UNDERSTANDING THE GENETIC BASIS OF SHATTERING IN PEARL MILLET

By

NIKEE SHRESTHA

Bachelor of Science in Agriculture

Tribhuvan University

Nepal

2018

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
In partial fulfillment of the requirements for
The Degree of
MASTER OF SCIENCE
May 2022

UNDERSTANDING THE GENETIC BASIS OF

SHATTERING IN PEARL MILLET

Thesis Approved:

Dr. Andrew Doust

(Thesis Advisor)

Dr. Ming Yang

Dr. Charles Chen

ACKNOWLEDGEMENTS

Thank you, my parents, Nanda Kumar Shrestha, and Rupa Shrestha who have always believed in me. Whoever I am today, I owe it to you.

I would like to thank my advisor Dr. Andrew Doust, for believing in me, making me a part of your team, and encouraging me to be the aspiring scientist that I am today. I genuinely appreciate your support and suggestions during my master's project. You have been an awesome, kind mentor and a father figure.

To the members of my committee, Dr. Ming Yang and Dr. Charles Chen for helping me in this project and giving me advice during my research. Thank you, Dr. Chen, for being a mentor and always being approachable. Your valuable suggestions and support during my graduate career have made this journey easy. Thank you for being so kind.

To my Doust lab members, Dr. Hu Hao, and Dr. John Hodge. Thank you, Hao, for always willing to help and answer my question without any hesitation. I have learned a lot from you. Thank you, John, for your support during master's. It was very nice to work in a lab, where we can turn into each other in need of help without a second thought.

To all the people in the department of PBEE, I am thankful for each one of you. Being away from home, in a new place, everyone struggles to find the sense of belonging. Thank you for making me feel that I belong here. Thank you for giving me opportunity to work as a TA in 1404 lab, which have developed my confidence, where I have also been one of the learners. Thank you, Lane, for being so awesome. I look up to you for your hard work, professionalism, inclusiveness and for making people feel heard and that they matter. I have found friends here, who I know, will be lifelong friends. Thank you for the emotional support.

To Kumar Shrestha, my biggest cheerleader and constant supporter. Thank you for your unconditional love.

Name: NIKEE SHRESTHA

Date of Degree: MAY 2022

Title of Study: UNDERSTANDING THE GENETIC BASIS OF SHATTERING IN PEARL MILLET

Major Field: PLANT BIOLOGY

Abstract: Over the last years, genes involved in several traits targeted during domestication have been studied in crops. Because shattering is an important domestication and agronomic trait, it has been intensively studied in crops such as rice, sorghum, *Setaria*, maize, wheat, and barley. However, shattering-related genes found in major cereal crops such as rice and wheat have not been validated in other crops. Additionally, recent transcriptomic analysis of abscission zone tissues from three grasses; a de-domesticated *Oryza* accession, and accessions of the wild species *Setaria viridis* and phylogenetically distant from each other have further supported the hypothesis of independent selection on genes for shattering between different grass species. However, it leaves the question of whether closely related genera might share similar shattering mechanisms. Hence, I have chosen to identify genomic regions associated with reduced shattering in pearl millet and compare them to identified genomic regions for shattering in the closely related genus, *Setaria,* as well as in the distantly related cereal species, rice. The wild relative, *Cenchrus americanus* ssp. *violaceum* (*monodii*), easily shatters by breaking at the base of the primary branch where the pedicel of the spikelet joins the rachis. Domesticated pearl millet, *Cenchrus americanus* ssp. *Americanus*, does not break at this location, making it non-shattering. A histological and SEM analysis of the shattering zone shows a unique indentation of the epidermis that is present from early development of the primary rachis branches in both domesticated and wild accessions. I crossed accessions of domesticated pearl millet and wild pearl millet; and created an F_2 population of 387 plants. Phenotyping of the F_2 and $F_{2:3}$ populations through a simple hand grasping method suggested that it followed a 15:1 segregation ratio ($p=0.223$) suggesting that two loci might be responsible for the non-shattering trait in pearl millet. I measured the force to detach the primary branch with a force gauge 28 days after heading, and mapped genetic loci associated with this trait using a high-density linkage map to identify quantitative trait loci associated with shattering. QTL mapping revealed a major QTL on chromosome 3 and a minor QTL on chromosome 5 associated with the shattering phenotype. This was confirmed by QTL analysis of the qualitative hand-shattering phenotyping trait identified in the F_2 and $F_{2:3}$ populations. Comparative genomics of QTL positions amongst grasses showed no conservation, which suggests, along with the histological details, that the shattering mechanism in wild millet is unique in grasses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

LITERATURE REVIEW

I. INTRODUCTION

Plant domestication is the process by which wild plants have evolved into crops through human selection. This selection is maintained during cultivation and results in phenotypic changes through time. Domestication occurs over various ranges of time and at different intensities (Harlan 1975), and it can be difficult to draw the lines between the initial domestication process and later improvement as crop cultivation spreads from the area of initial domestication (Doust, Mauro-Herrera, Hodge, & Stromski, 2017). It's astonishing to realize that every crop that is being cultivated today was once a wild plant and the domestication and improvement process is occurring continuously in the present for every crop.

Selection upon non-shattering forms of wild grasses was one of the earliest steps in domestication of cereal grains, as large-seeded varieties with low shattering and easy threshability are preferred by farmers, due to their high yield and easy handling. In wild cereal grasses, the dispersal units can disperse from the plant because they form a functional abscission layer, while in domesticated crops, dispersal is reduced because of either suppression in formation or loss of functionality of the abscission layer. It is likely that selection on shattering genes may have been unconscious, because grains that did not fall easily had a better chance of getting harvested (Li et al. 2006).

For most of the world's population, human and animal, seeds of crops serve as the major constituent of their diet. For efficient harvest and greater yield, it is important that crop varieties have low shattering characteristics. In crops, shattering can lead to 50-100% loss in yield (Clarke 1981), leading to shattering being one of the important agronomic traits economically in crop production. Shattering can also be a hindrance in crop breeding when using wild accessions, as the shattering trait can be linked to other desirable traits (Ji et al. 2006).

1. Diversity in shattering/abscission morphology in the grass family

Grasses shed their dispersal units through a process known as shattering. This process occurs because an abscission zone is formed between the parent plant and the dispersal unit. The abscission zone is usually a specialized cell layer, marking a point at which cells of the falling organ separate from its parent plant (Doust et al. 2014; Yu et al. 2020). Typical grass inflorescences have a main stalk called a rachis, which bears multiple primary branches. The primary branch can have further orders of branching, with the ultimate unit being the spikelet, which contains two glumes that subtend one to many florets (Figure.1).

Across the grass family, the position of the abscission zone (AZ) varies between species (Doust et al. 2014; Hodge and Kellogg 2017; Yu et al. 2020, Figure. 2). For example, the position of the abscission zone for rice is above the glumes, that of *Setaria* below the glumes, and barley in the rachis axis (Hodge and Kellogg 2016; Yu et al. 2020).

2. Genetic control of abscission zone formation in the grass family

In general, each cereal system that has been investigated has yielded genetic loci controlling shattering that were unique to that system, yet there has been debate over whether there is one or many genetic mechanisms to effect shattering. Paterson et al. (1995) and Lin et al. (2012) suggested that domestication in the shattering genes of cereals occurred due to mutations in orthologous loci, while Tsujimura et al. (2019) cautioned that there were likely multiple pathways responsible for seed shattering among cereal crops. The genetic locus *Shattering1* (*Sh1*) was found to contribute to domestication in seed dispersal in sorghum (Lin et al. 2012), and, for this locus, there was found to be parallel selection in rice, maize, and foxtail millet (Lin et al., 2012). In rice, the *OsSh1* locus is the ortholog of *Sh1*, in which a >4 kb insertion was found to reduce transcription levels and increase the shattering resistance phenotype. In maize, the ortholog of *Sh1* is on chromosome 1, where there is an insertion of an extremely large intron. The maize genome contains another copy of the *Sh1* gene on chromosome 5 (Lin et al. 2012). Doust et al. (2014) reported two QTL loci in *Setaria*; one on chromosome V and other on chromosome IX controlling seed shattering. The QTL locus found on chromosome IX contributed 35% of the variation in seed dispersal trait, and this region was found to be syntenic to the region in sorghum chromosome 1 which contains the *Sh1* locus (Lin et al. 2012, Odonkor et al. 2018). Furthermore, the *Sh1* ortholog found on chromosome IX has been found to be the main target for selection to reduce seed shattering, and a PIF/Harbinger MITE was found inserted in the 3' end of exon 2 for all analyzed cultivated accessions in *Setaria italica* and absent in wild accessions of *Setaria viridis* (Odonkor et al. 2018).

