DEVELOPING AND IMPLEMENTING GENOMIC RESOURCES FOR THE IMPROVEMENT OF MILLETS AND HEMP

by

MATTHEW JOHNSON

(Under Direction of Jason Wallace)

ABSTRACT

To increase the crop diversity of the worlds agricultural systems, expanded research and development efforts are needed to address many issues that keep various crops out of mainstream agricultural production systems. There are many potential crops, or groups of crops that could be studied, and it is important to choose plants that have the potential to give a valuable return on the investments that needs to be made for their improvement. Millets and hemp are some of the plants that have great potential to improve our cropping systems, but significant research is required to overcome their current limitations. Millets are a diverse group of grain crops that have great potential to deliver valuable yields as the world's climate continues to change and water scarcity increases. To further develop millet crops, genomic marker data sets were developed from accessions in the core collections of kodo millet, little millet, and proso millet. These marker sets were used to better understand the underlying genetics of these crops, and were provided as a public resource for future research efforts. For pearl millet, genomic selection was investigated as a potential approach to increase the speed at which pearl millet lines resistant to the parasitic weed *Striga* and the disease downy mildew can be developed. Hemp has historically been a major fiber crop, and has potential applications for food and medicine. In the

20th century the United States listed hemp as a schedule 1 drug and its production was completely outlawed, but in recent years it has been allowed to come back into production. We developed a genetic marker data set, and analyzed the chemical composition of twenty-two commercially available hemp accessions to better understand the genetic stability of the accessions that many farmers are working with. We also worked to better understand what are the current issues that need to be addressed by further research. The continued development of genomic resources for millets and hemp will give researchers the tools needed to quickly advance these crops and allow farmers to more fully utilize the potential of these crops to improve our agricultural systems.

INDEX WORDS: hemp; kodo millet; little millet; proso millet; pearl millet; orphaned crops; genomic resources

DEVELOPING AND IMPLEMENTING GENOMIC RESOURCES FOR THE IMPROVEMENT OF MILLETS AND HEMP

by

MATTHEW JOHNSON

B.S., California Polytechnic State University San Luis Obispo, 2016

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2020

DEVELOPING AND IMPLEMENTING GENOMIC RESOURCES FOR THE IMPROVEMENT OF MILLETS AND HEMP

by

MATTHEW JOHNSON

Major Professor:	Jason Wallace
Committee:	Zenglu Li
	Alexander Bucksch
	Tim Coolong
	Joseph Knoll

Electronic Version Approved:
Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
August 2020

TABLE OF CONTENTS

Page

CHAPTER

CHAPTER 1

INTRODUCTION TO MILLETS AND HEMP AND THEIR GENETIC AND GENOMIC RESOURCES (INTRODUCTION AND LITERATURE REVIEW)

Millets: Pearl, Kodo, Little, and Proso Millet

Millets are a diverse group of grain crops, many of which have been farmed for thousands of years. They serve an important role in food security in the developing world, and pose an opportunity to diversify global diets. Despite their importance to many communities, there have not been extensive investments in the development of genomic resources for many millet species (Saha et al., 2016). This lack of investment has greatly inhibited the improvement of these crops. Thankfully not all millets lack basic genomic resources: pearl millet (*Pennisetum glaucum* syn. *Cenchrus americanus*), foxtail millet (*Setaria italica*), and finger millet (*Eleusine coracana*) all have completed publically available reference genomes and established germplasm resources (Varshney et al., 2017; Zhang et al., 2012; Prasad 2017; Sehgal et al., 2015; Hittalmani et al., 2017; Jia et al., 2013; Bennetzen et al., 2012 and Hatakeyama et al., 2017). Unfortunately, many other millets have few, if any, genetic resources available (Goron and Raizada 2015).

Pearl millet is one of the millets that has had a relatively significant amount of work done on it. It is an important source of food and income to over 90 million people around the world (ICRISAT, 2019). It is planted on more than 75 million acres and accounts for roughly half of all millets grown every year (ICRISAT, 2019). The importance of pearl millet has continued to

grow in recent decades. In Africa, where the most pearl millet is produced, production has increased ~130% since the 1980's (ICRISAT, 2019).

Pearl millet has been a staple food in African since the Neolithic age (Bourlag, 1996; Harlan, 1992). Evidence points towards a domestication event around 4000 years ago (possibly as far back as 6000 years ago) in the west African Sahel region (Tostain, 1992; Fuller, 2003; Nuemann, 2003; Manning and Fuller, 2014). The cultivation and domestication of pearl millet started while the peoples of the region where still semi-sedentary (Mason et al., 2015). By 1700 BC (and possibly as early as 1900 BC) production of pearl millet had spread throughout much of Africa and over towards modern day India, near the Indus valley (Fuller, 2003; Boivin and Fuller, 2009; Manning et al., 2011). Today pearl millet is commonly grown all around the world as a grain and a forage crop, with Africa still being the most important production region focused on growing pearl millet for human consumption (FAO, 2018).

One of the reasons that Pearl millet is so valuable is its capacity to produce usable yields under conditions that are less than ideal. It is often grown in soil that is poorly suited for agricultural production (high pH, low organic matter, low nutrient levels, high soil aluminum, low soil moisture, high salinity) and in harsh environmental conditions (low rainfall and high temperatures). Pearl millet can even tolerate temperature hotter than 42 degrees Celsius during its reproductive stage, allowing it to produce its grain even if an untimely heat wave hits the crop (Gupta et al., 2015). It also has a genetically diverse group of accessions with significant ranges of maturity, allowing it to be grown throughout many areas of the world. Some accessions can mature in as little as two months and remain short (oasis-type pearl millets), while other accessions are photoperiod-dependent and can take up to five months to mature, and they grow

over 4.5 meters tall (Bilquez, 1963). Its capacity to grow in extreme and variable conditions make it such a valuable crop in places throughout Africa and India.

Beyond the value of its capacity to yield on marginal lands, pearl millet is a very healthy grain for human consumption. It has 8-19% protein, low levels of starch, and a high fiber concentration (Nambiar et al., 2011). It has also been shown to have higher levels of micronutrients than the other major grain crops (rice, wheat, maize, and sorghum) (Tako et al., 2015). Although it grown primarily as a grain for human consumption in Africa, the stover (leaves and stalks) of the plant is also an important food for livestock, and what is not fed to humans or animals often serves as fuel (Vadez et al., 2012; Mason et al., 2015).

Even though this crop is arguably the most important millet crop, its yields (~900 kg/ha) are still low compared to other major crops. A major contributing factor to the low yields are the dryland growing conditions pearl millet is commonly farmed under, but lack of needed research and development also plays a significant role. Although, it must be noted that plant breeders have made significant genetic gains in recent decades; from 1996-2013 there has been, on average, an increase of 24 kg/ha/year in India (Yadav and Rai, 2013).

The breeding work in Africa is also progressing well. In Africa, there is work being done to develop pearl millet lines that are resistant to the parasitic weed *Striga* (Kountche et al., 2013). *Striga* and several foliar diseases (downy mildew by *Sclerospora graminicola*, Pyricularia leaf spot by *Pyricularia grisea*, and rust *Puccinia substriata* var. *indica*) are major issues for pearl millet production in the developing world (Varshney et al., 2017). *Striga* is the most serious pest issue for pearl millet production in Africa, and while efforts are being made to breed resistance into pearl millet (Kountche et al., 2013) there is still much more work to be done for a resistant line to be developed and released (Tesso and Ejeta, 2011). There has been successful host plant

resistance bred into sorghum (*Sorghum bicolor* (L). This supports the idea that it is just a matter of time and work for resistances in pearl millet to be identified and implemented as well. The lack of commercially available resistant varieties has left farmers relying on production methods to deal with *Striga* infestations. Infestations build up with continual plantings of host plants (such as pearl millet) and when soil has low fertility (Andrews and Brammel-Cox, 1993). To fight against *Striga* loses there are many approaches farmers take: late plantings, low tillage, proper nutrient applications, and including legumes for intercropping have all been shown to reduce *Striga* infestations (Gworgwor et al., 1998; Andrews and Brammel-Cox, 1993; Hess and Ejeta, 1987; Carsky et al., 1994; Carson, 1988). Hand weeding is also a very common, but extremely labor intensive, method of weed control used (Hatcher and Melander, 2003). While *Striga* is the most serious pest issue, downy mildew (*Sclerospora graminicola*) is the most damaging disease of pearl millet in Africa, with yield loses of ~30% (Ndiaye, 2002). This disease can be dealt with through the planting of resistant varieties, seed treatments (which have been shown to work for ~35 days after planting), and removal of diseased plants (Hess et al., 2002; Thakur et al., 2011; Scheuring et al., 2002; Mbaye, 1992). There have been, and continues to be, great efforts put into breeding pearl millet plants that have increased yields and resistances, but even as the advances are made the yield potential of pearl millet is rarely achieved. This is because of the adverse conditions of planting and the lack of resources needed to supplement poor nutrient conditions in farmer's soils throughout the developing world (Maman et al., 2000; Payne, 1997).

An increased understanding of pearl millet genetics is essential to the development of improved varieties. Pearl millet has historically been classified into four races based off flower and seed characteristics. The races are *typhoides*, *nigritarum*, *globosum*, and *leonis*; however, not all pearl millet accessions perfectly fit into these four classifications (Brunken et al., 1977).

Using genomic relationships could give a better understanding of how to classify and group accessions. Pearl millet has a diploid genome (2n=14) with an estimated size of 1.76 Gb. With the recent development of a reference genome (Varshney et al., 2017), breeders will have a greater capacity to leverage genomic tools to develop pearl millet lines that will be resistant to the major diseases, and to better understand the taxonomic breakdown and genetics of this important plant.

While pearl millet is the most widely grown, and studied millet, there are other millets that have great cultural and regional significances. These other millets also have the potential to diversity our food system. Proso millet (*Panicum miliaceum* L.), little millet (*Panicum sumatrense*), and kodo millet (*Paspalum scrobiculatum* L.) are three such minor millet crops, and there has been relatively little work done with them to develop modern genetic resources (Upadhyaya et al., 2011, Upadhyaya et al., 2014, Kubesova et al., 2010). However, there has recently been an increase in the interest in developing resources for proso millet (e.g., Rajput and Santra, 2016, Habiyaremye et al., 2017, and Yue et al., 2016), and a reference genome was recently developed and released (Zou et al., 2019).

These three millets have been grown for thousands of years, with proso having the longest history. Proso millet was domesticated ~10,000 years ago and has had multiple proposed centers of origin: Northwest China (Bettinger et al., 2007, 2010a, b), Central China (Lu et al., 2009), and Inner Mongolia (Zhao, 2005). The most recent evidence is that there was either one domestication in China or a domestication in China and a separate domestication in Europe (Hunt et al., 2011). There is evidence pointing towards proso millet as the third-oldest cultivated grain (Habiyaremye et al., 2017; Upadhyaya et al., 2011). Proso millet was classified based on inflorescence morphology into five groups: *miliaceum*, *patentissimum*, *contractum*, *compactum,*

and *ovatum* (Prasada Rao et al., 1993). It is an allotetraploid (2n = 4x = 36) (Hunt et al., 2014; Saha et al., 2016).

Kodo millet was domesticated ~3000 years ago in India, and that has remained its most important production region (de Wet et al., 1983). It was classified based on panicle morphology into three races: *regularis*, *irregularis*, and *variabilis* (de Wet et al., 1983), and it is a tetraploid (2n = 4x = 40) (Saha et al., 2016)

Little millet was domesticated ~5,000 years ago in India (de Wet et al., 1983), and its major growing regions have historically been in India, Myanmar, Nepal, and Sri Lanka (Prasada Rao et al., 1993). It was classified into races and sub-races based on panicle morphology: *nana* (sub-races: *laxa* and *erecta*), and *robusta* (sub-races: *laxa* and *compacta*) (de Wet et al., 1983), and it is a tetraploid (2n = 4x = 36) (Saha et al., 2016).

These three millets have been greatly valued by many communities for their capacity to produce in hot and arid environments, and proso millet can even produce yields on only 330 mm of water in as few as 60 days (Dwivedi et al., 2012; Habiyaremye et al., 2017; and Shanahan et al., 1988). Proso millet has been the most widely grown of these three minor millets. It has historically been grown in the former Soviet Union and India (Roshevits, 1980), but has been produced on significant acreage in the United States as well (Baltensperger 2002). In recent years, a company in the Midwestern United States (Dryland Genetics) has been working on developing proso millet by attempting to capitalize on its low water requirements to develop a crop that can be grown in arid regions around the United States. This has the potential to be extremely valuable as water becomes a scarcer resource in the United States with the depletion of underground water reservoirs in the Midwest.

Likely because proso millet has been the most widely grown of these three minor millets it has the most extensive genetic resources, such as a reference genome (Zou et al., 2019). Most of the studies involving genetic markers in proso millet have been with less than 100 markers (reviewed in Habiyaremye et al., 2017), but there has been an extensive study that identified over 400,000 SNPs and 35,000 SSRs from the transcriptome of two proso millet accessions (Yue et al., 2016). A proso millet population-level analysis was conducted that used 100 SSRs and 90 proso millet accessions (Rajput et al. 2016), and it was found that there were some connections between geographic origin and genetic clustering. A QTL mapping study has also been done using 833 SNPs from GBS data (Rajput and Santra, 2016). Kodo millet has had far less genetic research conducted on it than has been done on proso millet. There has been a small set of molecular markers developed (M'Ribu and Hilu 1996), gene-specific primer sets (Kushwaha et al., 2014), and semi-targeted PCR amplification (Yadav et al. 2016). The latter of these studies found evidence that the 96 accessions they studied could be divided into 4 groups that showed little connection between geographic region and genetic relationships. Little millet has been the least studied of these three minor millets. There is one study that looked at specific single genes (Goron and Raizada 2015), and another that used a small set of RAPD markers who details were not described (M. S. Swaminathan Research Foundation 2000).

Developing genome wide genetic data for these three crops would have many uses. It could be used to see how closely related the various accessions of the core collections are. This understanding could be used to plan crosses that would maximize genetic diversity, which can help with breeding efforts. The data could also be used to predict phenotypes of accessions in the core collection, if a proper genomic selection model is built using high quality phenotyping subsets of the core collection. As reference genomes are developed for kodo millet and little

millet, genetic marker sequencing data that has been developed can be used in conjunction with the reference genomes to give an even larger marker data set for researchers to work with.

Millets are a diverse group of crops that play important roles as food sources for millions of peoples around the world, and they serve an especially important role in the developing world. For many years millets have not received the attention they deserve, but they are now coming into their own. As climates continue to change and peoples all over the world strive for a healthier more diverse diet, millets have the potential to play an important role in the worlds changing food systems. An important precursor to millets being able to play a more important role in the worlds food systems is the development of genetic resources. As mentioned previously there has historically been a lack of research on many millets, but in recent years many reference genomes and breeding projects have been started and completed. These resources (especially the reference genomes) have increased researcher's capacities to improve these crops. While there are still many millets that need reference genomes, and all these millets need more work done on them to fully bring them into the modern world, the work being done is beginning to show the potential these crops have.

Industrial Hemp

Hemp (*Cannabis sativa*) is a dioecious (separate male and female plants) annual plant, with some cases of plants living longer than a year under tropical conditions (Cherniak, 1982). It is believed to have been domesticated around 6000 years ago in China (Li, 1974) with some evidence giving a possible date as far back as 12,000 year ago (Fleming & Clarke, 1998; Merlin 2003). Nonetheless, it could be argued that hemp is only semi-domesticated, because it still has issues with seed shattering, which has become a major issue for seed production as the industry

moves towards bringing it to industrial scale (Schluttenhofer and Yuan 2017). Hemp has been widely used for its seed, fiber, and chemical profile for thousands of years (Russo, 2007), and in more recent times it has been used for various building and other industrial material. These materials include a concrete-like material, plastics, pressed woods, and industrial composites. The uses of hemp crop are strongly dependent on the variety grown. There are varieties for fiber, grain, and production of cannabis-specific chemicals. The fiber varieties have been bred for increased fiber production in the outer layers of the plant stems (bast fibers). Both male and female hemp plants are used in fiber production, but male plants are preferred because they produce a higher quality fiber (Hall et al., 2012). This fiber can be used to produce textiles, paper, rope, and many building composites (Amaducci et al., 2015). Female plants can produce grain and high levels of chemicals (mainly in the unpollinated flowers) that are of legal and economic interest. Hemp produces a very nutritious seed with ~30% oil, and ~30% protein (Callaway, 2004), that can be pressed to extract an oil that has a very healthy ratio of omega-6 to omega-3 fatty acids for human consumption (Galasso et al., 2016). Hemp oil has also successfully been used and marketed for cosmetics and fuel (Karus and Vogt 2004). The seed cake that is left after pressing the oil can be used to make protein-rich animal feed.

The chemicals of interest that are produced by *Cannabis sativa* plants --specifically delta-9-tetrahydrocannabinol, or THC--have caused contention in the past 100 years because of its psychoactive properties. This contention eventually led to all forms of *Cannabis sativa* being outlawed in the United States and much of the world, regardless of their THC content (Drug Policy Alliance, 2014). That ban began to change in the 2014 United States federal farm bill, as it allowed limited growing of low-THC ("industrial") hemp, and the 2018 farm bill has brought hemp (defined as any *Cannabis sativa* plant with less than 0.3% THC) almost completely back to

full production. Although hemp is not completely deregulated, thousands of farmers are filling out the needed paperwork with their states' Departments of Agriculture to begin growing it.

In 2018, US hemp acreage increased from ~25,000 acres the previous year to over 75,000 acres (2018 US Hemp Crop report), a 3-fold increase. Most these plants were grown to supply the demands of the US cannabidiol (CBD) market, valued at over $1 billion (The CBD Report 2018 Industry Outlook). CBD has become a popular health supplement that is taken to help with a variety of issues, such as anxiety, pain, depression, lack of appetite, and seizures. Most evidences for CBD's effectiveness in treating these issues are anecdotal, although there are some scientific studies on its efficacy (Perucca, 2017; Glass and Gilleece, 2019; and Hurd et. al., 2019). Recently, CBD has gained traction in the pharmaceutical industry. In late 2018, the FDA approved a drug called epidiolex (which has CBD as an active compound) for treatment of various kinds of seizures that disproportionately affect children, specifically Lennox-Gastaut syndrome and dravet syndrome.

Cannabis plants are often categorized by their CBD to THC ratios, and are generally broken into three types. Type I cannabis plants are THC dominant (they have much higher levels of THC than CBD). Type II cannabis plants produce CBD and THC in roughly a 1:1 ratio. Type III cannabis plants are CBD dominant (they have much higher levels of CBD than THC) (Lewis et al., 2018). More recently accessions have been developed that could be considered a type IV hemp, which is dominant in a molecule called cannabigerol (CBG), but these plants are not as widely available on the market as THC or CBD dominant strains. While THC and CBD are the most well-known of the cannabinoids there are over 100 other known cannabinoids produced in hemp plants (ElSohly and Slade, D. 2006). Cannabinoids are chemicals found in the *Cannabis sativa* plant, and are mainly produced in the trichomes of female flowers, with limited production

in other parts of the plants. *Cannabis sativa's* trichomes are generally classified into four type; capitate sessile, capitate stalked, non-glandular, and bulbous (Andre et al., 2016; Happyana et al., 2013). It has also been shown that these cannabinoids can also be produced in low levels by genetically modified yeast (Luo et al., 2019).

Cannabinoids are known to interact with the human body through cannabinoid receptors CB1 and CB2, and can cause a variety of effects depending on many factors (Matsuda et al., 1990; Munro et al., 1993). One factor which can impact the effect on the human body is that cannabinoids can occur either in an acid or neutral form depending on the presence (acid) or absence (neutral) of a carboxyl group (Taura et al., 2007). Often before being consumed or otherwise applied, cannabinoids are commonly changed from their acid forms to their neutral forms through the process of decarboxylation, which is the removal (usually through heating) of the carboxyl groups from the acids to change THCa and CBDa into THC and CBD respectively. Until THCa is decarboxylated it is not psychoactive. THCa and CBDa are both derived from the same precursor molecule, cannabigerolic acid (CBGa). The enzymes CBDa synthase and THCa synthase convert CBGa into their respective molecules (Sirikantaramas et al., 2004). The expression levels of CBDa synthase and THCa synthase is believed to be an important factor contributing to cannabinoid content and ratios, but the mechanisms controlling the expression of these two genes is not well understood (Laverty et al., 2019). With the levels of THC dictating the legal nature of plants addressing this issue is one of the most important issues for further studies to delve into, and proper utilization of genomic resources will be essential to properly understand the factors influencing such issues.

Hemp has a diploid genome (2n=20), with nine pairs of autosomal chromosomes and either an X/X or X/Y sex determining chromosome pair. The size of the hemp genome is

estimated to be 818 Mb for female plants and 843 Mb for male plants (Sakamoto et al., 1998). Sex determination in hemp is primarily controlled by genetics: female plants have an XX chromosome pair and males have an XY chromosome pair. However, factors beyond the genetics have been shown to play major roles in the control of hemp sex determination; environmental conditions (temperature, drought, photoperiod, etc.), plant hormones (gibberellic acid has been shown to cause female plants to produce male flowers), and ethylene inhibitors (colloidal silver and silver-thiosulfate have been shown to cause male flower formation on female plants) can affect the sexual phenotype of the plants (Hall et al., 2012; Mohan and Sett, 1982; Galoch, 2015; Lubell and Brand, 2018; Nelson, 1944; Eriksson et al., 2000). A further complication of sex determination in hemp is that even though hemp is naturally a dioecious plant, there are cultivars that have been bred to be monecious (hemp plants with both male and female flowers). These cultivars have an XX sex chromosome pair (Faux et al., 2014a) but produce plants that have both male and female flowers on them, with male flowers tending to be produced on the bottom of the inflorescences and female flowers near the top of the inflorescences (Small, 2015). Monecious crops are often preferred for grain production systems to maximize the number of plants that can produce grains while still ensuring that each flower is pollinated. In most cases outside of grain production it is best to have a physical separation between the male and female plants to avoid uncontrolled pollination, which can cause yield losses in cannabinoid production systems. Pollinated plants have been shown to lose ~50% of their potential cannabinoid yield (Meier and Mediavilla, 1998). Since there are no reliable ways to phenotypically distinguish male and female plants before flowering, genetic markers have been identified to assist in distinguishing plant sex while still in the vegetative growth stage (Peil et al. 2003). Markers are not a perfect solution for sex determination because as discussed

previously sex determination is not as simple as knowing the combination of sex chromosome. Plants can have predispositions to exhibit hermaphroditic tendencies, which is what is taken advantage of when selecting plants for monecious cultivars, because of this issue further studies identified QTL linked with sex phenotypes to try and better understand the mechanism of sex determination (Faux et al., 2014b; Faux et al., 2016). A better understanding of sex determination is especially important for the cannabinoid industries.

With pollen having so much potential to cause major losses for hemp farmers focusing on cannabinoid production, many farmer turn either to clones of known female plants, or to feminized seed. Hemp clones are rooted cuttings taken from the branches of the mature vegetative plants. When taking clones special care is taken to leave nodes on the branches for more clones to grow from, allowing for harvesting of additional clones every couple of weeks. Clones give immediate genetic consistency, but are more expensive and require much more space to produce an equivalent number of plants than can be attained from seed production. Feminized seeds are made when genetically female plants are treated with chemicals (often either gibberellic acid or an ethylene inhibitor) or purposefully stressed (rodelization) to induce male flower formation (Lubell and Brand, 2018; Mohan Ram and Jaiswal, 1972). The pollen produced from the female plants using any of these methods should always have an X chromosome, as such when it is used to pollinate a plant (which will be an XX female plant) all the resulting seed will be genetically female. When farmers use feminized seed or female clones from reputable suppliers the only pollen issues they should have to worry about are from hermaphrodites and pollen from other farmers. When talking about high quality feminized seeds, farmers and companies indicate that there should be no more than ~1 male plant out of 4000 plants (personal observation). If hermaphrodites are found in the fields they are rogued out.

Pollen from other farmers' fields can pose more of an issue because there is not an effective way for farmers to protect themselves through any cultural practices. Hemp pollen has been shown to be able to travel from fields over 100 kilometers away (Carinanos et al., 2004; Cabezudo et al., 1997). It is important to remember that pollen is only an issue for farmers who are growing hemp for cannabinoid production. Other farmers who are growing from grain or fiber will either not incur any losses from pollen (as is the case for fiber production) or depend on pollen being available (as is the case for grain production). For this issue to be dealt with hemp growers and regulatory agencies need to work together to better understand needed spatial and temporal separations to ensure all farmers can produce high quality product.

Once the plant is grown, it must be processed into a useable form. The simplest processing is when people choose to smoke the flowers. The flowers are taken from the plants, have the leaves trimmed off, and are dry cured. For the other uses of hemp much more must go into creating a usable product from the raw plant. When working with fiber plants, the processor must separate the outer part of the plant (the bast fiber) from the inner hurd (the woody part of the plant) and separate and clean off the excess plant material clinging to the fibers. The fibers can then go through a "cottonization" process which allows them to be interchangeable with cotton in textile factories (Khoathane et al., 2008). Grain can be harvested using standard combines, but special care needs to be taken to stop the hemp stalks and fibers from damaging the machines (Gusovius et al., 2016). Processing hemp plants for cannabinoids requires special equipment. There is no industry standard for plants grown for cannabinoid extraction, but it is common for the plants to be dried down and have the flowers taken from them. The flowers are then shipped to processing facilities that commonly used either ethanol or supercritical CO_2 extraction. These methods extract the organic compounds produced in plants. To further isolate

specific chemicals (such as CBD), chromatography and fractional distillation are used (Mechoulam and Hanus, 2000).

One of the major reasons that clones still play an important role in hemp industry, despite the availability of cheaper feminized seeds, is that there is a lack of genetically stable seed available to the farmers. When hemp began to be outlawed in the 1930's modern breeding was rapidly advancing many of the world's major crops, and because of the stigmatization and legal status of hemp it missed out on decades of research and development. A prominent hemp breeder (Dewey,1927) from the early 1900s was known to have created many valuable varieties for production in the United States. However, with the laws outlawing hemp, germplasm were destroyed and the varieties developed were lost (Small and Marcus. 2003). This setback has left the modern hemp farmers with less than ideal genetics to work with, but there are many organizations rushing to fix this issue. The USDA (2019) recently announced the development of a germplasm bank that will be based in Geneva, New York. This germplasm will serve as a vital resource for the genetic improvement of hemp. An important starting place for building a national germplasm collection will be working with other nations that devoted resources to maintaining or building their hemp germplasm over the past century. Russia's Vavilov Institute has maintained a large collection of over 500 accessions (Ranalli, 2004). Germany's Institute of Plant Genetics and Crop Plant Research Gatersleben (IKP) has built up a collection of ~55 accessions (Graner, 2017). There are also small collections in Turkey, Hungary, Italy, and Japan (Ranalli, 2004). It is estimated that there are ~1000 accessions of *Cannabis sativa* plants stored in various gene banks around the world, which is a small number compared to the hundreds of thousands for other crops, such as maize (Campbell et al., 2019).

Most of the breeding work in recent decades has been undertaken in Europe, Canada, and China, and it has focused on grain and fiber varieties (Ranalli, 2004). Today hemp research and breeding focuses have slightly expanded to now include three major targets: fiber, grain, and cannabinoids. Fiber breeding has traditionally been the major breeding focus, because it was historically the most commonly used material from the plants. Breeding efforts in the 20th century relied on collecting landraces from around Europe (especially Italy, Hungary, and Romania) and selecting for consistent photoperiodism, and height (Ranalli, 2004; Toth et al., 2020). Carmagnola is a landrace from Italy that served as the starting material from some of the most successful lines on the 20th century (Allavena, 1967). In the early days of hemp breeding selecting for fiber quality was extremely difficult because of the destructive nature of the test, but starting in the 1920's that changed. An *in vivo* method for testing fiber quality was developed that allowed breeders to test fiber quality and still make selection from that plant by working with small fiber samples taken from the bast fiber of the plants (Bredemann, 1924). The breeding process relied on carefully splitting plants open before flowering and testing their fiber content, and after that only letting the plants with the greatest fiber quality flower. This technique resulted in a three-fold increase in fiber production over a thirty-year period of research (Bredemann et al., 1961).

As mentioned earlier, breeding for grain has been focused on selecting for monecious plants (Schluttenhofer and Yuan, 2017). Hemp grain breeding is a relatively recent development, with the first plant bred purely for grain production being released in the 1990's (Ranalli, 2004). Selecting for grain size and overall yield can result in improved varieties, but the most important issue to solve for hemp grain production is seed shattering. Solving this issue will result in significant yield increases in a relatively short period of time. In other crops, such as rice, it has

been found that shattering was mostly solved long ago, and it was the result of selecting for plants with shattering genes inactivated (Yao et al., 2015). In hemp, this problem could potentially be addressed with mutagenesis to knock out genes (Sikora et al., 2011), followed by screening a large population of plants to identify any that have reduced or eliminated shattering. However, it is important to note that there is still no evidence that a single gene inactivation would have the same result in hemp. Also, direct gene editing does not yet have a fully developed method for hemp; full plant regeneration from callus tissues is the current bottleneck.

Breeding hemp for cannabinoids should have three major goals: increased total CBD, increased CBD to THC ratio, and increased flower dry weight yield. Since cannabinoids are measured in unpollinated female flowers, it can be a difficult trait to measure and select for within the same generation. One way to deal with this issue for photoperiod-sensitive plants is taking cuttings and keeping them in the vegetative state while the original mother plants flower. After taking the data from the mother plants, a breeder can go back to the cutting stocks and make crosses from those that were from the best performing plants. This technique has two major weaknesses: it cannot be used on non-photoperiod dependent (known as autoflowering) plants, and it still does not allow a breeder to take advantage of the genetic potential of cannabinoid production that comes from the male plants. To work with autoflowering plants and male plants, pedigrees or genomic selection could be used (Meuwissen et al., 2001). Another issue with breeding for cannabinoid content is the cost, time, and labor associated with measuring cannabinoid content. High-performance liquid chromatography (HPLC) and gas chromatography (GC) are two of the most common approaches used to measure cannabinoid content (De Backer et al., 2009). With the cost and labor limitation involved with using chromatography to phenotype cannabinoid content there is great value in exploring other options

for phenotyping. Near infrared spectroscopy (NIRS) has been investigated as a potential alternative and has been shown to have great potential in measuring the levels of a multitude of cannabinoids in hemp samples, including CBD and THC (Sanchez-Carnerero Callado et al., 2018). The proper application of this technology could give researchers, farmers, and law enforcement the tools needed to effectively and efficiently do their jobs. As the technology is further refined, a great potential application would be for farmers to be able to use drones paired with some other form of hyperspectral analysis to fly over their fields and monitor the crops, so harvest can occur at the optimal time where CBD is the highest while THC remains below or at the legal limit. It still must be further explored to see if the drones could effectively analyze such a small amount of THC in the plants (with the current federal limit being set a 0.3% total THC). As these tools and methods are better refined it can be expected that hemp breeders will be able to more rapidly develop this crop and work overcome the decades of its neglect. An increased understand of hemp genetics will be essential for bringing this crop into the modern world.

While *Cannabis sativa* genetics and genomic studies and resources are obviously not at the level of the world's major crops, there have been many great discoveries made and resources developed. The focuses of many *Cannabis* genetic studies have been on better understanding the genetic controls of cannabinoid production (Taura and Morimoto, 1995; de Meijer, 2003; Weiblen et al., 2015; de Meijer et al., 2009), terpene production (Booth et al., 2017), sex expression (Faux et al., 2016), fiber quality (van den Broeck et al., 2008), grain nutrition (Bielecka et al., 2014), the development reference and draft genomes (van Bakel et al., 2011; Laverty et al., 2019), and genetic diversity and population structure (Sawler et al., 2015; Lynch et al., 2016; Dufresnes et al., 2017). The studies on genetic diversity and population structure are especially interesting because of the issues they address. There has been a debate about the

speciation and types of *Cannabis sativa*, that has mainly been driven by phenotypic diversity (Campbell et al., 2019), with some people attempting to break *Cannabis sativa* into two additional species (*Cannabis indica* and *Cannabis ruderalis*). Population genetics analyses are a good way to better understand the issues that underline such debates. Sawler et al. (2015) used genotyping-by-sequencing (GBS) to study 81 marijuana samples and 43 hemp samples. The study used 14,031 SNPs to show that hemp and marijuana are genetically distinct, that there were some correlations between genetic structure of the samples and their reported *indica* or *sativa* ancestry, and that the names of the plants do not correlate to genetic relatedness.

An expanded population genetics study was conducted by Lynch et al. (2016) which used 340 *Cannabis* accessions and varieties. The researchers used fiber, grain, high cannabinoid drug-type, and feral plants to try and better understand how these types of hemp relate to each other and what groups could be identified within them. Within the high cannabinoid drug type *Cannabis* plants the researcher looked at two main groups: narrow leaflet drug-types (NLDT) and broad leaflet drug-types (BLDT), being distinguished by the size of the leaflets. It was found that hemp (grain and fiber plants), NLDT, and BLDT formed genetically distinct groups. They concluded that BLDT was found to have less genetic diversity than NLDR. However, the fact that they used less than half as many accessions of the BLDT compared to the NLDT could have also contributed to this observed lack of diversity. Dufresnes et al. (2017) used 13 simple-sequence-repeat (SSR) markers to analyze 1324 samples from 24 hemp accessions and 15 marijuana strains. They found that marijuana and hemp could be distinguished by their markers, and that an unknown strain could be classified based on the genetic markers. They concluded that using more than 13 markers would be essential to have a functional accuracy when attempting to distinguish strains for industry or law enforcement uses. All three of these

population genetic studies (Sawler et al., 2015; Lynch et al., 2016; Dufresnes et al., 2017) concluded that hemp and marijuana clustered into distinct groups based on their genetic markers, but the relationships between hemp, and the subgroups of marijuana (*indica* and *sativa*) were not as clear. From these studies, it could not be concluded if either of the groups showed a closer relationship to hemp. Since hemp is a legal definition, any *Cannabis* plant with proper selection and crosses can be made into a hemp plant, so trying to understand the relationships between hemp, *indica*, and *sativa* plants is highly dependent on the specific genetic heritage of the hemp accessions being studied and a somewhat arbitrary idea of what differentiates *sativa* and *indica* to begin with. Understanding genetic structures of various populations and accessions is an important step in being able to fully utilize the gene pools of a species of interest, but many other major milestones must be reach as well.

A major milestone for any crop is the development and release of a reference genome. For *Cannabis sativa,* there was a draft genome published in 2011, using a drug type strain (high THC levels) called Purple Kush. The draft genome was generated using Illumina sequencing and resulted in 136,290 scaffolds and an N50 of 16.2 kb (van Bakel et al., 2011). There is now a high-quality reference genome published, which used PacBio and Illumina sequencing that sequenced a female and male plant. The Pink Kush strain was used as the female for this study, and the grain variety Finola was used as the male for this study (Laverty et al., 2019). In the Finola sequencing there were 5,303 scaffolds, an N50 of 445.6 kb and it was sequenced to a depth of 98x. Pink Kush had 12,836 scaffolds, an N50 of 146 kb, and it was sequenced to a depth of 79x. PacBio sequencing was used in the reference genome to overcome the highly repetitive nature of the genome, which is likely made up of ~70% repetitive elements (Pisupati et al.,

2018). Having a high-quality reference genome allows researchers to better utilize modern genetic and genomic tools to improve plants.

Cannabis sativa is a plant that has garnered great interest for thousands of years because of its many uses. Today interest in cannabinoids is largely driving the interest in hemp and marijuana. As research and market demands continue to move this industry forward we can expect to see products from hemp becoming readily available to more and more consumers.

Genomic Tools in Plant Breeding

Over the past few decades, the world has rapidly moved into the information age. Generating data has become easier and cheaper, and this has changed every field of research, including genomics. In the early 2000's, sequencing the human genome cost $2.7 billion and in 2018 it cost about $1,500 dollars, this is 1.8 million times cheaper in a 15-year period (Cheifet, 2019). This rapid cost reduction in sequencing has not just changed the way humans are studied, it has had a large impact on the way we study plants. In a world where we need to produce more food on less land, the development of genomic tools has proven to be a valuable resource in increasing the worlds food security (Perez-de-Castro et al., 2012, Abberton et al., 2016). In 2005, the rice genome was sequenced using bacterial artificial chromosomes (BAC) and sanger sequencing (Sasaki, 2005). By 2007 the grape genome was released and this was done at a much lower cost, by using 454 sequencing and sanger sequencing (Velasco et al., 2007). Just two years later another improved method was implemented, in 2009 the cucumber genome was assembled using Illumina sequencing and sanger sequencing (Huang et al., 2009). These examples show a rapid change in the technologies and techniques used to sequence genomes in a short period of

time. The adoption of next generation sequencing technologies is what has given us over 260 publically available plant genomes today (Hu et al., 2018).

As implied in the name, reference genomes are tools researchers reference to improve their capacity to work with specific individuals or populations of a species that have a reference genome. Reference genomes can be important tools for a variety of tasks such as identifying genetic variations and discovering markers associated with specific traits.

Sections of DNA and individual base pairs can be classified into three general groups: repeats at a locus [microsatellites and simple sequence repeats (SSR)], insertions or deletions of segments (InDels), and single nucleotide polymorphisms (SNPs) (Weber and May, 1989; Ophir and Graur, 1997; David et al., 1998). These are the three major variant types, but there have been many molecular markers developed to identify the polymorphisms that are caused by these three (Gupta et al., 1999). Cost and reproducibility are the main driving forces behind choosing which molecular marker to use (Bernardo, 2008). As technology has changed, the markers of choice have as well. In the 1980s restriction fragment length polymorphisms (RFLPs) were commonly used in plant molecular genetics because of their reproducibility, but they fell out of favor because they were expensive and labor intensive (Lander and Botstein, 1989). In the last ten years SNPs have become the most commonly used option, which is due in part to the low cost of discovery, but also because SNPs are the most common form of genetic variation between individuals of the same species (Rafalski, 2002).

