D. (2015). Genome-wide association analysis of symbiotic nitrogen fixation in common bean. *Theoretical and Applied Genetics*, 128(10), 1999–2017. https://doi.org/10.1007/s00122-015-2562-5
- Lee, K.-J., & Lee, B.-W. (2013). Estimation of rice growth and nitrogen nutrition status using color digital camera image analysis. *European Journal of Agronomy*, 48, 57–65. https://doi.org/10.1016/j.eja.2013.02.011
- Li, Y., Chen, D., Walker, C. N., & Angus, J. F. (2010). Estimating the nitrogen status of crops using a digital camera. *Field Crops Research*, 118(3), 221–227. https://doi.org/10.1016/j.fcr.2010.05.011
- Liang, L., Di, L., Huang, T., Wang, J., Lin, L., Wang, L., & Yang, M. (2018). Estimation of Leaf Nitrogen Content in Wheat Using New Hyperspectral Indices and a Random Forest Regression Algorithm. *Remote Sensing*, 10(12), 1940. https://doi.org/10.3390/rs10121940
- Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, 15(1), 17. https://doi.org/10.1186/s13007-019-0402-3
- Matias, F. I., Caraza-Harter, M. V., & Endelman, J. B. (2020). FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials. *The Plant Phenome Journal*, 3(1), e20005. https://doi.org/10.1002/ppj2.20005

- Moghimi, A., Pourreza, A., Zuniga-Ramirez, G., Williams, L. E., & Fidelibus, M. W. (2020). A Novel Machine Learning Approach to Estimate Grapevine Leaf Nitrogen Concentration Using Aerial Multispectral Imagery. *Remote Sensing*, 12(21), 3515. https://doi.org/10.3390/rs12213515
- Moor, A., Carey, A., Hines, S., & Brown, B. (n.d.). Southern Idaho Fertilizer Guide. *University* of Idaho Extension.
- Mulvaney, R. L., Khan, S. A., & Ellsworth, T. R. (2009). Synthetic Nitrogen Fertilizers Deplete Soil Nitrogen: A Global Dilemma for Sustainable Cereal Production. *Journal of Environmental Quality*, 38(6), 2295–2314. https://doi.org/10.2134/jeq2008.0527
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7. https://www.frontiersin.org/articles/10.3389/fnbot.2013.00021
- Polania, J., Poschenrieder, C., Rao, I., & Beebe, S. (2016). Estimation of phenotypic variability in symbiotic nitrogen fixation ability of common bean under drought stress using 15N natural abundance in grain. *European Journal of Agronomy*, 79, 66–73. https://doi.org/10.1016/j.eja.2016.05.014
- Qin, Z., Myers, D. B., Ransom, C. J., Kitchen, N. R., Liang, S.-Z., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernandez, F. G., Franzen, D. W., Laboski, C. A. M., Malone, B. D., Nafziger, E. D., Sawyer, J. E., & Shanahan, J. F. (2018). Application of Machine Learning Methodologies for Predicting Corn Economic Optimal Nitrogen Rate. *Agronomy Journal*, 110(6), 2596–2607. https://doi.org/10.2134/agronj2018.03.0222
- Saberioon, M., Amin, M., Gholizadeh, A., & Ezrin, M. (2014). A Review of Optical Methods for Assessing Nitrogen Contents During Rice Growth. *Applied Engineering in Agriculture*, 30, 657–669. https://doi.org/10.13031/aea.30.10478
- Sankaran, S., Zhou, J., Khot, L. R., Trapp, J. J., Mndolwa, E., & Miklas, P. N. (2018). Highthroughput field phenotyping in dry bean using small unmanned aerial vehicle based multispectral imagery. *Computers and Electronics in Agriculture*, 151, 84–92. https://doi.org/10.1016/j.compag.2018.05.034
- Shearer, G., & Kohl, D. H. (1986). N2-Fixation in Field Settings: Estimations Based on Natural 15N Abundance. *Functional Plant Biology*, 13(6), 699–756. https://doi.org/10.1071/pp9860699
- Shi, P., Wang, Y., Xu, J., Zhao, Y., Yang, B., Yuan, Z., & Sun, Q. (2021). Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Computers and Electronics in Agriculture*, 180, 105860. https://doi.org/10.1016/j.compag.2020.105860
- Song, L., Guanter, L., Guan, K., You, L., Huete, A., Ju, W., & Zhang, Y. (2018). Satellite suninduced chlorophyll fluorescence detects early response of winter wheat to heat stress in the Indian Indo-Gangetic Plains. *Global Change Biology*, 24(9), 4023–4037. https://doi.org/10.1111/gcb.14302

- Stas, M., Van Orshoven, J., Dong, Q., Heremans, S., & Zhang, B. (2016). A comparison of machine learning algorithms for regional wheat yield prediction using NDVI time series of SPOT-VGT. 2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics), 1–5. https://doi.org/10.1109/Agro-Geoinformatics.2016.7577625
- Tobias, R. D. (1995). An Introduction to Partial Least Squares Regression.
- Turner, R. E., & Rabalais, N. N. (1994). Coastal eutrophication near the Mississippi river delta. *Nature*, 368(6472), 619–621. https://doi.org/10.1038/368619a0
- Turner, R. E., Rabalais, N. N., & Justic, D. (2008). Gulf of Mexico Hypoxia: Alternate States and a Legacy. *Environmental Science & Technology*, 42(7), 2323–2327. https://doi.org/10.1021/es071617k
- Yao, X., Huang, Y., Shang, G., Zhou, C., Cheng, T., Tian, Y., Cao, W., & Zhu, Y. (2015). Evaluation of Six Algorithms to Monitor Wheat Leaf Nitrogen Concentration. *Remote Sensing*, 7(11), 14939–14966. https://doi.org/10.3390/rs71114939
- Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., Feng, Z., & Kusnierek, K. (2020). Improving Unmanned Aerial Vehicle Remote Sensing-Based Rice Nitrogen Nutrition Index Prediction with Machine Learning. *Remote Sensing*, 12(2), 215–215. https://doi.org/10.3390/rs12020215
- Zheng, H., Li, W., Jiang, J., Liu, Y., Cheng, T., Tian, Y., Zhu, Y., Cao, W., Zhang, Y., & Yao, X. (2018). A Comparative Assessment of Different Modeling Algorithms for Estimating Leaf Nitrogen Content in Winter Wheat Using Multispectral Images from an Unmanned Aerial Vehicle. *Remote Sensing*, 10(12), 2026. https://doi.org/10.3390/rs10122026