In rice, three loci, *Sh4*, *qSh1* and *qSh3*, have been identified as the major loci for complete loss of seed shattering, by preventing the formation of an abscission layer (Tsujimura et al. 2019). *Sh4* was mapped in an F2 population derived from a cross between an Indica-type cultivar and its wild progenitor. The nucleotide substitution guanine/thymine (G/T) in the first exon of *Sh4* results in a mutated allele in all cultivars compared with alleles in wild rice, suggesting strong selection on this locus during the early domestication of rice (Ishikawa et al. 2017). Further study of near-isogenic lines with a wild genetic background carrying the *Sh4* alleles, however, showed strong seed shattering, suggesting that other genes also regulate formation of the abscission layer. For *qSh1*, a loss of function mutation was observed in japonica-type cultivars, suggesting selection in japonica-type cultivars after early stages of rice domestication while *qSh3* was shown to have undergone artificial selection in both Indica as well as Japonica cultivars (Ishikawa et al. 2017). However, these genes have not been found as targets for selection for shattering in other grasses.

In the domestication of wild wheat, seed dispersal traits based on rachis brittleness are contributed by dominant alleles of the *Br* locus. Mutations at the *Tg* and *q* loci favored the easy threshability of grains and prevent production of hulled grains in wheat. Spike disarticulation in wheat has two types; barrel type (B-Type) disarticulation which results from breakage at the lower side of the junction of the rachis and spikelet base, and wedge type (W-type) disarticulation in which rachis fractures at the upper side of the junction of the rachis and spikelet base, leaving rachis fragments attached below each spikelet. B-type disarticulation was mapped to the *Br1* gene on the short arm chromosome 3 while W-type disarticulation was mapped to the *Br2* gene on the long arm

of chromosome 3, suggestion that within wheat species, there are multiple pathways to domestication of the shattering trait. Comparative analysis indicates that orthologs of these loci are not detected in any of the other major cereals i.e., rice, sorghum, maize, and barley, contradicting the hypothesis of parallel domestication of shattering traits in cereal crops (Li and Gill 2006). More recent studies focusing on the genes involved in abscission zone development in rice, *Brachypodium* and *Setaria* found differences in their anatomy, cell wall structure, and gene expression (Yu et al. 2020), and found that no module of gene expression was completely conserved among the species, and that the developmental and positional patterns of gene expression were almost entirely different. As a further test of the orthologous loci hypothesis, I chose to examine pearl millet (*Cenchrus americanus* (L.) Morrone), a cereal that has a distinct abscission zone position, yet is closely related to the C4 model crop, foxtail millet (*Setaria italica)*. Pearl millet is close enough to foxtail millet (Doust et al. 2007) that I might expect there to be some conservation of genetic control of the abscission process.

3. Crop of study

Pearl millet is a C4, out-crossing grass, belonging to the subfamily Panicoideae of the family Poaceae. It is the sixth most important cereal crop, widely cultivated in more than 30 million ha in a total of more than 30 countries and five continents (Yadav and Rai 2013). It is especially utilized in the dry lands of the Sahel and in Somalia in Africa and in western parts of India as a subsistence crop (Baltensperger 2002). It has a short growing period, rapid growth, and high-water use efficiency, and its hardiness and drought tolerance have been important in its popularity in arid and semi-arid regions (Jukanti et al. 2016). In addition to its short growing period of 65-75 days to flower and

produce seed, its deep root system enables pearl millet to grow as well as produce in drought prone areas, with rainfall less than 600 mm, where crops like maize and sorghum fail to produce (Panaud 2006). Being the poorest of the poor farmer's crop, pearl millet can be and is being cultivated with low maintenance, fertility, and inputs.

From a nutritional point of view, pearl millet has a well-balanced composition of essential amino acids and proteins that are suitable for human consumption (Goswami et al. 1969). It is also high in protein, and a good source of fat, fiber, energy, and carbohydrate (Sawaya, Khalil, and Safi, 1984). Pearl millet is equivalent to rice and wheat as a good and cheap source of bioavailable trace metals like zinc and iron (Agte et al. 1999). Unlike sorghum (Gualtieri and Rapaccini 1990), pearl millet does not contain any hazardous components like tannins or prussic acids, and it can be used as a forage crop for hay and green chop by livestock farmers (Newman et al. 2010). In addition, pearl millet seed has been found to enable swine to gain slaughter weight faster in comparison to other feed (Gulia et al. 2007) and has also used as goat feed (Terrill et al. 1998) and poultry feed (Myers 2002).

Pearl millet has many characteristics that make its cultivation favorable under changing climates. Climate change scenarios predict an increase in global warming from (1.1 to 6.4)˚C during the 21st century, and anticipate a decrease in water availability in already drought prone areas, heavy precipitation in flood prone areas, increased growing seasons, and an increase in sea level, along with other drastic changes (Council 2010).It is estimated that by 2050, the world's population will increase by 34%, reaching 9.1 billion, and, to meet the food demand of this huge population, cereal production needs to increase from 2.1 billion tons today up to 3 billion tons (Alexandratos 2009). Pearl millet

is a crop that can substitute for other cereal crops like maize and sorghum, tolerating a wide range of salinity, very low input of nutrients as well as low levels of management (Zegada-Lizarazu and Iijima 2005). Even though pearl millet has a high tolerance to drought and low soil fertility, it also responds well to water and favorable soil conditions (Andrews and Kumar 1992), which is a good indication that there is scope for increased production in the USA (Terrill et al. 1998).

Pearl millet is a good candidate for coping with increasing global climate change, human population, and food security problems. However, very little research has been focused on its genetic improvement, and average production is only up to about 900 kg/ha. This is true even in India, which has the highest millet productivity and production, but where improvements in yield due to genetic gain over the last twenty-five years have only averaged around 24 kg/ha (Yadav and Rai 2013).

In this study, I have investigated the genetic loci controlling shattering, establishing how many regions control the traits, where those regions are located, and whether they correspond in position to chromosomal regions controlling shattering in other domesticated grasses. To do this I have created a wide cross between domesticated pearl millet and its wild progenitor (*Cenchrus americanus* ssp. *violaceum*), that I have genotyped using SNP markers. The F2 and F3 populations were then grown in greenhouse and field environments respectively to measure shattering phenotypes, before quantitative trait locus (QTL) analysis was performed, and comparative genomic analyses accomplished. Chapter 2 describes the creation of the mapping population and production of the map, and chapter 3 describes the phenotyping trials, QTL analysis, and comparative genomic analyses.

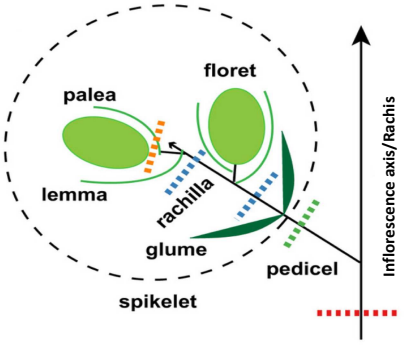

Figure 1 Basic structure of grass inflorescence (Yu et al. 2020).

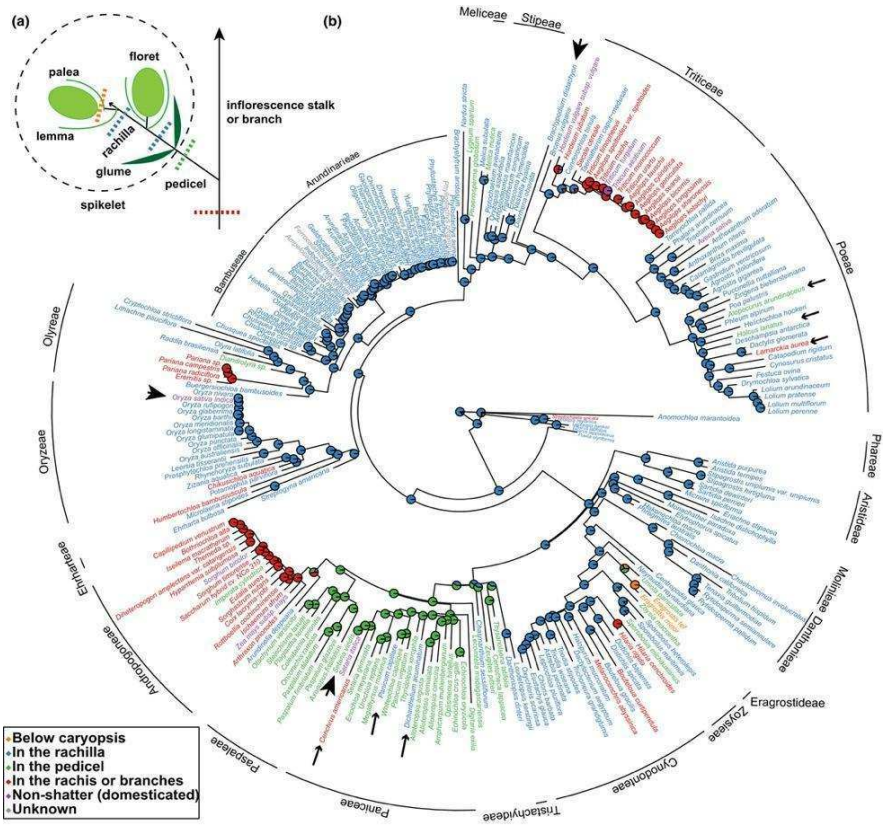

Figure 2 Phylogenetic tree of grass family where each color represents different position of abscission zone (Yu et al. 2020)

CHAPTER II

CONSTRUCTION OF HIGH-DENSITY LINKAGE MAP USING A GENOTYPING BY SEQUENCING APPROACH

I. INTRODUCTION

In this chapter I describe the creation of a mapping population from a wide cross between domesticated pearl millet and its wild ancestor and its subsequent use to create a genetic map of pearl millet. Previous mapping study for functional abscission zone suffered from the drawbacks of both small mapping population and low resolution of genotyping; with genetic map created with around 30 markers (Poncet et al. 2002). The greater map resolution is needed to examine whether the genes involved in domestication of pearl millet are the same as in the closely related crop foxtail millet, and in other grass domestications. A high-density genetic map is essential for quantitative trait loci mapping and eventual genetic improvement as it provides the opportunity for genetic dissection of quantitative trait loci, homology and synteny comparisons, provide direction for the completion of physical maps, and allows positional mapping of important genes (Moumouni et al. 2015; Yano et al. 1997).