When building a SNP marker data set, individual samples will go through some level of sequencing, and those sequences will often be aligned to a reference genome and using a pipeline, such as TASSLE's GBSv2 pipeline (Bradbury et al., 2007). The differences between the reference genome and the reads from individuals will be analyzed to identify SNPs to be used

in a marker data set. However, a reference genome is not required to generate a SNP marker data set, but having one increases the number of markers identified (Lu et al., 2013). It is important to note that bioinformatics is an essential component of being able to process large genomic data and gain useful insights (Moore et al., 2010; Batley and Edwards, 2016)

A common sequencing method used for discovering SNPs is genotyping by sequencing (GBS) (Elshire et al., 2011). GBS uses at least one, but often many, restriction enzyme to target specific locations along a genome for sequencing. Restriction enzymes recognize specific DNA base pair combinations as cut sites and will process the genome into smaller segments. This type of processing is cheap and yields highly reproducible results (Elshire et al., 2011). To simplify alignment problems specialized restriction enzyme combinations are chosen, specifically methylation sensitive restriction enzymes. Proper enzyme combinations can help avoid regions of genomes that are highly repetitive. This allows low copy regions to be more efficiently targeted (Gore et al., 2009; Gore et al., 2007). Three issues that should be kept in mind when working with GBS data sets are: missing data, sequencing errors, and paralogous sequences (Wallace and Mitchell, 2017). Missing data can result from the random nature of sequencing the DNA fragments. One way to deal with missing data is to increase the sequencing depth, and another is to filter out sites that are missing across many samples. Imputation, filling in missing data by making inferences from the available data, can also be used (Alipour et al., 2019). None of the imputation methods are perfect, but they have been shown to be quite useful (Chan et al., 2016). Imputation can be especially useful in association studies where missing data can make it more difficult to deduce proper associations (Chatterjee et al., 2009). It is best to use imputation after all other filtering has been done. If imputation is done prior to removing bad data, the imputation results will likely not be biologically valid, because the results will be based on

incorrect data. Sequencing errors are relatively rare and can be filtered out by removing rare alleles, often measured as minor allele frequency (MAF). An issue with this is that filtering out rare alleles can also remove true SNPs, especially when working with a diverse population. The risks of removing valid data need to be weighed against the risk of keeping false data. Paralogous sequences are those that have been duplicated at some point in evolutionary history in a plants genome, and because of ancient ploidy events and transposable elements they are quite common in many plant species (Vicient and Casacuberta, 2017; Gao et al., 2018; Panchy et al., 2016). Paralogs can cause issues for plant genetic studies, especially with DNA alignment (Choi et al., 2007). The simplest way to deal with these issues in GBS data is to look at sequencing depth and rates of heterozygosity. Paralogs tend to have higher sequencing depth and higher rates of heterozygosity than true SNPs (Wallace and Mitchell, 2017). All the previously mentioned steps work towards building up a genetic marker data set, but that is just a beginning of most experiments. The purpose of building a genetic marker set is using that data to better understand an individual, population, or species.

SNP markers can be used to better understand how individuals are genetically related to each other (Zheng et al., 2019). When using SNPs data to understand genetic relationships there are several important concepts to understand. I will focus on a select number of them.

Principle coordinates analysis (PCoA) [also known as multidimensional scaling (MDS)] and principle components analyses (PCA) are ways of visualizing similarities and dissimilarities of data, and have been shown to be very useful when working to visualize genetic relatedness (Rohlf, 2000; Peakall and Smouse, 2006). It is often better to use PCoA over PCA because PCoA can better handle missing data. When PCoA results are plotted, and paired with other data (such as geographic origin, subspecies designation, or race designation) it can be seen

if the other data is correlated with the genetic relationships among the individuals. This is done by seeing if assigned groups are clustered together, this would indicate the genetics play an important role in the designations. To validate the results from PCoA it can be helpful to also use SNP data to generate a **phylogenetic tree**. Phylogenetic trees should show the same relationships seen in PCA and PCoA, but generated in a different manner (Yang et al., 2018). If the results show the same patterns then that can work as a piece of evidence to show that they are both true. When attempting to interpret genomic data and the results from different analyses it is important to remember that what is shown is not always a true representation of biology and should be validated.

Population structure can cause issues in genetic studies and must be properly understood and accounted for to attain the best results (Sul et al., 2018). Both principle coordinate analyses and phylogenetic trees can play important roles in understanding population structure (Luo et al., 2019). The major issue that unaccounted population structure can cause is false associations, where the results of interest are confounded by the structure of the populations being studied. It is common to account for population structure in plant breeding studies to avoid such issues (Eltaher et al., 2018).

Linkage disequilibrium is the non-random association of alleles at two or more loci, and is affected by genetic recombination, physical distance, and population structure. Genetic recombination impact derives from the fact that the closer alleles of interest are to each other the more likely they are to be linked. This is because the only way to break the linkage is for recombination to happen between the alleles of interest, and the closer together they are the less likely that is to happen. Physical distance between alleles is closely tied with genetic recombination because physical distance is one of the major factors that affects how likely

genetic recombination is to occur between two loci. Population Structure can affect linkage disequilibrium because individuals in a population can have a greater chance of mating within their given population. Alleles that are not physically linked will still have a greater chance of being inherited together because they might be fixed within a given population. This goes back to the previously mentioned idea of accounting for population structure to avoid false associations.

Logarithm of odds (LOD) scores are a common statistical measurement used in conjunction with linkage disequilibrium analyses. A LOD Score is used to try and determine if two loci are linked. It works by taking the observable rates of recombination and comparing that to possible rate of recombination if there was no linkage. A LOD score of 3 (which has traditionally been used to determine significant association) would mean that the probability of linkage occurring in this instance is 1000 times greater than there being no linkage (Nyholt, 2000).

Genome wide marker data is extremely useful in plant genetics when it is paired with high quality phenotypic data. This pairing allows researchers to leverage genomic data to rapidly improve plants. The two areas I will focus on are genomic selection and genome wide association studies (GWAS).

Genomic selection is an analysis that uses markers, often at least a couple hundred of them, to select for traits of interest without the need to phenotype every generation (Voss-Fels et al., 2018). To use genomic selection a training population is grown out to be genotyped and phenotyped. From that a model is developed that predicts the phenotype from the genotype. That prediction is known as the genomic estimated breeding values (GEBV) (Lorenz et al., 2011). With the cost of phenotyping being much higher than the cost of genotyping for many traits, the

proper use of the GEBVs can save money and time by reducing the need to phenotype every generation and by not being restricted to specific locations or stresses to make selections (outside of the training population) (Heffner et al., 2009; Bernardo, 2010). There are many factors to consider (crop, population design, trait of interest, etc.) when choosing the statistical method used to develop the genomic selection model, the model chosen can have a statistically significant impact on the prediction accuracy (Daetwyler et al., 2013; Wang et al., 2019), although often the differences of accuracy between the models are quite small. A selection of commonly used genomic selection models are: RR-BLUP, GBLUP, Elastic net, and Random forest. In the ridge-regression best linear unbiased predictor (RR-BLUP) method, all loci are assumed to have a single effect variance in common between them (Endelman, 2011). This method uses a matrix of relatedness generated from genetic markers, and allows markers to have an unequal effect (Heffner et al., 2011). The genomic best linear unbiased predictor (GBLUP) uses a genomic relationship matrix (GRM), which has been shown to be highly accurate when used in predicting GEBVs (VanRaden, 2008). GBLUP can considered to be equivalent to RR-BLUP in most practical circumstances (Habier et al., 2007). Elastic net (EN) has shown to be a useful method when the number of predictors is far greater than the number of observations, and has been shown capable of dealing with many of the challenges involved in working with high dimensional data (Zou and Trevor, 2005). Random forest uses a regression model that relies on bootstrapping (random sampling with replacement) sample observations, building a "tree" of prediction models to develop the highest accuracy model, and can capture marker interactions in its model (Rutkoski et al., 2011; Jannink et al., 2010; Holliday et al., 2012). There are many more genomic selection models (Desta and Ortiz, 2014), which can make choosing a model difficult. When investing genomic selection's potential for a new crop and trait GBLUP has been

shown to be robust and works well with traits that have an unknown number of loci affecting them (Daetwyler et al., 2010). Beyond simply picking a model, it is important to determine if genomic selection is an economically advantageous option over traditional breeding. The main factors that must be considered are: the cost of genotyping and phenotyping, time to market, and differences in genetic gain per year over other methods (Lin et al., 2016). Once these factors are considered, an informed decision can be made to implement genomic selection or go with a more traditional breeding approach.

Quantitative trait loci (QTL) Mapping, or linkage mapping, is the association of a genomic region with a trait of interest. The region could even be as large as an entire chromosome. This has a lot more power than association mapping, but it lacks the finer resolution of association mapping. The resolution depends heavily on the amount of recombination that has happened in a population. For QTL mapping it often uses a population which is made when two parents are crossed and then the offspring are crossed among themselves. It is the limited number of generations (limiting recombination), which inhibits QTL mapping resolution (Borevitz and Nordborg, 2003). This can be improved by adding more generation of crosses when establishing the population for analysis (Balasubramanian, et al., 2009). Another limitation is that only traits that segregate between the original parents can be analyzed by this method. This can be dealt with, to a limited extent, by crossing multiple diverse accessions when developing the population to be studied (Kover et al., 2009).

Genome-wide association study (GWAS) is another method used in plant breeding that relies on genetic marker data and phenotypic data from the same plants. It is a type of association mapping that aims to discover significant associations between genetic markers and a trait of interest that has been measured in a large and genetically diverse population. It was originally

developed for applications in human genomic studies shortly after the publication of the human genome (Hirschhorn and Daly, 2005). GWAS are now commonly applied to many species for plant breeding research as well (Huang et al., 2012; Ranc et al., 2012; Zatybekov et al., 2017). The effectiveness of GWAS relies on the principles of linkage disequilibrium. This type of mapping operates under the assumption that in a population a marker tightly linked with a trait of interest in the past will remain tightly linked with that trait throughout a population, even after many generations. An advantage of GWAS over QTL mapping is its high resolution, meaning that it can narrow in on a small region associated with the traits of interest, but it has relatively little power. Having little power means that because most alleles are rare and since they are not always present it can be difficult to find strong association with those rare alleles and the trait of interest. This type of mapping is done in a general population instead of an F2 or backcrossing population, which is often done for QTL mapping. GWAS is preferable to QTL mapping for exploratory studies that are looking for useful genes to introgress into breeding lines (Korte and Farlow, 2013). Like any modern genomic analysis GWAS relies heavily on bioinformatics. GWAS and genomic selection are two tools that plant breeders have used in recent year, and recently it was even found the GWAS results can be used in conjunction with GS to improve accuracy of GS (Spindel et al., 2016).

The applications of genomics in plant breeding have revolutionized modern breeding programs, and have played an important role in the development of plants that can produce enough food to feed the world's continually growing demand. As resources continue to be developed and applied to new crops we will see the rapid development of many crops that had been previously ignored, and the further improved of crops that have already be pushed so far over the decades.

REFERENCES

"The CBD report 2018 Industry Outlook" Retrieved from https://newfrontierdata.com/product/cbdreport2018/

Abberton, M., Batley, J., Bentley, A., Bryant, J., Cai, H., Cockram, J., de Oliveira, A.C., Cseke, L.J., Dempewolf, H., De Pace, C., et al. 2016. Global agricultural intensification during climate change: A role for genomics. *Plant Biotechnol*. J.,14, 1095–1098

Alipour, H., Bai, G., Zhang, G., Bihamta, M. R., Mohammadi, V., & Peyghambari, S. A. 2019. Imputation accuracy of wheat genotyping-by-sequencing (GBS) data using barley and wheat genome references. PloS one, 14(1), e0208614. doi:10.1371/journal.pone.0208614

Allavena, D., 1967. CS, eine neue Sorte des zweih¨ausigen Hanfes.Fibra 12: 17–24.

Amaducci, S., Scordia, D., Liu, F. H., Zhang, Q., Guo, H., Testa, G., and Cosentino, S. L. 2015. Key cultivation techniques for hemp in Europe and China. Industrial Crops and Products, 68, 2–16.

Andre, C. M., Hausman, J.-F., and Guerriero, G. 2016. Cannabis sativa: The Plant of the Thousand and One Molecules. FRONTIERS IN PLANT SCIENCE, 7.

Andrews, D. J. and Brammel-Cox, P. J. 1993. Breeding cultivars for sustainable crop production in low input dryland agriculture in the tropics. In International Crop Science, 211–23 (Eds D. R. Burton, R. Shibles, R. A. Forsberg, B. L. Blad, K. H. Asay, G. M. Paulsen and R. F. Wilson). Madison, Wisconsin: Crop Science Society of America.

Balasubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Min Chul Kim, Maloof, J. N., et al., 2009. QTL Mapping in New Arabidopsis thaliana Advanced Intercross-Recombinant

Inbred Lines. PLoS ONE, 4(2), 1–8. https://doi-org.proxy-remote.galib.uga.edu/10.1371/journal.pone.0004318

Baltensperger, D.D. 2002. Progress with Proso, Pearl and Other Millets. Trends in new crops and new uses.

Batley, J. and Edwards, D. 2016. The application of genomics and bioinformatics to accelerate crop improvement in a changing climate. Curr. Opin. Plant Biol.30, 78–81

Bennetzen, J.L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, et al. 2012. Reference genome sequence of the model plant Setaria. Nat Biotechnol 30: 555-561. doi:10.1038/nbt.2196.

Bernardo R. 2010. Genomewide selection with minimal crossing in self-pollinated crops. Crop Sci 50: 624–627.

Bernardo, R. 2008. Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. CROP SCIENCE -MADISON-, (5), 1649

Bettinger, R.L., L. Barton and C. Morgan. 2010a. The origins of food production in north China: A different kind of agricultural revolution. Evolutionary Anthropology: Issues, News, and Reviews 19: 9-21.

Bettinger, R.L., L. Barton, C. Morgan, F. Chen, H. Wang, T.P. Guilderson, et al. 2010b. The transition to agriculture at Dadiwan, People's Republic of China. Current Anthropology 51: 703-714.

Bettinger, R.L., L. Barton, P.J. Richerson, R. Boyd, H. Wang and W. Choi. 2007. The transition to agriculture in northwestern China. Developments in Quaternary Sciences 9: 83-101.

Bielecka, M., F. Kaminski, I. Adams, H. Poulson, R. Sloan, Y. Li, T.R. Larson, T. Winzer, and I.A. Graham. 2014. Targeted mutation of D12 and D15 desaturase genes in hemp

produce major alterations in seed fatty acid composition including a high oleic hemp oil. Plant Biotechnol. J. 12(5):613–623. doi:10.1111/pbi.12167

Bilquez, A. F. 1963. Etude due mode d'hérédité de la précocité chez le mil pénicillaire (*Pennisetum typhoides* Stapf et Hubbard): I. Déterminisme génétique des diffénces de sensibilité a la longueur du jour existant entre les mils degroupe sanio et ceux du groupe souna (Study of *Pennisetum typhoides* Stapf et Hubbard): Genetic determination of day length sensitivity differences existing among millet in the sanio and souna groups). Agronomie Tropical18:1249–1253.

Boivin, N., and Fuller, D. 2009. Shell middens, ships and seeds: Exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula Journal of World Prehistory, 22(2), 113–180.

Booth, Judith K., Jonathan E. Page, and Jörg Bohlmann. 2017. Terpene synthases from Cannabis sativa. PloS one 12(3): E0173911. doi:10.1371/journal. pone.0173911

Borevitz, J.O. and Nordborg, M. 2003. The Impact of Genomics on the Study of Natural Variation in Arabidopsis. Plant Physiology, 132(2), 718.

Bourlag, N. (Ed.). 1996. Lost crops of Africa. Volume I. Grains.Washington: The National Academies Press.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Bredemann, G., 1924. Beitr¨age zur Hanfz¨uchtung II. Anslese faser-reicher. M¨annchen zur Befruchtung durch Faser-bestimmung ander lebeden Pflanze vor deu Bl¨ute. Angew Bot 6: 348–360.

Bredemann, G., K. Garber, W. Huhnke and R. von Sengbusch,1961. Die Z¨uchtung von mon¨ozischen und di¨ozischen, faser-ertragreichen Hanfsorten Fibrimon und Fibridia. Zeitschrift f¨urPflanzenz¨uchtung 46: 235–245.

Brunken, J., De Wet, J., & Harlan, J. 1977. The Morphology and Domestication of Pearl Millet. Economic Botany, 31(2), 163-174.

Cabezudo, B., Recio, M., Sanchez-Laulhe, J. M., Del Mar Trigo, M., Toro, F. J., and Polvorinos, F. 1997. Atmospheric Transportation of Marihuana Pollen from North Africa to the Southwest of Europe. ATMOSPHERIC ENVIRONMENT, (20), 3323.

Callaway, J. C. 2004. Hempseed as a nutritional resource: An overview. EUPHYTICA, (NO 1-2), 65.

Campbell, B., Zhang, D., McKay, J. K. 2019. Chapter 6: Hemp Genetics and Genomics. In: D.W. Williams, editor, Industrial Hemp as a Modern Commodity Crop, ASA CSSA and SSSA, Madison, WI. p. 94-108. doi:10.2134/industrialhemp.c6

Carinanos, P., Galan, C., Alcazar, P., and Dominguez, E. 2004. Analysis of the particles transported with dust-clouds reaching Cordoba, southwestern Spain. Archives of Environmental Contamination and Toxicology, 46(2), 141–146.

Carsky, R. J., Singh, I. and Ndikawa, R. 1994. Suppression of *Striga hermonthica* of sorghum using cowpea intercrop. Experimental Agriculture30:130–132.

Carson, A. G. 1988. Effect of intercropping sorghum and groundnuts on density of *Striga hermothicain* the Gambia. Tropical Pest Management34:97–101.

Chan, A.W., Hamblin, M.T., & Jannink, J.L. 2016. Evaluating Imputation Algorithms for Low-Depth Genotyping-By-Sequencing (GBS) Data. PLoS ONE, 11(8), 1–17. https://doi.org/10.1371/journal.pone.0160733

Chatterjee, N., Yi-Hau, C., Sheng L., & Raymond J.C. 2009. Analysis of Case-Control Association Studies: SNPs, Imputation and Haplotypes. Statistical Science, 24(4), 489.

Cheifet, B. 2019. Where is genomics going next?. Genome Biol 20, 17 doi:10.1186/s13059-019-1626-2

Cherniak, L. 1982. The Great Books of Cannabis vol. I, Book II. Damele Publishing, Oakland, CA.

Choi, I.-Y., Hyten, D. L., Matukumalli, L. K., Song, Q., Chaky, J. M., Quigley, C. V., et al., 2007. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics, 176(1), 685–696.

Daetwyler H.D., Calus M.P., Pong-Wong R., de Los Campos G., Hickey J.M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193: 347–365.

Daetwyler, H. D., Pong-Wong, R., Villanueva, B., & Woolliams, J. A. 2010. The impact of genetic architecture on genome-wide evaluation methods. Genetics, 185(3), 1021–1031. https://doi-org.proxy-remote.galib.uga.edu/10.1534/genetics.110.116855

David G. Wang, Jian-Bing Fan, Chia-Jen Siao, Anthony Berno, Peter Young, Ron Sapolsky, … Eric S. Lander.1998. Large-Scale Identification, Mapping, and Genotyping of Single-Nucleotide Polymorphisms in the Human Genome. Science, 280(5366), 1077.

De Backer, B., Debrus, B., Lebrun, P., Theunis, L., Dubois, N., Decock, L., … Charlier, C. 2009. Innovative development and validation of an HPLC/DAD method for the qualitative and quantitative determination of major cannabinoids in cannabis plant material. Journal of Chromatography B, 877(32), 4115–4124.

de Meijer, E.P.M., K.M. Hammond, and A. Sutton. 2009. The inheritance of chemical phenotype

in Cannabis sativa L.(IV): Cannabinoid-free plants. Euphytica 168(1):95–112. doi:10.1007/s10681-009-9894-7

de Meijer, E.P.M., M. Bagatta, A. Carboni, P. Crucitti, V.M.C. Moliterni, P. Ranalli, and G. Mandolino. 2003. The inheritance of chemical phenotype in Cannabis sativa L. Genetics 163(1):335–346.

de Wet, J.M.J., D.E. Brink, K.E.P. Rao and M.H. Mengesha. 1983. Diversity in Kodo millet, Paspalum scrobiculatum. Economic Botany 37: 159-163. doi:10.1007/bf02858779.

Desta, Z. A., and Ortiz, R. 2014. Genomic selection: genome-wide prediction in plant improvement. Trends in Plant Science, 19(9), 592–601. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.tplants.2014.05.006

Dewey, L.H., 1927. Hemp varieties of improved type are result of selection. Yearbook of the Department of Agriculture, U.S.A.:358–361.

Drug Policy Alliance, The DEA: Four Decades of Impeding and Rejecting Science, Multidisciplinary Associations for Psychedelic Studies, 2014

Dufresnes, C., C. Jan, F. Bienert, J. Goudet, and L. Fumagalli. 2017. Broad-scale genetic diversity of Cannabis for forensic applications. PloS one 12(1): E0170522. doi:10.1371/journal.pone.0170522

Dwivedi, S., H. Upadhyaya, S. Senthilvel, C. Hash, K. Fukunaga, X. Diao, et al. 2012. Millets: Genetic and genomic resources. Plant Breed. Rev. 35:247–375. doi:10.1002/9781118100509.ch5

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6: e19379.

ElSohly, M. A., and Slade, D. 2006. Chemical constituents of marijuana: The complex mixture of natural cannabinoids. LIFE SCIENCES, (5), 539.

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. 2018. Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Front. Genet. 9:76. doi: 10.3389/fgene.2018.00076

Endelman J.B., 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome.4(3):250–255. doi: 10.3835/plantgenome2011.08.0024.

Eriksson, M. E., Israelsson, M., Olsson, O., and Moritz, T. 2000. Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nature Biotechnology, 18(7), 784.

FAO. 2018. FAOSTAT data base. Available at http://faostat.fao.org/site/5676/DesktopDefault.aspx.

Faux, A.-M., Berhin, A., Dauguet, N., and Bertin, P. 2014a. Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica, 196(2), 183–197.

Faux, A.-M., Draye, X., Flamand, M.-C., Occre, A., and Bertin, P. 2016. Identification of QTLs for sex expression in dioecious and monoecious hemp (Cannabis sativa L.). Euphytica, 209(2), 357–376.

Faux, A., Bertin, P., and Piepho, H. -P. 2014b. Modelling approach for the quantitative variation of sex expression in monoecious hemp (Cannabis sativa L.). Plant Breeding, 133(6), 782–787.

Fleming, M. P., Clarke, R. C., Physical evidence for the antiquity of *Cannabis sativa*. L. J. Int. Hemp Assoc. 1998, 5, 280.v

Fuller, D. Q. 2003. African crops in prehistoric South Asia: A critical review. In K. Neumann, A. Butler, and S. Kahlheber (Eds.), Food, fuel, and fields: Progress in African archaeobotany (pp. 239–271). Cologne: Heinrich–Barth Institut.

Galasso, I., Russo, R., Mapelli, S., Ponzoni, E., Brambilla, I. M., Battelli, G., and Reggiani, R. 2016. Variability in Seed Traits in a Collection of *Cannabis sativa* L. Genotypes. FRONTIERS IN PLANT SCIENCE, 7.

Galoch, Elżbieta. 2015. The hormonal control of sex differentiation in dioecious plants of hemp (*Cannabis sativa*). The influence of plant growth regulators on sex expression in male and female plants. Acta Societatis Botanicorum Poloniae, (1–2), 153.

Gao, B., Chen, M., Li, X. et al. 2018. Evolution by duplication: paleopolyploidy events in plants reconstructed by deciphering the evolutionary history of VOZ transcription factors. BMC Plant Biol 18, 256. doi:10.1186/s12870-018-1437-8

Glass, M., and Gilleece, T. 2019. Cannabinoids: a possibility for pain management in the palliative cancer pathway. Poster session presented at Annual Radiotherapy Conference 2019, Newcastle, United Kingdom.

Gore M., Bradbury P., Hogers R., Kirst M., Verstege E., et al. 2007. Evaluation of target preparation methods for single-feature polymorphism detection in large complex plant genomes. Crop Sci 47(S2): S135–S148.

Gore M.A., Chia J-M., Elshire R.J., Sun Q., Ersoz E.S., et al. 2009. A first-generation haplotype map of maize. Science 326: 1115–1117.

Goron, T.L. and M.N. Raizada. 2015. Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. Frontiers in Plant Science 6. doi:10.3389/fpls.2015.00157.

Graner, A. 2017. Genebank department. IPK Gatersleben, Berlin, Germany. http://www.ipk-gatersleben.de/en/ genebank

Gupta, K., Varshney, R. K., Sharma, P. C., and Ramesh, B. 1999. Molecular markers and their applications in wheat breeding Plant Breeding, vol. 118, no. 5, pp. 369–390.

Gupta, S. K., Rai, K. N., Singh, P., Ameta, V. L., Gupta, S. K., Jayalekha, A. K., … Verma, Y. S. 2015. Seed set variability under high temperatures during flowering period in pearl millet (Pennisetum glaucum L. (R.) Br.). Field Crops Research, 171, 41–53. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.fcr.2014.11.005

Gusovius, H. J., Hoffmann, T., Budde, J., and Lühr, C. 2016. Still special? Harvesting procedures for industrial hemp. Landtechnik 71: 14–24.

Gworgwor, N. A., Anaso, A. B. and Turaki, Z. G. 1998. Integrated cultural practices for Striga control in millet-based cropping systems. In Pearl Millet in Nigerian Agriculture, 87–97 (Eds A. M. Emechebe, M. C. Ikwelle, O. Ajayi, M. Aminu-Kano and A. B. Anaso). Maiduguri, Nigeria: Ramadan Press.

Habier, D., Fernando, R. L., & Dekkers, J. C. M. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics, 177(4), 2389–2397.

Habiyaremye, C., J.B. Matanguihan, J. D'Alpoim Guedes, G.M. Ganjyal, M.R. Whiteman, K.K. Kidwell, et al. 2017. Proso Millet (Panicum miliaceum L.) and Its Potential for Cultivation in the Pacific Northwest, U.S.: A Review. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.01961.

Hall, J., Bhattarai, S. P., and Midmore, D. J. 2012. Review of Flowering Control in Industrial Hemp. JOURNAL OF NATURAL FIBRES, (1), 23.

Happyana, N., Agnolet, S., Muntendam, R., Van Dam, A., Schneider, B., and Kayser, O. 2013. Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR. Phytochemistry, 87, 51–59.

Harlan, J. 1992. Indigenous African agriculture. In C. Cowan and P. J. Watson (Eds.), The origins of agriculture: An international perspective (pp. 59–70). Washington: Smithsonian Press.

Hatakeyama, M., S. Aluri, M.T. Balachadran, S.R. Sivarajan, A. Patrignani, S. Grüter, et al. 2017. Multiple hybrid de novo genome assembly of nger millet, an orphan allotetraploid crop. DNA Res. 25:39–47. doi:10.1093/ dnares/dsx036

Hatcher, P. E. and Melander, B. 2003. Combining physical, cultural and biological methods: prospects for integrated on-chemical weed management strategies. Weed Research43:303–322.

Heffner E.L., Sorrells M.E., Jannink J.L. 2009. Genomic selection for crop improvement. Crop Sci 49: 1–12.

Heffner, E. L., J. Jannink, H. Iwata, E. Souza, and M. E. Sorrells. 2011. Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. Crop Sci. 51:2597-2606. doi:10.2135/cropsci2011.05.0253

Hess, D. E and Ejeta, G. 1987. Effect of cultural treatments on infestation of *Striga hermonthica* (Del.) Benth (*Scrophulariaceae*). In Proceedings of Fourth International Symposium on Parasitic Flowering Plants, 367–375 (EdsH. C. Weber and W. Forstreuter). Marburg, Germany: Philipps-Universitat.

Hess, D. E., Thakur, R. P., Hash, C. T., Sérémé, P. and Magill, C. W. 2002. Pearl millet downy mildew: problems and control strategies for a new millennium. In Sorghum and Millet Diseases, 37–41 (Ed J. F. Leslie). Ames, Iowa: IowaState Press.

Hirschhorn, J., and Daly, M. 2005. Genome-wide association studies for common diseases and complex traits. NATURE REVIEWS GENETICS, 6(2), 95–108. https://doi-org.proxy-remote.galib.uga.edu/10.1038/nrg1521

Hittalmani, S., H.B. Mahesh, M.D. Shirke, H. Biradar, G. Uday, Y.R. Aruna, et al. 2017. Genome and Transcriptome sequence of Finger millet (Eleusine coracana (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. BMC Genomics 18: 465. doi:10.1186/s12864-017-3850-z.

Holliday, J. A., Wang, T., & Aitken, S. 2012. Predicting Adaptive Phenotypes From Multilocus Genotypes in Sitka Spruce (*Picea sitchensis*) Using Random Forest. G3-GENES GENOMES GENETICS, 2(9), 1085–1093. https://doi-org.proxy-remote.galib.uga.edu/10.1534/g3.112.002733

Hu, H., Scheben, A., Edwards, D. 2018. Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline. Agriculture, 8, 75.

Huang, S. Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., et al. 2009. The genome of the cucumber, *Cucumis sativus* L. Nat. Genet.,41, 1275–1281

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., et al. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature Genetics, 44(1), 32. https://doi-org.proxy-remote.galib.uga.edu/10.1038/ng.1018

Hunt, H.V., F. Badakshi, O. Romanova, C.J. Howe, M.K. Jones and J.S. Heslop-Harrison. 2014. Reticulate evolution in Panicum (Poaceae): the origin of tetraploid broomcorn millet, P. miliaceum. J Exp Bot 65: 3165-3175. doi:10.1093/jxb/eru161.

Hunt, H.V., M.G. Campana, M.C. Lawes, Y.J. Park, M.A. Bower, C.J. Howe, et al. 2011. Genetic diversity and phylogeography of broomcorn millet (Panicum miliaceum L.) across Eurasia. Mol Ecol 20: 4756-4771. doi:10.1111/j.1365-294X.2011.05318.x.

Hurd, Y. L., Spriggs, S., Alishayev, J., Winkel, G., Gurgov, K., Kudrich, C., … Salsitz, E. 2019. Cannabidiol for the Reduction of Cue-Induced Craving and Anxiety in Drug-Abstinent Individuals With Heroin Use Disorder: A Double-Blind Randomized Placebo-Controlled Trial. The American Journal Of Psychiatry, appiajp201918101191. https://doi.org/10.1176/appi.ajp.2019.18101191

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23: 254-267.

ICRISAT, Pearl Millet. Retrieved from http://exploreit.icrisat.org/profile/Pearl Millet/178. date accessed 10/24/2019

Jannink, J.-L., Lorenz, A. J., & Iwata, H. 2010. Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics, 9(2), 166–177. https://doi-org.proxy-remote.galib.uga.edu/10.1093/bfgp/elq001

Jia, G., X. Huang, H. Zhi, Y. Zhao, Q. Zhao, W. Li, et al. 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nat Genet 45: 957-961. doi:10.1038/ng.2673.

Karus, M., and Vogt, D. 2004. European hemp industry: Cultivation, processing and product lines. EUPHYTICA, (NO 1-2), 7.

Khoathane, M. C., Vorster, O. C., and Sadiku, E. R. 2008. Hemp Fiber-Reinforced 1-Pentene/Polypropylene Copolymer: The Effect of Fiber Loading on the Mechanical and Thermal Characteristics of the Composites. JOURNAL OF REINFORCED PLASTICS AND COMPOSITES, (14), 1533.

Korte, A., & Farlow, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. PLANT METHODS, 9. https://doi-org.proxy-remote.galib.uga.edu/10.1186/1746-4811-9-29

Kountche, B. A., Hash, C. T., Dodo, H., Laoualy, O., Sanogo, M. D., Timbeli, A., … Haussmann, B. I. G. 2013. Development of a pearl millet Striga-resistant genepool: Response to five cycles of recurrent selection under Striga-infested field conditions in West Africa. Field Crops Research, 154, 82–90. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.fcr.2013.07.008

Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., et al., 2009. A Multiparent Advanced Generation Inter-Cross to Fine-Map Quantitative Traits in Arabidopsis thaliana. PLoS Genetics, 5(7), 1–15. https://doi-org.proxy-remote.galib.uga.edu/10.1371/journal.pgen.1000551

Kubesova, M., L. Moravcova, J. Suda, V. Jarosik and P. Pysek. 2010. Naturalized plants have smaller genomes than their non-invading relatives: a flow cytometric analysis of the Czech alien flora. PRESLIA -PRAHA-. 82: 81-96.

Kushwaha, H., K.W. Jillo, V.K. Singh, A. Kumar and D. Yadav. 2014. Assessment of genetic diversity among cereals and millets based on PCR amplification using Dof (DNA binding with One Finger) transcription factor gene-specific primers. Plant Systematics and Evolution 301: 833-840. doi:10.1007/s00606-014-1095-8.

Lander, E. S., & Botstein, D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps Genetics. 121. GENETICS -USA-, (4), 185.

Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., … Page, J. E. 2019. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Research, (1), 146.

Lewis, M. A., Russo, E. B., & Smith, K. M. 2018. Pharmacological foundations of Cannabis chemovars. Planta Medica, 84(4), 225–233.

Li, H. 1974. An Archaeological and Historical Account of Cannabis in China. Economic Botany, 28(4), 437-448.

Lin, Z., Wang, J., Cogan, N. O. I., Pembleton, L. W., Badenhorst, P., Forster, J. W., et al., 2016. Optimizing Resource Allocation in a Genomic Breeding Program for Perennial Ryegrass to Balance Genetic Gain, Cost, and Inbreeding. CROP SCIENCE, 57(1), 243–252. https://doi-org.proxy-remote.galib.uga.edu/10.2135/cropsci2016.07.0577

Lorenz A.J., Chao S., Asoro F.G., Heffner E.L., Hayashi T., Iwata H., et al. 2011. Genomic Selection in Plant Breeding. Advances in Agronomy, Vol. 110, pp 77–123.

Lu, F., Copenhaver, G.P., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, et al. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genetics 9: e1003215. doi:10.1371/journal.pgen.1003215.

Lu, H., J. Zhang, K.B. Liu, N. Wu, Y. Li, K. Zhou, et al. 2009. Earliest domesti- cation of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago. Proc. Natl. Acad. Sci. USA 106:7367–7372. doi:10.1073/ pnas.0900158106

Lubell, J. D., and Brand, M. H. 2018. Foliar Sprays of Silver Thiosulfate Produce Male Flowers on Female Hemp Plants. HORTTECHNOLOGY, 28(6), 743–747.

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., and Zhang, Y. 2019. Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. Nature, (7746), 123. https://doi-org.proxy-remote.galib.uga.edu/10.1038/s41586-019-0978-9

Luo, Z., Brock, J., Dyer, J. M., Kutchan, T., Schachtman, D., Augustin, M., et al., 2019. Genetic Diversity and Population Structure of a Camelina sativa Spring Panel. FRONTIERS IN PLANT SCIENCE, 10. https://doi-org.proxy-remote.galib.uga.edu/10.3389/fpls.2019.00184

Lynch, R.C., D. Vergara, S. Tittes, K. White, C.J. Schwartz, M.J. Gibbs, T.C. Ruthenburg, K. deCesare, D.P. Land, and N.C. Kane. 2016. Genomic and chemical diversity in Cannabis. Crit. Rev. Plant Sci. 35(5–6): 349-363. doi:10.1080/07352689.2016.1265363

M'Ribu, H.K. and K.W. Hilu. 1996. Application of random amplified polymorphic DNA to study genetic diversity in Paspalum scrobiculatum L. (Kodo millet, Poaceae). Genetic Resources and Crop Evolution 43: 203-210. doi:10.1007/bf00123272

Ma, C., Blackwell, T., Boehnke, M., & Scott, L. J. 2013. Recommended Joint and Meta-Analysis Strategies for Case-Control Association Testing of Single Low-Count Variants. GENETIC EPIDEMIOLOGY, (6), 539.

Maman, N., Mason, S. C. and Sirifi, S. 2000. Influence of variety and management level on pearl millet production in Niger. I. Grain yield and dry matter accumulation. African Crop Science Journal8:25–34.

Manning, K., and Fuller, D. 2014. Early millet farmers in the Lower Tilemsi Valley, northeastern Mali. In C. Stevens, S. Nixon, M. Murray, and D. Fuller (Eds.), The archaeology of African plant use (pp. 73–82). Walnut Creek: Left Coast Press.

Manning, K., Pelling, R., Higman, T., Schwenniger, J.-L., & Fuller, D. 2011. 4500-year-old domesticated pearl millet (Pennisetum glaucum) from the Tilemsi Valley, Mali: New insights into an alternative domestication pathway. Journal of Archaeological Science, 38,312–322.

Mason, S. C., Maman, N., and Pale, S. 2015. Pearl Millet Production Practices in Semi-Arid West Africa: A Review. EXPERIMENTAL AGRICULTURE, 51(4), 501–521. https://doi-org.proxy-remote.galib.uga.edu/10.1017/S0014479714000441

Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C., and Bonner, T. I. 1990. Structure of a cannabinoid receptor and functional expression of the cloned cDNA. Nature, (6284), 561.

Mbaye, D. F. 1992. Les maldies du mil au Sahel: Etat des connaissances et propositions de lutte. In La Lutte Intégrée Contre les Ennemis des Cultures Vivrières dans le Sahel (Pearl millet diseases in the Sahel: State of knowledge and management proposals). Montrouge, France: John Libby Eurotext.

Mechoulam, R., and Hanus, L. 2000. A historical overview of chemical research on cannabinoids. CHEMISTRY AND PHYSICS OF LIPIDS, (1–2), 1.

Meier C. and Mediavilla V. 1998. Factors influencing the yield and the quality of essential hemp oil (*Cannabis sativa* L.). J. of the International Hemp Association, 5, n°1, 16-20.

Merlin, M. D., Archaeological Evidence for the Tradition of Psychoactive Plant use in the Old World Economic Botany 2003, 57, 295.

Meuwissen T. H. E., Hayes B. J., Goddard M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157 1819–1829.

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Mohan Ram, H. Y., and Jaiswal, V. S., 1972. Induction of Male Flowers on Female Plants of Cannabis sativa by Gibberellins and Its Inhibition by Abscisic Acid. Planta, 105(3), 263.

Mohan Ram, H. Y., and Sett, R. 1982. Reversal of Ethephon-Induced Feminization in Male Plants of Cannabis sativa by Ethylene Antagonists. Zeitschrift Für Pflanzenphysiology, 107(1), 85–89.

Moore, J.H.; Asselbergs, F.W.; Williams, S.M. 2010. Bioinformatics challenges for genome-wide association studies. Bioinformatics. 26, 445–455.

Munro, S., Thomas, K. L., and Abu-Shaar, M. 1993. Molecular characterization of a peripheral receptor for cannabinoids. Nature, (6441), 61.

Nambiar, V. S., Dhaduk, J. J., Neha Sareen, Tosha Shahu, & Rujuta Desai. 2011. Potential functional implications of pearl millet (Pennisetum glaucum) in health and disease. Journal of Applied Pharmaceutical Science, 1(10), 62–67.

Ndiaye, S. 2002. Amelioration du controle du mildiou. Resultats du Projet P3: 1992–1997 (Improved control of mildew: Results of Project P3). Niamey, Niger: Réseau Ouest et Centre Afríaín de Recherche sur le Mil.

Nelson, Clarence H. 1944. Growth Responses of Hemp to Differential Soil and Air Temperatures. Plant Physiology, 19(2), 294.

Neumann, K. 2003. The late emergence of agriculture in sub-Saharan Africa: Archaeobotanical evidence and ecological considerations. In K. Neumann, A. Butler, and S. Kahlheber

(Eds.), Fuel, foods and fields: Progress in African archaeobotany (pp. 71–92). Africa Praehistorica 15). Köln: Heinrich-Barth Institute.

Nyholt, D. R. 2000. All LODs are not created equal. American journal of human genetics, 67(2), 282–288. doi:10.1086/303029

Ophir, R., & Graur, D. 1997. Patterns and rates of indel evolution in processed pseudogenes from humans and murids. GENE -AMSTERDAM-, (1/2), 191.

Panchy, N., Lehti-Shiu, M., & Shiu, S. H. 2016. Evolution of Gene Duplication in Plants. Plant physiology, 171(4), 2294–2316. doi:10.1104/pp.16.00523

Payne, W. A. 1997. Managing yield and water use of pearl millet. Agronomy Journal89:481–490.

Peakall R. and Smouse P.E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Not 6: 288–295.