For pearl millet, the first linkage map was constructed by Liu et al. (1994) using 181 RFLP markers which covered total map distance of 303 centiMorgans (cM) with average map distance of 2 cM and a few of the map distances between two markers more

than 20 cM. Another map, used for comparative genomics with *Setaria* and rice, revealed the complex relationship between the grasses with possible duplication events in the pearl millet (Devos et al. 2000). An integrated consensus genetic map using four different crosses was produced where 353 RFLP and 65 SSR markers covered a total distance of 473 cM with a few of the map distances between two markers being more than 30 cM (Qi et al. 2004). Both maps were created using F2 generation lines. Following that, a consensus map using a total of 176 expressed sequence tags and SSR markers-based was created using an F7 recombinant inbred population which spanned 899 cM map distance (Rajaram et al. 2013). Similarly, a genetic map based on 321 DArTs and SSR markers with a total distance of 1148 cM was produced using F7 recombinant inbred lines. In addition, a high-density linkage map using an F2 mapping population of 93 progenies based on 2809 SNP markers produced through genotyping-by-sequencing method was created, that spanned total distance of 640 cM (Moumouni et al. 2015). Previous maps created either had low density markers throughout the linkage groups or had more than 20-30 cM distance between two markers. In addition to that, genetic maps with high density linkage maps spanned less distance than the previously created ones. Thus, the production of high-density linkage maps with uniform coverage remains a challenge.

Genotyping-by-sequencing (GBS) has proven its efficiency to produce thousands of reproducible SNP markers without the need of prior characterization of candidate loci unlike RFLPs, SSR and ESTs (Moumouni et al. 2015). The GBS approach is cheap and capable of being multiplexed and has been successfully used to produce libraries of in many species. GBS takes advantage of the sequencing of the ends of sequences cut by restriction enzymes to reduce genomic complexity and enhance the discovery of shared

SNPs across individuals (Baird et al. 2008). It incorporates the principle of genome simplification by only sequencing restriction site associated (RAD) markers, usually by high-throughput Illumina sequencing of barcoded fragments. With an efficient barcoding system for each fragment, it is possible to use a multiplexing approach and reduce the cost of sequencing. In addition, the single well digestion of genomic DNA and adapter ligation is simple and efficient (Elshire et al. 2011). SNPs are the most common type of genetic variation found, generally due to point mutations. These markers are found to be most reliable because of their stability relative to other markers as well as their abundance and uniformity throughout the genome (Reshma and Das 2021). The GBS approach has been used to find tens of thousands to hundreds of thousands of SNP markers across the genome in many species, including finger millet (Qi et al. 2018), rice (Yadav et al. 2019), sorghum (Kong et al. 2018), and maize (Tomkowiak et al. 2019). In the present study, I aimed to use the GBS approach to produce SNP markers in the F2 mapping population and eventually create a high-density linkage map which can be further used for quantitative trait loci mapping of the shattering trait.

II. METHODS

1. Creation of mapping population:

The mapping population consists of 400 F_2 progeny derived from the cross between domesticated *C. americanus* ssp. *glaucum*, accession Tift 23DB, and wild *C. americanus* ssp. *monodii*, accession Tift 5120. Tift 23DB has recently had its genome sequenced (Varshney et al. 2017) and has been selfed for many years to maintain its homozygosity,

while Tift 5210 is a highly heterozygous wild accession. Both accessions were received from Wayne Hanna and Katrien Devos, University of Georgia, Athens.

For creation of the mapping population, Tift 23DB was used as a female parent, and Tift 5120 as a male parent. Tift 23DB was bagged after one third of the inflorescence had emerged out of the sheath, and before the stigmas had exerted, to prevent out-crossing. After some days, when the styles were fully exerted and ready to receive pollen, pollen from a bagged head of Tift 5120 was transferred to the receptive bagged head of Tift 23DB and thoroughly shaken to dust pollen over the receptive stigmas. The pollinated inflorescence was then bagged until maturation and seed harvesting (Figure. 3).

Forty of the resulting putative F_1 seed were grown in a greenhouse at OSU in Summer 2020, with average temperatures of 23 °C and photoperiod 12-14 hours, to screen for F_1 plants that showed the requisite shattering traits. At least one inflorescence on each F_1 plant was bagged before the exertion of the stigmas to promote selfing. The F1 plant that was chosen to create the F2 mapping population was freely shattering at the primary abscission zone and hard to thresh, characteristics of the wild parent and therefore markers for a hybrid plant. Confirmation of a F1 plant was done by PCR amplification, using differences in the length of the *Teosinte branched1* (*Tb1*) gene that are caused by the insertion of a miniature inverted-repeat transposable element locus downstream of the *Tb1* gene (Dussert et al. 2013). There is an insertion of MITE of approximately 300 bp at the 3' UTR region of *Tb1* gene in Tift 23DB; domesticated pearl millet while Tift 5120; wild type lacks this insertion.

I planted a total of 400 plants from seeds harvested from our chosen F_1 plant. Plants were planted on March 17, 2020, inside the greenhouse of OSU located at the western road inside USDA premises. Out of 400 F2 plants, 13 failed to germinate. One inflorescence per F_2 plant was bagged allowing selfing for further grow-outs and the creation of recombinant inbred lines. Throughout the growing period, I recorded multiple growth parameters including shattering.

2. DNA Extraction:

Around 5 cm of leaf samples was collected from all 1-month-old F2 plants. Leaf samples were kept inside 1.5 ml tubes and frozen by dipping into liquid nitrogen. 3 metal beads of 4 mm were added to the frozen sample tubes, and the leaf tissue ground in a bead beater. DNA was extracted using the CTAB method adapted from Healey et al. (2014). Quantification was done using a nanodrop (ND-1000) spectrometer and Qubit DNA fluorescence meter. Samples were checked for DNA degradation by running on a 0.7% agarose gel. Further detection of quality was also done by digesting DNA samples with restriction enzymes and running the samples again on 0.7% agarose gel (Figure 4).

3. GBS library Preparation and Sequencing:

Genotyping using the GBS approach consists of two major steps: library preparation and sequencing. GBS library preparation was done by the Devos lab at the University of Georgia, using a protocol developed by Qi et al. (2018), adapted from Elshire et al. (2011). Briefly, each individual F2 sample was double digested with restriction enzymes PstI and MspI. A common Y-adapter was annealed to the MspI end and sample specific barcoded adapters annealed to PstI ends.

Fragments with a length smaller than 300 base pairs were removed using AMPure-like beads (Baird et al. 2008). The remaining fragments were then amplified using the polymerase chain reaction with an extension time of 15 sec to ensure that only fragments smaller than ~800 bp are amplified. After PCR amplification, the concentration of DNA in each library was assessed using the Qubit 2.0 Fluorometer and Qubit™ dsDNA HS assay kit (Thermo Fisher Scientific). Following individual library preparation, 5 ul of genomic dsDNA from each library was pooled into a single tube, before small DNA fragments such as primers were removed with AMPure-like beads. Pooled libraries were paired end sequenced on the Illumina next generation sequencing platform. Once sequence data was received, sequence pre-processing and filtering was conducted using UGBS-Flex pipeline (Qi et al. 2018). Brief discussion of pre-processing and filtering of the reads are discussed below. The steps involved in this process are illustrated in Figure 5 and 6.

4. Filtering Raw Sequence Data for preprocessing:

After retrieving the raw sequences, sequence quality was checked with FastQC version 0.11.9 (Wingett and Andrews 2018). A read was retained if 1) it had perfectly matched one of the barcodes and PSTI restriction site, 2) was not an adapter/adapter dimer, 3) did not contain any N's in its first few base sequences. Reads were then de-multiplexed according to their barcodes into separate files using the module "Process_Radtags" within the "Stacks" program (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013). Following the demultiplexing of the reads. "FASTX_trimmer" from the "FASTX Toolkit" package (Bolger, Lohse, & Usadel, 2014) was used to remove adaptor/barcode

sequences along with the restriction sites. The last five base pairs of each fragment read that were more likely to contain errors with lower quality than 20 were removed. Along with this, additional bases were removed from the 3' end of the forward read to make them all the same length.

5. SNP/indel calling and filtering:

To detect the informative SNPs from the raw data after filtering, filtered and pre-processed reads were aligned using "Unified Genotyper 3.4" (DeLuca et al. 2012) against the whole genome reference sequence of pearl millet (Varshney et al. 2017). Biallelic SNPs with allele frequency between 10 and 90% were retained for further processing. SNPs with minor allele frequency less than 10% were removed as those likely represented sequencing errors or rare alleles. The allelic depth information provided in the .vcf file was then used to score the allelic status of each locus. Loci with a total allelic depth (AD) less than 8 were scored as missing data points (-). Loci with a $AD_{\text{reference allele}}/AD_{\text{alternate allele}}$ (AD_R/AD_A) ratio ≥ 10 was scored as A (homozygous for domesticated parent allele) and AD_R/AD_A ratio ≤ 0.10 as B (homozygous for wild parent allele), $10> AD_R/AD_A >4$ as D (ambiguous A or heterozygous (H)) and $0.25> AD_R/AD_A >0.1$ as C (ambiguous B or heterozygous (H)). Loci with other ratios were scored as H. Scoring for genotyping was done using the script "SNP_genotyper.py ". Following the genotyping scoring, co-segregating markers were removed using script "SNP_cosegregating.py" (Qi et al. 2018). Samples with more than 20% missing data were removed as well as SNPs that were missing in more than 20% of the samples were removed as well. Following this

filtering, chi-square tests were done for each marker and significantly distorted markers ($P < 1 X 10^{-10}$) removed.

6. Construction of the genetic map:

After pre-processing and filtering the SNP markers, I used the final retained SNP markers to construct the genetic map using the Haldane function. Initially as described in Peng et al. (2018), I used Mstmap (Wu et al. 2008) for creating raw framework maps. Following the establishment of initial map order, finalization was done using Mapmaker (Lander et al. 1987). As MAPMAKER is limited to ordering ~100 markers at a time, each linkage group was split into small subgroups of 60-100 markers. The Haldane map function was employed to convert recombination frequency to centiMorgans (cM). Double recombination events were identified for each linkage group, and then removed in a way that at least 80% of the SNP markers were retained in each linkage group. Python scripts and a pipeline developed by Devos lab were used to ripple the marker order and finalize the map at the end (Qi et al. 2018).

III. RESULTS

1. Genotyping by Sequencing:

Over 989 million raw reads were generated from libraries consisting of the domesticated parent, the F1 hybrid and 384 F2 lines using an Illumina NextSeq High Output Flow Cell. The average length and GC content of raw reads was 151 bp and 46% respectively. Using the uGBS pipeline we selected 10,326 biallelic SNP markers from the libaries. Co-

segregated markers were removed, leaving 8267 SNP markers, which were then filtered based on conformity to the expected 1:2:1 (AA:AB:BB) segregation ratio. Segregation distortion was identified in several chromosomes, with the most severe in chromosome 3, with only 45 SNP markers passing the chi-square filtering at p-value of e^{-10}. For this chromosome, the threshold p-value of e^{-10} had to be changed to e^{-30} to have enough markers to map, resulting in the retainment of many severely distorted markers for the genetic map. Another constraint for our genotypes was lack of recombination across the individuals for most of the linkage groups, again extremely severe for chromosome 3 (figure7).

After removing as many markers with severe segregation distortion as possible, I retained a total of 5585 SNP markers evenly distributed among seven chromosomes. I then removed SNP markers with more than 20% missing data, finally retaining a total of 3925 SNP markers uniformly distributed throughout the seven linkage groups. LG1 had the greatest number of SNP markers while LG4 had the fewest. However, a large proportion of these markers were redundant, resulting in a total of 1802 unique loci (Table 1).

2. Genetic Mapping

The first draft of the genetic map identifying the order of markers in each chromosome was done by MSTMAP (Supplementary data). Following the order of markers provided by MSTMAP, I was able to produce final genetic linkage map using MAPMAKER, which exhibited a total map distance of 1391.5 cM with average interval distance of 0.35 cM between the markers (Figure 8, Table 1).

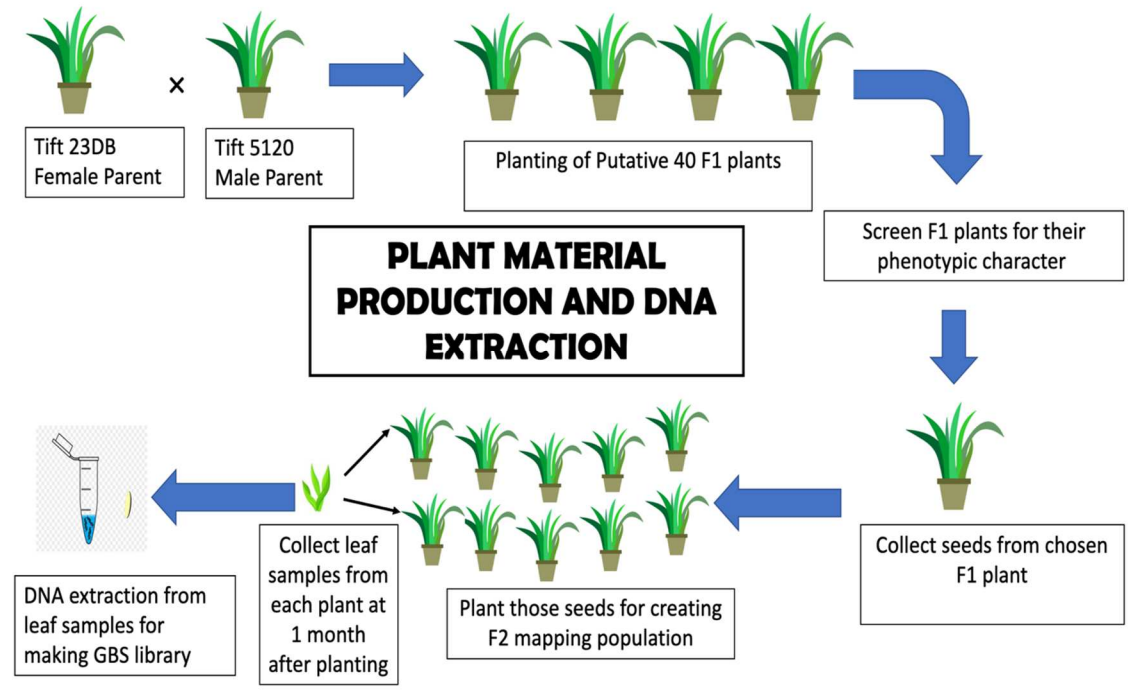

Figure 3 Flow diagram showing process of plant material production and DNA extraction

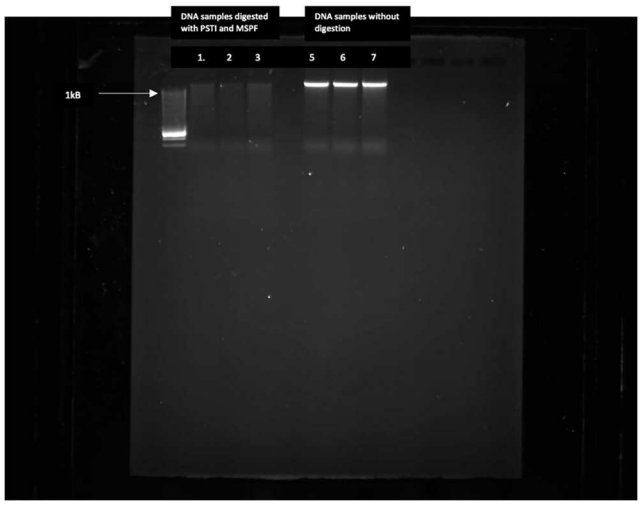

Figure 4 DNA samples extracted using CTAB method. (1), (2) and (3) lane: DNA samples double digested with restriction enzymes. (5), (6) and (7) lane: DNA samples without digestion.