Peil, A., Flachowsky, H., Schumann, E., and Weber, W. E. 2003. Sex-linked AFLP markers indicate a pseudoautosomal region in hemp (Cannabis sativa L.). Theoretical and Applied Genetics, 107(1), 102–109.

Perez-de-Castro, A.M., Vilanova, S., Cañizares, J., Pascual, L., M Blanca, J., J Diez, M., Prohens, J., Picó, B. 2012. Application of genomic tools in plant breeding. *Curr. Genom*.,13, 179–195

Perucca E. 2017. Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last?. Journal of epilepsy research, 7(2), 61–76. doi:10.14581/jer.17012

Pisupati, R., Vergara, D., and Kane, N. C. 2018. Diversity and evolution of the repetitive genomic content in Cannabis sativa. BMC Genomics, 19(1), 156.

Prasad, M. 2017. The Foxtail Millet Genome. doi:10.1007/978-3-319-65617-5.

Prasada Rao, K., J. De Wet, V. Gopal Reddy and M. Mengesha. 1993. Diversity in the small millets collection at ICRISAT. Advances in small millets. Edited by KW Riley, SC Gupta, A. Seetharam, and JN Mushonga. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, India.

Pucher, A., Sy, O., Sanogo, M. D., Angarawai, I. I., Zangre, R., Ouedraogo, M., … Haussmann, B. I. G. 2016. Combining ability patterns among West African pearl millet landraces and prospects for pearl millet hybrid breeding. Field Crops Research, 195, 9–20. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.fcr.2016.04.035

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. AMERICAN JOURNAL OF HUMAN GENETICS, (3), 559.

Rafalski, J. A. 2002. Novel genetic mapping tools in plants: SNPs and LD-based approaches. PLANT SCIENCE -LIMERICK-, (3), 329.

Rajput, S.G. and D.K. Santra. 2016. Evaluation of genetic diversity of proso millet germplasm available in the United States using simple-sequence repeat markers. Crop Science 56: 2401-2409.

Ranalli, P. 2004. Current status and future scenarios of hemp breeding. EUPHYTICA, (NO 1-2), 121.

Ranc, N., Munos, S., Xu, J., Le Paslier, M.-C., Chauveau, A., Bounon, R., et al. 2012. Genome-Wide Association Mapping in Tomato (Solanum lycopersicum) Is Possible Using Genome Admixture of Solanum lycopersicum var. cerasiforme. G3-GENES GENOMES GENETICS, 2(8), 853–864. https://doi-org.proxy-remote.galib.uga.edu/10.1534/g3.112.002667

Rohlf, F.J. 2000. NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System Version 2.1. Exeter Publishing Setauket, New York.

Roshevits, R.Y. 1980. Grasses: an introduction to the study of fodder and cereal grasses.

Russo, E. B. 2007. History of Cannabis and its preparations in saga, science, and sobriquet. Chem. Biodivers. 4: 1614–1648.

Russo, E. B. 2011. Taming THC: potential Cannabis synergy and phytocannabinoid-terpenoid entourage effects. Br. J. Pharmacol. 163: 1344–1364.

Rutkoski, Jessica E., Elliot L. Heffner, and Mark E. Sorrells. 2011. Genomic Selection for Durable Stem Rust Resistance in Wheat. Euphytica 179 (1): 161–73. doi:10.1007/s10681-010-0301-1.

Saha, D., M.V.C. Gowda, L. Arya, M. Verma and K.C. Bansal. 2016. Genetic and Genomic Resources of Small Millets. Critical Reviews in Plant Sciences 35: 56-79. doi:10.1080/07352689.2016.1147907.

Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H., and Satoh, S. 1998. Characterization: Genome Sizes and Morphology of Sex Chromosomes in Hemp (Cannabis sativa L.). CYTOLOGIA- INTERNATIONAL JOURNAL OF CYTOLOGY, (4), 459.

Sanchez-Carnerero Callado, C., Nunez-Sanchez, N., Casano, S., and Ferreiro-Vera, C. 2018. The potential of near infrared spectroscopy to estimate the content of cannabinoids in Cannabis sativa L.: A comparative study. Talanta. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.talanta.2018.07.085

Sasaki, T. 2005. The map-based sequence of the rice genome. Nature 436, 793–800 doi:10.1038/nature03895

Sawler, J., J.M. Stout, K.M. Gardner, D. Hudson, J. Vidmar, L. Butler, J.E. Page, and S. Myles. 2015. The genetic structure of marijuana and hemp. PloS one 10(8): E0133292. doi:10.1371/journal.pone.0133292

Scheuring, J. F., Katilé, S. O. and Kollo, I. A. 2002. Boosting pearl millet yields with Apron Plus and Apron Star seed treatments. In Sorghum and Millet Diseases, 47–49 (Ed J. F. Leslie). Ames, Iowa: Iowa State Press.

Schluttenhofer, C., and Yuan, L. 2017. Challenges towards Revitalizing Hemp: A Multifaceted Crop. Trends in Plant Science, 22(11), 917–929.

Seghatoleslami, M. J., Kafi, M., and Majidi, E. (n.d.). Effect of Drought Stress at Different Growth Stages on Yield and Water Use Efficiency of Five Proso Millet (Panicum Miliaceum L.) Genotypes. PAKISTAN JOURNAL OF BOTANY, 40(4), 1427–1432.

Sehgal, D., S.K. Parida, L. Skot, R. Singh, R.K. Srivastava, S.P. Das, et al. 2015. Exploring Potential of Pearl Millet Germplasm Association Panel for Association Mapping of Drought Tolerance Traits. Plos One 10: e0122165. doi:10.1371/journal.pone.0122165.

Shanahan, J., R. Anderson and B. Greb. 1988. Productivity and water use of proso millet grown under three crop rotations in the Central Great Plains. Agronomy journal 80: 487-492.

Sikora, P., Chawade, A., Larsson, M., Olsson, J., and Olsson, O. 2011. Mutagenesis as a Tool in Plant Genetics, Functional Genomics, and Breeding. International Journal of Plant Genomics, 2011, 1–13.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., and Taura, F. 2004. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. The Journal Of Biological Chemistry, 279(38), 39767–39774.

Small, E. 2015. Evolution and Classification of Cannabis sativa (Marijuana, Hemp) in Relation to Human Utilization. Botanical Review, 81(3), 189–294.

Small, E., and D. Marcus. 2003. Tetrahydrocannabinol levels in hemp (Cannabis sativa) germplasm resources. Econ. Bot. 57(4):545–558. doi:10.1663/0013-0001(2003)057[0545:TLIHCS]2.0.CO;2

Spindel, J. E., Begum, H., Akdemir, D., Collard, B., Redona, E., Jannink, J. L., & McCouch, S. 2016. Genome-wide prediction models that incorporate de novo GWAS are a powerful new tool for tropical rice improvement. HEREDITY -LONDON-, (4), 395.

Sul J.H., Martin L.S., Eskin E. 2018. Population structure in genetic studies: Confounding factors and mixed models. PLoS Genet 14(12): e1007309. https://doi.org/10.1371/journal.pgen.1007309

Tako, E., Reed, S. M., Budiman, J., Hart, J. J., & Glahn, R. P. 2015. Higher iron pearl millet (Pennisetum glaucum L.) provides more absorbable iron that is limited by increased polyphenolic content. Nutrition Journal, (1). https://doi-org.proxy-remote.galib.uga.edu/10.1186/1475-2891-14-11

Taura, F., and S. Morimoto. 1995. First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis. J. Am. Chem. Soc. 117(38):9766– 9767. doi:10.1021/ja00143a024

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., and Morimoto, S. 2007. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. FEBS Letters, 581(16), 2929–2934.

Tesso, T. and Ejeta, G. 2011. Integrating multiple control options enhances Striga management and sorghum yield on heavily infested soils. Agronomy Journal103:1464–1471.

Thakur, R. P., Veeranki, P. R. and Sharma, R. 2011. Influence of dosage, storage time and temperature on efficacy of netalazyl-treated seed for the control of pearl millet downy mildew. European Journal of Plant Pathology129:353–359.

Tostain, S. 1992. Enzyme diversity in pearl millet (*Pennisetum glaucum* L.). Theoretical and Applied Genetics, 83,733–742.

Toth, J. A., Stack, G. M., Cala, A. R., Carlson, C. H., Wilk, R. L., Crawford, J. L., Viands, D. R., Philippe, G., Smart, C. D., Rose, J. K. C., & Smart, L. B. 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. GCB Bioenergy, 12(3), 213–222. https://doi.org/10.1111/gcbb.12667

U.S. Hemp Crop Report. (2018). Retrieved from https://www.votehemp.com/u-s-hemp-crop-report/

Upadhyaya, H., S. Dwivedi, S. Singh, S. Singh, M. Vetriventhan and S. Sharma. 2014. Forming core collections in barnyard, kodo, and little millets using morphoagronomic descriptors. Crop Science 54: 2673-2682

Upadhyaya, H., S. Sharma, C. Gowda, V.G. Reddy and S. Singh. 2011. Developing proso millet (Panicum miliaceum L.) core collection using geographic and morpho-agronomic data. Crop and Pasture Science 62: 383-389.

USDA. 2019. Filling a Hemp Germplasm Chasm (a News Story Versified). https://www.usda.gov/media/radio/daily-newsline/2019-08-15/filling-hemp-germplasm-chasm-news-story-versified

Vadez, V., Hash, T., Bidinger, F. R., & Kholova, J. 2012. II.1.5 Phenotyping pearl millet for adaptation to drought. Frontiers In Physiology, 3, 386. https://doi-org.proxy-remote.galib.uga.edu/10.3389/fphys.2012.00386

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., and Page, J. E. 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biology (Online Edition), R102.

van den Broeck, H. C., Maliepaard, C., Ebskamp, M. J. M., Toonen, M. A. J., and Koops, A. J. 2008. Differential expression of genes involved in C1 metabolism and lignin biosynthesis in wooden core and bast tissues of fiber hemp (Cannabis sativa L.). Plant Sci. 174, 205–220.

VanRaden, P. M. 2008. Efficient Methods to Compute Genomic Predictions. Journal of Dairy Science, 91(11), 4414–4423. https://doi-org.proxy-remote.galib.uga.edu/10.3168/jds.2007-0980

Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, et al. 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nature Biotechnology 35: 969-976. doi:10.1038/nbt.3943.

Velasco, R., Zharkikh, A., Troggio, M., Cartwright, D.A., Cestaro, A., Pruss, D., Pindo, M., Fitzgerald, L.M., Vezzulli, S., Reid, J., et al. 2007. A high-quality draft consensus sequence of the genome of a heterozygous grapevine variety. PLoS ONE.,2, e1326.

Vicient, C. M., & Casacuberta, J. M. 2017. Impact of transposable elements on polyploid plant genomes. Annals of botany, 120(2), 195–207. doi:10.1093/aob/mcx078

Voss-Fels, K. P., Cooper, M., & Hayes, B. J. 2019. Accelerating crop genetic gains with genomic selection. TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik, 132(3), 669–686.

Wallace, J.G. and Mitchell, S.E. 2017. Genotyping-by-sequencing. Curr. Protoc. Plant Biol. 2:64-77. doi: 10.1002/cppb.20042

Wang, X., Miao, J., Chang, T., Xia, J., An, B., Li, Y., et al., 2019. Evaluation of GBLUP, BayesB and elastic net for genomic prediction in Chinese Simmental beef cattle. PLoS ONE, 14(2), 1–14. https://doi-org.proxy-remote.galib.uga.edu/10.1371/journal.pone.0210442

Weber, J. L. and May, P. E.. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. American Journal of Human Genetics, vol. 44, no. 3, pp. 388–396.

Weiblen, G.D., J.P. Wenger, K.J. Craft, M.A. ElSohly, Z. Mehmedic, E.L. Treiber, and M.D. Marks. 2015. Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytol. 208(4):1241–1250.

Yadav, O.P. & Rai, K.N. 2013 Genetic improvement of pearl millet in India. Agric. Res. 2, 275–292.

Yadav, Y., G.R. Lavanya, S. Pandey, M. Verma, C. Ram and L. Arya. 2016. Neutral and functional marker based genetic diversity in kodo millet (*Paspalum scrobiculatum* L.). Acta Physiologiae Plantarum 38. doi:10.1007/s11738-016-2090-1.

Yang, H., Dong, L., Wang, H., Liu, C., Liu, F., & Xie, C. 2018. A simple way to visualize detailed phylogenetic tree of huge genome-wide SNP data constructed by SNPhylo. Journal of Integrative Agriculture, 17(9), 1972–1978. https://doi-org.proxy-remote.galib.uga.edu/10.1016/S2095-3119(18)62023-4

Yao, N., Wang, L., Yan, H., Liu, Y., and Lu, B.-R. 2015. Mapping quantitative trait loci (QTL) determining seed-shattering in weedy rice: evolution of seed shattering in weedy rice through de-domestication. Euphytica, 204(3), 513–522.

Yue, H., L. Wang, H. Liu, W. Yue, X. Du, W. Song, et al. 2016. De novo Assembly and Characterization of the Transcriptome of Broomcorn Millet (Panicum miliaceum L.) for Gene Discovery and Marker Development. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.01083.

Zatybekov, A., Abugalieva, S., Didorenko, S., Gerasimova, Y., Sidorik, I., Anuarbek, S., & Turuspekov, Y. 2017. GWAS of agronomic traits in soybean collection included in breeding pool in Kazakhstan. BMC Plant Biology, 17(Suppl 1), 179. https://doi-org.proxy-remote.galib.uga.edu/10.1186/s12870-017-1125-0

Zhang, G., X. Liu, Z. Quan, S. Cheng, X. Xu, S. Pan, et al. 2012. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. Nature Biotechnology 30: 549-554. doi:10.1038/nbt.2195.

Zhao, Z. 2005. Palaeoethnobotany and its new achievements in China. Kaogu (Archaeology) 7: e49

Zheng, Z., Sun, Z., Fang, Y. 2018. et al. Genetic Diversity, Population Structure, and Botanical Variety of 320 Global Peanut Accessions Revealed Through Tunable Genotyping-by-Sequencing. Sci Rep 8, 14500. doi:10.1038/s41598-018-32800-9

Zou, C., L. Li, D. Miki, D. Li, Q. Tang, L. Xiao, et al. 2019. The genome of broomcorn millet. Nature Communications 10: 436. doi:10.1038/s41467-019-08409-5.

Zou, H. and Trevor, H. 2005. Regularization and Variable Selection via the Elastic Net. Journal of the Royal Statistical Society. Series B (Statistical Methods)

CHAPTER 2

GENOME-WIDE POPULATION STRUCTURE ANALYSES OF THREE MINOR MILLETS: KODO MILLET, LITTLE MILLET, AND PROSO MILLET

[1] Matthew Johnson, Santosh Deshpande, Mani Vetriventhan, Hari D Upadhyaya, Jason G. Wallace. Published in *The Plant Genome*: DOI: 10.3835/plantgenome2019.03.0021, printed with open copyright permission from the publisher (CC-BY).

Abstract

Millets are a diverse group of small-seeded grains that are rich in nutrients but have received relatively little advanced plant breeding research. Millets are important to smallholder farmers in Africa and Asia because of their short growing season, good stress tolerance, and high nutritional content. To advance the study and use of these species, we present a genome-wide marker datasets and population structure analyses for three minor millets: kodo millet (*Paspalum scrobiculatum*), little millet (*Panicum sumatrense*), and proso millet (*Panicum miliaceum*). We generated genome-wide marker data sets for 190 accessions of each species with genotyping-by-sequencing (GBS). After filtering, we retained between 161 and 165 accessions of each species, with 3461, 2245, and 1882 single-nucleotide polymorphisms (SNPs) for kodo, little, and proso millet, respectively. Population genetic analysis revealed 7 putative subpopulations of kodo millet and 8 each of proso millet and little millet. To confirm the accuracy of this genetic data, we used public phenotype data on a subset of these accessions to estimate the heritability of various agronomically relevant phenotypes. Heritability values largely agree with the prior expectation for each phenotype, indicating that these SNPs provide an accurate genome-wide sample of genetic variation. These data represent one of first genome-wide population genetics analyses, and the most extensive, in these species and the first genomic analyses of any sort for little millet and kodo millet. These data will be a valuable resource for researchers and breeders trying to improve these crops for smallholder farmers.

Abbreviations:

CDS, coding sequences; ESTs, expressed sequence tags; FAO, Food and Agriculture Organization; GBS, Genotyping-by-Sequencing; ICRISAT, International Crops Research

Institute for the Semi-Arid Tropics; MDS, multidimensional scaling; PCR, Polymerase chain reaction; RAPD, Random Amplification of Polymorphic DNA; SNP, single nucleotide polymorphism

Introduction

It is estimated that there are up to 7,000 cultivated crop species in the world (Khoshbakht and Hammer 2008), yet major breeding and research efforts have focused on just a small number of these (Hammer et al., 2001). Many crops that have been ignored by modern research are still essential to the local communities that have relied on them for thousands of years, and they have potential for diversifying cropping systems around the world (Naylor et al., 2004). These crops are also genetic resources to increase global food security as climate changes and resources (land, water, fertilizers) become more limited. The ability to generate inexpensive, genome-wide data can quickly bring some of these crops into the modern genomics era (Varshney et al 2012).

Millets are a diverse group of small-seeded grains that have been largely overlooked by modern genetics research. Although pearl millet (*Pennisetum glaucum* syn. *Cenchrus americanus*), foxtail millet (*Setaria italica*), and finger millet (*Eleusine coracana*) all have complete genome sequences and well-established germplasm resources (Varshney et al., 2017; Zhang et al., 2012; Prasad 2017; Sehgal et al., 2015; Hittalmani et al., 2017; Jia et al., 2013; Bennetzen et al., 2012 and Hatakeyama et al., 2017), most other millets have few if any resources available (Goron and Raizada 2015). These crops are often important for smallholder farmers, especially in southeast Asia and Sub-Saharan Africa (FAO; http://www.fao.org/docrep/W1808E/w1808e0e.htm).

Proso millet (*Panicum miliaceum* L.), little millet (*Panicum sumatrense*), and kodo millet (*Paspalum scrobiculatum* L.) are three minor millets with limited modern genetic resources (Table 2.1), although proso millet has recently seen a significant increase in genomic studies (e.g., Rajput and Santra, 2016, Habiyaremye et al., 2017, and Yue et al., 2016), including a reference genome (Zou et al., 2019). Various germplasm repositories maintain collections of these species (Goron and Raizada 2015); the current study focuses on those held by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), which maintains collections with 849, 473, and 665 accessions of proso, little, and kodo millet, respectively. These collections have been assessed for morphological and agronomic traits, and representative core collections have been created for each of them (Upadhyaya et al., 2014, 2011). These millets are hardy C_4 grasses (Upadhyaya et al., 2014, and Brown, 1999) with nutritional content on par with or superior to the major grains (Vetriventhan and Upadhyaya, 2018; Mengesha 1966, Saleh et al., 2013, Kalinova and Moudry 2006).

Proso Millet

Proso millet is believed to have been domesticated ~10,000 years ago. There have been multiple centers of origins proposed: Northwest China (Bettinger et al., 2007, 2010a,b), Central China (Lu et al., 2009), and Inner Mongolia (Zhao, 2005), with the most recent evidence pointing towards either one domestication event in China or one in China and one in Europe (Hunt et al., 2011). It is the third-oldest cultivated cereal after wheat and barley (Habiyaremye et al., 2017; Upadhyaya et al., 2011). Proso millet is valued for its low water requirements (330 mm) and short growing season (60 days) (Habiyaremye et al., 2017; Shanahan et al., 1988). Proso millet varieties are classified into five races based on inflorescence morphology:

miliaceum, patentissimum, contractum, compactum, and ovatum (de Wet et al., 1983). Proso millet is tetraploid (2n = 4x = 36) (Saha et al., 2016), with an allotetraploid origin (Hunt et al., 2014). Proso millet has historically be the most widely grown of the three millets studied here, with cultivation concentrated in the former Soviet Union and India (Roshevits, 1980), and with significant production in the United States (Baltensperger 2002).

Proso millet genetic diversity has been investigated with a variety of genetic markers, most of them at a very small scale (<100 markers; reviewed in Habiyaremye et al., 2017). Recently, however, the most extensive genetic analysis in proso millet identified over 400,000 SNP markers and 35,000 SSRs from the transcriptomes of two proso millet accessions (Yue et al., 2016). Population-level analyses were made by Rajput et al. (2016) using 100 SSRs and 90 proso millet accessions, it was found that there were some connections between genetic clustering and geographic origin. There has also been QTL mapping using 833 SNPs from GBS data (Rajput and Santra, 2016), and a just-published reference genome (Zou et al., 2019).

Kodo Millet

Kodo millet was domesticated in India around 3,000 years ago, and India has historically been the major center of cultivation (de Wet et al., 1983). Kodo millet accessions have been classified into three races based on panicle morphology: regularis, irregularis, and variabilis (de Wet et al., 1983; Prasada Rao et al., 1993). Kodo millet is tetraploid (2n = 4x = 40) (Saha et al., 2016), and it is valued for its ability to produce consistently in hot, drought-prone arid and semi-arid land (Dwivedi et al., 2012). A few sets of molecular markers have been developed for kodo millet based on RAPD markers (M'Ribu and Hilu 1996), gene-specific primer sets (Kushwaha et al., 2014), and semi-targeted PCR amplification (Yadav et al. 2016); the latter study concluded

that the 96 accessions they used could be divided into four groups that showed little connection between geographic region and genetic relationship of the accessions. There have been no truly genome-wide datasets on this species before now.

Little Millet

Little millet was domesticated 5,000 years ago in India (de Wet et al., 1983). It has historically been grown mainly in India, Mayanmar, Nepal, and Sri Lanka (Prasada Rao et al., 1993). Little millet accessions have been classified into 2 races based on panicle morphology: nana and robusta, with two subraces per race (laxa and erecta for nana, and laxa and compacta for robusta) (de Wet et al., 1983; Prasada Rao et al., 1993). Little millet is tetraploid (2n = 4x = 36) (Saha et al., 2016). Like kodo millet, little millet can give consistent yields on marginal lands in drought-prone arid and semi-arid regions, and it is an important crop for regional food stability (Dwivedi et al., 2012). Little millet is arguably the least studied of these three millets, and we are unaware of any molecular markers developed for it outside of specific single genes (Goron and Raizada 2015) and a small set of RAPD markers whose details were not described (M. S. Swaminathan Research Foundation 2000).

Expanding genetic resources

As mentioned above, the genetic and genomic resources of proso millet, kodo millet, and little millet are very limited (Saha et al., 2016; Goron and Raizada 2015), although the situation for proso millet, at least, is improving. The main focuses of the resources have been centered on the accessions stored in gene banks, expression sequence tags (ESTs), and complete coding sequences (CDSs).

Core collections for each of these species were created several years ago, consisting of ~10% of ICRISAT's collection for each species (Upadhyaya et al., 2011,2014). These collections have been assessed for morpho-agronomic traits but no genetic data has been available for them. Our goal in this project was to generate genome-wide marker data on each of these species to enable population genetic analysis and empower breeders and researchers to make more informed decisions about germplasm selection and representation. We genotyped 190 accessions of each species, including the entire core collections of each, using genotyping-by-sequencing (Elshire et al. 2011) to generate the most comprehensive population genetics resource for each of these species to date. These data dramatically expand the genomic resources available to each of these crops and will help make better use of them moving forward.

Materials and Methods

Plant materials

One-hundred ninety accessions (see Supplemental Data) were taken from the ICRISAT genebank for each millet species. These included the accessions from the core collection for proso millet, kodo millet and little millet (Upadhyaya et al., 2011, 2014). The remaining samples were chosen to broadly sample the available diversity based on the cluster information used to create the core collections.

DNA Extraction and sequencing

Seedlings of each accession were grown at the ICRISAT research station in Patancheru, India, in 2015. DNA was extracted from 55-day old plants, with one plant from each accession used for extraction, by the modified CTAB method (Mace et al., 2003), lyophilized, and shipped

to the Genomic Diversity Facility at Cornell University for genotyping-by-sequencing (GBS) (Elshire et al., 2011). GBS library preparation followed standard methods (Wallace and Mitchell 2017) using the PstI restriction enzyme. For kodo millet, 190 samples plus 2 blanks were multiplexed into a single lane for sequencing on an Illumina HiSeq 2500 with single-end 100 bp sequencing. For little and proso millet the procedure was similar, except that samples were multiplexed into 2 lanes of 95 samples plus 1 blank. The raw sequence data for these samples is available at the Sequence Read Archive, accessions PRJNA494158.

SNP calling

Single-nucleotide polymorphisms (SNPs) were called from raw sequencing data using the TASSEL (Bradbury et al., 2007) GBS v2 SNP pipeline. Since this pipeline uses alignment to a reference genome and none of these species has such a reference, we included slight alterations to make use of the UNEAK filter (Lu et al., 2013) for reference-free alignment of tags to each other. All code for this analysis is available in supplemental data and at https://github.com/wallacelab/2018_minor_millets.

Filtering

Low-quality SNPs were removed by filtering the raw genotype data to remove sites with >50% missing data, minor allele frequency <5%, and >20-25% heterozygosity. (The heterozygosity filter removes false SNPs due to paralogs misaligning in polyploid species (Wallace et al 2014). Samples with >50% missing data across all remaining sites were also filtered out.

The above filters reduced total missing data to <20% in each species (9% in kodo, 16.8% in little, 13% in proso). We tested imputation with LinkImpute (Money et al., 2015) to reduce this further; however, the imputed data did not result in any appreciable differences (Figures 2.3-2.5) in downstream analyses, so all analyses were conducted with the non-imputed data.

Phylogeny

Phylogenetic networks were constructed using the NeighborNet method (Bryant and Moulton, 2003) in SplitsTree4 V4.14.4 (Huson and Bryant, 2006). Genotype data was converted to SplitsTree-compatible NEXUS files by first using TASSEL version 5.2.29 (Bradbury et al., 2007) to export as a PHYLIP (interleaved) format, which was then converted to NEXUS format using Alter (Glez-Peña et al., 2010) (http://www.sing-group.org/ALTER/). The resulting files were manually edited to change the data type to "dna" and replace all colons (:) with underscores (_), at which point the files were loaded into SplitsTree for network creation. Dendograms were also generated using TASSEL. A cladogram was generated using a neighbor joining cluster method, and archaeopteryx (Han & Zmasek, 2009) was used to generate the final dendograms.

Population Structure determination

Population structure analysis was performed with FastStructure v1.0 (Anil et al., 2014), with the number of potential populations (k) varying from 1 to 15. The optimum population size was determined by chooseK and results were visualized with Distruct, both parts of the FastStructure software package. Default parameters were used for all programs. Samples were assigned to a population if they had at least 60% membership in that population. (This cutoff is arbitrary, but using a higher cutoff did not appreciably change population membership (data not

shown.) Fst values were calculated for comparisons between all populations in each species as additional validation (see Supplemental Data)

Plotting Accessions by Geographic Data

Geographic data on kodo millet and little millet was plotted in python using basemap (Hunter 2007) and the known GPS coordinates, then colored by subpopulation. Since proso millet had only the country of origin for some accessions, its map was created manually in Inkscape 0.91 (https://inkscape.org) by placing dots (colored by subpopulation) in the country of origin for each accession.

Principal coordinate analysis

Genetic principal coordinates were calculated using multidimensional scaling (MDS) analysis in TASSEL V5.2.29 (Bradbury et al., 2007).

Heritability analysis

Phenotype data for heritability analysis was taken from public data on the kodo and little millet core collections (Upadhyaya et al., 2014) and proso millet phenotype data was provided by collaborators at IRCISAT (see Supplemental Data). Phenotypes were fit as part of a mixed linear model in TASSEL (Bradbury et al., 2007) with a kinship matrix as the only covariate, using default parameters. Narrow-sense heritability (h^2) was estimated as the ratio of genetic variance to total variance in the model.

Software

The following software and packages were used as part of this analysis:

- SplitsTree4: V4.14.4: Huson and Bryant, 2006
- Faststructure: 1.0: https://rajanil.github.io/fastStructure/ Raj et al., 2014
- TASSEL 5.2.29: http://www.maizegenetics.net/tassel Bradbury et al., 2007
- PLINK 2.0: http://zzz.bwh.harvard.edu/plink/ Purcell 2007
- Inkscape 0.91: https://inkscape.org/release/0.91/ 2015
- GNU Parallel 20141022: Tange 2011
- Matplotlib 2.2.2: Hunter 2007
- Pandas 0.17.: McKinney 2010
- NumPy 1.11.0: Oliphant 2006
- Python 3.5.2
- VCF Tools 0.1.13 (Danecek et al., 2011)

Results and Discussion

Genome-wide marker sets

Genotype data was generated using GBS (Elshire et al. 2011) using PstI restriction digestion. This resulted in ~12,000-14,000 raw SNPs per species (Table 2.2), although many of these are probably artifacts due to misalignments. Low-quality SNPs were removed by filtering the raw genotype data to remove sites with >50% missing data, minor allele frequency <5%, and >20-25% heterozygosity. (The heterozygosity filter removes false SNPs due to paralogs misaligning in polyploid species (Wallace et al 2014)). Samples with >50% missing data across all remaining sites were also filtered out. The final, filtered genotype data (see Supplemental

Data) consists of 3,461 SNPs across 165 accessions of kodo millet, 2,245 SNPs across 165 accessions of little millet, and 1,882 SNPs across 161 accessions of proso millet. (Table 2.2)

Proso millet

Population structure analysis of proso millet with fastStructure (Raj et al., 2014) grouped the 161 post-filtering samples into 8 putative subpopulations (Figure 2). Some of the subpopulations consist almost entirely of "pure" individuals of that subpopulation, implying strong separation from the other subpopulations (e.g., groups 7 and 8), while others (e.g., groups 2 through 5) show significant admixture among the subpopulations. We observed that small changes in fastStructure parameters would significantly alter the division and number of subpopulations in this admixed group (data not shown), further indicating that the divisions among the highly-admixed subpopulations are weak and should be interpreted with caution.

A phylogenetic network of these samples (Figure 2.1) mirrors this, where most of the subpopulations group together, although subpopulations 6 and 8 are strongly separated from the rest. Divisions among the main group of populations appear weak, as evidenced by the large number of alternative splits ("webbing") in the phylogenetic web and the fact that two populations (2 and 4) are split into 2-3 groups across the phylogeny. The genetic principal coordinates of these samples (Figure 2.2) indicate a similar pattern, where most samples cluster together but with subpopulations 6 and 8 distinct.

While the race data was not available for proso millet in this study there is strong evidence that race is not a good indicator of genetic relatedness among accessions (Vetriventhan and Upadhyaya 2018). Existing core collections were designed with the races of the species

being a central focus. Core collections could likely be improved by using population structure and genomic data to improve the genetic diversity of the collections.

The plotting of the 109 accessions with known geographic location supports the population analysis (Figure 2.6).

The reference genome was published after the bulk of the work for this paper had been done, and when analyses were rerun using the reference genome there were no significant difference found from the original analysis (Figure 2.7)

Kodo millet

Population structure analysis places the 165 kodo millet samples into 7 putative subpopulations (Figure 2.2). Most accessions cluster together, but 2 subpopulations (5 and 7) are strongly separated from the rest Figures (2.1,2.2). The separation of subpopulation 7 from the other samples drives the largest genetic differences in this species, with 67.5% of the genetic variation along this axis (PC1 in Figure 2.2).

These patterns of population structure do not correlate with existing race designations (Figure 2.8), implying that existing race designations do not strongly correlate to genetic groupings. Similar to proso millet, this implies that core collections of kodo millet could be improved by using population and genomic data instead of race designation.

Only 17 of the kodo millet accessions had known geographic origins, all of them in India. Plotting these origins geographically supports the population analysis by showing clustering of population 4 (Figure 2.9), although more samples would be needed to completely confirm this interpretation.

Little millet

The 165 filtered accessions of little millet were grouped into 8 putative subpopulations by fastStructure (Figure 2.2). Little millet appears to have weaker population structure than the other two species (compare the variance explained by each principal coordinate, Figure 2.2). Subpopulation 5 is relatively well separated from the others, followed by subpopulations 6 and 7, and then the other subpopulations being relatively close together. (These subpopulation designations would shift around with minor changes in fastStructure parameters, implying they are very weakly separated; data not shown.) Similar to the other two species, little millet population structure did not correlate with existing race designations (Figure 2.10). Plotting the 29 accessions with known geographic location supports the population analysis by showing distinct clustering's of populations 4 and 5 (Figure 2.11).

Validation of Genotype Data via Estimated Heritability

To validate these genetic data, we used public phenotype data on the kodo millet and little millet core collections (Upadhyaya et al.,2014), and proso millet phenotypic data that was provided by ICRISAT (see Supplemental Data). These data were used to estimate narrow-sense heritability (h^2) for flowering time and plant height (Tables 2.3). Our expectation was that true genotype data should result in moderate to high heritability values for some phenotypes, especially ones (such as flowering time and plant height) that are known to have a strong genetic component.

Heritability was estimated using a mixed linear model in TASSEL (see Methods) to obtain estimated variance components. Little millet traits exhibited heritability from 0.209 to 0.807, while kodo millet traits were slightly lower, ranging from 0.0 to 0.505 (see Supplemental

Data). There were no other phenotypes provided for proso millet other than what was included in table 2.3. In all three species, flowering time and plant height were some of the most heritable traits (Table 2.3). These traits are all known to be under strong genetic control in other grass crops (e.g., Buckler et al 2009; Peiffer et al 2014; Mace et al 2013; Morris et al 2013; Ma et al 2016; Alqudah et al 2013). Since poor-quality SNPs should show little to no relationship with phenotype, these results imply that the SNP datasets we have generated accurately represent the genetic variation within these populations and can be used for real-world breeding applications.

Conclusions

The results from this study represent one of the first genome-wide analyses for proso millet, and the first for kodo millet, and little millets, three "orphan" crops that are important for food security in developing nations. We have identified thousands of SNPs in each of these species that accurately capture the population structure of each, as indicated by the geographic correlations and estimates of narrow-sense heritability. Our analyses can be used as a foundation for further exploration into the genetics of these species, including selecting appropriate breeding materials and identifying priority populations for further collection and curation.

Existing core collections were designed with the races of the species being a central focus. The evidence strongly implies that race is not a good determiner of genetic relatedness, and as such the core collections could likely be improved by using modern genomic data to improve the genetic diversity of the collections.

For both major and minor crops, obtaining genetic data is now (almost) trivial, even in species with highly complex genomes and no prior history of genetic analysis. As the price of DNA sequencing continues to drop, more and more orphan species will have genotype data

available. The major question going forward will be how to best deploy these data to benefit breeders and growers. Given that the price of phenotyping is often the limiting factor in many studies (Cobb et al., 2013), finding ways to deploy genomic prediction and/or high-throughput phenotyping for orphan crops will likely be the next major step to democratize modern genomics for the developing world.