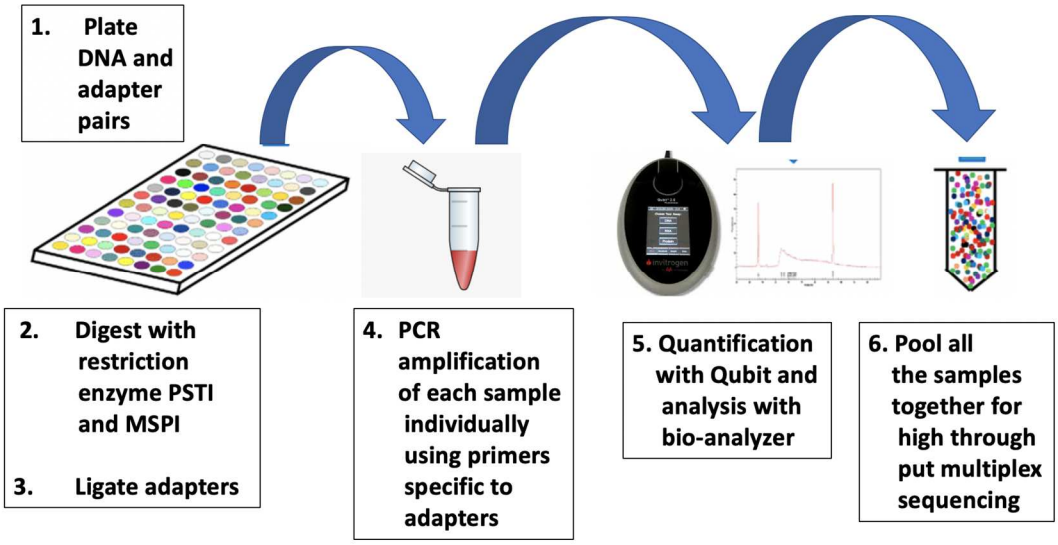

Figure 5 Basic principle of GBS library preparation (Elshire et al. 2011).

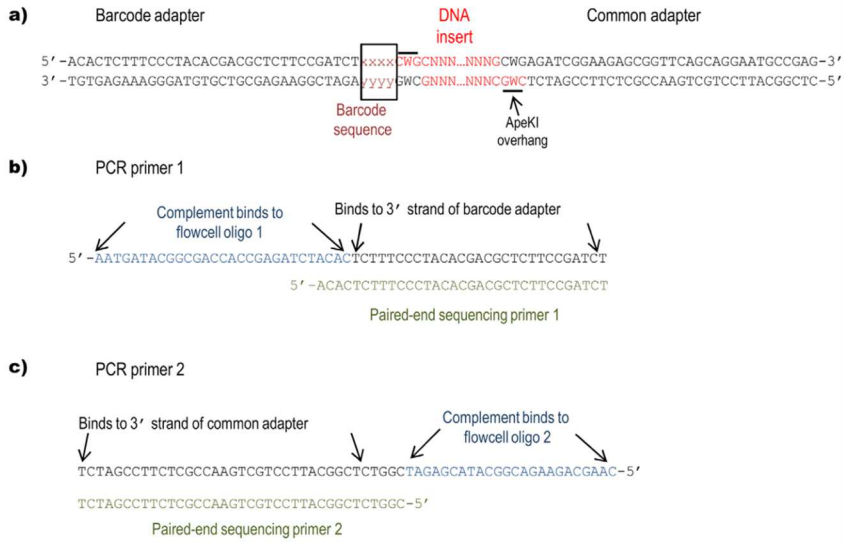

Figure 6 Steps in producing the DNA libraries. a) NNNN represents the restriction fragment, xxxx and yyyy represents barcode sequence, instead of APeKI overhang, we will have PSTI overhang in the barcode adaptor site while MsPI overhang in the common adaptor site.

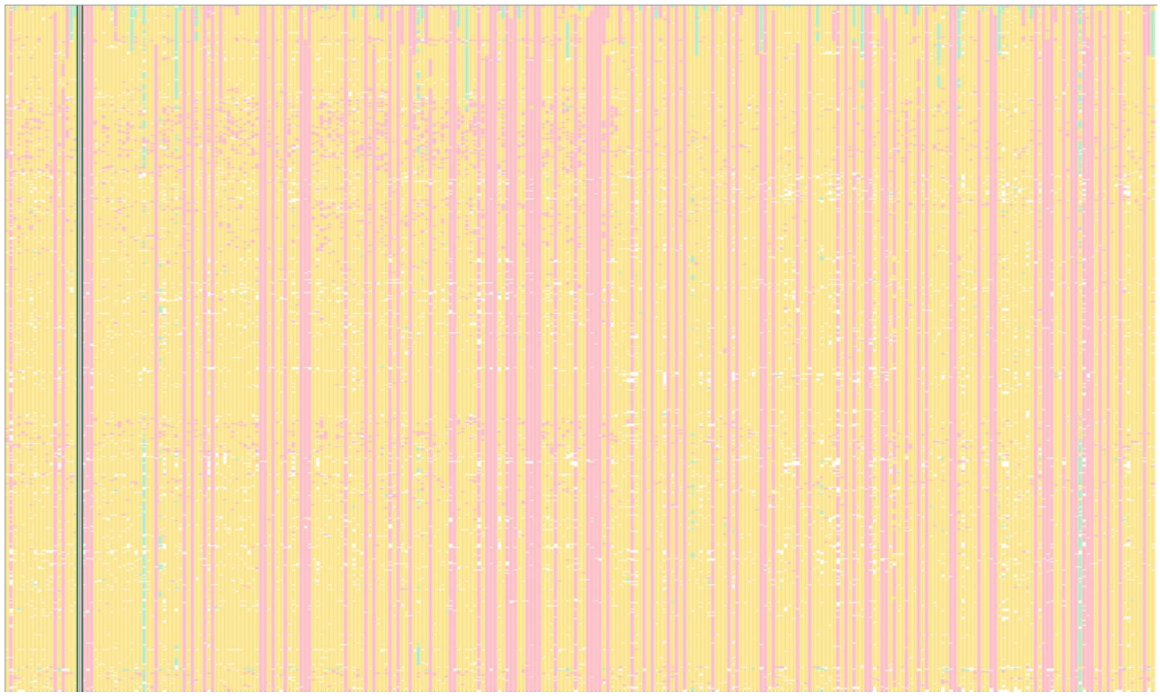

Figure 7 Snapshot of linkage group 3 where each column represents F2 individuals, and each row represents individual SNP markers. Red; allele A (homozygous for domesticated parent), yellow; allele H (heterozygous) and green; allele B (homozygous for wild allele).

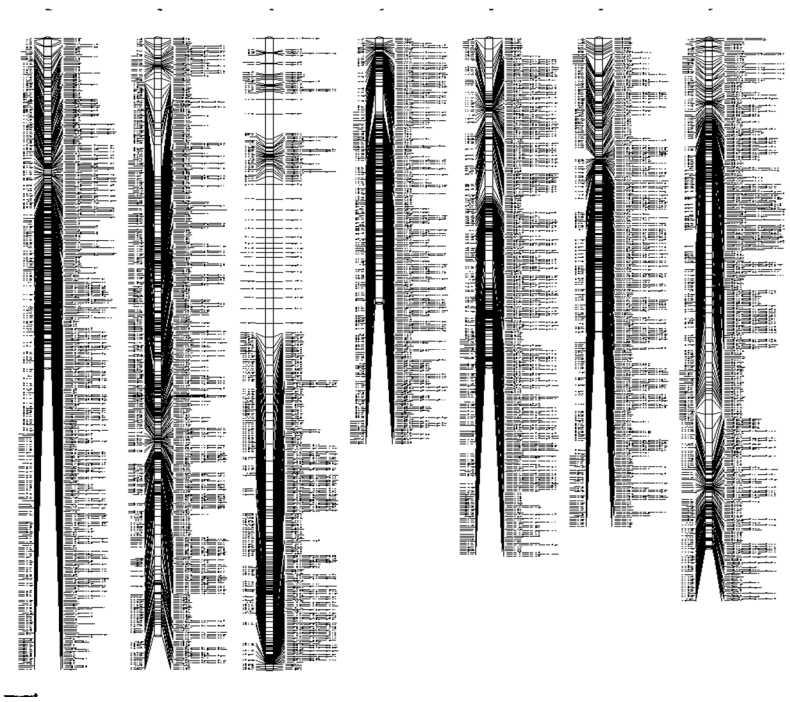

Figure 8 High-density genetic linkage map created using 3925 SNP markers

ChroGroupID	TotalSNPMarkers	TotalDistance(cM)	AverageDistance(cM)	NumberOfUniqueLocis
Chr01	679	155.5	0.229	282
Chr02	666	281.2	0.422	511
Chr03	437	297.3	0.68	190
Chr04	434	124.5	0.286	174
Chr05	612	155.2	0.25	201
Chr06	529	137.7	0.26	196
Chr07	568	240.1	0.421	248
Total	3925	1391.5	/	1802

Table 1 Basic information about linkage groups and SNP markers

IV. DISCUSSION

High resolution maps are required for gene discovery for traits of interest. Various attempts have been made to create a linkage map for pearl millet. However, early genetic maps had a low density of markers with gaps of more than 20-30 cM between two markers, while other genetic maps with high densities of markers were shorter than previous maps (Moumouni et al. 2015). In addition, previously created linkage maps were based on a small mapping population. Hence, I decided to use the genotyping-by-sequencing approach for producing high numbers of reproducible and stable SNP markers in a large mapping population of 285 F2 plants. SNP markers generated from the current population will be useful in genetic assessment of diversity panels, genome wide association studies, genome mapping, QTL mapping, and marker assisted selection, as well as breeding experiments.