Table 2.1: Genetic resources including: Total ICRISAT Accessions (http://genebank.icrisat.org/Default, accessed November 2018), Proso millet core collection (Upadhyaya et al., 2011, Kodo and little millet (Upadhyaya et al., 2014), ploidy information (Upadhyaya et al., 2011 and 2014), and estimated genome size for proso millet (Kubesova et al., 2010)

Crop	Total ICRISAT Accessions	Accessions in ICRISAT Core Collection	Expressions Sequence Tags	Complete Coding Sequences	Ploidy Level	Estimated genome size (Mbp)
Proso Millet	849	106	195	7	2n=4x=36	1020.5
Kodo Millet	665	75	29	0	2n=4x=40	unknown
Little Millet	473	55	12	0	2n=4x=36	unknown

Table 2.2: Number of SNPs and accessions before and after filtering

Species	**Raw SNPs**	**Filtered SNPs**	**Raw accessions**	**Filtered accessions**
Kodo Millet	12995	3461	190	165
Little Millet	14473	2245	190	165
Proso Millet	12839	1882	190	161

Table 2.3: Estimated heritability's for minor millets

Species	**Trait**	**Heritability**
Little	Days to 50% flowering	0.807
Little	Plant height (cm)	0.664
Proso	Days to 50% flowering	0.69
Proso	Plant height (cm)	0.615
Kodo	Days to 50% flowering	0.505
Kodo	Plant height (cm)	0.293

Table 2.4: Accession population assignments

Kodo	Population	Proso	Population	Little	Population
IPs216	Pop_2	IPm361	Pop_6	IPmr1020	Pop_5
IPs159	Pop_6	IPm2100	Pop_3	IPmr1060	Pop_1
IPs286	Pop_7	IPm2205	unknown	IPmr875	Pop_6
IPs63	unknown	IPm2660	Pop_1	IPmr850	Pop_8
IPs892	Pop_3	IPm277	Pop_5	IPmr716	Pop_4
IPs264	Pop_4	IPm2682	Pop_5	IPmr889	Pop_6
IPs716	Pop_3	IPm2745	Pop_5	IPmr858	Pop_6
IPs854	Pop_3	IPm2861	Pop_8	IPmr724	Pop_6
IPs914	Pop_4	IPm2555	Pop_5	IPmr945	Pop_2
IPs729	Pop_4	IPm2171	unknown	IPmr984	Pop_8
IPs368	Pop_1	IPm2718	Pop_5	IPmr825	Pop_4
IPs13	Pop_7	IPm2675	Pop_8	IPmr712	Pop_7
IPs411	Pop_3	IPm2816	Pop_5	IPmr772	Pop_4
IPs582	Pop_6	IPm1586	Pop_4	IPmr386	Pop_4
IPs68	Pop_4	IPm2227	Pop_4	IPmr462	unknown
IPs91	Pop_4	IPm2535	Pop_2	IPmr442	Pop_8
IPs850	unknown	IPm1608	Pop_4	IPmr862	Pop_6
IPs806	Pop_1	IPm2101	Pop_3	IPmr998	unknown
IPs429	Pop_6	IPm2529	Pop_2	IPmr493	Pop_8
IPs793	Pop_3	IPm55	unknown	IPmr711	Pop_1
IPs280	Pop_7	IPm2076	Pop_7	IPmr1022	Pop_5
IPs744	Pop_3	IPm2544	Pop_2	IPmr905	unknown
IPs620	Pop_3	IPm2767	Pop_5	IPmr492	Pop_8
IPs258	unknown	IPm2904	Pop_6	IPmr811	Pop_4
IPs147	Pop_1	IPm2900	Pop_8	IPmr828	Pop_8
IPs830	Pop_4	IPm2629	Pop_4	IPmr840	Pop_5
IPs776	Pop_4	IPm2812	Pop_6	IPmr127	Pop_4
IPs279	Pop_7	IPm102	unknown	IPmr762	Pop_2
IPs883	Pop_3	IPm2765	unknown	IPmr808	Pop_4
IPs110	Pop_6	IPm448	Pop_6	IPmr814	Pop_4
IPs874	Pop_3	IPm2686	Pop_5	IPmr992	Pop_7
IPs630	Pop_6	IPm2108	Pop_3	IPmr999	Pop_7
IPs238	unknown	IPm2872	Pop_8	IPmr840	Pop_5
IPs60	Pop_1	IPm2700	Pop_6	IPmr708	Pop_1
IPs575	Pop_4	IPm2155	Pop_4	IPmr864	Pop_6
IPs828	Pop_4	IPm104	unknown	IPmr881	Pop_6
IPs222	Pop_4	IPm16	Pop_4	IPmr939	unknown

IPs759	Pop_6	IPm2573	Pop_2	IPmr884	Pop_6
IPs654	unknown	IPm2733	unknown	IPmr760	Pop_8
IPs702	Pop_2	IPm301	Pop_6	IPmr1058	Pop_6
IPs152	Pop_1	IPm2844	Pop_8	IPmr1075	Pop_4
IPs837	unknown	IPm2685	Pop_4	IPmr964	unknown
IPs388	unknown	IPm2077	Pop_7	IPmr843	Pop_8
IPs49	Pop_4	IPm2710	Pop_5	IPmr934	unknown
IPs418	Pop_6	IPm2501	Pop_3	IPmr1069	Pop_4
IPs769	Pop_2	IPm2282	Pop_3	IPmr987	Pop_3
IPs595	Pop_4	IPm2802	Pop_6	IPmr982	Pop_1
IPs640	Pop_1	IPm2594	Pop_2	IPmr774	Pop_2
IPs803	Pop_6	IPm2817	Pop_6	IPmr725	Pop_1
IPs671	Pop_2	IPm66	Pop_6	IPmr800	unknown
IPs684	Pop_7	IPm1591	Pop_4	IPmr817	Pop_4
IPs239	unknown	IPm2687	unknown	IPmr41	Pop_2
IPs319	Pop_3	IPm2016	Pop_5	IPmr730	Pop_1
IPs777	Pop_2	IPm2805	Pop_3	IPmr1005	Pop_8
IPs255	unknown	IPm1604	Pop_4	IPmr996	Pop_7
IPs597	Pop_3	IPm2758	Pop_5	IPmr733	Pop_2
IPs814	Pop_3	IPm2699	Pop_5	IPmr1043	Pop_6
IPs287	Pop_5	IPm283	Pop_2	IPmr1049	Pop_6
IPs593	Pop_4	IPm2782	Pop_8	IPmr841	Pop_5
IPs875	Pop_3	IPm2850	Pop_8	IPmr1035	Pop_6
IPs773	Pop_1	IPm2108	Pop_3	IPmr3	Pop_2
IPs236	Pop_2	IPm2186	Pop_3	IPmr963	Pop_2
IPs318	Pop_3	IPm2135	Pop_5	IPmr926	Pop_2
IPs181	Pop_4	IPm2783	Pop_6	IPmr944	Pop_2
IPs47	Pop_4	IPm2288	Pop_7	IPmr978	unknown
IPs585	Pop_1	IPm2187	Pop_3	IPmr1026	unknown
IPs741	Pop_2	IPm2735	unknown	IPmr878	Pop_6
IPs735	Pop_4	IPm2858	Pop_8	IPmr940	unknown
IPs133	Pop_6	IPm2688	Pop_5	IPmr787	Pop_8
IPs282	Pop_7	IPm2797	Pop_6	IPmr365	Pop_4
IPs862	Pop_3	IPm1603	Pop_7	IPmr974	unknown
IPs594	Pop_4	IPm2789	Pop_7	IPmr845	Pop_8
IPs686	Pop_4	IPm2069	Pop_7	IPmr740	Pop_4
IPs730	Pop_4	IPm99	Pop_4	IPmr887	Pop_6
IPs648	unknown	IPm982	unknown	IPmr771	Pop_8
IPs240	Pop_2	IPm2846	Pop_8	IPmr737	Pop_1
IPs199	Pop_3	IPm362	Pop_6	IPmr417	Pop_4

IPs857	Pop_4	IPm2826	Pop_2	IPmr989	Pop_7
IPs645	Pop_1	IPm2273	Pop_2	IPmr1039	Pop_6
IPs913	Pop_4	IPm2694	Pop_4	IPmr719	unknown
IPs178	Pop_3	IPm2278	unknown	IPmr393	Pop_4
IPs207	Pop_4	IPm2716	Pop_3	IPmr62	Pop_4
IPs292	Pop_5	IPm2575	unknown	IPmr1025	Pop_5
IPs9	Pop_3	IPm2510	Pop_2	IPmr1066	Pop_4
IPs628	Pop_3	IPm2601	Pop_2	IPmr449	Pop_4
IPs928	Pop_5	IPm381	Pop_6	IPmr877	Pop_6
IPs614	Pop_4	IPm2698	unknown	IPmr465	unknown
IPs713	unknown	IPm2665	Pop_5	IPmr452	unknown
IPs243	Pop_4	IPm1536	Pop_5	IPmr1021	Pop_5
IPs228	unknown	IPm2889	Pop_8	IPmr920	Pop_4
IPs244	Pop_3	IPm2854	Pop_8	IPmr718	Pop_1
IPs329	Pop_3	IPm2053	Pop_5	IPmr805	Pop_8
IPs186	Pop_4	IPm2639	Pop_4	IPmr904	Pop_5
IPs155	unknown	IPm2209	Pop_5	IPmr854	Pop_6
IPs674	Pop_3	IPm2257	Pop_4	IPmr1037	Pop_6
IPs694	Pop_5	IPm1554	Pop_4	IPmr713	Pop_6
IPs588	Pop_7	IPm1606	Pop_4	IPmr754	Pop_4
IPs795	unknown	IPm2650	Pop_4	IPmr852	Pop_8
IPs275	Pop_7	IPm2792	Pop_3	IPmr759	Pop_4
IPs245	Pop_4	IPm2537	Pop_2	IPmr847	unknown
IPs176	Pop_3	IPm2198	Pop_3	IPmr700	Pop_6
IPs69	Pop_4	IPm2064	Pop_5	IPmr837	Pop_5
IPs908	unknown	IPm2005	Pop_5	IPmr1070	Pop_4
IPs151	Pop_3	IPm2824	Pop_6	IPmr897	Pop_6
IPs634	Pop_1	IPm2577	Pop_2	IPmr706	Pop_1
IPs415	Pop_3	IPm2004	Pop_5	IPmr786	unknown
IPs838	Pop_4	IPm388	Pop_6	IPmr868	Pop_6
IPs584	Pop_2	IPm2661	Pop_1	IPmr851	Pop_4
IPs745	Pop_4	IPm2680	Pop_5	IPmr927	Pop_2
IPs158	Pop_2	IPm1598	Pop_4	IPmr866	Pop_6
IPs669	Pop_7	IPm2784	Pop_2	IPmr1016	Pop_5
IPs100	Pop_3	IPm2832	unknown	IPmr427	Pop_2
IPs855	Pop_4	IPm2727	Pop_3	IPmr812	Pop_4
IPs796	Pop_3	IPm2780	Pop_5	IPmr838	Pop_5
IPs689	Pop_7	IPm2656	Pop_1	IPmr414	Pop_4
IPs147	Pop_1	IPm2270	Pop_1	IPmr739	Pop_4
IPs213	Pop_4	IPm2721	Pop_2	IPmr988	Pop_7

IPs99	unknown	IPm2689	Pop_5	IPmr1074	Pop_4
IPs858	Pop_4	IPm9	Pop_6	IPmr985	Pop_3
IPs83	Pop_3	IPm2875	Pop_8	IPmr980	Pop_2
IPs870	Pop_2	IPm2087	unknown	IPmr855	Pop_6
IPs261	Pop_4	IPm2592	unknown	IPmr1065	Pop_4
IPs77	Pop_4	IPm2730	unknown	IPmr972	Pop_2
IPs227	unknown	IPm2095	Pop_3	IPmr901	Pop_1
IPs339	Pop_4	IPm2813	Pop_5	IPmr913	Pop_4
IPs274	Pop_7	IPm2122	Pop_5	IPmr936	Pop_2
IPs293	Pop_4	IPm1621	Pop_4	IPmr986	Pop_1
IPs835	Pop_3	IPm2809	Pop_6	IPmr794	Pop_2
IPs606	unknown	IPm2821	Pop_1	IPmr391	Pop_4
IPs775	unknown	IPm2124	Pop_7	IPmr1018	Pop_5
IPs160	Pop_3	IPm2847	Pop_8	IPmr917	Pop_4
IPs587	Pop_5	IPm2753	Pop_5	IPmr894	unknown
IPs714	Pop_2	IPm2164	Pop_7	IPmr768	Pop_2
IPs695	Pop_5	IPm2866	Pop_6	IPmr721	Pop_6
IPs44	Pop_3	IPm1609	Pop_4	IPmr1063	Pop_4
IPs98	Pop_4	IPm2532	unknown	IPmr983	Pop_2
IPs182	Pop_3	IPm2540	Pop_2	IPmr1017	Pop_5
IPs919	Pop_3	IPm384	Pop_7	IPmr753	unknown
IPs208	Pop_1	IPm2267	Pop_3	IPmr699	Pop_7
IPs766	Pop_6	IPm366	Pop_6	IPmr1042	Pop_6
IPs653	Pop_6	IPm2119	Pop_3	IPmr1000	Pop_7
IPs929	Pop_5	IPm2517	Pop_4	IPmr790	Pop_2
IPs827	unknown	IPm2635	Pop_4	IPmr993	Pop_7
IPs358	Pop_4	IPm2522	unknown	IPmr770	Pop_8
IPs212	Pop_2	IPm62	Pop_6	IPmr1008	Pop_1
IPs592	Pop_7	IPm2587	Pop_2	IPmr1057	unknown
IPs710	Pop_2	IPm27	Pop_2	IPmr995	unknown
IPs891	Pop_3	IPm2755	Pop_5	IPmr758	Pop_8
IPs641	Pop_1	IPm2552	Pop_2	IPmr867	Pop_1
IPs618	Pop_4	IPm2880	Pop_8	IPmr776	unknown
IPs709	Pop_4	IPm2620	Pop_6	IPmr743	Pop_4
IPs881	Pop_1	IPm2093	Pop_5	IPmr778	unknown
IPs699	Pop_4	IPm2140	Pop_4	IPmr842	Pop_4
IPs383	unknown	IPm2123	Pop_7	IPmr859	Pop_6
IPs824	Pop_2	IPm2230	Pop_4	IPmr1036	Pop_6
IPs622	Pop_3	IPm2507	unknown	IPmr853	unknown
IPs21	Pop_4	IPm2811	unknown	IPmr849	Pop_8

IPs749	Pop_2	IPm2236	unknown	IPmr1040	Pop_6
IPs254	Pop_2	IPm2903	Pop_2	IPmr977	Pop_2
IPs5	Pop_7	IPm2083	Pop_7	IPmr1064	Pop_4
IPs4	Pop_7	IPm2769	Pop_2	IPmr1002	Pop_7
IPs599	Pop_3			IPmr925	unknown
IPs737	unknown			IPmr991	Pop_7
IPs23	Pop_3			IPmr773	Pop_2
IPs627	unknown			IPmr501	Pop_8

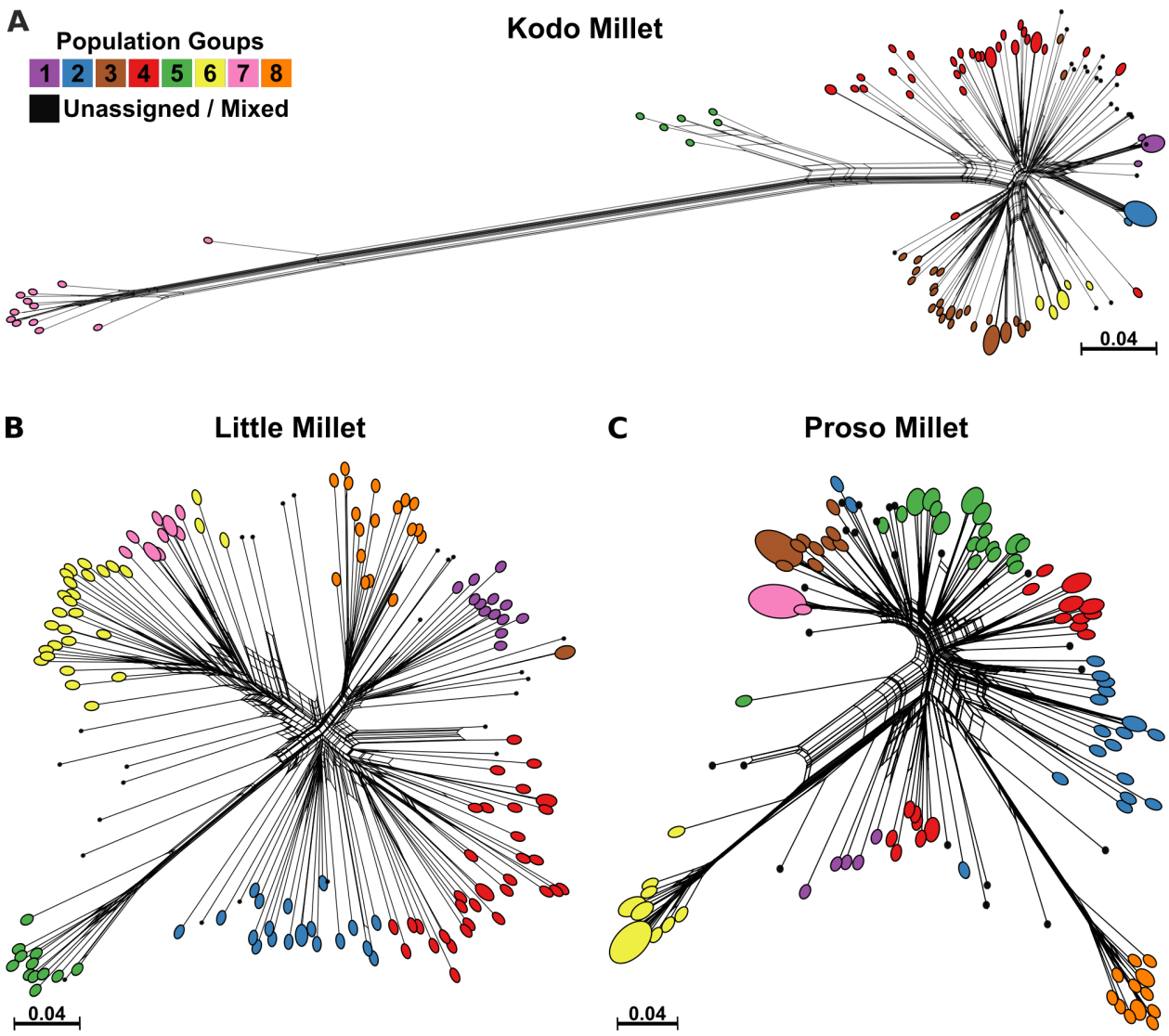

Figure 2.1
Phylogenetic trees for Kodo Millet (A), Little Millet (B), and Proso Millet (C), colored by population calls from fastStructure for accessions that fit into a population with at least 60% fit. Each accession has one line that ends in a point and has branching to show alternative branching options. Where multiple points of the same population grouped closely together a single oval with a size proportional to the number of accession it encompasses was used.

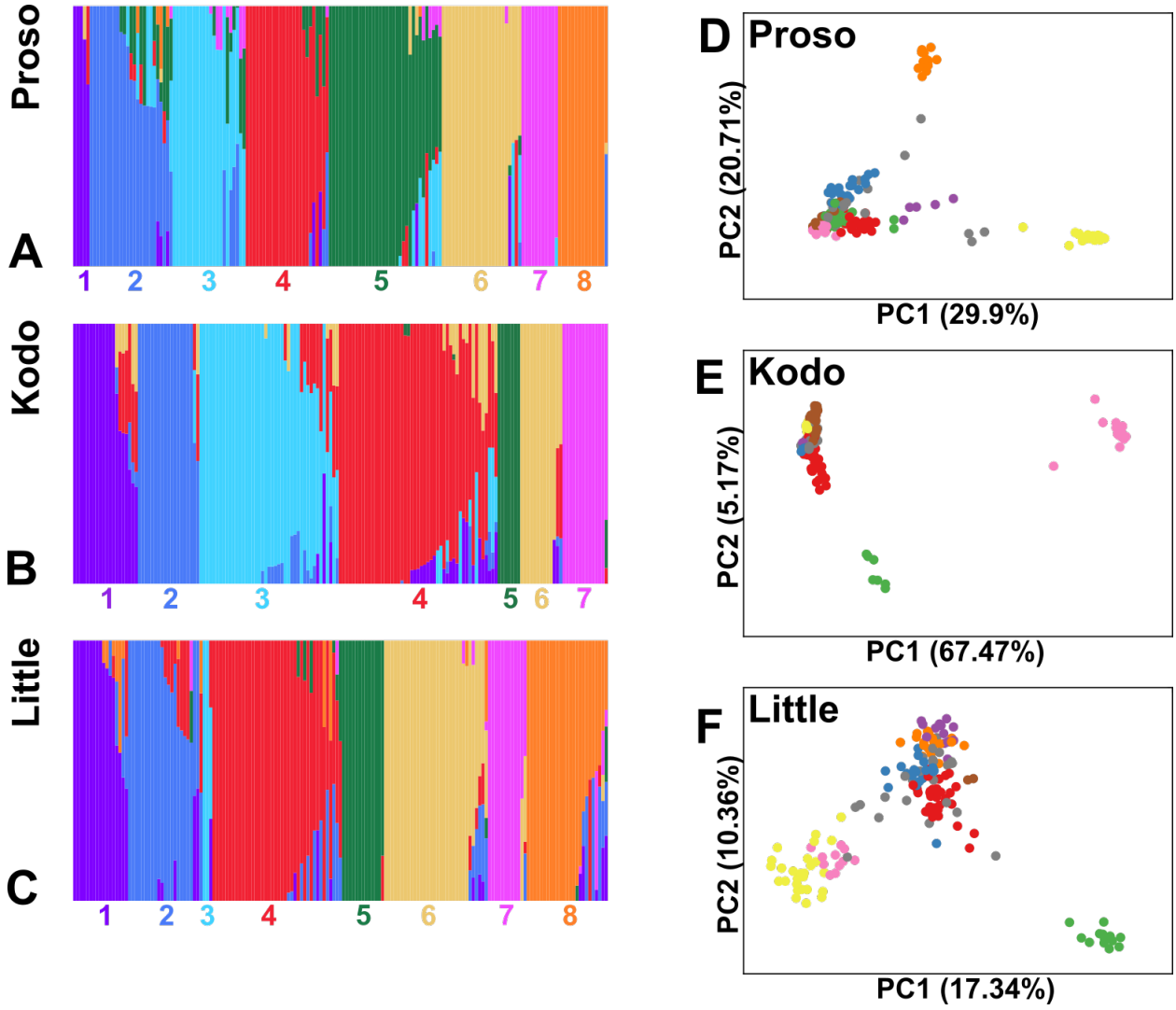

Figure 2.2
A-C Visual representation of how well individual accessions fit into putative populations from distruct, a part of the fastStructure program. The vertical bars each represent a single accession, and the colors correspond to populations as listed on the bottom of each subfigure (Proso-A, Kodo-B, and Little-C). D-F Principle coordinate analysis based off genetic distances and colored by putative populations for Proso Millet (A), Kodo Millet (B), and Little Millet (C).

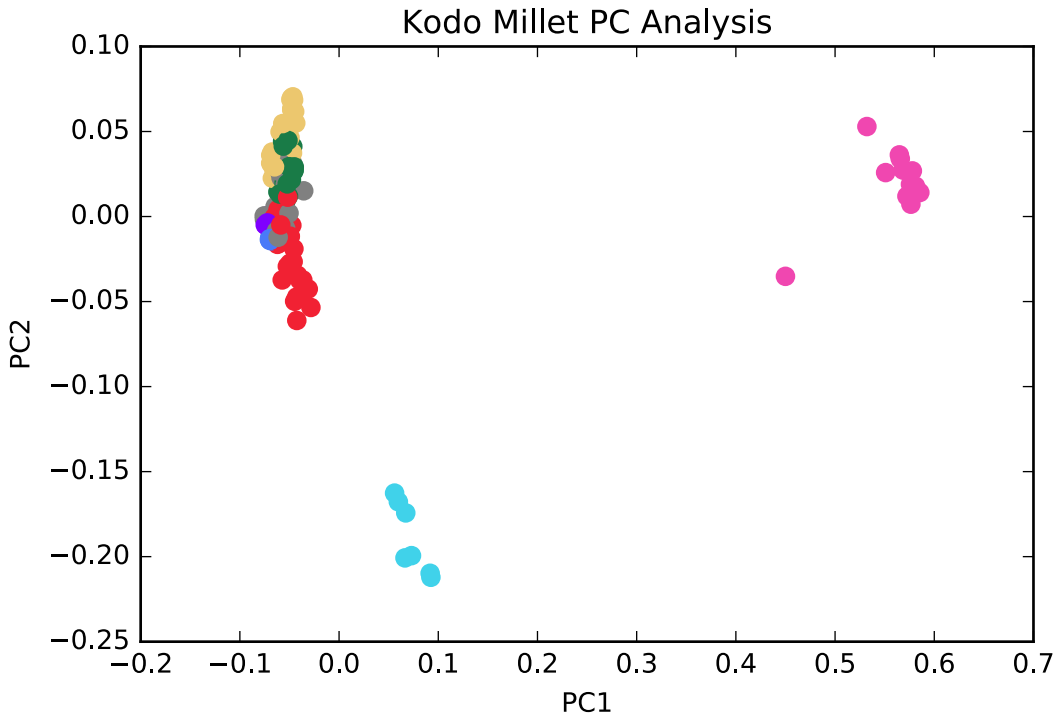

Figure 2.3
Principal coordinate analysis for kodo millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Figure 2E).

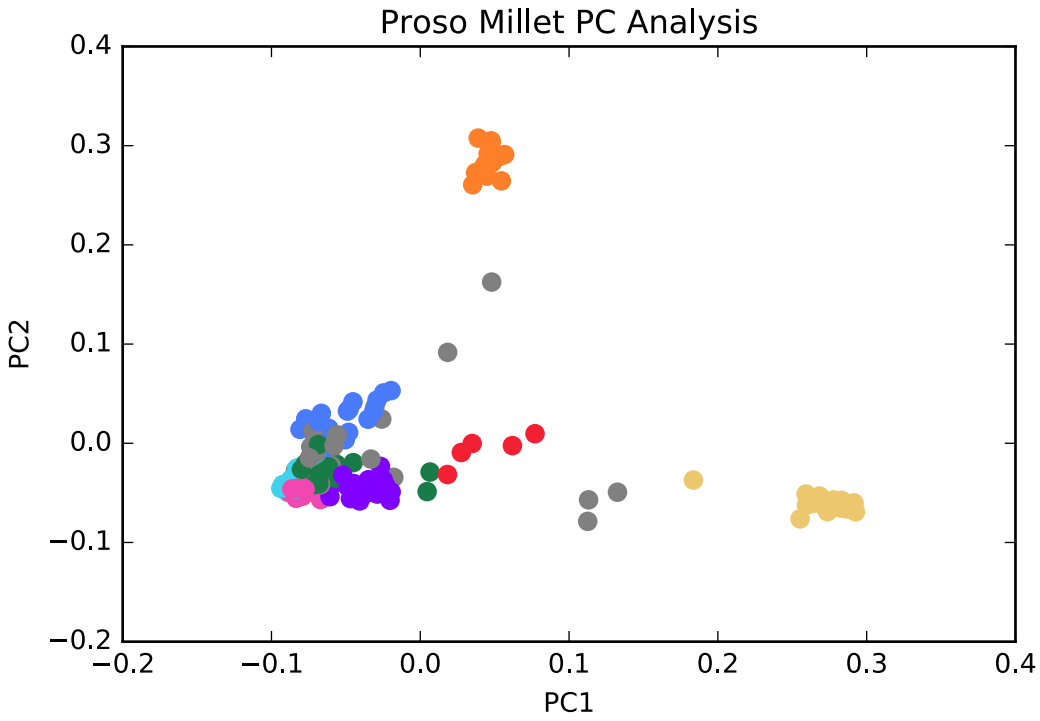

Figure 2.4
Principal coordinate analysis for little millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Figure 2D).

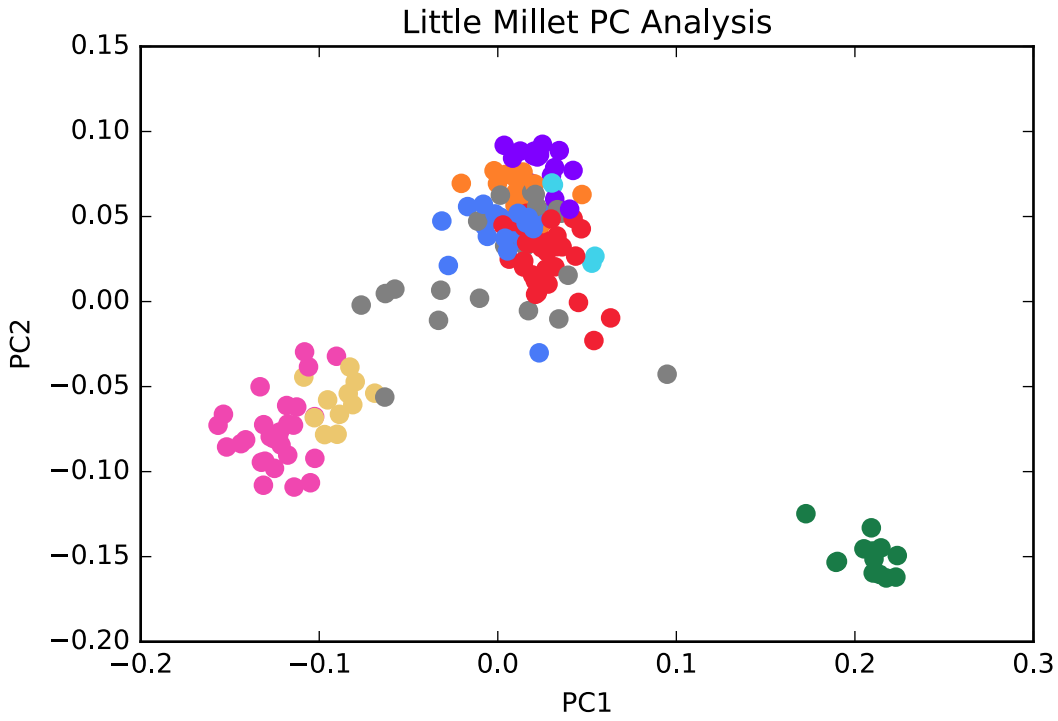

Figure 2.5
Principal coordinate analysis for kodo millet using imputed genotype data. Population structure and groupings are almost identical to those made with unimputed data (main text Figure 2F).

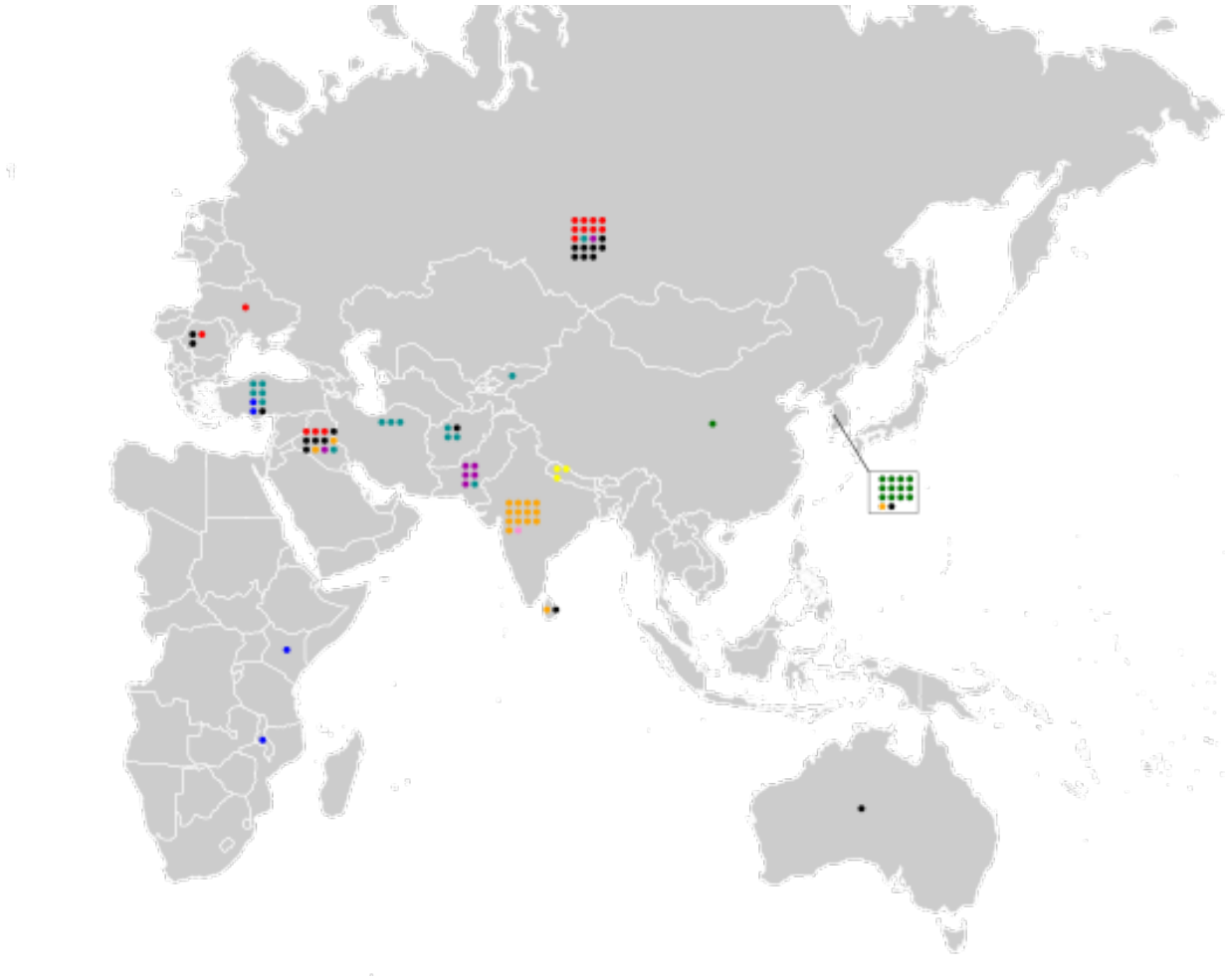

Figure 2.6
Plotting of Proso millet accessions by country of origin. Within-country coordinates were not available, so points within each country are displayed as a grid. Points are colored by subpopulation (see main text Figure 2); black indicates the accession didn't fit into any given population.

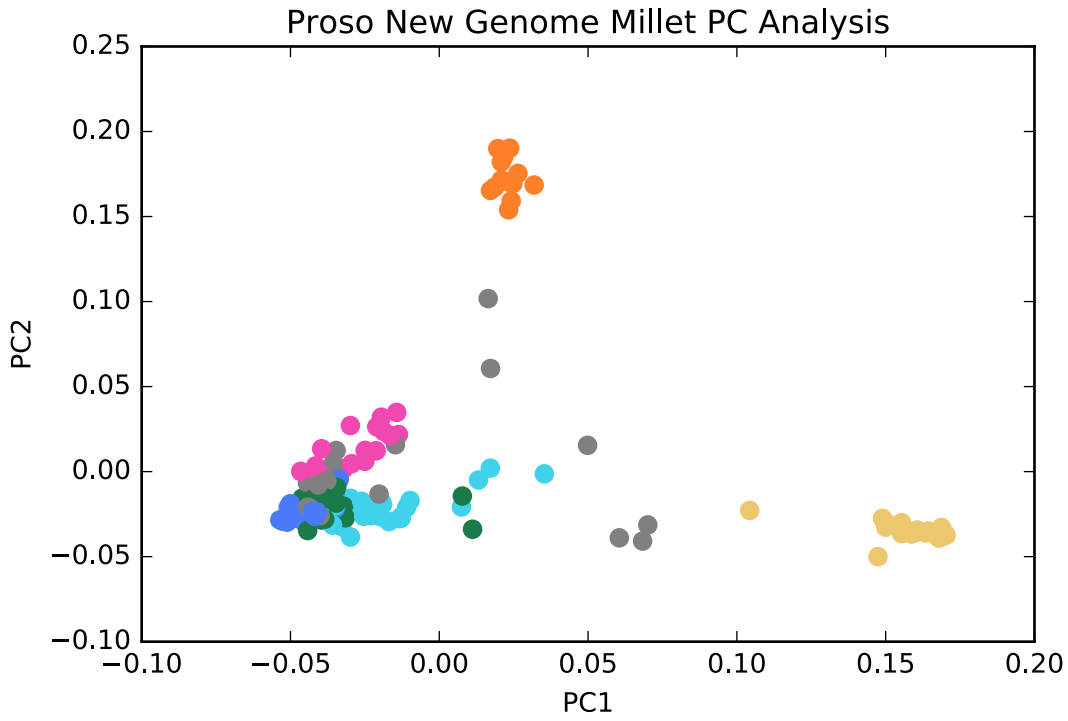

Figure 2.7
Principal Coordinate analysis of Proso millet using the recently published reference genome (Zou et al., 2019). These results are consistent with those done without a reference genome (main text Figure 2D).

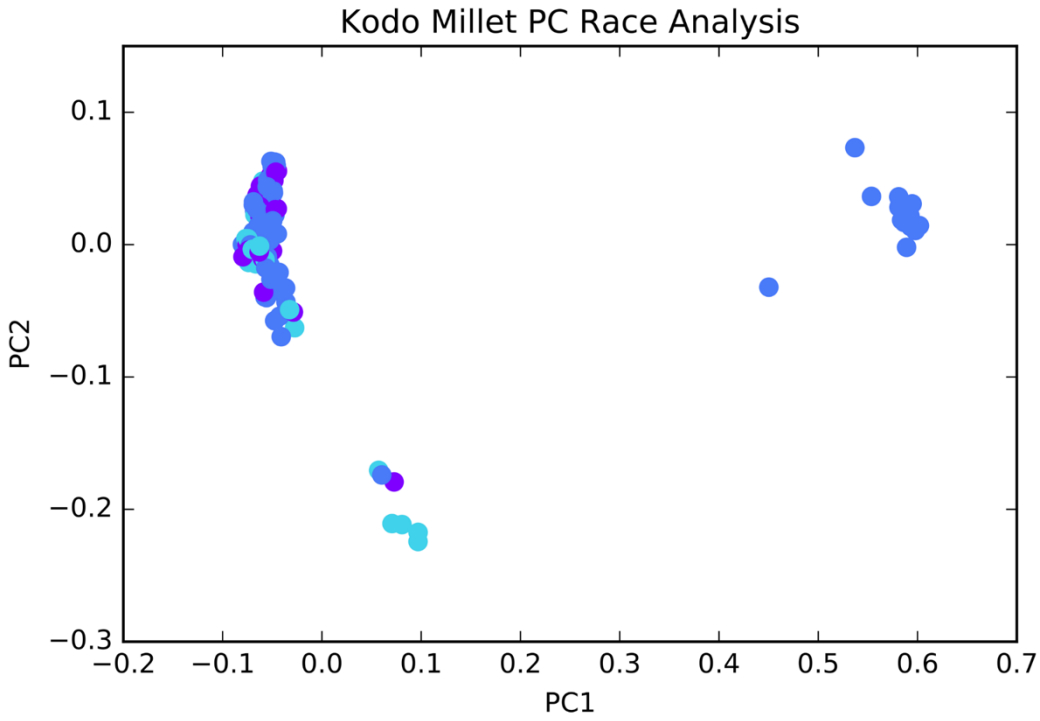

Figure 2.8
PC analysis of kodo millet using race to color the accessions. Races do not correspond with genetic clusters, indicating that they do not reflect genetic relationships.

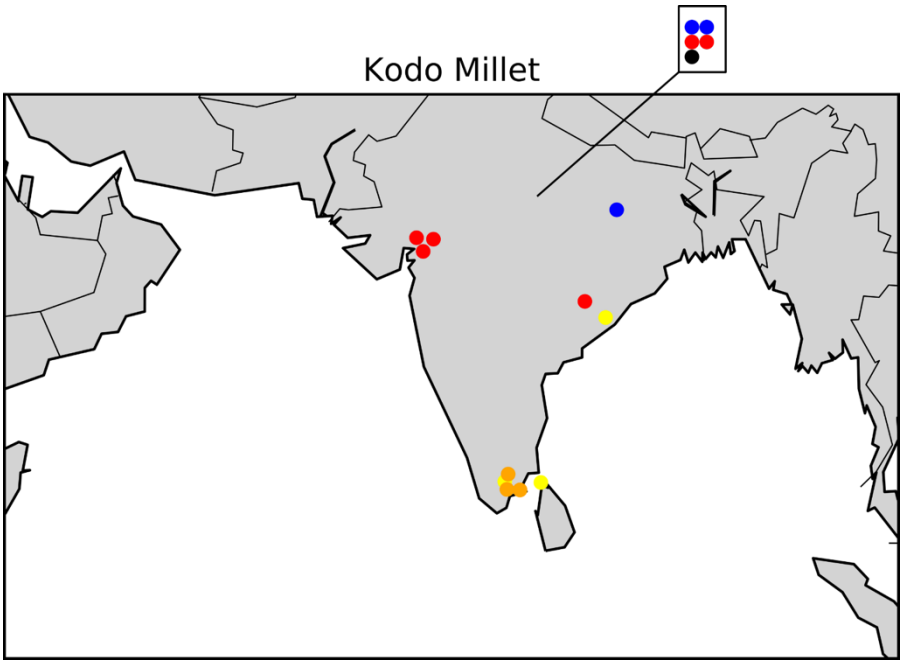

Figure 2.9
Kodo millet accessions plotted by GPS coordinates. Only a few kodo millet accessions had GPS data available, though those that do are relatively consistent with subpopulations clustering by region of origin. Points are colored by subpopulation (see main text Figure 2); black indicates the accession didn't fit into any given population.

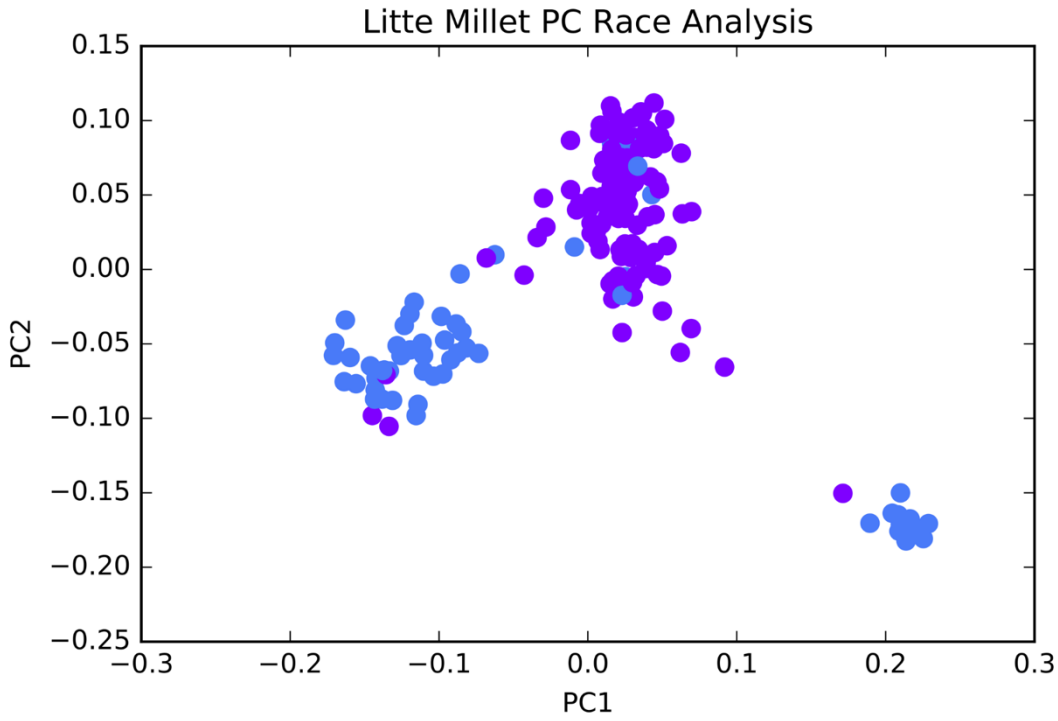

Figure 2.10
PC analysis of Little millet using race to color the accessions. Race does not show the diversity of genetic clusters, and as such it is not a good reflection of genetic relationships.

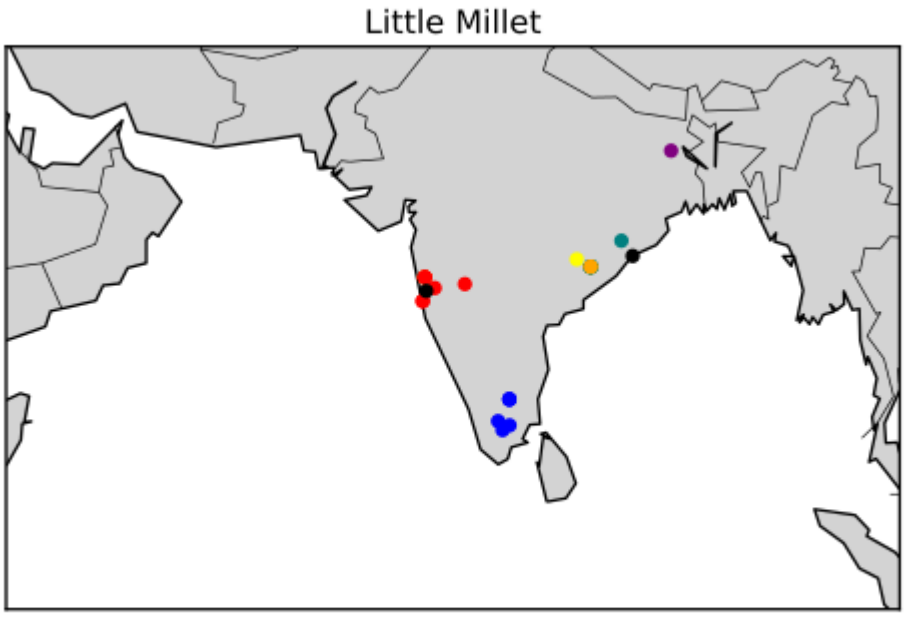

Figure 2.11
Little millet accessions plotted by GPS coordinates. Only a few little millet accessions had GPS data available, though those that do are relatively consistent with subpopulations clustering by region of origin. Points are colored by subpopulation (see main text Figure 2); black indicates the accession didn't fit into any given population.

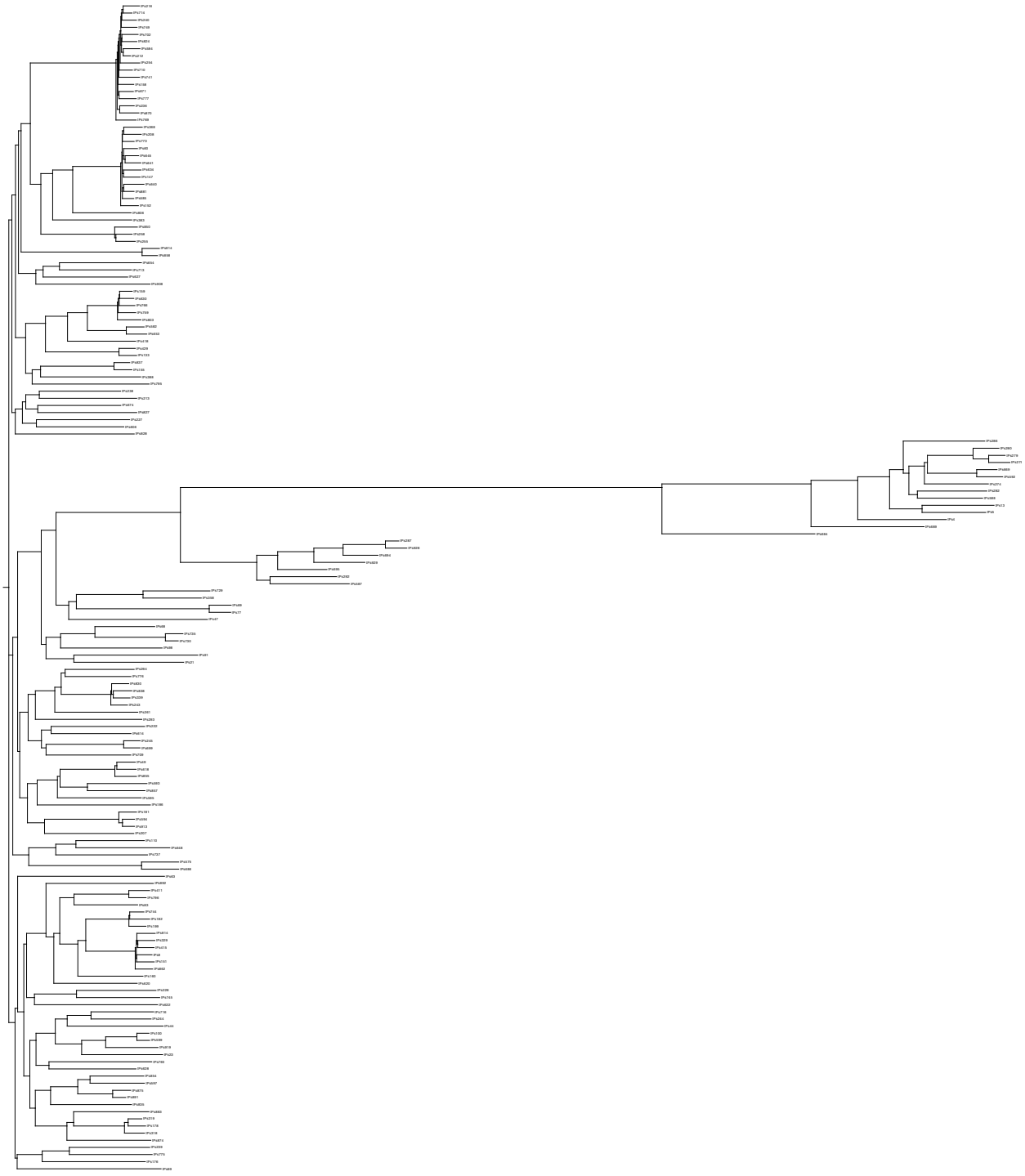

Figure 2.12: Kodo Millet Dendogram
A neighbor-joining tree of kodo millet was constructed in TASSEL based on genetic polymorphisms and exported using Archaeopteryx. Samples show clustering consistent with other kodo millet figures (2.1 & 2.2)

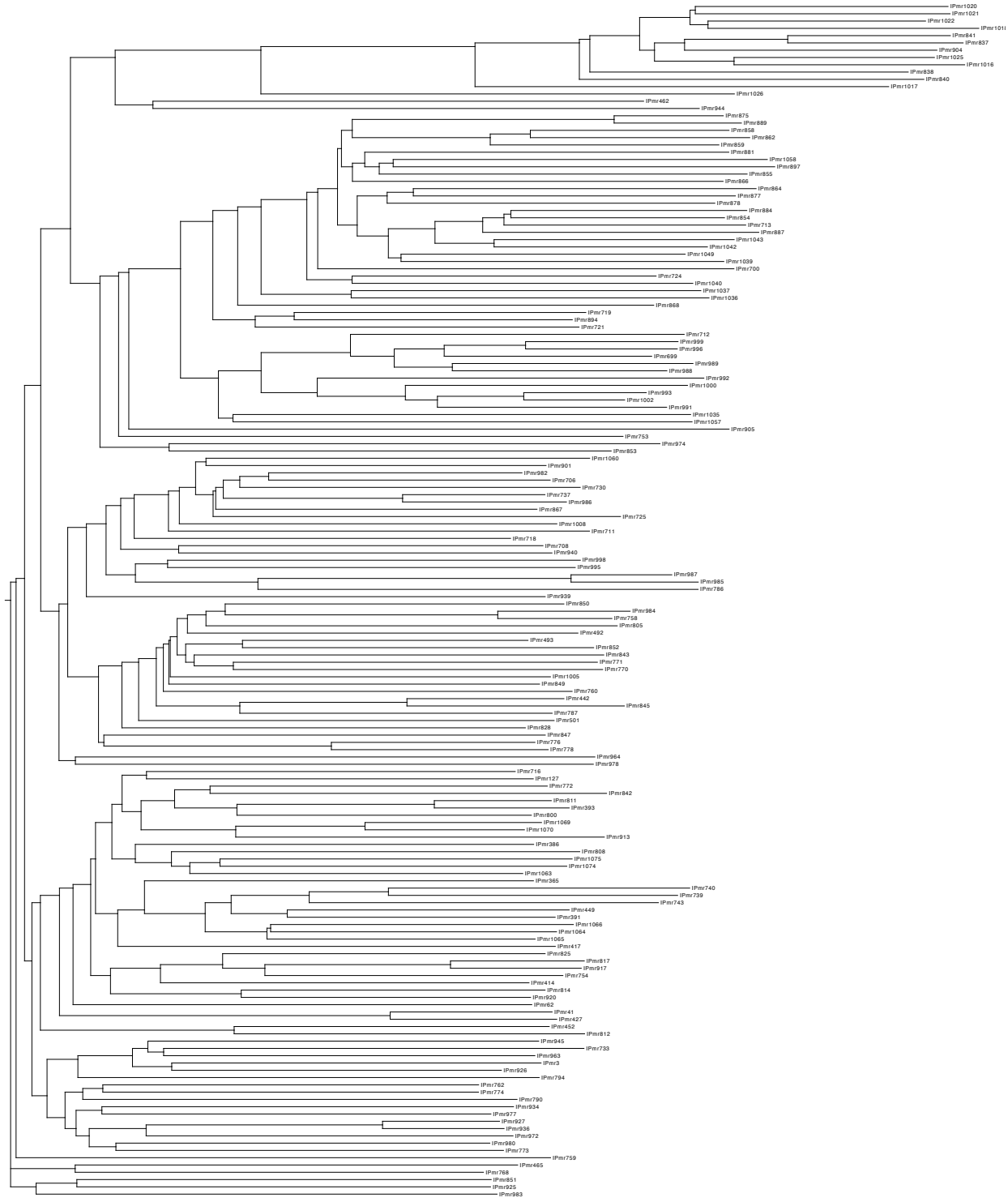

Figure 2:13: Little Millet Dendogram
A neighbor-joining tree of little millet was constructed in TASSEL based on genetic polymorphisms and exported using Archaeopteryx. Samples show clustering consistent with other little millet figures (2.1 & 2.2)

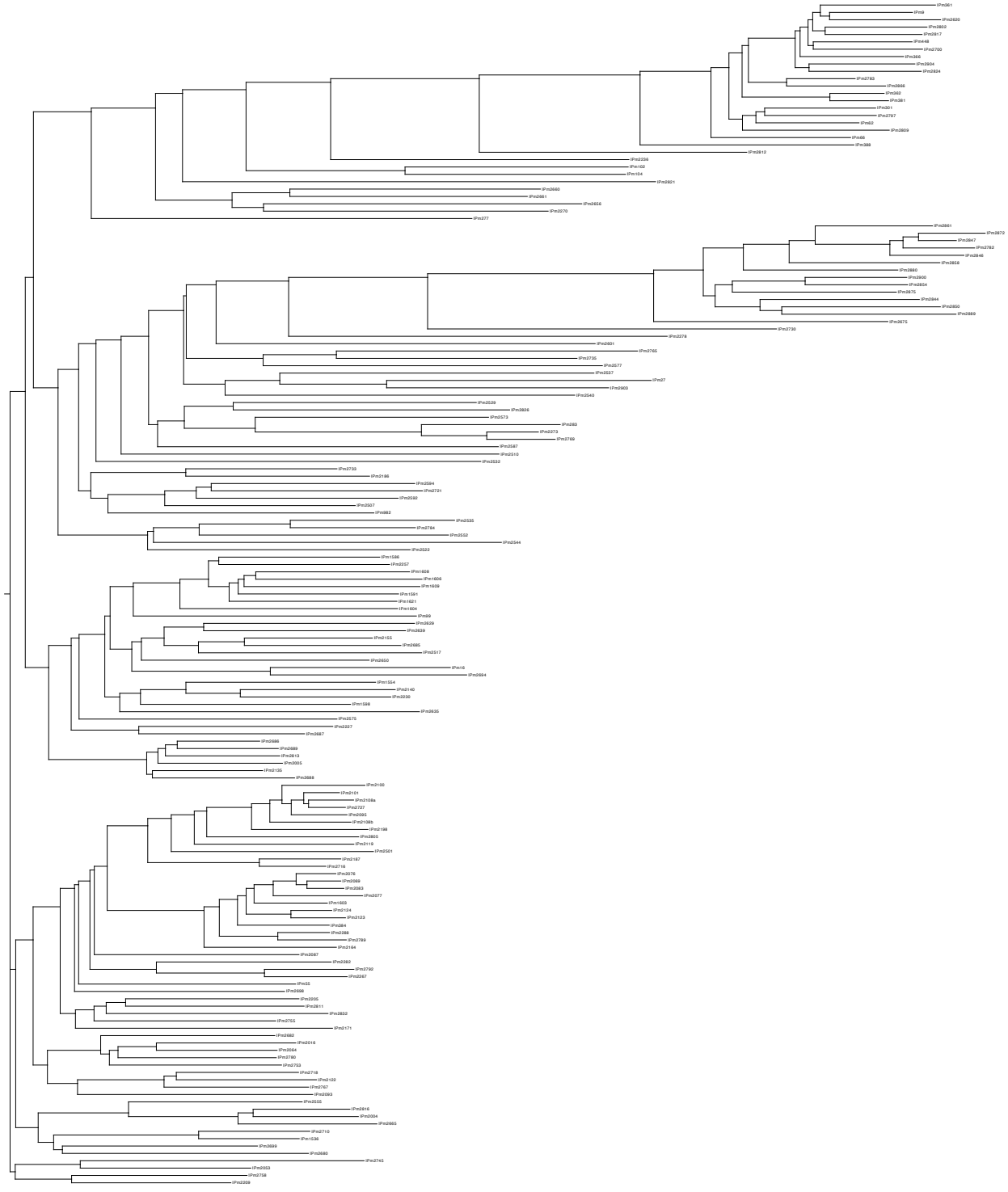

Figure 2:14: Proso Millet Dendogram
A neighbor-joining tree of proso millet was constructed in TASSEL based on genetic polymorphisms and exported using Archaeopteryx. Samples show clustering consistent with other proso millet figures (2.1 & 2.2)

REFERENCES

Alqudah, A.M. and T. Schnurbusch. 2014. Awn primordium to tipping is the most decisive developmental phase for spikelet survival in barley. Functional Plant Biology 41: 424. doi:10.1071/fp13248.

Baltensperger, D.D. 2002. Progress with Proso, Pearl and Other Millets. Trends in new crops and new uses.

Bennetzen, J.L., J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, et al. 2012. Reference genome sequence of the model plant Setaria. Nat Biotechnol 30: 555-561. doi:10.1038/nbt.2196.

Bettinger, R.L., L. Barton and C. Morgan. 2010a. The origins of food production in north China: A different kind of agricultural revolution. Evolutionary Anthropology: Issues, News, and Reviews 19: 9-21.

Bettinger, R.L., L. Barton, C. Morgan, F. Chen, H. Wang, T.P. Guilderson, et al. 2010b. The transition to agriculture at Dadiwan, People's Republic of China. Current Anthropology 51: 703-714.

Bettinger, R.L., L. Barton, P.J. Richerson, R. Boyd, H. Wang and W. Choi. 2007. The transition to agriculture in northwestern China. Developments in Quaternary Sciences 9: 83-101.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Brown, R.H. 1999. Agronomic implications of C4 photosynthesis. C4 plant biology: 473-507.

Bryant, D. and V. Moulton. 2003. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21: 255-265. doi:10.1093/molbev/msh018.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, et al. 2009. The Genetic Architecture of Maize Flowering Time. Science 325: 714-718. doi:10.1126/science.1174276.

Cobb, J.N., G. DeClerck, A. Greenberg, R. Clark and S. McCouch. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype–phenotype relationships and its relevance to crop improvement. Theoretical and Applied Genetics 126: 867-887. doi:10.1007/s00122-013-2066-0.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, et al. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22: 961-975. doi:10.1016/j.tplants.2017.08.011.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools, Bioinformatics, 27:2156-2158, https://doi.org/10.1093/bioinformatics/btr330

de Wet, J.M.J., D.E. Brink, K.E.P. Rao and M.H. Mengesha. 1983. Diversity in Kodo millet,Paspalum scrobiculatum. Economic Botany 37: 159-163. doi:10.1007/bf02858779.

Dwivedi, S., H. Upadhyaya, S. Senthilvel, C. Hash, K. Fukunaga, X. Diao, et al. 2012. Millets: Genetic and genomic resources. Plant Breed.: 247-375.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6: e19379.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS one 9: e90346.

Glez-Peña, D., D. Gómez-Blanco, M. Reboiro-Jato, F. Fdez-Riverola and D. Posada. 2010. ALTER: program-oriented conversion of DNA and protein alignments. Nucleic Acids Research 38: W14-W18. doi:10.1093/nar/gkq321.

Goron, T.L. and M.N. Raizada. 2015. Genetic diversity and genomic resources available for the small millet crops to accelerate a New Green Revolution. Frontiers in Plant Science 6. doi:10.3389/fpls.2015.00157.

Habiyaremye, C., J.B. Matanguihan, J. D'Alpoim Guedes, G.M. Ganjyal, M.R. Whiteman, K.K. Kidwell, et al. 2017. Proso Millet (Panicum miliaceum L.) and Its Potential for Cultivation in the Pacific Northwest, U.S.: A Review. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.01961.

Han, M. V., & Zmasek, C. M. 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 10, 356.

Hammer, K., J. Heller and J. Engels. 2001. Monographs on underutilized and neglected crops. Genetic Resources and Crop Evolution 48: 3-5. doi:10.1023/a:1011253924058.

Hittalmani, S., H.B. Mahesh, M.D. Shirke, H. Biradar, G. Uday, Y.R. Aruna, et al. 2017. Genome and Transcriptome sequence of Finger millet (Eleusine coracana (L.) Gaertn.) provides insights into drought tolerance and nutraceutical properties. BMC Genomics 18: 465. doi:10.1186/s12864-017-3850-z.

Hunt, H.V., F. Badakshi, O. Romanova, C.J. Howe, M.K. Jones and J.S. Heslop-Harrison. 2014. Reticulate evolution in Panicum (Poaceae): the origin of tetraploid broomcorn millet, P. miliaceum. J Exp Bot 65: 3165-3175. doi:10.1093/jxb/eru161.

Hunt, H.V., M.G. Campana, M.C. Lawes, Y.J. Park, M.A. Bower, C.J. Howe, et al. 2011. Genetic diversity and phylogeography of broomcorn millet (Panicum miliaceum L.) across Eurasia. Mol Ecol 20: 4756-4771. doi:10.1111/j.1365-294X.2011.05318.x.

Hunter, J.D. 2007. Matplotlib: A 2D Graphics Environment. Computing in Science and Engineering, Comput. Sci. Eng.: 90. doi:10.1109/MCSE.2007.55.

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23: 254-267.

Inkskape [Computer software]. 2015. Retrieved from https://inkscape.org/release/0.91/

Jia, G., X. Huang, H. Zhi, Y. Zhao, Q. Zhao, W. Li, et al. 2013. A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nat Genet 45: 957-961. doi:10.1038/ng.2673.

Kalinova, J. and J. Moudry. 2006. Content and quality of protein in proso millet (Panicum miliaceum L.) varieties. Plant Foods for Human Nutrition 61: 43.

Khoshbakht, K. and K. Hammer. 2008. How many plant species are cultivated? Genetic resources and crop evolution 55: 925-928.

Kubesova, M., L. Moravcova, J. Suda, V. Jarosik and P. Pysek. 2010. Naturalized plants have smaller genomes than their non-invading relatives: a flow cytometric analysis of the Czech alien flora. PRESLIA -PRAHA-. 82: 81-96.

Kushwaha, H., K.W. Jillo, V.K. Singh, A. Kumar and D. Yadav. 2014. Assessment of genetic diversity among cereals and millets based on PCR amplification using Dof (DNA binding

with One Finger) transcription factor gene-specific primers. Plant Systematics and Evolution 301: 833-840. doi:10.1007/s00606-014-1095-8.

Lu, F., Copenhaver, G.P., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, et al. 2013. Switchgrass Genomic Diversity, Ploidy, and Evolution: Novel Insights from a Network-Based SNP Discovery Protocol. PLoS Genetics 9: e1003215. doi:10.1371/journal.pgen.1003215.

Lu, H., J. Zhang, K.-b. Liu, N. Wu, Y. Li, K. Zhou, et al. 2009. Earliest domestication of common millet (Panicum miliaceum) in East Asia extended to 10,000 years ago. Proceedings of the National Academy of Sciences 106: 7367-7372.

M'Ribu, H.K. and K.W. Hilu. 1996. Application of random amplified polymorphic DNA to study genetic diversity in Paspalum scrobiculatum L. (Kodo millet, Poaceae). Genetic Resources and Crop Evolution 43: 203-210. doi:10.1007/bf00123272.

Ma, X., F. Feng, H. Wei, H. Mei, K. Xu, S. Chen, et al. 2016. Genome-Wide Association Study for Plant Height and Grain Yield in Rice under Contrasting Moisture Regimes. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.01801.

Mace, E.S., K.K. Buhariwalla, H.K. Buhariwalla and J.H. Crouch. 2012. A high-throughput DNA extraction protocol for tropical molecular breeding programs. Plant Molecular Biology Reporter 21: 459-460. doi:10.1007/bf02772596.

Mace, E.S., C.H. Hunt and D.R. Jordan. 2013. Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. Theoretical and Applied Genetics 126: 1377-1395. doi:10.1007/s00122-013-2059-z.

McKinney W. 2010. Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56.

Mengesha, M.H. 1966. Chemical composition of teff (Eragrostis tef) compared with that of wheat, barley and grain sorghum. Economic Botany 20: 268-273.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., and Myles, S. (2015). LinkImpute: Fast and Accurate Genotype Imputation for Nonmodel Organisms. G3 (Bethesda, Md.), 5(11), 2383-90. doi:10.1534/g3.115.021667

Morris, G.P., P. Ramu, S.P. Deshpande, C.T. Hash, T. Shah, H.D. Upadhyaya, et al. 2012. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences 110: 453-458. doi:10.1073/pnas.1215985110.

M. S. Swaminathan Research Foundation. 2000. Ninth annual report Centre for Research on Sustainable Agricultural and Rural Development.

Naylor, R.L., W.P. Falcon, R.M. Goodman, M.M. Jahn, T. Sengooba, H. Tefera, et al. 2004. Biotechnology in the developing world: a case for increased investments in orphan crops. Food Policy 29: 15-44. doi:10.1016/j.foodpol.2004.01.002.

Oliphant, T.E. 2006. A guide to NumPy, USA: Trelgol Publishing.

Peiffer, J.A., M.C. Romay, M.A. Gore, S.A. Flint-Garcia, Z. Zhang, M.J. Millard, et al. 2014. The Genetic Architecture Of Maize Height. Genetics 196: 1337-1356. doi:10.1534/genetics.113.159152.

Prasad, M. 2017. The Foxtail Millet Genome. doi:10.1007/978-3-319-65617-5.

Prasada Rao, K., J. De Wet, V. Gopal Reddy and M. Mengesha. 1993. Diversity in the small millets collection at ICRISAT. Advances in small millets. Edited by KW Riley, SC Gupta, A. Seetharam, and JN Mushonga. Oxford and IBH Publishing Co. Pvt. Ltd., New Delhi, India.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

Raj, A., M. Stephens and J.K. Pritchard. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197: 573-589.

Rajput, S.G. and D.K. Santra. 2016. Evaluation of genetic diversity of proso millet germplasm available in the United States using simple-sequence repeat markers. Crop Science 56: 2401-2409.

Roshevits, R.Y. 1980. Grasses: an introduction to the study of fodder and cereal grasses.

Saha, D., M.V.C. Gowda, L. Arya, M. Verma and K.C. Bansal. 2016. Genetic and Genomic Resources of Small Millets. Critical Reviews in Plant Sciences 35: 56-79. doi:10.1080/07352689.2016.1147907.

Saleh, A.S., Q. Zhang, J. Chen and Q. Shen. 2013. Millet grains: nutritional quality, processing, and potential health benefits. Comprehensive Reviews in Food Science and Food Safety 12: 281-295.

Sehgal, D., S.K. Parida, L. Skot, R. Singh, R.K. Srivastava, S.P. Das, et al. 2015. Exploring Potential of Pearl Millet Germplasm Association Panel for Association Mapping of Drought Tolerance Traits. Plos One 10: e0122165. doi:10.1371/journal.pone.0122165.

Shanahan, J., R. Anderson and B. Greb. 1988. Productivity and water use of proso millet grown under three crop rotations in the Central Great Plains. Agronomy journal 80: 487-492.

Hatakeyama, M., S. Aluri, M.T. Balachadran, S.R. Sivarajan, A. Patrignani, S. Grüter, et al. 2017. Multiple hybrid de novo genome assembly of nger millet, an orphan allotetraploid crop. DNA Res. 25:39–47. doi:10.1093/ dnares/dsx036

Tange, O. 2011. GNU Parallel - The Command-Line Power Tool. The USENIX Magazine, February 2011:42-47.

Upadhyaya, H., S. Dwivedi, S. Singh, S. Singh, M. Vetriventhan and S. Sharma. 2014. Forming core collections in barnyard, kodo, and little millets using morphoagronomic descriptors. Crop Science 54: 2673-2682.

Upadhyaya, H., S. Sharma, C. Gowda, V.G. Reddy and S. Singh. 2011. Developing proso millet (Panicum miliaceum L.) core collection using geographic and morpho-agronomic data. Crop and Pasture Science 62: 383-389.

Varshney, R.K., J.M. Ribaut, E.S. Buckler, R. Tuberosa, J.A. Rafalski and P. Langridge. 2012. Can genomics boost productivity of orphan crops? Nature Biotechnology 30: 1172-1176.

Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, et al. 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nature Biotechnology 35: 969-976. doi:10.1038/nbt.3943.

Vetriventhan, M, and Upadhyaya, H. D. 2018. Diversity and trait-specific sources for productivity and nutritional traits in the global proso millet (Panicum miliaceum L.) germplasm collection. Crop Journal, Vol 6, Iss 5, Pp 451-463 (2018), (5), 451. https://doi.org/10.1016/j.cj.2018.04.002

Wallace, J.G. and S.E. Mitchell. 2017. Genotyping-by-Sequencing. 64-77. doi:10.1002/cppb.20042.

Yadav, Y., G.R. Lavanya, S. Pandey, M. Verma, C. Ram and L. Arya. 2016. Neutral and functional marker based genetic diversity in kodo millet (Paspalum scrobiculatum L.). Acta Physiologiae Plantarum 38. doi:10.1007/s11738-016-2090-1.

Yue, H., L. Wang, H. Liu, W. Yue, X. Du, W. Song, et al. 2016. De novo Assembly and Characterization of the Transcriptome of Broomcorn Millet (Panicum miliaceum L.) for Gene Discovery and Marker Development. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.01083.

Zhang, G., X. Liu, Z. Quan, S. Cheng, X. Xu, S. Pan, et al. 2012. Genome sequence of foxtail millet (Setaria italica) provides insights into grass evolution and biofuel potential. Nature Biotechnology 30: 549-554. doi:10.1038/nbt.2195.

Zhang, Z., E. Ersoz, C.-Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42: 355-360. doi:10.1038/ng.546.

Zhao, Z. 2005. Palaeoethnobotany and its new achievements in China. Kaogu (Archaeology) 7: e49.

Zou, C., L. Li, D. Miki, D. Li, Q. Tang, L. Xiao, et al. 2019. The genome of broomcorn millet. Nature Communications 10: 436. doi:10.1038/s41467-019-08409-5.

CHAPTER 3

THE USE OF GENOMIC SELECTION TO IMPROVE RESISTANCE TO *STRIGA* AND DOWNY MILDEW IN PEARL MILLET (*CENCHRUS AMERICANUS* (L.) M.)

[1] Matthew S. Johnson, [1] Boubacar A. Kountche, Charles Thomas Hash, Yves Vigouroux, Salim Al-Babili, Jason Wallace, Bettina I. G Haussmann. To be submitted to *The Plant Genome*.

Abstract

Pearl millet (*Cenchrus americanus* (L.) M.) is an important food source for millions of people around the world, especially in Sub-Saharan Africa. It can produce valuable yields in extremely harsh environments, but its production can be greatly reduced by factors such as the parasitic weed *Striga* and the disease downy mildew. Such stresses can cause issues of food insecurity for farmers and communities that rely on this crop. In the regions where pearl millet is most often grown it can be difficult for farmers to afford needed herbicides and fungicides, which leaves them often relying on manual labor. Breeding improved varieties resistant to *Striga* and downy mildew would give farmers a tool to better fight against these issues. We used genome-wide marker data and phenotyping data from multiple field locations to investigate the potential of genomic selection to increase the speed and efficiency of breeding pearl millet varieties resistant to *Striga* and downy mildew. Our results showed genomic prediction accuracies of ~35%, which indicate genomic selection could be a valid approach for resistance breeding that would allow for more generations per year and would not require disease and pest pressure outside of training generations.

Abbreviations:

ASNPC, area under the *Striga* number progress curve; DM, incidence of downy mildew; FLO, flowering time; GBS, genotyping-by-sequencing; GS, genomic selection; HYD, panicle yield; QTL, quantitative trait loci; S1, Sadoré *Striga* infected field 1; S2, Sadoré *Striga* infected field 1

Introduction

Pearl millet (*Cenchrus americanus* (L.) M.) is the most widely grown millet in the world (Kalaisekar et al., 2017). It is an important food source to millions of people in the developing world, and has been an important food crop in Africa since the Neolithic age (Bourlag, 1996; Harlan, 1992). There is good evidence pointing towards a domestication event over 4000 years ago (possibly as far back as 6000 years ago) in the west African Sahel region (Tostain, 1992; Fuller, 2003; Nuemann, 2003; Manning and Fuller, 2014). Pearl millet production spread throughout large parts of Africa and over to India by 1700 BC, and possibly as early as 1900 BC. In India, production was focused in the north-western part of India, below the Indus Valley (Fuller, 2003; Boivin and Fuller, 2009; Manning et al., 2011). Today pearl millet is commonly grown all around the world as a grain and a forage crop, with Africa still being the most important production region focused on growing pearl millet for human consumption (FAO, 2018).

Roughly 90 million people rely on pearl millet as an essential component of their diet and income (ICRISAT, 2019). Most of these people live in the poor rural regions of the developing world. Pearl millet is planted on more than 75 million acres each year around the world, and accounts for roughly half of all millet grown every year (ICRISAT, 2019). Production in Africa has increased ~130% since the 1980's (ICRISAT, 2019). Part of pearl millet's importance is its hardy nature, because it can survive in difficult environments and growing conditions. It is often grown in poor soil that has high pH, low organic matter, low nutrients, high aluminum levels, low moisture, and/or high salt content. These factors make the soil not well suited for agricultural production. The environmental conditions are also often less than ideal, with low rainfall and high temperatures. In Africa, pearl millet is grown primarily as a grain crop for

human consumption, but other parts of the plant (leaves and stalks) are important food sources for livestock. What is not fed to humans or animals often is used as fuel (Vadez et al., 2012; Mason et al., 2015).

Aside from the benefit its stress tolerances can give to farmers, pearl millet is also an extremely healthy grain for human consumption. It has protein levels of 8-19%, low levels of starch, high fiber, and higher levels of micronutrients than major grain crops such as rice, wheat, maize, and sorghum. (Nambiar et al., 2011; Tako et al., 2015). Pearl millet can be grown over a wide range of geographic regions because of its ability to withstand harsh conditions and the variability in maturity times found among landraces and conserved in gene banks.

Pearl millet yields (~900 kg/ha) are low when compared to other major crops, partly because of the poor conditions the crop is farmed under. Nonetheless, plant breeders have made significant genetic gains in recent decades, and improved agronomic practices have also contributed to increasing yields. For example, from 1996-2013 there has been on average an increase of 24 kg/ha/year yield increase in India (Yadav and Rai, 2013). In Africa, there are important breeding effort going into increasing resistances to diseases and parasitic weeds. One of the major focuses of these efforts is to breed resistance to the parasitic weed *Striga* (Kountche et al., 2013). *Striga* and several foliar diseases (downy mildew by *Sclerospora graminicola*, Pyricularia leaf spot by *Pyricularia grisea*, and rust *Puccinia substriata* var. *indica*) are major issues for pearl millet production in the developing world (Varshney et al., 2017). *Striga* is the most serious weed issue for pearl millet production in Africa, and it can become especially problematic when a host crop (such as pearl millet) is planted several years in a row in low-fertility soil (Andrews and Brammel-Cox, 1993). Management techniques can be used to partly control *Striga* and includes late plantings, low tillage, proper nutrient applications, intercropping

legumes with pearl millet, and labor-intensive hand-weeding (Gworgwor et al., 1998; Andrews and Brammel-Cox, 1993; Hess and Ejeta, 1987; Carsky et al., 1994; Carson, 1988, Hatcher and Melander, 2003).

Downy mildew (*Sclerospora graminicola*) is the most damaging disease of pearl millet in Africa and can result in yield loses of ~30% (Ndiaye, 2002). There are various ways to deal with this disease, such as: planting resistant varieties, seed treatments (which have been shown to work for ~35 days after planting), and removal of diseased plants (Hess et al., 2002; Thakur et al., 2011; Scheuring et al., 2002; Mbaye, 1992).

Many of the management techniques to control these pests are labor-intensive or involve inputs farmers cannot often afford. Genetic resistance to *Striga* and downy mildew could present an opportunity for farmers to have access to an affordable method to reduce crop loss from these pests, and one tool that could potentially help breed genetic resistance into varieties is genomic selection. Genomic selection uses genome-wide marker data paired with phenotypic data to train a model that can be used to predict phenotypes from genomic data (Bernardo & Yu 2007; Heffner et al., 2009). Models are created using a training population where there is both phenotypic and genomic data on the individual plants. These models are used for a few generations to predict phenotypes, and eventually the model is retrained to maintain high accuracies (Heffner et al., 2011). Genomic selection can allow for more breeding cycles per year and is not limited to evaluations in specific locations or environments (outside of the training and retraining generations). The application of genomic selection has already been shown to be a useful tool to develop *Striga* resistant maize lines (Badu-Apraku et al., 2019), where yields were improved under heavy *Striga* pressure by more than 50% after three cycles of selection. The genomic selection resulted in increased yields associated with improved root lodging resistance,

plant ear height, and *Striga* resistance. To breed for *Striga* and downy mildew resistances using traditional selections, *Striga* and/or downy mildew pressures must be present in some manner. Ideally each plant would be under equal stress and exposure so the genetics resistance can be properly accounted for, but that is not a feasible scenario for most breeding programs. The inconsistencies of the pressure can make it difficult to accurately select for plants with genetic resistance. Genomic selection allows for selections to be made without the presence of the stresses of interest. For these reasons, *Striga* and downy mildew are ideal candidate traits for breeders to apply genomic selection to.

Materials and Methods

Plant material and phenotypic data

Plants for this study came from a *Striga*-resistant genepool population consisting of 400 full-sib families (FS) representing the fifth cycle (C_5) of phenotypic recurrent selection. A detailed description of how these lines were generated can be found in Kountche et al. (2013). In short, these 400 FS families were generated from an initial group of 64 landraces that were screened for resistances to *Striga* and downy mildew. That initial population then went through 5 cycles of recurrent selection, and that resulting population was used to make controlled crosses to generate the 400 FS families used in this experiment. The C_5-Full-sib population was screened for its reaction (resistance/susceptibility) to *Striga* during the 2011 rainy season at two test locations. Based on seed availability, the 400 FS families were divided into two sets of 200 FS. The first set of 200 entries was used to conduct a multi-location experiments at Sadoré (S1-field) (Niger) and Cinzana (Mali). The second set, with limited seed quantities, was phenotyped only at Sadoré (Niger) in an independent *Striga*-infested field (S2-field). The check plants used in the

experiment consisted of the seven original parental varieties, the five experimental varieties, and eight agronomically elite open-pollinated varieties. Each experiment thus comprised a total of 220 entries. *Striga* resistance evaluation of the C_5-Full-sib population was performed under naturally *Striga*-infested field at Cinzana (Mali). At Sadoré, artificially *Striga*-infested fields were used, with about 50,000 *Striga* seeds per m² applied for the field's infestation. The entries in each experiment were planted in a randomized incomplete block design with four replications, with the checks added to each block as controls. The experimental design and field management were described in Kountche et al. (2013).

Observations were recorded on *Striga* resistance measured by the area under the *Striga* number progress curve (ASNPC) (e.g. Haussmann et al., 2000). ASNPC is calculated using counts of striga infestations in a plot from multiple dates, and is analogous to the Area Under the Disease Progress Curve (AUDPC) used for plant diseases. In addition, pearl millet flowering time (FLO), panicle yield (HYD) (g m^{-2}), and the incidence of downy mildew (DM) were further evaluated. Data analyses were performed using PLABSTAT v3A and R packages (Utz, 2005; R core, 2018). The adjusted means (log transformation and square root transformation) of each of the 400 full-sib family were generated to deal with highly skewed data. After transformation, most phenotypes showed a more normal distribution that was less likely to violate the assumptions of the regression models used in our analyses.

Genotypic data

Total genomic DNA was isolated from two-week-old seedlings of each full-sib family. DNA was extracted from pooled leaf tissues of 16 seedlings using a modified CTAB/β-mecaptoethanol protocol (Mariac et. al., 2006). DNA samples were sent to the Institute for

Genomic Diversity (Cornell, USA) for genotyping-by-sequencing (GBS) (Elshire et al., 2011) with restriction enzyme ApeKI. The Burrows-Wheeler Aligner (BWA) (Li and Durbin 2009) was then used to align tags to the Pearl millet reference genome (Varshney et al., 2017). SNPs were called using the Tassel GBS v2 SNP pipeline (Glaubitz et al., 2014).

Low-quality SNPs were filtered out by removing any with >40% missing data, >25% heterozygosity (likely due to misalignment of paralogs), and minor allele frequency <0.5%. Eight taxa with nearly 100% missing data were also filtered out. Following filtering, imputation was done using Beagle (Browning & Browning 2016).

Association analysis

A genome wide association analysis (GWAS) was done using a mixed-linear-model (MLM) in TASSEL, Version 5.2.52 (Bradbury et al., 2007). Population structure was accounted for using the first five genetic principle coordinates (also calculated in TASSEL using the "MDS" plugin). Empirical p-values were generated using permuted hapmaps that had the plant names randomly shuffled 1000 times run through the same analysis. The most significant p-value from each of the 1000 permuted runs were used to build a null distribution of what we would expect to see by chance alone. The p-values from the actual run were compared against that distribution, and any in the top 5% of the permuted p-value distribution were considered to be significant. (That is, empirical $p \leq 0.05$.)