Segregation distortion occurred across the linkage groups, with severe distortion in LG3, significantly biased in favor of the Tift 23DB allele. Segregation distortion and a relative lack of recombination is a common phenomenon in pearl millet crosses and has been reported in almost all earlier mapping studies (Punnuri et al. 2016; Qi et al. 2004; Rajaram et al. 2013; Supriya et al. 2011). LG3 exhibited stronger distortion of marker segregation with alleles significantly biased towards the female parent (Punnuri et al. 2016). One reason for the segregation distortion might be the considerable genetic differentiation of the parents used in the hybridization, one of which was the wild accession, and another was domesticated, improved accession. It has been suggested that two parents whose genomes have significantly diverged might exhibit segregation distortion in the hybrid progenies, due to rapid elimination of an allele or the whole

genome of one parent (Parvathaneni et al. 2013; Yang et al. 2014). While severe segregation distortion is common in pearl millet, it has also been reported in other grasses such as wheat and rice (Yang et al. 2014).

Our map has more non-redundant markers than the most recent high density linkage map (Moumouni et al. 2015) and spans a total of 1391.5 cM (Supriya et al. 2011). The map coverage achieved in the present study is quite uniform with an average distance of 0.36 (±0.2) cM between two SNP markers. This interval is smaller than the value of 2.1 (±0.6) cM in a previous high density linkage map created in pearl millet through the genotyping by sequencing approach (Moumouni et al. 2015). The current map also had no mapping distance of more than 10 cM, unlike those previously reported (Supriya et al. 2011; Moumouni et al. 2015). A high-density linkage map with good coverage can be used for further quantitative trait loci mapping in our study.

CHAPTER III

QUANTITATIVE TRAIT LOCI MAPPING COMBINED WITH COMPARATIVE ANALYSIS TO IDENTIFY CANDIDATE GENE FOR REDUCED SHATTERING IN PEARL MILLET

I. INTRODUCTION

Over the last few years, genes involved in several traits targeted during domestication have been studied in crops. Because shattering is an important domestication and agronomic trait, it has been intensively studied in crops such as rice, sorghum, *Setaria*, maize, wheat, and barley (Lin et al. 2012). Researchers have hypothesized that parallel selection of genes occurs during the domestication process for reduced shattering (CLARKE, 1981). However, genes found in individual major cereal crops such as rice and wheat have not been identified in other crop systems (Tsujimura et al. 2019). Additionally, recent transcriptomic analysis of abscission zone tissues from three grasses; a de-domesticated *Oryza sativa* accession, and accessions of the wild species *Setaria viridis* and *Brachypodium distachyon* have further supported the hypothesis of independent selection on genes for shattering between different grass species (Yu et al., 2020). However, Yu et al. (2020) tested three grasses that were phylogenetically distant from each other (in three major tribes of the Poaceae), leaving

open the question of whether closely related genera might share similar shattering mechanisms. To test this hypothesis, the objective of this chapter is to identify the genomic regions associated with reduced shattering in pearl millet and compare them to identified genomic regions for shattering in the closely related genus, *Setaria,* as well as another important cereal species, rice. We will also examine any genomic regions found to identify any other candidate genes.

The approach taken in this study is to search for quantitative trait loci, which essentially are regions of the genome that are highly correlated with differences in the phenotype of interest. The key to quantitative trait loci mapping is creating a linkage/genetic map using recombination and segregation analysis within a population. QTL mapping links the linkage map and the phenotype data of trait of interest through regression analysis to help identify the most likely genomic region that explains the variation in the trait of interest of a population. In agriculture, QTL mapping has been used intensively in various crops to map the QTL regions associated with agronomically important traits such as disease resistance (Mutschler et al. 1996), stress related genes (Sanchez et al. 2002), marker assisted selection (Chandra and Pandey 2017), fine mapping (Fan et al., 2021) and positional mapping of specific genes (Jagodic and Stridh 2014).

In pearl millet, Quantitative Trait Loci (QTL) loci for the presence of an abscission layer and other domestication traits such as a functional abscission layer had been localized to linkage groups 6 (Poncet et al. 2000, 2002). However, the genetic dissection of domestication traits in those studies was hampered by both the low resolution of the genetic maps and the small size of the segregating mapping populations. Hence, I am using the high-density genetic marker map from chapter II to achieve the

objective of this chapter. The trait I are interested in pearl millet is the primary abscission zone, at the base of the primary branches, that gives rise to the shattering phenotype of the wild ancestor. In the domesticated parent, the shattering phenotype is highly reduced because the primary branches do not detach from the inflorescence axis, and the spikelets are more usually detached from above the glumes during the threshing process. Hence, before crossing two parents to create the mapping population, genetic map and phenotyping for quantitative trait loci mapping, our lab was interested in in-depth study of position and structure of the abscission zone of pearl millet and its uniqueness relative to identified abscission zones in other grass species such as *Setaria* and rice. To do this, John Hodge, a student in the Doust lab, performed histology and Hao Hu, a post-doc in the lab, performed scanning electron microscope (SEM) analysis to identify the position and mode of initiation of the abscission zone. I hypothesized that due to reduced/non shattering phenotype in the domesticated parent for the primary abscission zone that there won't be any formation of an abscission layer in domesticated accession at all.

After assessing the histology of the phenotype, I used forward genetics to identify the genetic loci underlying our trait of interest (Peters et al. 2003).

II. METHODS

1. Histology sectioning and SEM:

Primary branches along one inflorescence at different time points from each accession were harvested beginning at 30 days after planting. Samples were fixed in FAA (formalin-acetic acid-ethanol) and dehydrated in an ethanol series. Samples were embedded in paraffin to enable thin sections to be cut through serial sectioning of 10 μm

thickness using a Leica microtome. Sections were then stained with safranin and fast green using standard protocols (Hodge and Kellogg 2014). Images of the sections were taken using a Leica microscope.

2. Phenotyping of F2 Population:

After identifying the presence of the abscission zone at the point where the primary branch joins the rachis, further phenotyping analysis was done. The F2 population of 387 plants were phenotyped at 27 days after planting for shattering with two methods of phenotyping. These two methods were used for retaining quantitative and qualitative data. For quantitative data, I measured breaking tensile strength of primary branch-rachis junction using force gauge instrument. An inflorescence was cut off the plant at 27 days after heading and hung upside down from the force gauge. The primary branch and spikelet as a whole unit were pulled down one at a time using forceps. For a single inflorescence, I pulled 20 primary branches and recorded tensile strength in grams (gF) required to detach each of those primary branches. At the end, an average of 20 records was used as a breaking tensile strength measurement, an index for shattering for each F2 plant. Caryopsis maturation and senescence occur roughly around 27 days after inflorescence emergence in pearl millet, suggesting that this time frame would be adequate for characterizing the shattering.

For qualitative data, I used a simple visual way of scoring the phenotype through a hand grasping method. Again, a single inflorescence from each F2 plant at 28 days after panicle emergence was grasped by hand and if the primary branch along with spikelet as

a whole unit detached easily then it was scored as shattering, otherwise it was scored as non-shattering.

For the $F_{2:3}$ population, I planted a total of 387 families of 10 plants each with three replications in a randomized block design during the period of May to September 2021 in the field at the Cimarron Valley Research Station located at the North Perkins, Oklahoma. However, due to excessive rainfall causing water logging, I had very low germination, eventually having 150 plants from 357 F2:3 lines survive. At least 5 plants of each F2:3 families were then sampled for phenotyping where freely and easily shattering was characterized as 0, non-shattering was characterized as 2 and families with a segregating population with a mix of both along with intermediate shattering characters were characterized as 1.

3. QTL mapping:

Using the genetic map from chapter II, with phenotypic measurements of shattering in each line, I performed quantitative trait loci mapping using composite interval mapping (CIM) in Windows QTL Cartographer 2.5 (Silva et al. 2012) as well as standard interval mapping (IM) in R/QTL (Broman, Wu, Sen, & Churchill, 2003) with quantitative data from the F2 generation. I used model 6 (composite interval mapping) with step wise forward and backward regression, a 10 cM window size, and a mapping step size of 1 cM. For qualitative data from the F2 and F2:3 generation, I performed quantitative trait loci mapping using standard interval mapping using the "binary model" in R/QTL (Broman et al. 2003). The logarithm of odds (LOD) threshold that defined a significant QTL was calculated based on 1000 permutations and a significance level of p=0.05. To

find candidate genes underlying the QTL, QTL regions delineated by the most distal flanking markers, along with peak markers with LOD score above the threshold value were located on the Tift 23DB genome sequence assembly v1.1 (Varshney et al. 2017).