Genomic prediction

Genomic prediction was performed using basic GBLUP (Clark and van der Werf, 2013) analysis in TASSEL (Bradbury et al., 2007). Accuracy was assessed with 1000 iterations of 10-fold cross-validation.

Results

GBS genotyping resulted in 233,481 raw SNP markers, and 64,707 markers after filtering. To validate these markers, we estimated the narrow-sense heritability at the three locations. (Table 3.1). Most of the traits showed reasonable heritability in the range of 0.4 to 0.5, with some heritability ratings approaching 0.8 or 0.9 in some traits. Flowering time heritability results were lower than we would normally expect, possibly from the *Striga* pressure in the field, which would have stressed the plants. (Stresses on plants have been shown to reduce trait heritability; Sanad et. al., 2019.)

Many of the phenotypes had highly non-normal distributions, which can degrade the accuracy of mixed linear models because they violate key assumptions. (McCulloch and Searle, 2006). We tested both square-root (SQ) and log transformations (LN) to improve the distribution of the data, and for most phenotypes, the log transformation gave the highest genomic heritability (Table 3.1). The transformations were shown to be especially important for the downy mildew data from Cinzana. The untransformed data showed an unreasonably high heritability (nearly 100%; table 3.1) that is not biologically accurate for this trait, but after transforming the data reasonable heritabilities were generated.

We then performed genomic selection on these traits using the GBLUP method (Clark and van der Werf, 2013) (Figure 3.2). Prediction accuracies using the most accurate data

transformations for *Striga* resistance ranged from 21.07% to 48.87% across the locations. The Sadoré fields had higher accuracies (32.47% & 48.87%) than the Cinzana field (21.07%). The Cinzana field relied on naturally occurring *Striga* pressure instead of the artificial inoculations that the Sadoré fields had. This could explain some of the lower accuracies because of the inconsistent pest pressure that would have been present throughout the field. Prediction accuracies using the most accurate data transformations for downy mildew resistance ranged from 27.17% to 38.98% across the locations. As with the *Striga* results, the two Sadoré fields accuracies (35.26% & 38.98%) were higher than the Cinzana accuracy (27.17%). This trend can be seen with all the traits between the Sadoré locations and Cinzana except for flowering time between Cinzana and Sadoré 1 (Figure 3.2).

While our attempt at identifying QTL strongly associated with resistances to *Striga* and downy mildew were not successful, it is possible that with a mapping population instead of a breeding population the results would have shown sites significantly associated with resistances. A mapping population would be specifically designed to be segregating for the traits of interest, while the purpose of the breeding population was just to improve the resistances to *Striga* and downy mildew. The genetic architecture of the traits can also impact the effectiveness of GWAS, since for traits that are made up of many small effect alleles, it can be difficult for GWAS to properly identify them. This is because of the relatively small impact each allele has on the traits of interest. For traits like that genomic selection can be a better approach.

Discussion

With the decreasing cost of generating genomic data there is a growing interest in using genomic data to predict phenotypes for germplasm collections (Crossa et al., 2017). Phenotypic

predictions using genomic data relies on a large sample of the collection (ideally, all of it), phenotyping a subset, and using statistical analyses to build predictive phenotype models. Genomic predictions can be especially useful for traits that are work-, time-, and location-restrictive to phenotype, such as *Striga* and downy mildew resistances (Oldenbroek and van der Waaij, 2015). It has already been shown in maize that genomic selection is a useful tool for breeding for *Striga* resistance (Badu-Apraku et al., 2019). In this study, we tested the accuracy of genomic selection in predicting *Striga* and downy mildew resistance in Pearl Millet. It has been estimated that genomic selection will outperform traditional marker assisted selection for traits with heritability greater than or equal to 0.2 (Heffner et al., 2009 & 2010; Zhong et al., 2009; Poland et al., 2012; Muleta et al., 2019). It is important to note that trait heritability greatly influences the maximum genomic prediction accuracies (Zhang et al., 2019). Our results, with heritability and prediction accuracies above 0.2, indicate that genomic selection is likely a valid approach in breeding for resistances to *Striga* and downy mildew in pearl millet. The prediction accuracies were in line with other genomic selection studies (Daetwyler et al., 2013). A limitation of traditional breeding for *Striga* and downy mildew resistance is that the disease and parasitic plant pressure must be present to properly phenotype for resistance to these issues. Beyond just the presence of the disease and pest pressure it is also important that each plant is equally stressed so that proper selection can be done to improve the genetic resistance, but ensuring that each plant is equally stressed would be extremely labor and capital intensive. That makes being able to select based upon genotypic data an appealing approach (Taylor, 2014), though the training set (and periodic retraining of the model) would still require disease/parasite presence. Applying genomic selection to pearl millet *Striga* and downy mildew resistance would allow more breeding cycles per year. This would be done by taking advantage of controlled

environments for multiple generations a year, and by breeding in locations for more grow cycles because the weed and disease pressure are not needed to be present (Li and Dungey, 2018). Periodic retraining of the model would allow breeders to ground-truth predictions and monitor the progress of resistance in their breeding population.

Conclusion

Proper application of genomic selection for traits such as downy mildew and *Striga* resistance would allow for faster development of resistant pearl millet lines. By allowing breeding in the absence of disease and parasite pressure the speed of development can be increased and the labor involved can be significantly reduced. This study has shown that there is strong potential for using genomic selection to be able to increase our capacity to breed for resistance to these pressures.

Table 3.1: The narrow sense heritability of the traits and the transformed data

	Cinzana	**Sadoré 1**	**Sadoré 2**
Trait	**Heritability**		
DM	0.999	0.323	0.694
SQ	0.554	0.386	0.840
LN	0.409	0.391	0.907
FLOdap	0.488	0.417	0.413
SQ	0.482	0.424	0.442
LN	0.465	0.429	0.463
HYD	0.411	0.480	0.308
SQ	0.445	0.531	0.273
LN	0.476	0.587	0.559
ASNPC	0.462	0.767	0.523
SQ	0.457	0.574	0.513
**LN*	0.513	0.425	0.530
ASVPC	0.434	0.747	0.854
SQ	0.458	0.549	0.663
**LN*	0.519	0.384	0.527

SQ: Square root transformation; *LN*: Log transformation

*Raw values were divided by 10 prior to applying the log transformation

Table 3.2: Description of the trait and data transformation abbreviations

Trait	**Description**
DM	Number of hills with downy mildew infestation
FLOdap	50% flowering time
HYD	Panicle Weight (g m^{-2})
ASNPC	Area under Striga number progress curve
ASVPC	Area under Striga severity progress curve
SQASN	Square root transformation of ASNPC
LNASN1	Log transformation of ASN10
SQASV	Square root transformation of ASVPC
LNASV1	Log transformation of ASV10

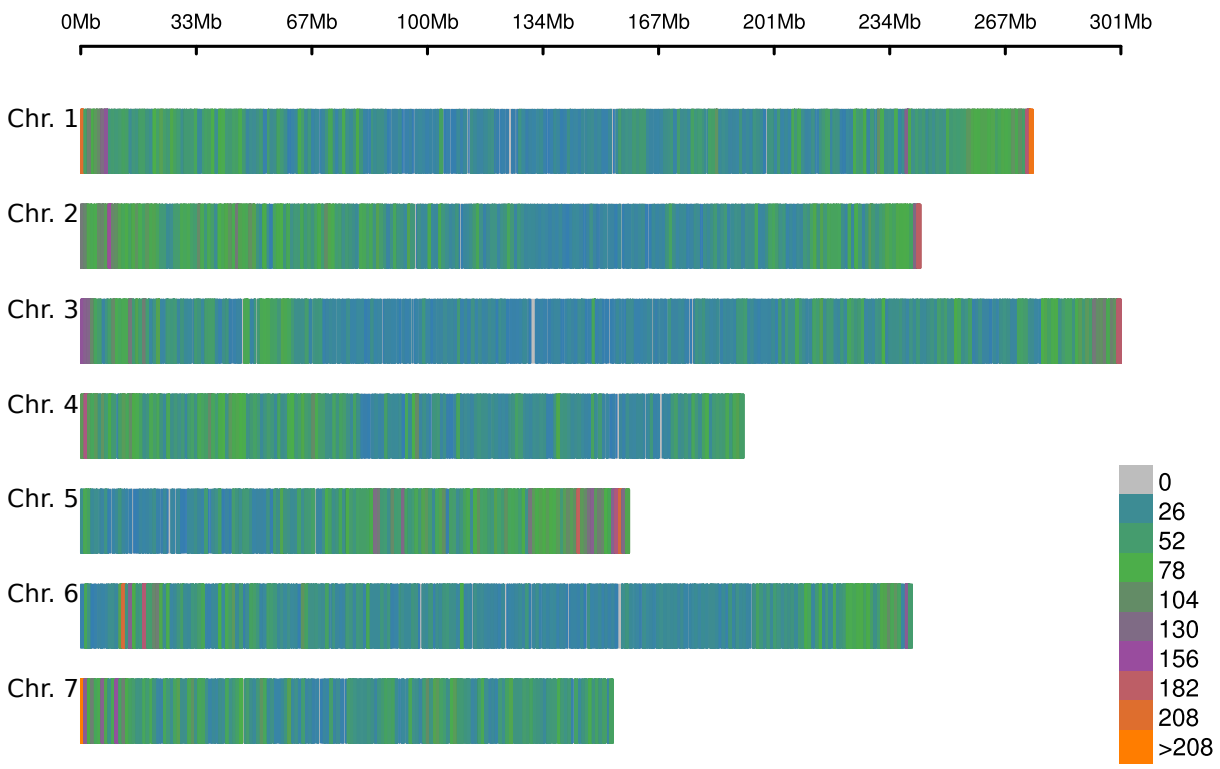

Figure 3.1
Genome wide SNP density plot of pearl millet generated using CM plot R package (LiLin, 2020). Each vertical bar represents a 1 Mb region of the genome, colored based on the number of SNPs in a 1 Mb region. SNPs are distributed across the entire genome, with denser regions near the outer edges of the chromosomes.

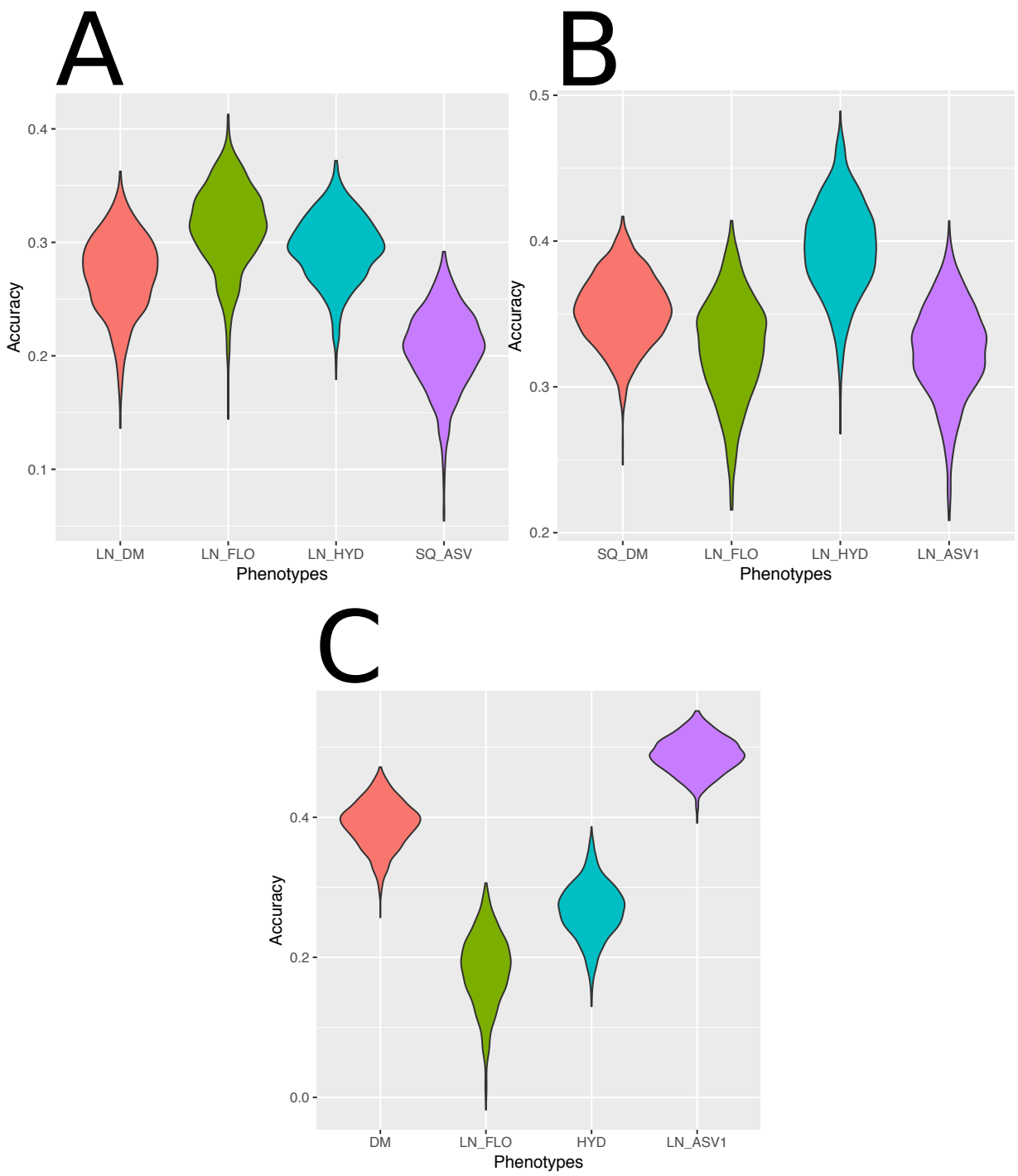

Figure 3.2
Violin plot of the top genomic selection accuracies of downy mildew (DM), flowering time (FLO), panicle weight (HYD), and striga resistance (ASV) using either the log transformation (LN) or square root transformation (SQ). For Cinzana (A), Sadoré 1 (B) and, Sadoré 2 (C) Average prediction accuracy for Cinzana (A) ranges from ~0.21 to 0.31, for Sadoré 1 (B) ranged from ~0.32 to 0.39, and for Sadoré 2 (C) ranged from ~0.18 to 0.48

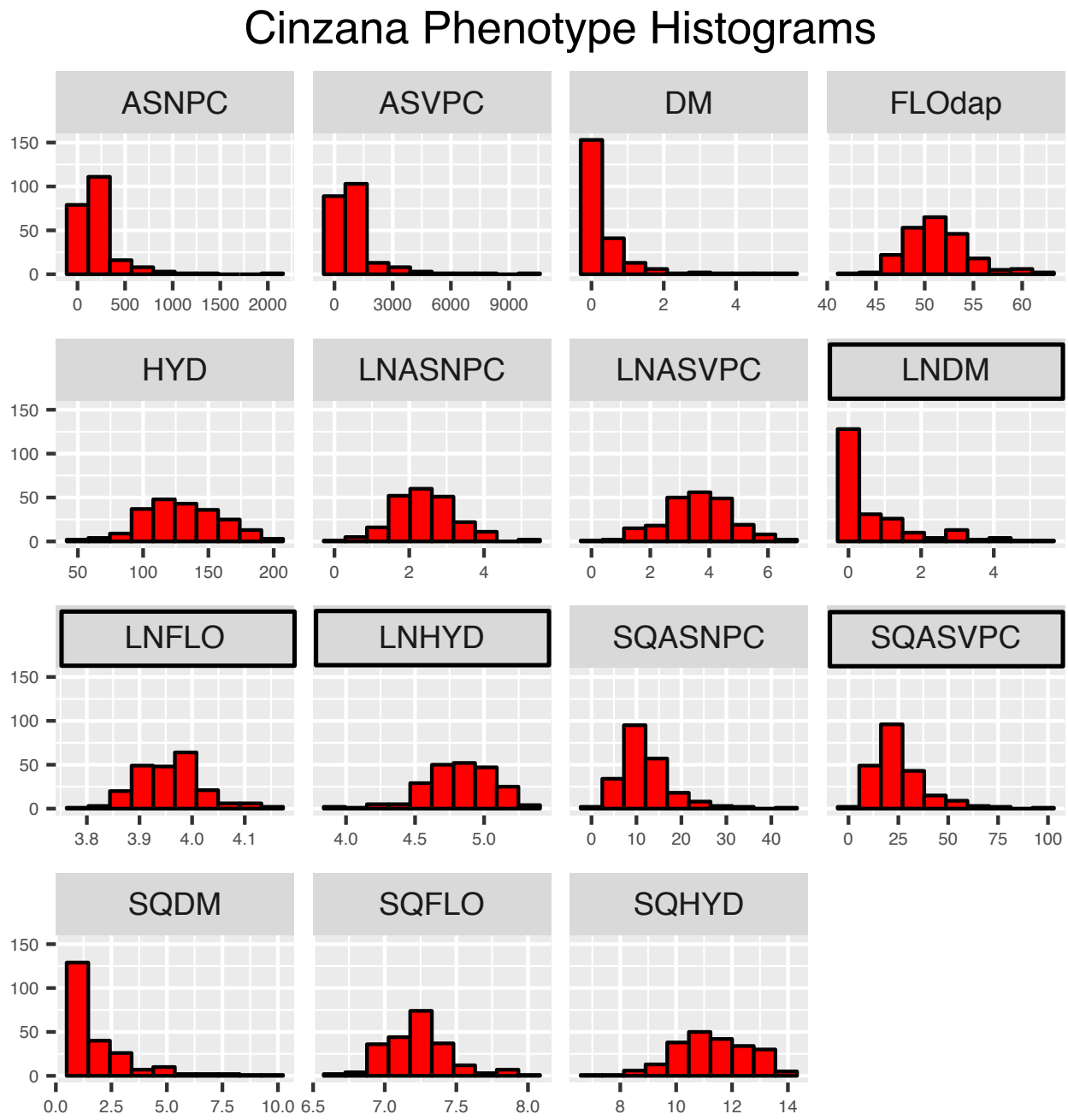

Figure 3.3

Distributions of pearl millet phenotypes and transformations of those phenotypes at the Cinzana field. Black boxes indicate which phenotypes were used in the final analyses.

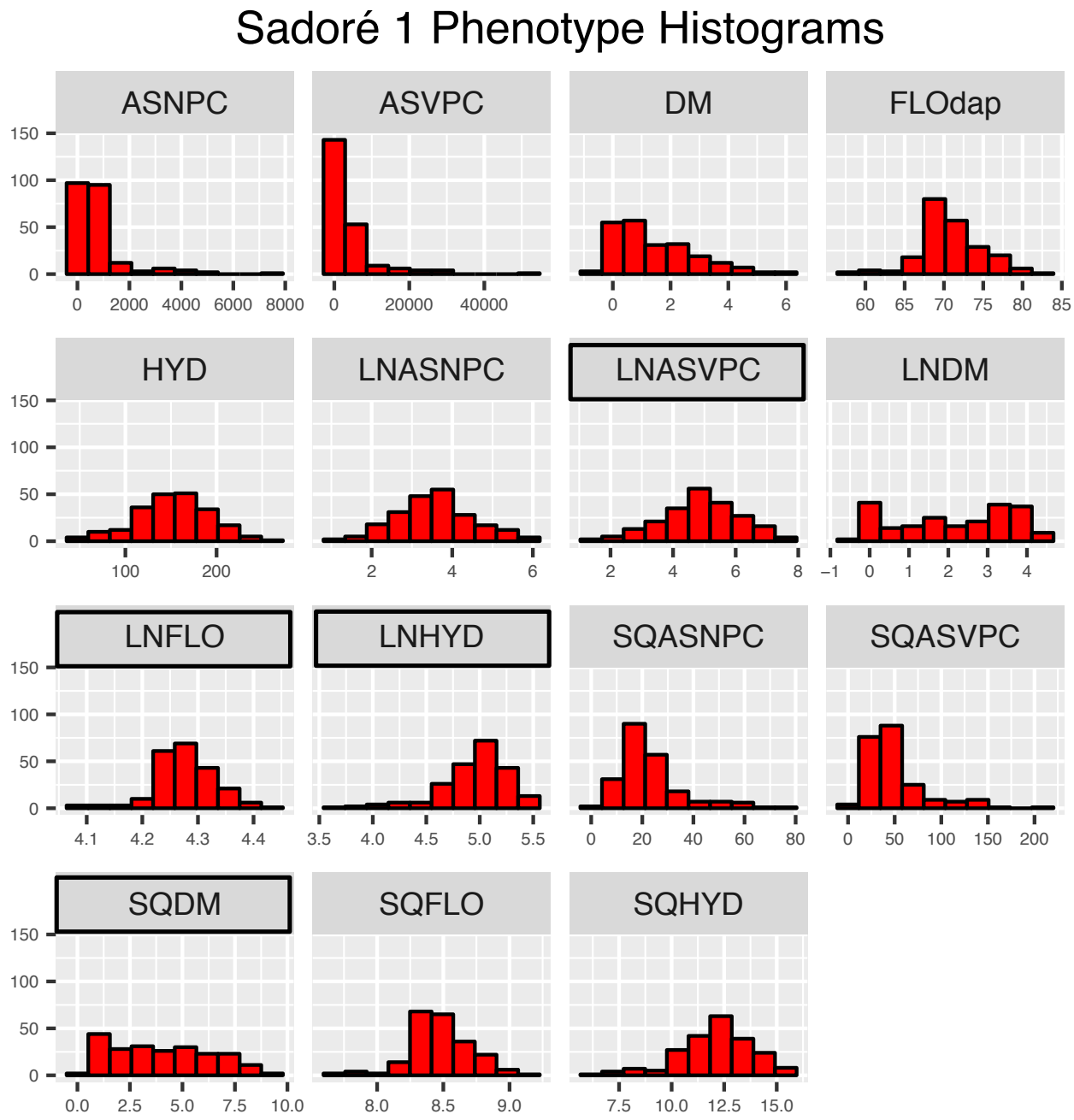

Figure 3.4

Distributions of pearl millet phenotypes and transformations of those phenotypes at the Sadoré 1 field. Black boxes indicate which phenotypes were used in the final analyses.

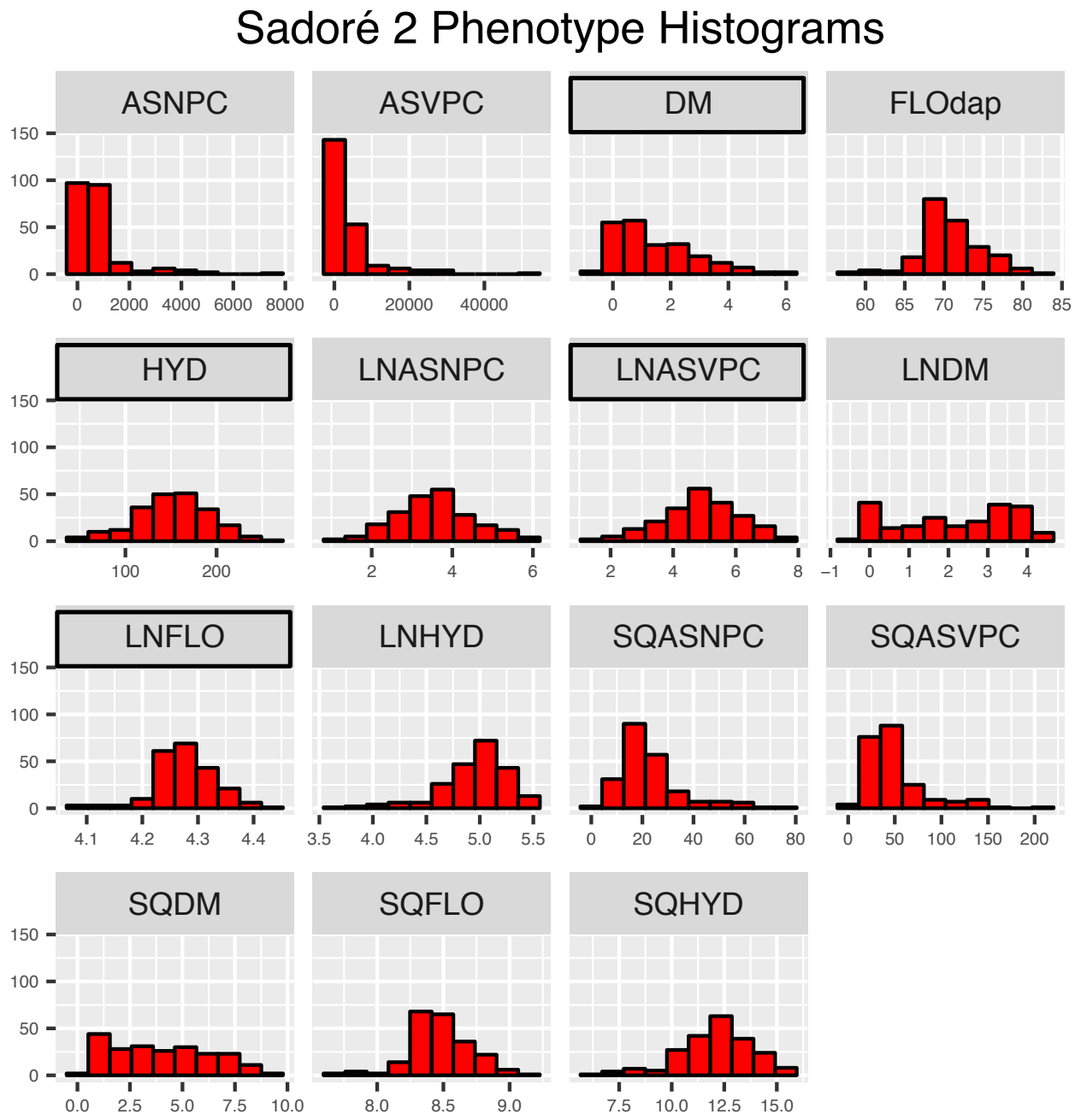

Figure 3.5

Distributions of pearl millet phenotypes and transformations of those phenotypes at the Sadoré 2 field. Black boxes indicate which phenotypes were used in the final analyses.

REFERENCES

Andrews D. J, Rajewski J.F and Anand Kumar K. 1993. Pearl millet – a new feed grain crop. Pages 198–208in New crops (Janick J and Simon JE, eds.). New York, USA: John Wiley & Sons.

Andrews, D. J. and Brammel-Cox, P. J. 1993. Breeding cultivars for sustainable crop production in low input dryland agriculture in the tropics. In International Crop Science, 211–23 (Eds D. R. Burton, R. Shibles, R. A. Forsberg, B. L. Blad, K. H. Asay, G. M. Paulsen and R. F. Wilson). Madison, Wisconsin: Crop Science Society of America.

Badu-Apraku, B., Talabi, A. O., Fakorede, M. A. B., Fasanmade, Y., Gedil, M., Magorokosho, C., & Asiedu, R. 2019. Yield gains and associated changes in an early yellow bi-parental maize population following genomic selection for Striga resistance and drought tolerance. BMC Plant Biology, 19(1), 129. https://doi.org/10.1186/s12870-019-1740-z

Badu-Apraku, B., Talabi, A. O., Fakorede, M. A. B., Fasanmade, Y., Gedil, M., Magorokosho, C., & Asiedu, R. 2019. Yield gains and associated changes in an early yellow bi-parental maize population following genomic selection for Striga resistance and drought tolerance. BMC Plant Biology, 19(1), 129

Bernardo R, Yu J. 2007. Prospects for genome wide selection for quantitative traits in maize. Crop Sci. 47(3):1082–90. 17.

Bilquez, A. F. 1963. Etude due mode d'hérédité de la précocité chez le mil pénicillaire (*Pennisetum typhoides* Stapf et Hubbard): I. Déterminisme génétique des diffénces de sensibilité a la longueur du jour existant entre les mils degroupe sanio et ceux du groupe

souna (Study of *Pennisetum typhoides* Stapf et Hubbard): Genetic determination of day length sensitivity differences existing among millet in the sanio and souna groups). Agronomie Tropical18:1249–1253.

Boivin, N., and Fuller, D. 2009. Shell middens, ships and seeds: Exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula Journal of World Prehistory, 22(2), 113–180.

Bourlag, N. (Ed.). 1996. Lost crops of Africa. Volume I. Grains.Washington: The National Academies Press.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Browning, B. L., and Browning, S. R. 2016. Genotype imputation with millions of reference samples. Am J Hum Genet 98:116-126. doi:10.1016/j.ajhg.2015.11.020

Carsky, R. J., Singh, I. and Ndikawa, R. 1994. Suppression of *Striga hermonthica* of sorghum using cowpea intercrop. Experimental Agriculture30:130–132.

Carson, A. G. 1988. Effect of intercropping sorghum and groundnuts on density of *Striga hermothicain* the Gambia. Tropical Pest Management34:97–101.

Clark, S. A., and van der Werf, J. 2013. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. METHODS IN MOLECULAR BIOLOGY -CLIFTON THEN TOTOWA-, 321.

Crossa, J., P. Pérez-Rodríguez, J. Cuevas, O. Montesinos-López, D. Jarquín, G. de los Campos, et al. 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends in Plant Science 22: 961-975. doi:10.1016/j.tplants.2017.08.011.

Daetwyler H.D., Calus M.P., Pong-Wong R., de Los Campos G., Hickey J.M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193: 347–365.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6: e19379.

Eltaher, S., Sallam, A., Belamkar, V., Emara, H. A., Nower, A. A., Salem, K. F. M., et al. 2018. Genetic diversity and population structure of F3:6 Nebraska winter wheat genotypes using genotyping-by-sequencing. Front. Genet. 9:76. doi: 10.3389/fgene.2018.00076

FAO. 2018. FAOSTAT data base. Available at http://faostat.fao.org/site/5676/DesktopDefault.aspx.

Fuller, D. Q. 2003. African crops in prehistoric South Asia: A critical review. In K. Neumann, A. Butler, and S. Kahlheber (Eds.), Food, fuel, and fields: Progress in African archaeobotany (pp. 239–271). Cologne: Heinrich–Barth Institut.

Fusi, N., Lippert, C., Lawrence, N. D., and Stegle, O. 2014. Genetic Analysis of Transformed Phenotypes.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. PloS one 9: e90346.

Gupta, S. K., Rai, K. N., Singh, P., Ameta, V. L., Gupta, S. K., Jayalekha, A. K., … Verma, Y. S. 2015. Seed set variability under high temperatures during flowering period in pearl millet (Pennisetum glaucum L. (R.) Br.). Field Crops Research, 171, 41–53. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.fcr.2014.11.005

Gworgwor, N. A., Anaso, A. B. and Turaki, Z. G. 1998. Integrated cultural practices for Striga control in millet-based cropping systems. In Pearl Millet in Nigerian Agriculture, 87–97 (Eds A. M. Emechebe, M. C. Ikwelle, O. Ajayi, M. Aminu-Kano and A. B. Anaso). Maiduguri, Nigeria: Ramadan Press.

Harlan, J. 1992. Indigenous African agriculture. In C. Cowan and P. J. Watson (Eds.), The origins of agriculture: An international perspective (pp. 59–70). Washington: Smithsonian Press.

Hatcher, P. E. and Melander, B. 2003. Combining physical, cultural and biological methods: prospects for integrated on-chemical weed management strategies. Weed Research43:303–322.

Haussmann, B.I.G., Hess, D.E., Welz, H.G., Geiger, H.H., 2000.Improved methodologies for breeding Striga-resistant sorghums. Field Crops Res. 66, 195–211.

Heffner EL, Jannink JL, Sorrells ME. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. Plant Genome.;4(1):65–75.

Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. 2010. Plant breeding with genomic selection: gain per unit time and cost. Crop Sci. 50(5):1681–90

Heffner EL, Sorrells ME, Jannink JL. 2009. Genomic selection for crop improvement. Crop Sci. 49(1):1–12.

Hess, D. E and Ejeta, G. 1987. Effect of cultural treatments on infestation of *Striga hermonthica* (Del.) Benth (*Scrophulariaceae*). In Proceedings of Fourth International Symposium on Parasitic Flowering Plants, 367–375 (EdsH. C. Weber and W. Forstreuter). Marburg, Germany: Philipps-Universitat.

Hess, D. E., Thakur, R. P., Hash, C. T., Sérémé, P. and Magill, C. W. 2002. Pearl millet downy mildew: problems and control strategies for a new millennium. In Sorghum and Millet Diseases, 37–41 (Ed J. F. Leslie). Ames, Iowa: IowaState Press.

ICRISAT, Pearl Millet. Retrieved from http://exploreit.icrisat.org/profile/Pearl Millet/178. date accessed 10/24/2019

Kalaisekar, A., Padmaja, P. G., Bhagwat, V. R., & Patil, J. V. 2017. Insect pests of millets: systematics, bionomics, and management. Elsevier/Academic Press.

Kanai, M., Tanaka, T., & Okada, Y. 2016. Empirical estimation of genome-wide significance thresholds based on the 1000 Genomes Project data set. Journal of human genetics, 61(10), 861–866. doi:10.1038/jhg.2016.72

Kountche, B. A., Hash, C. T., Dodo, H., Laoualy, O., Sanogo, M. D., Timbeli, A., … Haussmann, B. I. G. 2013. Development of a pearl millet Striga-resistant genepool: Response to five cycles of recurrent selection under Striga-infested field conditions in West Africa. Field Crops Research, 154, 82–90. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.fcr.2013.07.008

Kountche, B. A., Hash, C. T., Dodo, H., Laoualy, O., Sanogo, M. D., Timbeli, A., … Haussmann, B. I. G. 2013. Development of a pearl millet Striga-resistant genepool: Response to five cycles of recurrent selection under Striga-infested field conditions in West Africa. Field Crops Research, 154, 82– 90. doi:10.1016/j.fcr.2013.07.008.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. BIOINFORMATICS -OXFORD-, (14), 1754. doi:10.1093/bioinformatics/btp324

Li, Y., & Dungey, H. S. 2018. Expected benefit of genomic selection over forward selection in conifer breeding and deployment. PLoS ONE, 13(12), 1–21.

LiLin, Y. (2020). CMplot. Retrieved from https://github.com/YinLiLin/R-CMplot

Maman, N., Mason, S. C. and Sirifi, S. 2000. Influence of variety and management level on pearl millet production in Niger. I. Grain yield and dry matter accumulation. African Crop Science Journal8:25–34.

Manning, K., and Fuller, D. 2014. Early millet farmers in the Lower Tilemsi Valley, northeastern Mali. In C. Stevens, S. Nixon, M. Murray, and D. Fuller (Eds.), The archaeology of African plant use (pp. 73–82). Walnut Creek: Left Coast Press.

Manning, K., Pelling, R., Higman, T., Schwenniger, J.-L., & Fuller, D. 2011. 4500-year-old domesticated pearl millet (Pennisetum glaucum) from the Tilemsi Valley, Mali: New insights into an alternative domestication pathway. Journal of Archaeological Science, 38,312–322.

Mariac C., Luong V., Kapran I., Mamadou A., Sagnard F., Deu M., et al. 2006. Diversity of wild and cultivated pearl millet accessions (Pennisetum glaucum [L.] R. Br.) in Niger assessed by microsatellite markers. Theor. Appl. Genet. 114 49–58. doi:10.1007/s00122-006-0409-9

Mason, S. C., Maman, N., and Pale, S. 2015. Pearl Millet Production Practices in Semi-Arid West Africa: A Review. EXPERIMENTAL AGRICULTURE, 51(4), 501–521. https://doi-org.proxy-remote.galib.uga.edu/10.1017/S0014479714000441

Mbaye, D. F. 1992. Les maldies du mil au Sahel: Etat des connaissances et propositions de lutte. In La Lutte Intégrée Contre les Ennemis des Cultures Vivrières dans le Sahel (Pearl

millet diseases in the Sahel: State of knowledge and management proposals). Montrouge, France: John Libby Eurotext.

McCulloch, C. E., Searle, S. R. 2006. Generalized, Linear, and Mixed Models. Journal of the American Statistical Association, 101(476), 1724.

Muleta, K. T., Pressoir, G., Morris, G. P. 2019. Optimizing genomic selection for a sorghum breeding program in Haiti: a simulation study. G3. Genes Genomes Genet. 9, 391–401. doi: 10.1534/g3.118.200932

Nambiar, V. S., Dhaduk, J. J., Neha Sareen, Tosha Shahu, & Rujuta Desai. 2011. Potential functional implications of pearl millet (Pennisetum glaucum) in health and disease. Journal of Applied Pharmaceutical Science, 1(10), 62–67.

Ndiaye, S. 2002. Amelioration du controle du mildiou. Resultats du Projet P3: 1992–1997 (Improved control ofmildew: Results of Project P3). Niamey, Niger: Réseau Ouest et Centre Afríaín de Recherche sur le Mil.

Neumann, K. 2003. The late emergence of agriculture in sub-Saharan Africa: Archaeobotanical evidence and ecological considerations. In K. Neumann, A. Butler, and S. Kahlheber (Eds.), Fuel, foods and fields: Progress in African archaeobotany (pp. 71–92). Africa Praehistorica 15). Köln: Heinrich-Barth Institute.

Oldenbroek, K. and van der Waaij, L. 2015. Textbook Animal Breeding and Genetics for BSc students. Centre for Genetic Resources The Netherlands and Animal Breeding and Genomics Centre, 2015. Groen Kennisnet: https://wiki.groenkennisnet.nl/display/TAB/

Payne, W. A. 1997. Managing yield and water use of pearl millet. Agronomy Journal89:481–490.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. 2012. Genomic selection in wheat breeding using genotyping-by-sequencing. Plant Genome 5, 103–113. doi: 10.1105/tpc.012526

R Core Team.2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

Sanad, M., Smertenko, A., & Garland-Campbell, K. A. 2019. Differential Dynamic Changes of Reduced Trait Model for Analyzing the Plastic Response to Drought Phases: A Case Study in Spring Wheat. *Frontiers in plant science*, *10*, 504. doi:10.3389/fpls.2019.00504

Scheuring, J. F., Katilé, S. O. and Kollo, I. A. 2002. Boosting pearl millet yields with Apron Plus and Apron Star seed treatments. In Sorghum and Millet Diseases, 47–49 (Ed J. F. Leslie). Ames, Iowa: Iowa State Press.

Senthilvel, S., Jayashree, B., Mahalakshmi, V., Kumar, P. S., Nakka, S., Nepolean, T., & Hash, C. 2008. Development and mapping of simple sequence repeat markers for pearl millet from data mining of expressed sequence tags. BMC Plant Biology, 8, 119. https://doi.org/10.1186/1471-2229-8-119

Tako, E., Reed, S. M., Budiman, J., Hart, J. J., & Glahn, R. P. 2015. Higher iron pearl millet (Pennisetum glaucum L.) provides more absorbable iron that is limited by increased polyphenolic content. Nutrition Journal, (1). https://doi-org.proxy-remote.galib.uga.edu/10.1186/1475-2891-14-11

Taylor, J. F. 2014. Implementation and accuracy of genomic selection. Aquaculture, 420–421(Supplement 1), S8–S14.

Tesso, T. and Ejeta, G. 2011. Integrating multiple control options enhances Striga management and sorghum yield on heavily infested soils. Agronomy Journal103:1464–1471.

Thakur, R. P., Veeranki, P. R. and Sharma, R. 2011. Influence of dosage, storage time and temperature on efficacy of netalazyl-treated seed for the control of pearl millet downy mildew. European Journal of Plant Pathology129:353–359.

Tostain, S. 1992. Enzyme diversity in pearl millet (*Pennisetum glaucum* L.). Theoretical and Applied Genetics, 83,733–742.

Utz, H.F. 2005. PLABSTAT. Plant breeding statistical program. Version 3A. Institute of Plant Breeding, Seed Science and Population Genetics, University of Hohenheim, Stuttgart, Germany.

Vadez, V., Hash, T., Bidinger, F. R., & Kholova, J. 2012. II.1.5 Phenotyping pearl millet for adaptation to drought. Frontiers In Physiology, 3, 386. https://doi-org.proxy-remote.galib.uga.edu/10.3389/fphys.2012.00386

Varshney, R.K., C. Shi, M. Thudi, C. Mariac, J. Wallace, P. Qi, et al. 2017. Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. Nature Biotechnology 35: 969-976. doi:10.1038/nbt.3943.

Yadav, O.P. & Rai, K.N. 2013 Genetic improvement of pearl millet in India. Agric. Res. 2, 275–292.

Zhang, H., Yin, L., Wang, M., Yuan, X., & Liu, X. (2019.). Factors Affecting the Accuracy of Genomic Selection for Agricultural Economic Traits in Maize, Cattle, and Pig Populations. FRONTIERS IN GENETICS, 10.

Zhong, S., Dekkers, J. C., Fernando, R. L., Jannink, J. L. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182, 355–364. doi: 10.1371/journal.pgen.0030004

CHAPTER 4

GENOMIC AND CHEMICAL DIVERSITY OF COMMERCIALLY AVAILABLE INDUSTRIAL HEMP ACCESSIONS

[1] Matthew S. Johnson, Jason Wallace. To be submitted to *The Plant Genome*

Abstract

High consumer demand for cannabidiol (CBD) has made industrial hemp (*Cannabis sativa*) an extremely high-value crop for farmers to grow. Demand for and interest in this crop have caused the industry to develop faster than research and development has been able to keep up. This has resulted in the sale of many hemp accessions which lack stability, and proper regional testing. These issues can cause farmers to lose their crop because of weak growth and high THC. In this industry, the name given to a marketed hemp accession is thought to be meaningless, with plants of the same name from different companies showing different phenotypes, the most concerning of which is the variability in the THC content. Understanding the genetic diversity between and within commercially available accessions will play an important role in understanding the quality of the commercially available accessions being sold. Most of hemp accessions in this study showed a lack of genetic stability, but there were some that showed genetic consistency. Many of the accessions grown from seed showed evidence that they contained copies of the THCa synthase gene, leading to very high levels of THC produced in the female flowers. Trials and trust are needed to properly identify viable accessions and seed sources, but beyond simply understanding if a commercial accession is viable it is important to work toward developing improved cultivars for farmers.

Introduction

Hemp (*Cannabis sativa)* is a dioecious annual plant (Cherniak, 1982) that is believed to have been domesticated around 6000 years ago in China with some evidence of use as far back as 12,000 years ago (Li, 1974; Fleming & Clarke, 1998; Merlin 2003). It has been used for fiber, seeds, and its chemical profile for thousands of years (Russo, 2007). Today the plant is also used

in the production of biofuels, plastics, and building composites. The specific use of hemp is dependent on the variety grown. There are distinct fiber, grain, and cannabinoids varieties, and some varieties have been created to serve dual purposes. The fiber varieties have been selected to have tall thin stalks, with a large amount of the biomass being concentrated in the outer layers of the plant stem, known as bast fiber. While both male and female plants are used in fiber production males are preferred because of the higher quality fiber that they produce, and to make the best quality fiber the plants are harvested before full maturity (Hall et al., 2012). The fiber is used to produce fabrics, paper, ropes, and building composites (Amaducci et al., 2015). While male plants are best for fiber production female plants are essential for grain and cannabinoid production. Hemp seed is a high protein high oil seed (Callaway, 2004) with an oil that has a ratio of omega-6 to omega-3 fatty acids that is ideal for human consumption (Galasso et al., 2016). Hemp seed oil has also been used for cosmetics, paints, soaps, and fuels (Karus and Vogt 2004). Despite the many different uses of *Cannabis sativa,* it is only legally differentiated into hemp and marijuana, and that is only based on the concentration of total tetrahydrocannabinol (THC) (total THC is delta-9-THC + [THCa *.877]). In the United States, plants with all parts of the plant containing less than 0.3% total THC are legally considered hemp, and plants with any part of the plant containing over 0.3% total THC are legally considered marijuana (Hemp Farming Act of 2018). The use of THC as a psychoactive drug resulted in governments around the world outlawing *Cannabis sativa* use and production in the 20th century, and this ban applied to all production of *Cannabis sativa* regardless of the concentration of THC (Drug Policy Alliance, 2014). In 2014, the United States federal farm bill allowed limited growing of industrial hemp plants, and in 2018 the farm bill brought hemp almost back into full production. It was in these bills that the 0.3% THC limit for hemp was set. It is important to note that any

Cannabis sativa plant will be considered hemp until the concentration of the total THC exceeds 0.3%.

While THC is the most important chemical for legal considerations, cannabidiol (CBD) has become a chemical of economic interest for industrial hemp, with an estimated US market value of over one billion dollars of CBD products sold each year (2018 US Hemp Crop report). CBD is marketed to help with anxiety, pain, depression, sleep, and seizures. Most support for this compound is anecdotal, although there are some studies supporting its usefulness (Perucca, 2017; Glass and Gilleece, 2019; and Hurd et. al., 2019). The most significant, and well-studied, application of CBD is its ability to help control seizures. In 2018, the FDA approved the CBD based drug epidiolex, which helps in the management of Lennox-Gastaut syndrome and dravet syndrome.

THC and CBD are used to categorize *Cannabis sativa* plants, specifically the ratios of CBD to THC. Plants are generally broken into type I, type II, and type III plants. Type I plants are THC dominant with little to no CBD, type II plants have roughly a 1:1 ratio of CBD to THC, and type III plants are CBD dominant with very small levels of THC (Lewis et al., 2018). These cannabinoids are generally found in their acid forms in the plant (THCa and CBDa), and through time and heat they are decarboxylated into THC and CBD. They are produced from the same precursor molecule cannabigerol-acid (CBG) (Taura et al., 2007). The CBDa synthase enzyme converts CBGa into CBDa, and the THCa synthase enzyme converts CBGa into THCa (Sirikantaramas et al., 2004). While these are the two most common cannabinoids, there are over 100 other cannabinoids for in *Cannabis sativa* plants (ElSohly and Slade, D. 2006). All known cannabinoids are produced in the highest concentrations in the trichomes of female flowers, with production throughout other parts of the plant in much lower concentrations (Turner et al., 1981).

Cannabis sativa has a diploid genome (2n=20) with nine pairs of autosomal chromosomes and an XY genetic sex determination system. The size of the genome is estimated to be 818 Mb for females and 843 Mb for male plants (Sakamoto et al., 1998). In general plants with a female phenotype will have an XX sex chromosome pair and plants with a male phenotype will have an XY pair. Other factors such as environmental conditions, plant hormones, and ethylene inhibitors can affect the sexual phenotype of the plants, resulting in a genetically female plant which is phenotypically male (Hall et al., 2012; Mohan and Sett, 1982; Galoch, 2015; Lubell and Brand, 2018; Nelson, 1944; Eriksson et al., 2000). Although naturally dioecious with some off-type sex expressions resulting from various conditions there have also been cultivars bred to be monecious. These monecious plants have an XX sex chromosome pair and produce both male and female flowers on one plant (Faux et al., 2014). The male flowers tend to be produced on the bottom of the inflorescences and the female flowers are produced near the top of the inflorescences (Small, 2015). These monecious crops are used for grain production systems to maximize the number of female flowers in the field while ensuring significant pollinations of the flowers.

Outside of seed production for grain or plant multiplication it is best to avoid pollination. In cannabinoid production, pollinated plants have been shown to lose ~50% of their potential cannabinoid yield (Meier and Mediavilla, 1998). To avoid pollination issues farmers often work with clones of known female plants or use feminize seeds. Clones are taken from the branches of plants kept in a vegetative state by keeping them under at least 16 hours of light. Feminized seeds are produced from genetically female plants pollinated by genetically female plants that have had male flower production induced through chemical applications (gibberellic acid or ethylene inhibitors) so that pollination results in seeds that all are genetically female (Lubell and

Brand, 2018; Mohan Ram and Jaiswal, 1972). While these seeds will be XX females there will still be cases of male flower production (hermaphrodites) in the field. Farmers expect to find no more than 1 plant out of 4000 plants to produce male flowers (personal observation). When hermaphrodites are found, they are rogued from the fields.

Although the genomic resources available to hemp are obviously not at the level of the world's major crops, there have been many great advances made for studying the genetics of cannabinoid production (Taura and Morimoto, 1995; de Meijer, 2003; Weiblen et al., 2015; de Meijer et al., 2009), sex expression (Faux et al., 2016), fiber quality (van den Broeck et al., 2008), diversity and population structures (Sawler et al., 2015; Lynch et al., 2016; Dufresnes et al., 2017), and reference and draft genomes (van Bakel et al., 2011; Laverty et al., 2019).

With an increasing interest from farmers and consumers, planted hemp acreage has been increasing in the united states every year since the 2014 farm bill. With this growing interest, there is a great need for practical research to answer the questions needed to give farmers and consumers the information to make well educated choices on how to proceed with the planting and us of this crop.

In this study, we surveyed the genetic and chemical diversity of twenty-two commercially available hemp accessions. The sequencing was done to better understand genetic similarities between individual plants within an accessions and plants of different accessions. The chemical composition was studied to better understand the viability of these hemp accessions for on farm production, and to see if they the actual results were close to what the suppliers claimed they would be.

Materials and Methods

Plant material and cannabinoid collection

The plants used in this study were made up of twenty-two accessions commercially available in the United States, with accessions for seed, fiber, and cannabinoid production all represented. Twenty of the accessions were from seeds and two of them were clonally propagated. The clonal plants were produced from a single plant for each accession to ensure that each replicate was an exact genetic copy. The clones were taken by cutting stems seven inches long, trimming off all leaves and growing points except the top growing point and leaves. The stems were dipped in cloneX rooting solutions (https://www.growthtechnology.com/product/clonex/) to promote rooting. The accessions included both feminized and non-feminized seeds. The experiment was set up as a randomized complete block design with eight replicates. Prior to planting, the seeds were soaked in water for 12 hours to initiate germination. Fifteen seeds were planted into each pot to attempt to ensure that each plot would have a plant. The plants were grown in half gallon pots. Two weeks after planting, thinning of plants began and continued until each pot had only one plant. Plants were fertilized twice a week using a diluted 20-20-20 (1000ppms) and a diluted micronutrient mixture (Jackpot Micronutrient Mixture; 500 ppm). To maintain the plants in a controlled vegetative state, growth conditions were kept under an 18:6 light-dark cycle (18 hours of light and 6 hours of darkness per day) in greenhouses at the University of Georgia (Athens, Georgia). 52 days after sowing the plants were moved from the 18:6 light-dark cycle and they were placed into a flower room with a 12:12 light-dark cycle to initiate flowering. Any plants that showed male flowers were removed from the room to attempt to eliminate any pollination events. The female plants were kept in the flowering room for 12 weeks. After the twelve weeks, the panicles in the top six inches of each plant were harvested, trimmed of excess leaf material,

and placed in paper bags. For drying the bags holding the floral material were placed in a drying chamber. The drying chamber circulated 35° Celsius air.

Cannabinoid Potency Analysis

Chemical analysis was done using Shimadzu's LC2030C high-performance liquid chromatography (HPLC) machine. The components used for the HPLC analysis were the NexLeaf CBX column (2.7 µm, 4.6 x 150mm; part number: 220-91525-70), NexLeaf CBX guard column (part number: 220-91525-72), eleven-cannabinoid standard mix (part number: 220-91239-21), and high sensitivity method solvents A (0.085% phosphoric acid in water) and B (0.085% phosphoric acid in acetonitrile) (part number: 220-91394-81). The flow rate was 1.5 ml/min with a gradient of solvent B of 70% (initial) to a final concentration of solvent B of 95% (8 minutes). A guard column temperature of 35° Celsius was maintained by an internal oven. Injection volumes of 5µL were used. Standard curves were generated for each target cannabinoid with minimum correlation coefficients (R^2) of 0.999 over the six concentration levels (0.5ppm, 1ppm, 5ppm, 10ppm, 50ppm, and 100ppm).

Flower samples were weighed with a target weight of ~200 mg. The specific weights for each sample were entered into the Shimadzu software for calculation of accurate cannabinoid results. The flower samples were placed in 20 ml of methanol and agitated for 3 minutes to extract the cannabinoids into the solution. 1 ml of the supernatant was passed through 0.22 um filter, 50 ul of that supernatant were put into 950 ul of methanol. This resulted in a 400x dilution of the flower samples. These dilutions were then run through the HPLC and the quantification reports were generated showing the percentage by dry weight of the detectable cannabinoid. One sample per plant was analyzed with the HPLC to generate the cannabinoid potency data.

Genomic Data

Ten leaf punches were taken from each plant. The tissue samples were sent to LGC Genomics for DNA extraction and genotyping-by-sequencing (GBS) (Elshire et al., 2011) with restriction enzyme MsII. 150 bp paired-end reads were generated using Illumina NextSeq V500/550. Libraries were demultiplexed using the Illumina bcl2fastq software (version 2.17.14). SNPs were called using the genome analysis toolkit (GATK) haplotype caller (McKenna, et al., 2010).

After being called, low quality SNPs were filtered to remove sites that had > 30% missing data, >20% heterozygosity, and minor allele frequency <5%.

Diversity Analysis

A phylogenetic network was generated using NeighborNet method (Bryant and Moulton, 2003) in SplitsTree4 V4.14.4 (Huson and Bryant, 2006). Genotype data was converted to SplitsTree-compatible NEXUS file by first using TASSEL version 5.2.29 (Bradbury et al., 2007) to export as a PHYLIP (interleaved) format, which was then converted to NEXUS format using Alter (Glez-Peña et al., 2010) (http://www.sing-group.org/ALTER/). The resulting file was manually edited to change the data type to "dna", at which point the file was loaded into SplitsTree for network creation. A dendogram was also generated using TASSEL. A cladogram was generated using a neighbor joining cluster method, and archaeopteryx (Han & Zmasek, 2009) was used to generate the final dendograms.

Genomic Prediction

Genomic prediction was performed using basic GBLUP (Clark and van der Werf, 2013) analysis in TASSEL (Bradbury et al., 2007), as was done for pearl millet (see Chapter 3). Accuracy was assessed with 1000 iterations of 10-fold cross-validation.

Seed Source

The accessions were sourced from various companies (Table 5.2). It is important to note since many accessions are passed between groups it is impossible to say if these companies were the original sources of these accessions, or if they have done any selections on them to differentiate these accessions from wherever they might have originally received them.

Results

There were 176 pots planted, and three of the pots had zero seeds germinate. After the plants were put into flowering there were 148 female plants identified and 25 male plants. Eleven of the accessions had male plants, but none of those accessions made claims of being feminized seed. Eleven of the female plants had flowers that did not properly develop so they were not analyzed in the HPLC. There were 137 flower samples tested on the HPLC. THC levels in the flowers ranged from non-detectable levels up to 11.08% THC. Twelve plants produced THC without any CBD, with three of those plants producing flowers that were more than 1% THC. All of those plants were fiber plants. Fifteen of the plants produced flowers that were more than 1% THC, with seven of those producing flowers that were greater than 4% THC and one plant that had flowers material that was more than 10% THC (Figure 5.1). Of the fifteen plants that produced flowers with more than 1% THC eleven accessions were represented (AbacusxBB#1,

Chardonnay, Cherry_Wine, Berry_Blossom, Oregon_Melon, Baox, Cherry_Wine, Otto_II, Ka'uXXX, Wife) (Figure 5.1). 89 of the plants produced flowers with more than the legal limit of 0.3% THC (Figure 5.1). CBD levels in the flower ranged from non-detectable up to 16.66% (Figure 5.2). The fiber varieties showed the lowest levels of CBD, while the clonal varieties (BaoxSP_07 & LifterSP_01) and two seeded accessions (Cherry/Wu & Chardonnay) showed consistently high cannabinoid levels (Figure 5.2). The CBD to THC ratios varied from ~0 (for the plants that produced no CBD) up to ~28:1 (Figure 5.3). There were some plants that produced no detectable levels of THC and as a result could not have a CBD to THC ratio. These plants produced very low levels of CBD (Table 5.1). GF_BHWH only had one female plant develop, and that plant produced no detectable THC, so there is no information on the potential CBD to THC ratio for this accession. Twenty-eight of the plants had ratios that indicate the potential presence of a copy of the THCa synthase gene (Figure 5.3). Twelve of the accessions came with a certificate of analysis (COA). The COAs are supposed to show what level of cannabinoid production should be expected from the plant. For farmers, the value of their crop is nothing if it does not stay below 0.3% total THC, and if the crop is within the legal limit for THC then the farmer is paid based on the concentration of CBD. For farmers to accurately model their potential returns they need accurate information on the potential CBD concentration of the accessions they plant. When the provided COAs were compared to the mean results from the HPLC analysis almost all the accessions had much higher THC than their COAs showed (Figure 5.4) and all the accessions had lower CBD than their COAs showed (Figure 5.5).

GBS genotyping resulted in 8,105 markers after filtering. These markers were used to generate a phylogenetic tree (Figures 5.6 & 5.7). All the accessions where shown to closely cluster with each other, with no individual accessions showing large divergence from the group

(Figure 5.6). While none of the accessions showed large divergences, there were clusters of specific types of accessions. The clonal accessions grouped together and showed little within-accession variation (Figure 5.7); most of the within accessions variations, for the clonal accessions, are likely due to GBS genotyping errors, such as heterozygotes being called as homozygotes (Wallace and Mitchell, 2017). The fiber accessions(BHWH_Chinese_Fiber_Hemp, BHWH_For_Fiber, Ka'uXXX) also clustered together well with only a few individual plants showing any divergence (Figure 5.7). While most of the CBD accessions grown from seed showed little consistency, two of the accessions, Abacus Early Bird and Abacus Early Bird 2, clustered together tightly and showed little within-accession variation (Figure 5.6). One of the most interesting results was that two of the accessions that had the same name (baox), but were from different companies did not show any strong relationship to each other. This indicates that plants with the same names from different sources might not have any more genetic similarities with each other than any other accessions (Figure 5.7).

We attempted to use genomic selection to predict cannabinoid content, similar to how the striga analysis was done for pearl millet in Chapter 3. However, the 1000, 10-fold cross-validations indicated that none of the models had sufficient accuracy to make them usable, as all accuracies were less than 2% (data not shown). When we limited the training data to just work with clones and accessions which showed high levels of stability the accuracies increased (maximum 6.78%), but still not to a usable level.

Discussion

The current federal regulations have created a fine line between growing a *Cannabis sativa* plant that is a legal hemp plant and one that is an illegal marijuana plant. Inconsistencies

in plant genetics can greatly complicate the already arduous process of growing legal product, by making it difficult to predict when a field needs to be harvested to avoid going over the 0.3% federal total THC limit. Some farmers grow clonally propagated plants to ensure consistency, but currently many farmers are choosing to plant seeds over clones because of the high cost of clones compared to seeds. Beyond the legal considerations, inconsistencies in the genetics and seed feminizations can make it very difficult to produce hemp profitably (Meier and Mediavilla, 1998).

Plant Feminization

Many farmers have had to deal with seeds that were improperly feminized or not feminized at all. This has resulted in lawsuit and millions of lost capital and potential revenue (Associated Press, 2019). While this is a major issue for many farmers, none of the accessions used in this experiment that were sold as feminized seeds had any male flowers produced. However, it is important to note that a larger number of plants would need to be grown to verify that these accessions have a good feminization rate. For completely non-feminized seed it can be expected to have half the plants as male and half as female, and feminized seed should all be genetically female (XX) with a low number of male flowers produced. The rates at which genetically female plants produce male flowers is not well understood but is controlled by many genetic factors (Moliterni et al., 2004; Toth et al., 2020). Farmers consider a seed lot to be well feminized if less than 1 in 4000 plants produce male flowers (personal observation).

High Levels of THC in Supposedly Low-THC Accessions

The most concerning results from this experiment were the number of plants that produced excessively high levels of THC. Any plant producing CBD can be expected to produce THC, as it has been shown that the CBDa synthase gene naturally produces low levels of THC (Zirpel et al., 2018). However, if THC is produced in equal or greater concentrations than CBD produced, then that is indicative of the presence of a THCa synthase gene (Toth et al., 2020). Eleven of the twenty-two accessions showed evidence of an active THCa synthase gene (Table 5.1 & Figure 5.3), even though all were supposed to be low-THC varieties. Four of the accessions that showed evidence of the THCa synthase gene (BHWH_Chinese_Fiber_Hemp, BHWH_For_Fiber, Ka'uXXX, and Ka'uXX) had individual plants with no measurable CBD, indicating the possibility that there is no CBDa synthase gene present. Three of the accessions with plants that produced THC and no CBD were fiber accessions (Ka'uXX is the offspring of a fiber plant and a high cannabinoid plant), and while the purpose of the plant does not change the federal total THC limit, it is important to note that these fiber accessions would be either harvested prior to floral maturity (Hall et al., 2012), or if produced for planting seed would have cannabinoid productions greatly reduced (Meier and Mediavilla, 1998). These considerations mean that the fiber accessions likely pose little risk to anyone farming them, but the contamination of other seed lots by these fiber accessions could introduce the THCa synthase gene into plant populations where it could cause serious harm. While these fiber accessions pose little immediate risk, the likely presence of THCa synthase genes in the commercially available accessions sold for CBD production do pose an immediate and a major risk to farmers. The evidence shows that planting seed accessions can pose a large risk for the farmers because of the likelihood of THCa synthase genes being present in the seed lot (28/121 seeded plants tested

showed evidence of a THCa synthase gene). It is still important to note that not all seed accessions showed issues, so with small scale trials to test seed and supplier quality a farmer will likely be able to find a seed supplier that will sell stable, high quality seeds.

CBDa Synthase Produces THC

The presence of a THCa synthase gene will almost certainly ensure that a crop is not in compliance with federal regulations, but CBD accessions without THCa synthase genes will still likely produce above the federal limit of THC. As has been previously mentioned, the CBDa synthase gene will produce low levels of THC, which explains why almost all the plants tested showed some level of THC (table 5.1). The plants which produced the highest levels of CBD all produced too much THC at full floral maturity. Even though most plants will produce more THC than is federally allowed, farmers can still work with these types of plants. By regularly testing plants during flowering, farmers can harvest their plants prior to the levels of THC increasing beyond the legal limit.

Genetic Relationships

The results of the genetic analysis of the twenty-two accessions showed that most of them lacked genetic consistency (Figure 5.6). These genetic inconsistencies can result in the phenotypic inconsistencies and the associated problems that can come along with them. However, not all the accessions showed issues with genetic consistencies. The clonal, fiber, and Abacus Early Bird accessions all showed good evidence of within-accession genetic consistency. These again shows that clones are a safe option for farmers to go with, and that some seed accessions have genetic consistency, but most do not. While the clonal accessions did properly

cluster, and showed little within-accession variation, the grouping was not as tight as would normally be expected for clonal replicates. An increased sequencing depth is likely needed to properly deal with the issue of miscalled heterozygous sites. One of the most interesting results from the diversity analysis was that plant accessions with the same name did not show any type of genetic clustering. This implies that they are no more genetically related to each other than to any of the other accessions. While this issue and the general lack of genetic consistency might cause issues for farmers the increased genetic diversity is a valuable opportunity for plant breeders.

Genomic Selection

Although our attempt at genomic selection did not show high predictive accuracies, there are many things which could be done to improve the chances of success for future attempts at genomic selection in hemp. A larger training population would likely improve the accuracies. Another option would be using clonal replicates of all the plants. By taking clones from all the plants grown from seed, each genotype could have true replicates. This data could be used to get a better representation of what each plant's genetic potential for cannabinoid production would be, hopefully resulting in an improved accuracy. Future studies investigating the use of genomic selection for hemp will potentially be very useful. This is because taking advantage of the genetic diversity in the male plants for traits that are of interest in the female plants is not simple through traditional methods.

Conclusion

The hemp industry is experiencing many growing pains associated with its rapid development in the last six years. There are issues with genetic stability, economic viability, and

governmental regulations. Despite these issues the market continues to grow year after year, and interest in this crop continues to expand. With the support that this crop receives from consumers and the support it is beginning to receive from a wide range of researchers, there is a great opportunity for hemp to play an increasingly important role in a wide range of industries. It is essential that research on this crop continues so that farmers and consumers can have confidence in the crops that are grown and products that are produced from them. As excitement in this crop continues to grow it is important to remember to temper that excitement with caution. Hemp is not a miracle plant that can solve all the world's issues, but it is a crop with great potential that can serve an important role in improving the health and environmental stability of both individuals and industry.

Table 5.1: HPLC results of the CBD and THC percentage by dry weight of the female flower material.

Accession	PlantID	THC%	CBD%
Abacus_Early_Bird	HDGS_99	0.39	7.85
Abacus_Early_Bird	HDGS_5	0.39	7.23
Abacus_Early_Bird	HDGS_75	0.36	7.09
Abacus_Early_Bird	HDGS_121	0.24	5.14
Abacus_Early_Bird	HDGS_180	0.27	4.96
Abacus_Early_Bird	HDGS_29	0.28	4.7
Abacus_Early_Bird	HDGS_165	0.22	4.66
Abacus_Early_Bird	HDGS_53	0.24	4.3
Abacus_Early_Bird_2.0	HDGS_70	0.44	8.56
Abacus_Early_Bird_2.0	HDGS_15	0.36	6.37
Abacus_Early_Bird_2.0	HDGS_169	0.31	6.16
Abacus_Early_Bird_2.0	HDGS_97	0.28	5.33
Abacus_Early_Bird_2.0	HDGS_35	0.3	5.3
Abacus_Early_Bird_2.0	HDGS_86	0.28	5.09
AbacusxBB#1	HDGS_145	0.72	15.56
AbacusxBB#1	HDGS_89	0.36	6.82
AbacusxBB#1	HDGS_187	11.08	3.45
AbacusxBB#1	HDGS_25	1.49	1.28
AbacusxBB#1	HDGS_7	0.83	0.82
AbacusxBB#1	HDGS_107	0	0.74

AbacusxBB#1	HDGS_136	0	0.32
AbacusxBB#1	HDGS_58	0	0.18
Baox	HDGS_11	0.52	10.27
Baox	HDGS_157	0.45	9.7
Baox	HDGS_181	0.46	9.22
Baox	HDGS_140	0.42	8.59
Baox	HDGS_112	0.34	6.56
Baox	HDGS_77	4.05	4
BaoxSP_07	HDGS_102	0.62	13.05
BaoxSP_07	HDGS_31	0.64	12.66
BaoxSP_07	HDGS_131	0.49	12.48
BaoxSP_07	HDGS_63	0.59	12.3
BaoxSP_07	HDGS_173	0.54	11.64
BaoxSP_07	HDGS_148	0.54	11.6
BaoxSP_07	HDGS_6	0.53	10.11
BaoxSP_07	HDGS_82	0.46	9.08
Berry_Blossom	HDGS_109	0.36	7.53
Berry_Blossom	HDGS_61	0.27	6.94
Berry_Blossom	HDGS_149	4.72	5.17
Berry_Blossom	HDGS_48	0.18	4.88
Berry_Blossom	HDGS_90	0.13	2.84
Berry_Blossom	HDGS_175	0	0.19
Berry_Blossom	HDGS_123	0	0.14

BHWH_Chinese_Fiber_Hemp	HDGS_186	0.27	0.28
BHWH_Chinese_Fiber_Hemp	HDGS_108	0.45	0
BHWH_Chinese_Fiber_Hemp	HDGS_127	0.15	0
BHWH_Chinese_Fiber_Hemp	HDGS_18	0.2	0
BHWH_Chinese_Fiber_Hemp	HDGS_46	0.66	0
BHWH_Chinese_Fiber_Hemp	HDGS_93	0.46	0
BHWH_For_Fiber	HDGS_191	0.6	0
BHWH_For_Fiber	HDGS_68	0.34	0
BHWH_For_Fiber	HDGS_9	0.68	0
C4	HDGS_49	0.22	5.63
C4	HDGS_137	0.16	4.08
C4	HDGS_32	0.14	3.45
C4	HDGS_150	0	1.57
C4	HDGS_21	0	0.92
C4xBB#1	HDGS_114	0.33	8.01
C4xBB#1	HDGS_87	0.23	5.88
C4xBB#1	HDGS_141	0.15	3.56
C4xBB#1	HDGS_66	0.14	3.47
C4xBB#1	HDGS_154	0.13	3.29
C4xBB#1	HDGS_190	0.11	2.63
C4xBB#1	HDGS_22	0.12	2.36
C4xBB#1	HDGS_41	0.1	2.13
Chardonnay	HDGS_12	0.73	15.76

Chardonnay	HDGS_72	0.62	12.51
Chardonnay	HDGS_124	0.55	11.21
Chardonnay	HDGS_91	0.53	10.94
Chardonnay	HDGS_176	0.53	10.8
Chardonnay	HDGS_152	0.49	10.56
Chardonnay	HDGS_106	0.48	9.81
Chardonnay	HDGS_42	5.41	7.03
Cherry_original	HDGS_57	0.25	5.65
Cherry_original	HDGS_34	0.17	3.99
Cherry_original	HDGS_129	0.12	2.82
Cherry_original	HDGS_147	0.09	2.03
Cherry_original	HDGS_98	0	0.22
Cherry_original	HDGS_3	0	0.19
Cherry_original	HDGS_185	0	0.14
Cherry_original	HDGS_78	0	0
Cherry_Wine	HDGS_122	0.59	12.58
Cherry_Wine	HDGS_52	5.21	8.63
Cherry_Wine	HDGS_166	0.33	6.17
Cherry_Wine	HDGS_85	0.22	5.38
Cherry_Wine	HDGS_174	2.85	4.99
Cherry_Wine	HDGS_24	0.1	1.79
Cherry/Wu	HDGS_101	0.69	15.46
Cherry/Wu	HDGS_95	0.61	13.46

Cherry/Wu	HDGS_135	0.6	12.96
Cherry/Wu	HDGS_43	0.62	12.8
Cherry/Wu	HDGS_13	0.54	10.81
Cherry/Wu	HDGS_158	0.47	10.05
Cherry/Wu	HDGS_54	0.47	9.08
GF_BHWH	HDGS_163	0	1.93
Ka'uXX	HDGS_105	0.48	10.29
Ka'uXX	HDGS_50	0.5	9.97
Ka'uXX	HDGS_130	0.31	8.18
Ka'uXX	HDGS_2	0.69	0
Ka'uXXX	HDGS_37	0.36	0.65
Ka'uXXX	HDGS_143	2.57	0
Ka'uXXX	HDGS_156	2.7	0
Ka'uXXX	HDGS_59	1.56	0
LifterSP_01	HDGS_26	0.79	16.66
LifterSP_01	HDGS_8	0.61	12.23
LifterSP_01	HDGS_125	0.52	11.16
LifterSP_01	HDGS_164	0.52	11.09
LifterSP_01	HDGS_67	0.53	10.89
LifterSP_01	HDGS_116	0.49	10.51
LifterSP_01	HDGS_79	0.47	9.86
LifterSP_01	HDGS_188	0.46	9.84
Oregon_Melon	HDGS_40	0.37	6.85

Oregon_Melon	HDGS_19	0.28	4.83
Oregon_Melon	HDGS_80	0.17	4.63
Oregon_Melon	HDGS_62	0.17	4.06
Oregon_Melon	HDGS_179	0.14	3.88
Oregon_Melon	HDGS_138	4.37	2.59
Otto	HDGS_142	0.65	13.13
Otto	HDGS_182	0.53	11.48
Otto	HDGS_117	0.56	11.41
Otto	HDGS_69	0.54	11.38
Otto	HDGS_92	0.5	9.5
Otto	HDGS_160	0.48	9.44
Otto	HDGS_33	0.45	9.05
Otto	HDGS_16	0.11	1.87
Otto_II	HDGS_104	0.67	13.98
Otto_II	HDGS_132	0.45	9.43
Otto_II	HDGS_171	0.35	9.06
Otto_II	HDGS_155	0.21	5.87
Otto_II	HDGS_83	4.41	5.64
Otto_II	HDGS_36	1.77	2.99
Wife	HDGS_172	0.58	11.86
Wife	HDGS_47	0.51	8.69
Wife	HDGS_74	0.16	3.95
Wife	HDGS_167	1.29	1.63

Wife	HDGS_4	1.58	1.49
Wife	HDGS_126	0.93	0.92

Table 5.2: Accession origins

Accession	**Source**
Ka'uXXX	Earth Matters Hemp (https://hawaiicannabis.org/earth-matters-farm/)
Ka'uXX	Earth Matters Hemp (https://hawaiicannabis.org/earth-matters-farm/)
Cherry (original)	Blank Land Botanicals (https://blacklandsbotanicals.org/)
Wife	Blank Land Botanicals (https://blacklandsbotanicals.org/)
Abacus_Early_Bird	Colorado CBD (https://www.coloradocbdseed.com/)
BaoxSP_07	Cross Creek Hemp (https://crosscreekhemp.com/)
AbacusxBB#1	Blank Land Botanicals (https://blacklandsbotanicals.org/)
LifterSP_01	Cross Creek Hemp (https://crosscreekhemp.com/)
BHWH_For_Fiber	Bulk Hemp Warehouse (https://www.bulkhempwarehouse.com/)
Berry Blossom	Blank Land Botanicals (https://blacklandsbotanicals.org/)
Baox	GaXtracts (https://www.gaxtracts.com/)
Chardonnay	GaXtracts (https://www.gaxtracts.com/)
Cherry/Wu	GaXtracts (https://www.gaxtracts.com/)
Otto II	Blank Land Botanicals (https://blacklandsbotanicals.org/)
Abacus_Early_Bird_2.0	Colorado CBD (https://www.coloradocbdseed.com/)
Otto	GaXtracts (https://www.gaxtracts.com/)
BHWH_Chinese_Fiber_Hemp	Bulk Hemp Warehouse (https://www.bulkhempwarehouse.com/)
Oregon Melon	GaXtracts (https://www.gaxtracts.com/)
GF_BHWH	Bulk Hemp Warehouse (https://www.bulkhempwarehouse.com/)
C4	Blank Land Botanicals (https://blacklandsbotanicals.org/)
C4xBB#1	Blank Land Botanicals (https://blacklandsbotanicals.org/)
Cherry Wine	Blank Land Botanicals (https://blacklandsbotanicals.org/)

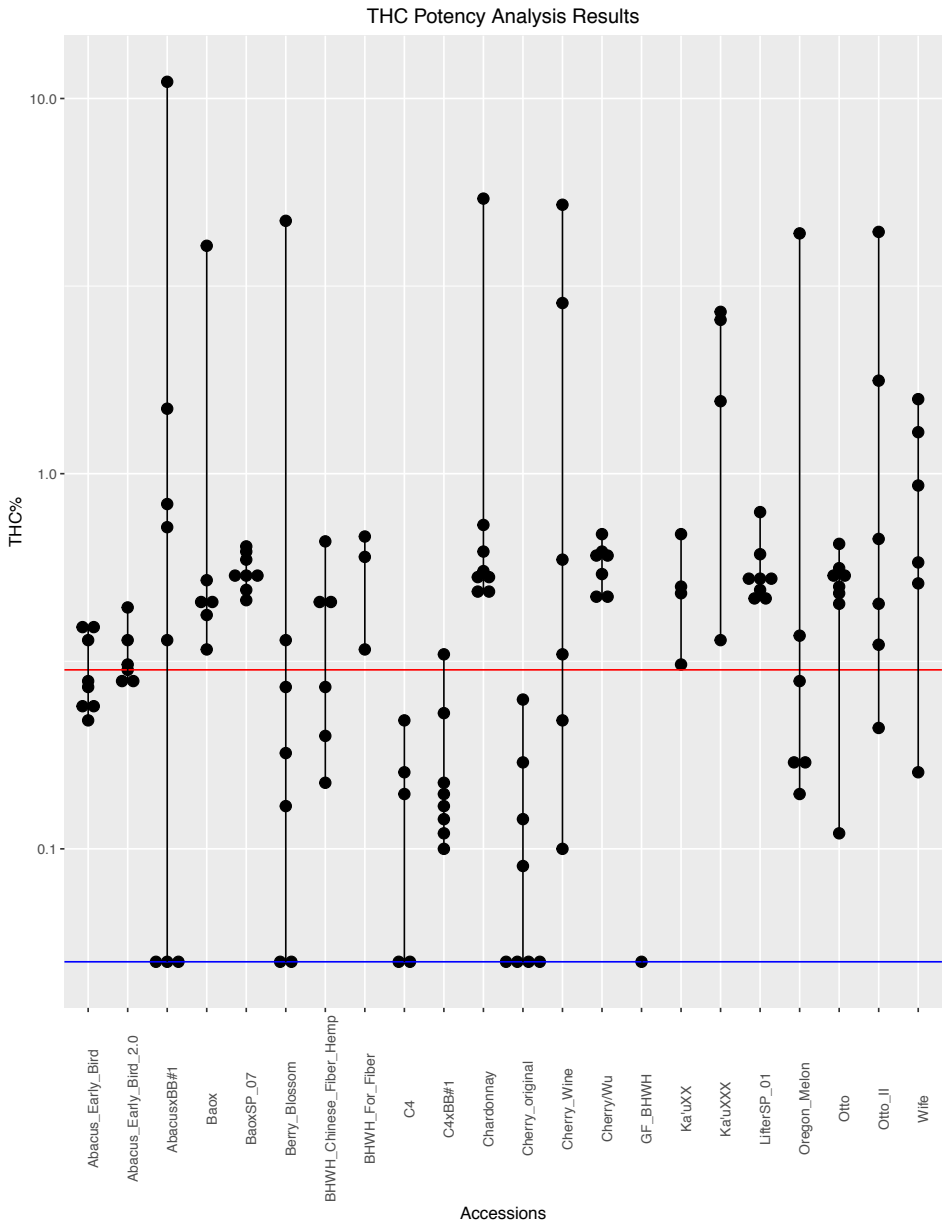

Figure 5.1: THC HPLC results
THC potency results generated from HPLC runs. THC levels were measured in % of THC by dry weight. The data was log transformed and values of zero changed to 0.05 to allow for plotting; vertical black lines show the range within each accession and are included for clarity. The red line indicates the legal limit of THC (0.3%) all points above this line show plants that were above the legal limit. Points on the blue line show the plants that had no detectable levels of THC. Almost all accessions included at least some plants with unacceptably high THC levels, and in some accessions all plants were too high.