4. Comparative genomic and candidate gene analysis:

To compare QTL regions among panicoid grasses, syntenic dot plots were generated between *Cenchrus* and *Setaria* and between *Cenchrus* and rice using the SynMap module in CoGe (Lyons et al., 2008). *Cenchrus* genome version 1.1 was uploaded to CoGe along with its annotation while *Setaria* genome version 2.1 and *Oryza sativa* genome version 4.0 are already available in CoGE. I configured CoGe to assign gene pairs to classes based on their Ks values and thus to determine the orthologous syntenic regions among genomes. Syntenic regions in *Setaria* and rice aligning with our identified major QTL were then used to compare QTL. Further, syntenic regions in *Setaria* and rice were scanned for published and well-characterized shattering QTLs, and protein sequences of known and well characterized shattering genes were used in BLASTP searches against *Cenchrus* transcript database to identify the position of corresponding orthologs in the *Cenchrus* genome. The location of the top hit for each candidate gene on the *Cenchrus* genomic sequence was then compared to that of QTL regions.

III. RESULTS

1. Histology and SEM analysis:

After harvesting primary branches at different time points for histology analysis, I found that in both accessions an indentation was visible at the primary branch-rachis junction at

the same time (figure 9) contradicting our initial hypothesis regarding the absence of the abscission zone in domesticated accessions. Histology revealed that development of abscission zone initiates at the similar time, which is visible from 15 days before anthesis. In both accessions, an indentation was formed from the very beginning of inflorescence development; at the junction where the primary branch joins the main stem. However, the domesticated accession has a thick and elongated pedicel while the wild accession has a comparatively thin and short pedicel.

Further SEM analysis was done to characterize the function of abscission zone to closely characterize the difference between shattering phenotype between two parents. However, SEM analysis revealed that there is a clear difference on the surface of the abscission layer among the parents. In wild accessions, detachment of primary branch leaves a comparatively smooth surface of the abscission zone. In contrast, the surface of the abscission zone on the pedicel tissue of domesticated inflorescence appeared torn and ripped, sometimes the abscission cup itself completely lost integrity, leaving a torn surface leaving parenchymatous cells protruding (figure 9). Both histology and SEM analysis provided evidence about the position of abscission zone and revealed clearer distinction of function of abscission zone and degree of shattering phenotype between two parents.

2. Screening of F1 plants

Due to heterozygosity of '*monodii'* which was used as the male parent, I found variation among most of the F1 plants. Since the female parent was assumed homozygous, any F1 plant which showed differences in their traits from Tift 23DB, the female parent, was

considered as a successful cross. I recorded traits such as plant height (cm), number of tillers, presence/absence of shoot hairs and shattering related traits. For these traits, I found variation among F1 plants.

3. Plant Height(cm):

Among the 40 F1 plants, plant height (cm) at maturity was found to be an average of 124.9 cm (standard deviation = 40.28 cm) while the minimum height was found as 50 cm and maximum was 211 cm. (Figure 10).

4. Number of tillers at maturity:

I recorded the total tillers of all F1 plants at maturity. I found the average mean of total tillers among F1 plants as 11 with standard deviation of 4.76, where the minimum was 5 as shown in (Figure 11) below.

5. Shattering:

Degree of shattering and threshability was recorded for the F1 plants. Among the total 40 F1 plants, 24 plants were not found to be non-shattering. These were the plants which showed the similar traits as the domesticated accession Tift23DB; and hence, concluded to be selfed. The remaining 16 plants showed shattering and were non-threshable, with seeds coming off with bristles and glumes attached (Figure 12).

6. Shoot hairs and Nodal roots:

In addition, I recorded traits such as degree of presence of shoot hairs in the F1 plants, for which I found that 30 out of 40 plants were found to have shoot hairs all over the leaves, whereas there were plants which had only few shoot hairs present on the nodal region

(Figure. 13). I also recorded the trait of nodal roots, of which, out of a total of 40 F1 plants, 23 plants did not contain nodal roots while 17 plants did have nodal roots (Figure 14). Most of the plants which appeared to be domesticated (unsuccessful cross, i.e., selfed) did not have such nodal roots.

Among the F1 plants, I chose one F1 plant that showed traits different from the domesticated parent in terms of the traits I recorded. Particularly, the plant I chose for F2 population was 171 cm tall, had 10 tillers, an abundance of axillary branches, presence of nodal roots, and specifically had the shattering trait at the primary abscission zone.

7. Phenotypic variation of F2 mapping population:

Among 400 plants planted from seeds of the chosen F1 plants, 13 plants failed to germinate leaving a total of 387 F2 plants for mapping.

I measured the plant height and number tillers of the F2 population at 1 month after planting. Plants showed segregation for plant height as well as for number of tillers at the 1-month-old stage. Plant height at 1 month was found to be normally distributed with an average mean of 9.2 cm and standard deviation of 2.19 cm (Figure 15).

8. Shattering phenotype:

I measured the shattering phenotype through two different methods.

8.1. By Hand visual method:

Through a hand visual method, for each F2 plant, I recorded either “shattering” or “non-shattering”. Out of 285 F2 plants, I found out that 13 F2 plants were found to be non-shattering while 272 F2 plants were found to be shattering (Figure 16). A chi-square test

revealed that F2 plants follow the 15:1 Mendel's segregation ratio (p=0.223) implying that two genomic loci might be responsible for shattering phenotype in the pearl millet.

8.2. By Breaking tensile strength method:

Following the hand grasping method, on the same day, inflorescence of each F2 plant at 27 days after flowering were taken to measure the breaking tensile strength required to pull off the whole primary branch unit breaking at the point where it joins the rachis. Among the F2 plants, minimum breaking tensile strength was found to be 0.65 gF while the maximum was found to be 159.25 gF. On average, the F2 line had 26.65 gF with its median at 18.7, first quartile at 7.75 gF and third quartile at 35.35 gF. The distribution of BTS raw data is shown below in figure 17. F2 lines characterized as non-shattering had an average of 102.91 gF while F2 lines characterized as shattering had an average of 21.37 gF.

I also performed Pearson's correlation coefficient test between two phenotyping methods. Test revealed that there is a significant correlation between two phenotyping methods ($p<0.001$) with correlation coefficient value of 0.72 (Figure 18). Since the correlation was significantly positive and measurement of BTS data would be extremely time consuming for thousands of $F_{2:3}$, I decided to characterize $F_{2:3}$ plants through a simple hand grasping method which would be used as phenotype data to further confirm the phenotype characterization of the **F_2 mapping population.**

In F2:3 population, a total of 152 families were successful at germination of at least five lines for successful phenotyping. Among them, in 35% of the F2:3 families, all the lines displayed freely shattering, in 52% of the F2:3 families, lines were mix of both

shattering, and non-shattering and finally, in 11% of F2:3 families, all the lines germinated were found to be non-shattering.

9. QTL analysis:

Two seed shattering QTLs were identified that explained variation in the breaking tensile strength (BTS) data from F_2 plants (Figure 19, Table 2). Similarly, two QTLs were associated with the shattering characterization achieved through hand grasping methods of both F_2 and $F_{2:3}$ plants (Figure 20, Table 2). QTLs identified through both methods overlapped as in both cases, chromosome 3 and chromosome 5 was found to be associated with the variation in the data (figure 21).

In the F2 population, for breaking tensile strength data, QTLs found on the chromosome 3 and chromosome 5 explained 21.57% ed 13.68% variation in the phenotypic data respectively. The QTL region on chromosome 3 was located in between 207757021-299415938 bp with peak marker position at 260396484 bp. Similarly, the QTL region on chromosome 5 was delineated to the region 60776619-147490303 bp with peak marker position at 116928697 bp. For both QTLs, chromosome 3 and 5, the domesticated allele led to increase in mean breaking tensile strength of the primary branch with additive effect of 0.498 and 0.513 respectively.

Similarly, in the F2 population, for simple hand grasping method phenotype, QTL on chromosome 3 was delineated to the region 20620825-300052462 bp with peak marker position at 273684987 bp. Similarly, QTL at chromosome 5 was located at the region 81419087- 152287762 bp with peak marker position at the 43422339 bp. In congruence with QTLs associated with BTS data, in both QTLs identified on chromosome 3 and 5,

the domesticated allele led to non-shattering phenotype. Using a simple hand grasping method in the $F_{2:3}$ population, QTL on the same location on chromosomes 3 was identified. QTL at chromosome 3 was delineated at the region of 20620825-298708044 bp with peak marker locus at the 236314895 bp. In all three cases, chromosome 3 explained most of the phenotypic variance with significant LOD score, determining it as a major QTL.

10. Location of SH1 ortholog in the *Cenchrus* genome:

BLASTP search of *Setaria* SH1 ortholog Sevir.9G153200 revealed that the *Cenchrus* SH1 ortholog is located on chromosome 2 spanning with physical location of 27006164-27010214 bp with 95.906% identity in protein sequence (figure23, Table 3). This gene, a major locus controlling shattering in *Setaria*, does not lie either in the major QTL or in the minor QTL for shattering identified in pearl millet.