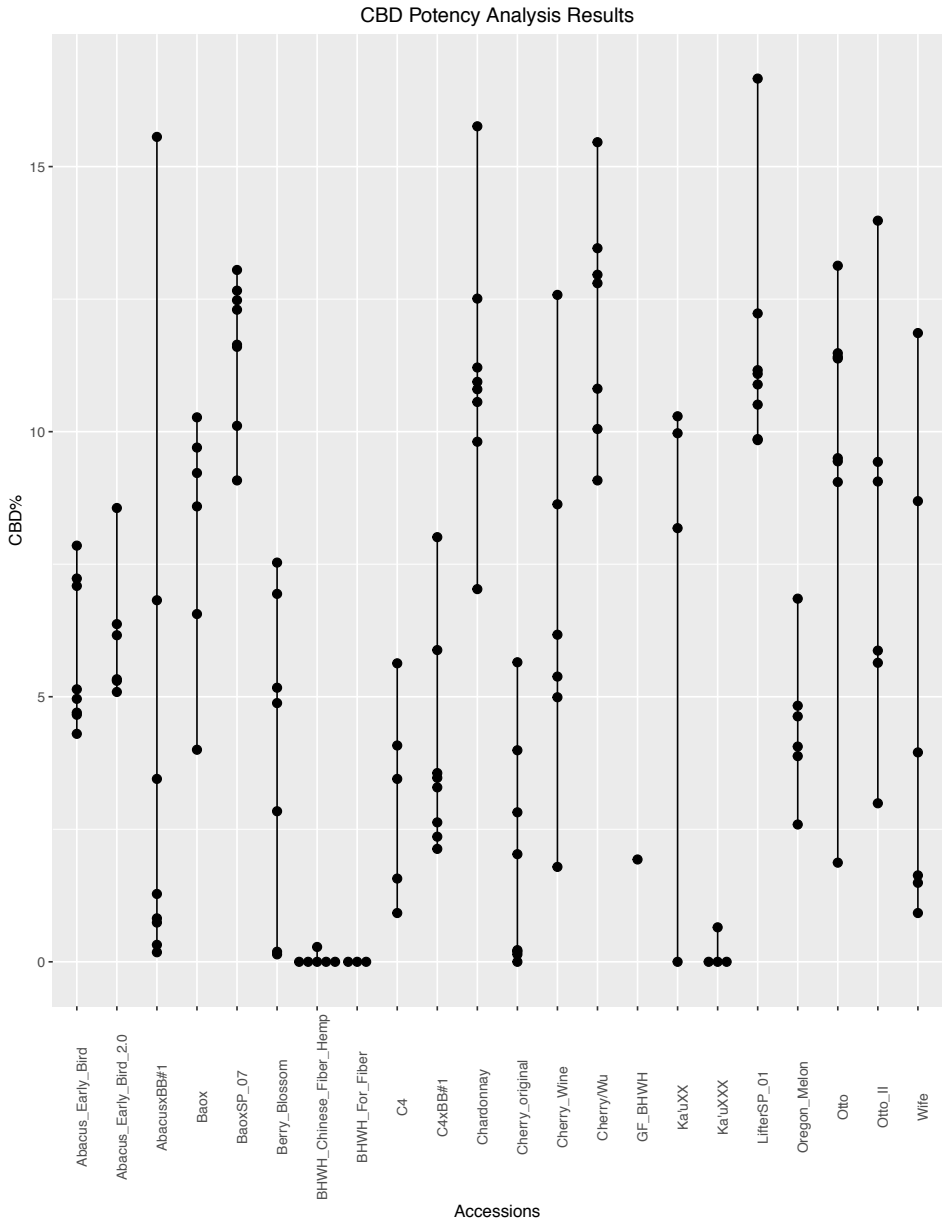

Figure 5.2: CBD HPLC Results
CBD potency results generated from HPLC runs, displayed as in Figure 5.1. CBD levels were measured in % of CBD by dry weight.

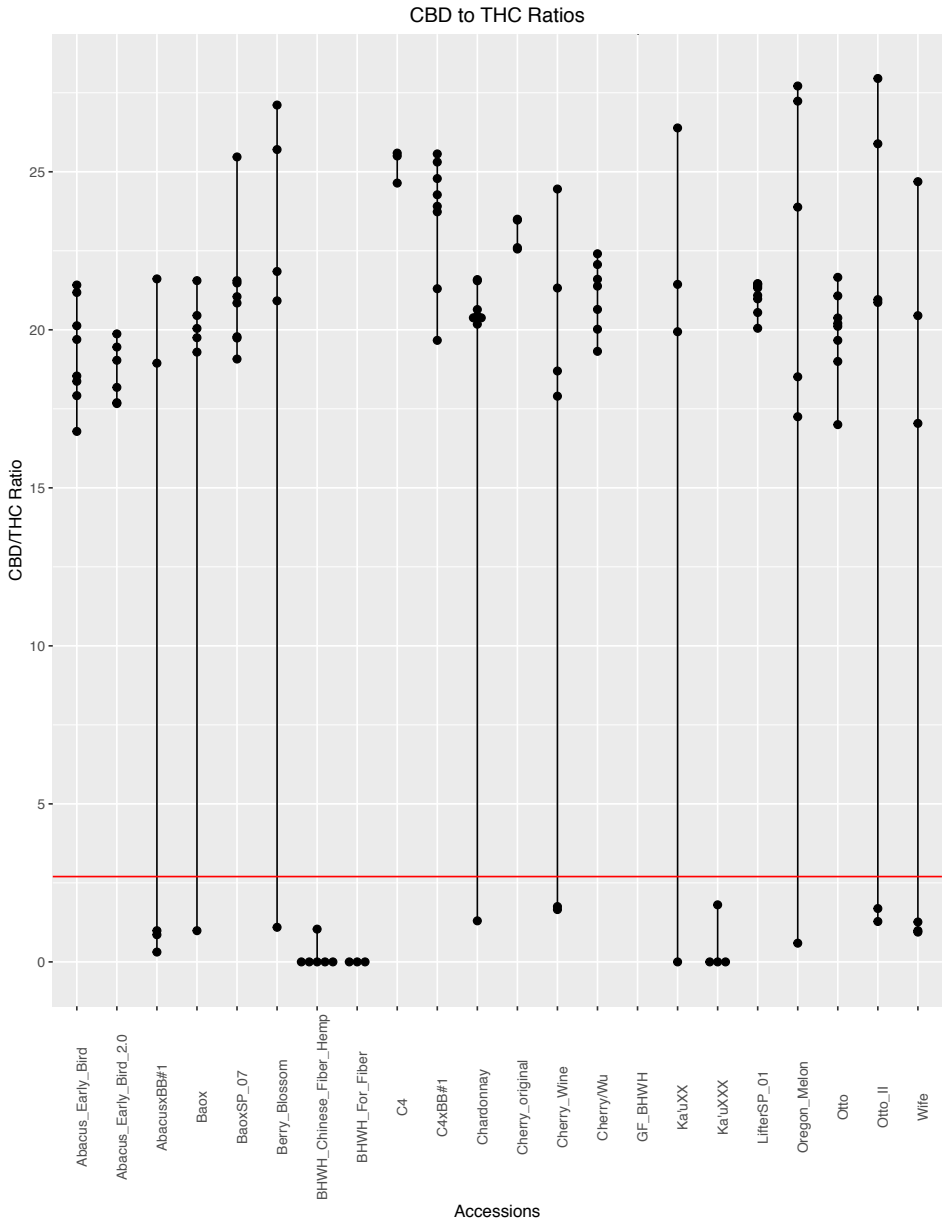

Figure 5.3: Ratios CBD/THC
Plot of the ratios of CBD to THC for plants with a detectable THC level. The points below the red line indicate plants that likely had a copy of the THCa synthase gene.

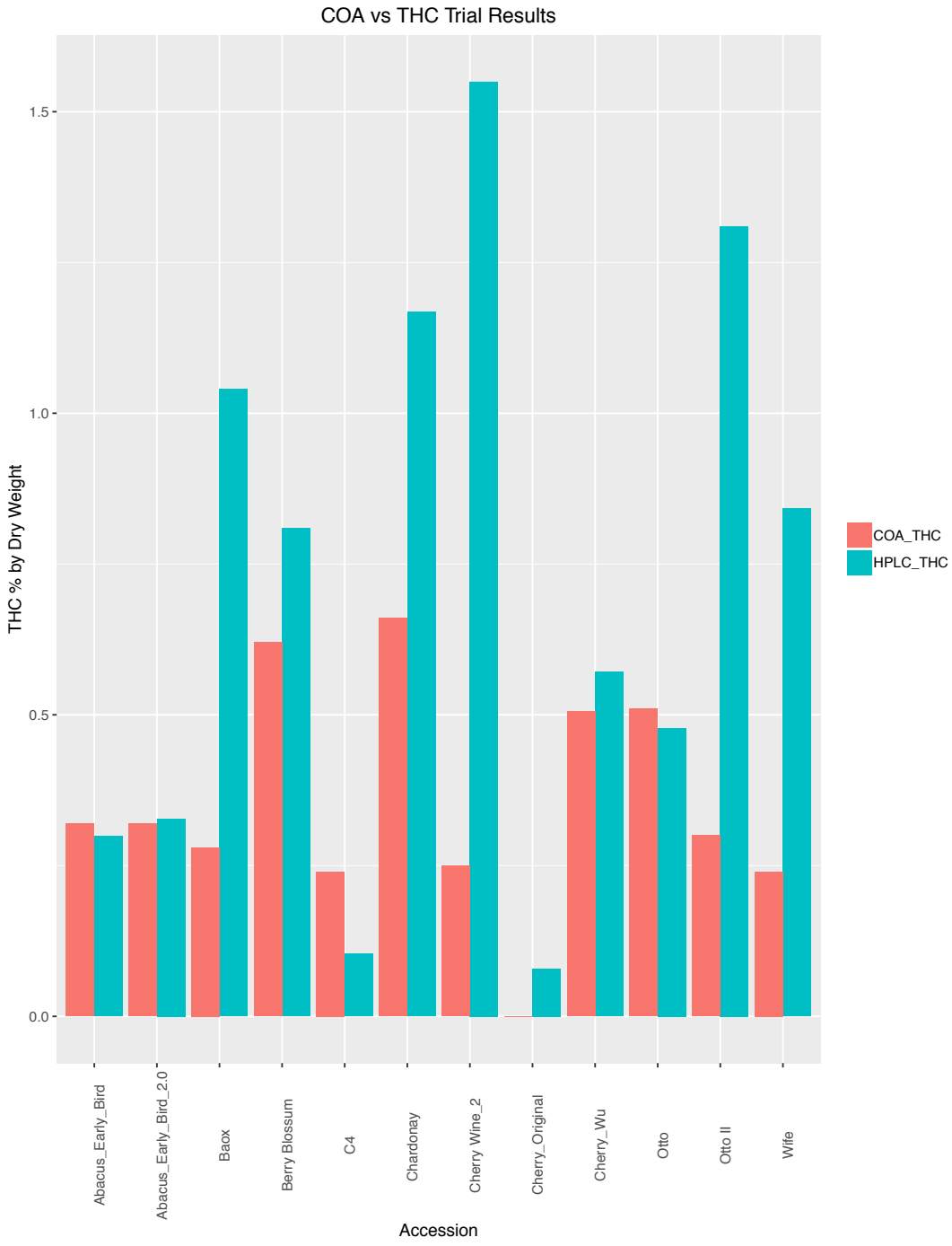

Figure 5.4: THC COA vs HPLC
Comparison of the mean THC % by dry weight from the plants grown for this study compared to the claimed results provided by the certificate of analysis from seed supplier.

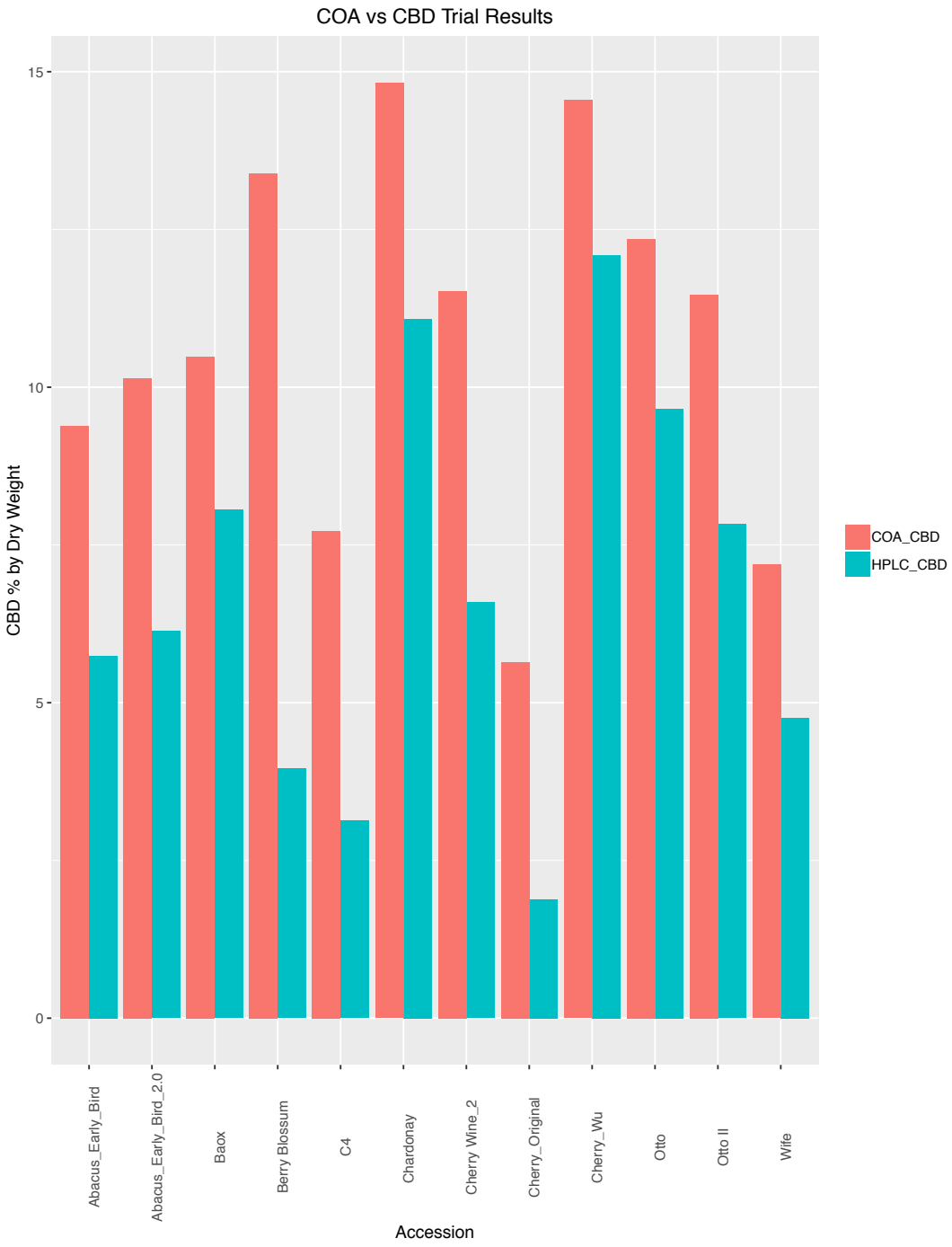

Figure 5.5: CBD COA vs HPLC
Comparison of the mean CBD % by dry weight from the plants grown for this study compared to the claimed results provided by the certificate of analysis from seed supplier.

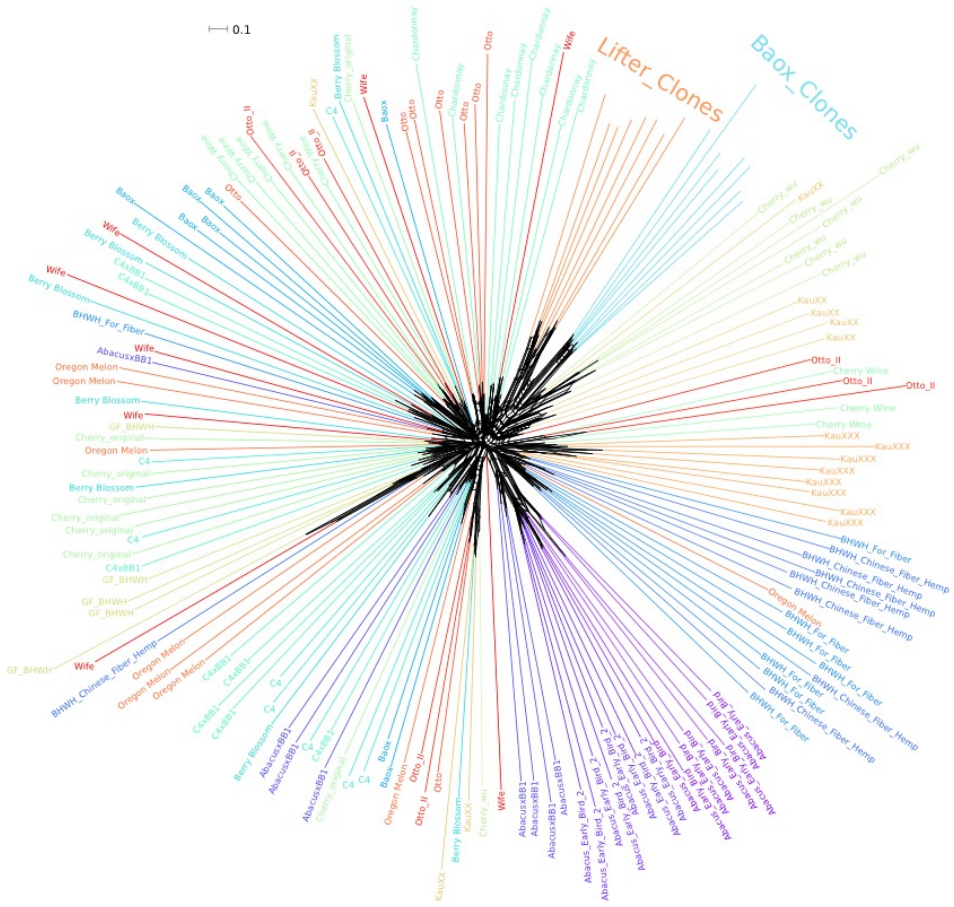

Figure 5.6: Phylogenetic tree
Phylogenetic tree for the 173 plants from twenty-two hemp accessions. Each tip is a single plant, and branching indicates alternative options. The lines are colored and labeled by accession.

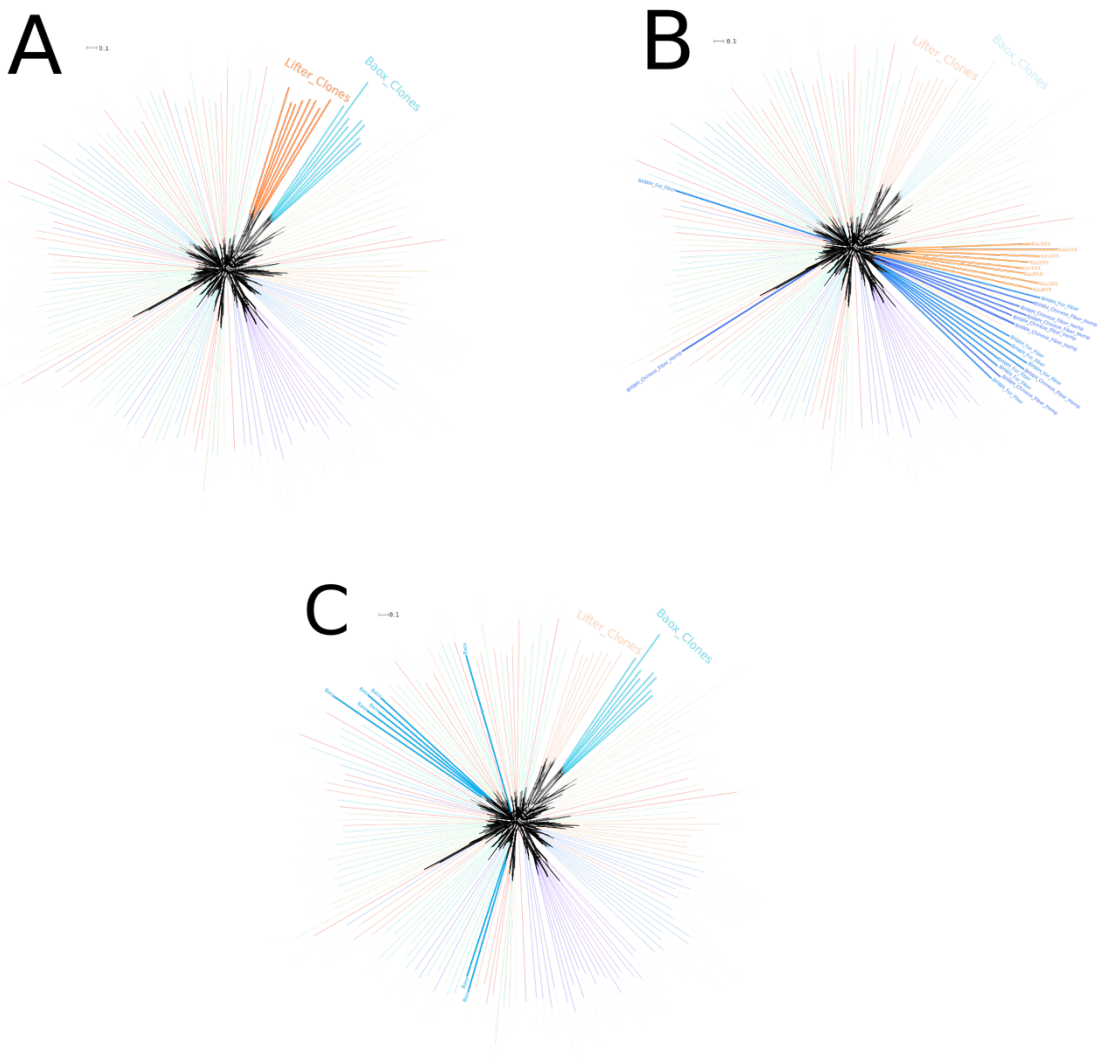

Figure 5.7: Clones highlighted phylogenetic tree
Phylogenetic trees focusing on the clonal accessions (A), fiber accessions (B), and accessions with the same name (Baox) from different countries (C).

Figure 5.8: Hemp Denodogram
A neighbor-joining tree of commercially available hemp accessions was constructed in TASSEL based on genetic polymorphisms and exported using Archaeopteryx. Samples show clustering consistent with other hemp figure (5.6)

REFERENCES

Associated Press (2019). Oregon hemp seed company sued for $44 million after crops allegedly failed. Retrieved May 21, 2020, from https://www.statesmanjournal.com/story/news/2019/09/28/oregon-company-sued-44-million-allegedly-selling-worthless-hemp-seeds/3805696002/

Amaducci, S., Scordia, D., Liu, F. H., Zhang, Q., Guo, H., Testa, G., and Cosentino, S. L. 2015. Key cultivation techniques for hemp in Europe and China. Industrial Crops and Products, 68, 2–16.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23: 2633-2635.

Bryant, D. and V. Moulton. 2003. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. Molecular Biology and Evolution 21: 255-265. doi:10.1093/molbev/msh018.

Callaway, J. C. 2004. Hempseed as a nutritional resource: An overview. EUPHYTICA, (NO 1-2), 65.

Cherniak, L. 1982. The Great Books of Cannabis vol. I, Book II. Damele Publishing, Oakland, CA.

Clark, S. A., and van der Werf, J. 2013. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. METHODS IN MOLECULAR BIOLOGY -CLIFTON THEN TOTOWA-, 321.

de Meijer, E.P.M., K.M. Hammond, and A. Sutton. 2009. The inheritance of chemical phenotype in Cannabis sativa L.(IV): Cannabinoid-free plants. Euphytica 168(1):95–112. doi:10.1007/s10681-009-9894-7

de Meijer, E.P.M., M. Bagatta, A. Carboni, P. Crucitti, V.M.C. Moliterni, P. Ranalli, and G. Mandolino. 2003. The inheritance of chemical phenotype in Cannabis sativa L. Genetics 163(1):335–346.

Drug Policy Alliance, The DEA: Four Decades of Impeding and Rejecting Science, Multidisciplinary Associations for Psychedelic Studies, 2014

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PloS one 6: e19379.

ElSohly, M. A., and Slade, D. 2006. Chemical constituents of marijuana: The complex mixture of natural cannabinoids. LIFE SCIENCES, (5), 539.

Eriksson, M. E., Israelsson, M., Olsson, O., and Moritz, T. 2000. Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. Nature Biotechnology, 18(7), 784.

Faux, A.-M., Berhin, A., Dauguet, N., and Bertin, P. 2014. Sex chromosomes and quantitative sex expression in monoecious hemp (*Cannabis sativa* L.). Euphytica, 196(2), 183–197.

Faux, A.-M., Draye, X., Flamand, M.-C., Occre, A., and Bertin, P. 2016. Identification of QTLs for sex expression in dioecious and monoecious hemp (Cannabis sativa L.). Euphytica, 209(2), 357–376.

Fleming, M. P., Clarke, R. C., Physical evidence for the antiquity of *Cannabis sativa*. L. J. Int. Hemp Assoc. 1998, 5, 280.

Galasso, I., Russo, R., Mapelli, S., Ponzoni, E., Brambilla, I. M., Battelli, G., and Reggiani, R. 2016. Variability in Seed Traits in a Collection of *Cannabis sativa* L. Genotypes. FRONTIERS IN PLANT SCIENCE, 7.

Galoch, Elżbieta. 2015. The hormonal control of sex differentiation in dioecious plants of hemp (*Cannabis sativa*). The influence of plant growth regulators on sex expression in male and female plants. Acta Societatis Botanicorum Poloniae, (1–2), 153.

Glass, M., and Gilleece, T. 2019. Cannabinoids: a possibility for pain management in the palliative cancer pathway. Poster session presented at Annual Radiotherapy Conference 2019, Newcastle, United Kingdom.

Hall, J., Bhattarai, S. P., and Midmore, D. J. 2012. Review of Flowering Control in Industrial Hemp. JOURNAL OF NATURAL FIBRES, (1), 23.

Han, M. V., & Zmasek, C. M. 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinformatics, 10, 356.

Hurd, Y. L., Spriggs, S., Alishayev, J., Winkel, G., Gurgov, K., Kudrich, C., … Salsitz, E. 2019. Cannabidiol for the Reduction of Cue-Induced Craving and Anxiety in Drug-Abstinent Individuals With Heroin Use Disorder: A Double-Blind Randomized Placebo-Controlled Trial. The American Journal Of Psychiatry, appiajp201918101191. https://doi.org/10.1176/appi.ajp.2019.18101191

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. Molecular biology and evolution 23: 254-267.

Karus, M., and Vogt, D. 2004. European hemp industry: Cultivation, processing and product lines. EUPHYTICA, (NO 1-2), 7.

Laverty, K. U., Stout, J. M., Sullivan, M. J., Shah, H., Gill, N., Holbrook, L., … Page, J. E. 2019. A physical and genetic map of Cannabis sativa identifies extensive rearrangements at the THC/CBD acid synthase loci. Genome Research, (1), 146.

Lewis, M. A., Russo, E. B., & Smith, K. M. 2018. Pharmacological foundations of Cannabis chemovars. Planta Medica, 84(4), 225–233.

Li, H. 1974. An Archaeological and Historical Account of Cannabis in China. Economic Botany, 28(4), 437-448.

Lubell, J. D., and Brand, M. H. 2018. Foliar Sprays of Silver Thiosulfate Produce Male Flowers on Female Hemp Plants. HORTTECHNOLOGY, 28(6), 743–747.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. GENOME RESEARCH, 9, 1297.

Meier C. and Mediavilla V. 1998. Factors influencing the yield and the quality of essential hemp oil (*Cannabis sativa* L.). J. of the International Hemp Association, 5, n°1, 16-20.

Merlin, M. D., Archaeological Evidence for the Tradition of Psychoactive Plant use in the Old World Economic Botany 2003, 57, 295.

Mohan Ram, H. Y., and Jaiswal, V. S., 1972. Induction of Male Flowers on Female Plants of Cannabis sativa by Gibberellins and Its Inhibition by Abscisic Acid. Planta, 105(3), 263.

Mohan Ram, H. Y., and Sett, R. 1982. Reversal of Ethephon-Induced Feminization in Male Plants of Cannabis sativa by Ethylene Antagonists. Zeitschrift Für Pflanzenphysiology, 107(1), 85–89.

Moliterni, V. M. C., Cattivelli, L., Ranalli, P., & Mandolino, G. 2004. The sexual differentiation of Cannabis sativa L.: A morphological and molecular study. EUPHYTICA, NO 1-2, 95.

Nelson, Clarence H. 1944. Growth Responses of Hemp to Differential Soil and Air Temperatures. Plant Physiology, 19(2), 294.

Petit, J., Salentijn, E., Paulo, M. J., Thouminot, C., van Dinter, B. J., Magagnini, G., Gusovius, H. J., Tang, K., Amaducci, S., Wang, S., Uhrlaub, B., Müssig, J., & Trindade, L. M. 2020. Genetic Variability of Morphological, Flowering, and Biomass Quality Traits in Hemp (Cannabis sativa L.). Frontiers in plant science, 11, 102. https://doi.org/10.3389/fpls.2020.00102

Perucca E. 2017. Cannabinoids in the Treatment of Epilepsy: Hard Evidence at Last?. Journal of epilepsy research, 7(2), 61–76. doi:10.14581/jer.17012

Russo, E. B. 2007. History of Cannabis and its preparations in saga, science, and sobriquet. Chem. Biodivers. 4: 1614–1648.

S.2667S. 2667 — 115th Congress: Hemp Farming Act of 2018." www.GovTrack.us. 2018. March 9, 2020 https://www.govtrack.us/congress/bills/115/s2667

Sakamoto, K., Akiyama, Y., Fukui, K., Kamada, H., and Satoh, S. 1998. Characterization: Genome Sizes and Morphology of Sex Chromosomes in Hemp (Cannabis sativa L.). CYTOLOGIA- INTERNATIONAL JOURNAL OF CYTOLOGY, (4), 459.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., and Taura, F. 2004. The gene controlling marijuana psychoactivity: molecular cloning and heterologous expression of Delta1-tetrahydrocannabinolic acid synthase from Cannabis sativa L. The Journal Of Biological Chemistry, 279(38), 39767–39774.

Small, E. 2015. Evolution and Classification of Cannabis sativa (Marijuana, Hemp) in Relation to Human Utilization. Botanical Review, 81(3), 189–294.

Taura, F., and S. Morimoto. 1995. First direct evidence for the mechanism of. DELTA. 1-tetrahydrocannabinolic acid biosynthesis. J. Am. Chem. Soc. 117(38):9766– 9767. doi:10.1021/ja00143a024

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., and Morimoto, S. 2007. Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. FEBS Letters, 581(16), 2929–2934.

Toth, J. A., Stack, G. M., Cala, A. R., Carlson, C. H., Wilk, R. L., Crawford, J. L., Viands, D. R., Philippe, G., Smart, C. D., Rose, J. K. C., & Smart, L. B. 2020. Development and validation of genetic markers for sex and cannabinoid chemotype in Cannabis sativa L. GCB Bioenergy, 12(3), 213–222. https://doi.org/10.1111/gcbb.12667

Turner, J. C., Hemphill, J. K., & Mahlberg, P. G. 1981. Interrelationships of glandular trichomes and cannabinoid content. II. Developing vegetative leaves of Cannabis sativa L. (Cannabaceae). Bulletin on Narcotics, 33(3), 63–71.

U.S. Hemp Crop Report. (2018). Retrieved from https://www.votehemp.com/u-s-hemp-crop-report/

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., and Page, J. E. 2011. The draft genome and transcriptome of Cannabis sativa. Genome Biology (Online Edition), R102.

van den Broeck, H. C., Maliepaard, C., Ebskamp, M. J. M., Toonen, M. A. J., and Koops, A. J. 2008. Differential expression of genes involved in C1 metabolism and lignin biosynthesis

in wooden core and bast tissues of fiber hemp (Cannabis sativa L.). Plant Sci. 174, 205–220.

Wallace, J.G. and Mitchell, S.E. 2017. Genotyping-by-sequencing. Curr. Protoc. Plant Biol. 2:64-77. doi: 10.1002/cppb.20042

Weiblen, G.D., J.P. Wenger, K.J. Craft, M.A. ElSohly, Z. Mehmedic, E.L. Treiber, and M.D. Marks. 2015. Gene duplication and divergence affecting drug content in Cannabis sativa. New Phytol. 208(4):1241–1250.

Zirpel, B., Kayser, O., & Stehle, F. 2018. Elucidation of structure-function relationship of THCA and CBDA synthase from Cannabis sativa L. JOURNAL OF BIOTECHNOLOGY, 284, 17–26. https://doi-org.proxy-remote.galib.uga.edu/10.1016/j.jbiotec.2018.07.031

CHAPTER 5

CONCLUSIONS

There are tens of thousands of edible plants that humans can eat, but less than 20 of those plants are used to satisfy most of the worlds agricultural needs. These major crops have been the focus of most of the research and development conducted in agriculture over the past 100 years, and have seen significant increases in many valuable traits. The remainder of the plants that have been farmed around the world, but were not part of the research efforts that brought about our modern agricultural system, have seen little to no improvement in most of their traits. These crops, known as orphaned crops, have played important roles in the agricultural systems around the world, but tended to only have regional significance. They have often been farmed under the same systems for centuries, but in recent years these orphaned crops have garnered more interest from both consumers and researchers. With the increased investments in research and development, some of these crops are starting to move away from the title of orphaned crops, and move toward the title of modern crops.

The crops that were studied in this work--kodo millet (*Paspalum scrobiculatum* L.), little millet (*Panicum sumatrense)*, proso millet (*Panicum miliaceum* L.), pearl millet (*Pennisetum glaucum* syn. *Cenchrus americanus)*, and hemp (*Cannabis sativa* L.)--have in many ways been left behind by the modern developments of the world's agricultural systems. However, pearl millet and hemp are beginning to receive the attention needed for them to pull into the modern age, and proso millet has had some significant work done on it as well. Little millet and kodo

millet are still being largely ignored. Regardless of how much or how little work has been done on these five crops, all of them still need more work conducted to further develop the resources available for them. Each of the studies in this project worked to further develop these crops and bring them a few more steps towards a full integration into our modern agricultural systems.

The pearl millet study explored the possibility of using genomic selection to develop varieties with increased resistances to *Striga* (a parasitic weed) and downy mildew. Pearl millet is a significant source of income and food to about ninety-million people around the world, but has received relatively little work towards research and development compared to the major crops in the world. Despite the limited resources put into pearl millet, it is widely grown in the developing world because of its capacity to produce yields on marginal lands. *Striga* and downy mildew are two stresses which cause significant losses for pearl millet yield. In our study, we investigated genomic selection as a possible breeding approach to improve resistances to these stresses. Breeding populations, which had been phenotyped for resistances to *Striga* and downy mildew, were genotyped and analyzed. We discovered that genomic selection has good potential to be a viable breeding approach for developing varieties with improved resistances to *Striga* and downy mildew. The implementation of genomic selection for these traits has many potential benefits over traditional selections. It could increase the number of generations that a breeder can work with each year by allowing for selections in a greenhouse. Along with increasing the number of generations per year, selections can be done without the presence of either downy mildew or *Striga*. Our results will need to be validated through implementation in the breeding program, but it is our hope that this will improve the speed at which resistant varieties can be developed.

The minor millets (kodo millet, little millet, and proso millet) projected focused on developing resources for the improvement of these crops, and worked to better understand the underlying genetic relationships. Kodo millet and little millet have been almost entirely ignored by the modern developments in agriculture. Proso millet has received some focus, but still is far behind other crops in terms of research and development. For this project, genomic marker data sets were developed from samples of the core collections of each of the millet species. These marker data sets can serve as genomic resources for other researchers looking to study and improve these crops. We used these marker data sets to understand the population structure and how that related to geographic origin of the various core collection accessions. It was discovered that the accessions which showed genetic clustering tended to be from similar geographic origins. The data were also used to see how well the race designation correlated with genetic relatedness, and it was shown that the race designations showed no correlation to genetic relatedness. These small advances, paired with other research, continue to slowly push all these minor millets into the modern world, but there is still a long way still left to go.

The hemp study focused on identifying issues that need to be addressed, and improving the information of our germplasm. Legal hemp cultivation restarted in the US in 2014, and since then farmers have had many issues cultivating this crop. For researchers to solve the issues causing problems for hemp farmers, properly identifying the sources of these issues is an essential first step. In our study, we showed that there is a lack of genetic consistency in many commercially available hemp accessions, and that this lack of genetic consistency correlated with a lack of consistency of cannabinoid levels. The cannabinoid variances represent the most significant issue for farmers. We showed that this lack stability has left significant proportions of the studied accessions with THCa expression consistent with the presence of a copy of the THCa

synthase gene. For hemp to be legally sold it must undergo government-mandated testing, and if one of the plants that a government representative selects for testing carries a copy of the THCa synthase gene a farmer is going to likely run into multiple issues. The first issue is that farmer will lose their entire crop and have no way to recoup the sunken cost. The second issue is that the farmer risks being labeled as negligent (the federal government may consider a farmer negligent if they grow plants with more than 0.5% total THC) and lose their opportunity to farm hemp. The third issue is that depending on the farmer's local law enforcement agencies, they could be at risk for drug prosecution. This is because, if there are plants that test very high in THC (in our study we found one plant with over 10% THC) the farmer could potentially be prosecuted for production of marijuana.

While keeping in line with federal regulations is the most important goal when farming hemp, other inconsistencies can cause major issues for farmers. When an accessions lack consistencies of important agronomic traits, such as flowering time and plant architecture, it can become much more difficult for farmers to successfully farm this crop. Many modern field crop accessions have been bred to have consistent flowering time to allow for a single harvest. Those field crops have also been selected for plant architectures that allow for mechanized harvesting. Developing improved varieties is the best way to protect farmers from the issues discussed previously. The genotyping of twenty-two accessions from our germplasm increases our capacity to breed improved varieties. This understanding of their genetic relatedness gives us a better chance of quickly stabilizing traits of interest. With the interest in the crop continuing to grow it is likely that the resources needed to improve this crop will continue being invested, and we expect to see rapid and significant improvement in many of the valuable traits of hemp.

It took the better part of a century for the major crops of the world agriculture systems to become what they are today. For orphaned crops, it doesn't need to take so long. Our current understanding of plant genetics and our modern plants breeding tools gives us the potential to rapidly improve orphaned crops and bring them more fully into the modern agricultural systems. However, it must be considered that for an investment to be worth making there must be a return that is greater than the value of the investment. When looking at investing into orphaned crops their values must be considered. It is not possible to invest vast resources into developing the potential of every orphaned crop so specific crops must be selected. While this means that there will likely always be a significant number of orphaned crops left any advances made to improve the diversity of available crops will be good. Every plant geneticist knows that there is inherit value in increased diversity. Increasing the diversity of crops available in modern cropping systems will allow us to better handle the worlds changing climates, and give us greater access to a more diverse food, fiber, and biomass base. Beyond the diversification of our food systems these crops have the potential to increase farmers' incomes and play significant roles in the development of new industries. Making the case for the values of orphaned crops will only result in investments if the individuals and organizations with the capital to invest can be convinced. Currently hemp is receiving significant private investments because of the potential market returns that businesses are seeing, but many orphaned crops do not have so much interest from the public and businesses. To improve these other crops, it is likely that governmental and non-governmental organizations will need to increase their investments into these crops. For this time must also be a consideration, because if these groups wait until an increased diversity of our agricultural systems is an undeniable need, because of the changing situations around the world, it will be too late. Researchers around the world need to keep working with and pushing the

organizations involved in agricultural research to fund the studies needed to improve these orphaned crops, thus diversifying the plants that constitute our world's agricultural systems.