11. Comparison of QTL regions and location of known shattering candidate genes:

To look for other candidate genes and conservation of shattering loci across grass, I compared identified QTLs of pearl millet with *Setaria* and rice. I retained physical locations of the QTL regions identified from previous studies of *Setaria* and rice (Doust et al. 2014; Odonkor et al. 2018) and examined if identified QTLs were syntenic to the QTLs found in pearl millet. Syntenic dot plots show that there is less genome rearrangement between pearl millet and *Setaria* than between pearl millet and rice (Figure 22 and 23). Genomic regions were investigated in detail between *Cenchrus* QTLs

identified chromosomes with the corresponding *Setaria* and rice chromosomes (figure 22,24).

Two major syntenic blocks of a segment of chromosome 3 were found in *Setaria* chromosome 1 with conserved genomic order while another segment was identified in *Setaria* chromosome 7 with inverted genomic order arrangement. Neither the QTL region for shattering nor any candidate genes were identified in these syntenic blocks. In addition, a major syntenic block of segment in chromosome 5 was identified in *Setaria* chr. 9 with inverted genomic arrangement. In *Setaria* chr. 9, a major previous QTL region associated with shattering has been identified (Doust et al. 2014). However, this QTL region did not co-localize within the syntenic region implying non-conservation of shattering loci between pearl millet and its closely related grass; *Setaria*.

Similarly in rice, two major syntenic blocks for major QTL chromosome 3 were identified in chromosome 2 and 4 with the similar pattern of genomic region with a segment in chromosome 2 while inverted genomic pattern arrangement with a segment in chromosome 4. In chromosome 4 of rice, six QTL regions were previously identified from independent studies co-localizing with each other. Narrowing down and fine mapping of these QTL regions had resulted in the cloning and characterization of Sh4 gene (Li et al. 2006). Similarly, two major syntenic blocks for minor QTL on the chromosome 5 were identified in rice chromosome 3 and 6, where a syntenic relationship of similar genomic arrangement was identified in chromosome 6 but inverted in chromosome 3. QTLs for shattering have been identified in chromosome 3 of rice but not within the syntenic block between pearl millet and rice.

Along with comparative genomics, BLASTP searches for known candidate genes revealed that no known genes can be identified either in the major QTL on the chromosome 3 or the minor QTL on the chromosome 5 (figure23).

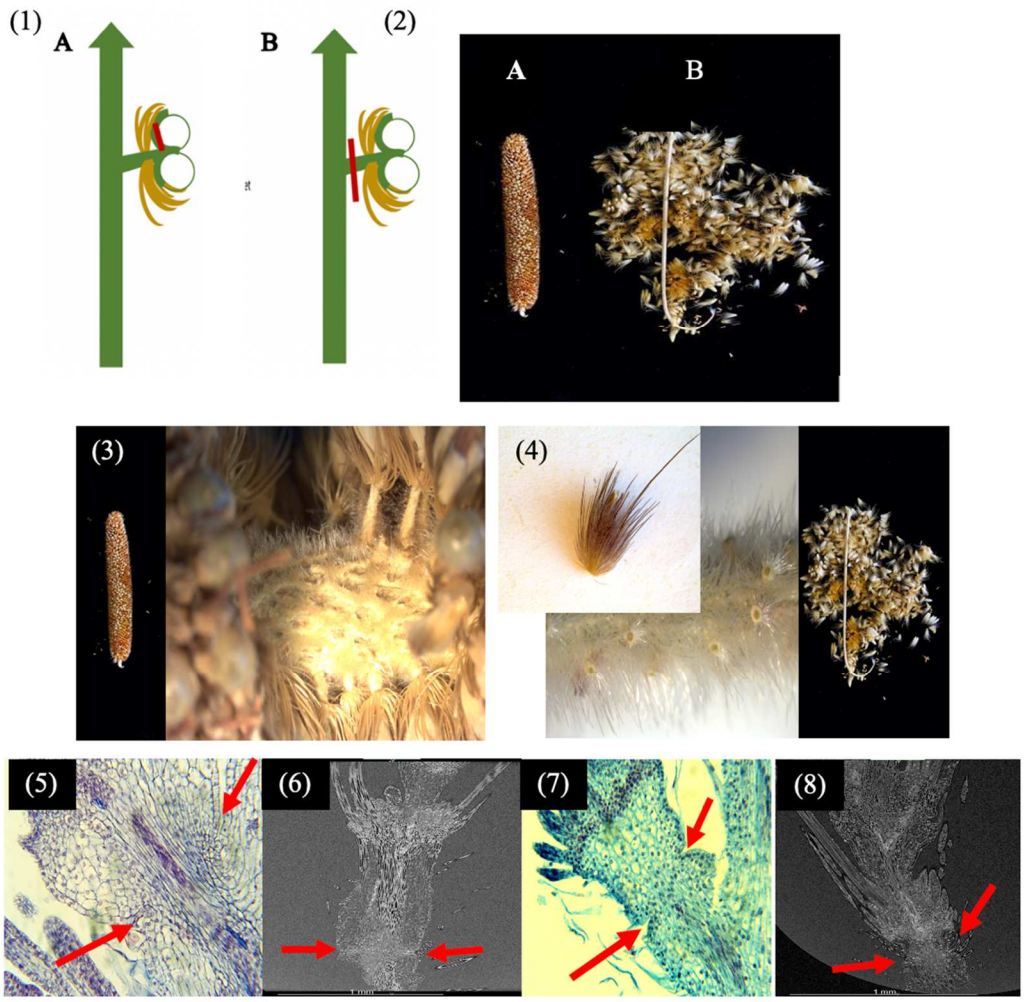

Figure 9 (1). Position of abscission zone (AZ, red lines) in (A) non-shattering and easily thresh-able domesticated parent Tift 23DB whose seeds come off above the glume and (B) shattering and hard to thresh wild parent Tift 5120 whose abscission zone is located at the base of primary branch. (2). Non-shattering phenotype in domesticated parent Tift 23DB (A) and freely and easy shattering phenotype in wild parent Tift 5120 (B). (3).

Primary branch pulled off from inflorescence leaving torn surface behind in domesticated parent Tift 23DB and (4) smooth surface seen at primary branch-rachis junction in wild parent Tift 5120. Histology Section with red arrows showing position of abscission zone in Tift 23DB (5 and 6) and in Tift 5120 (7 and 8). Red arrows showing position of abscission zone in Tift 23DB (5 and 6) and in Tift 5120 (7 and 8).

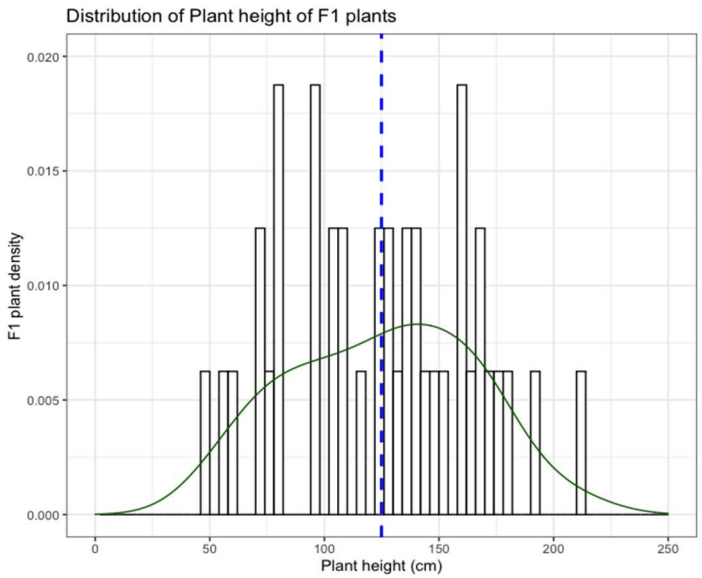

Figure 10 Plant height was found to be variable among the 40 putative F1 plants where blue dashed line represents the mean plant height among F1 plants.

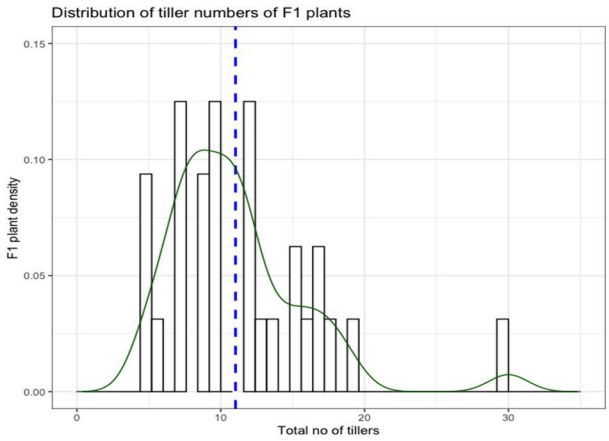

Figure 11 Total number of tillers at maturity was found to be variable among the putative F1 plants where blue dashed line represents the average number of tillers among plants.

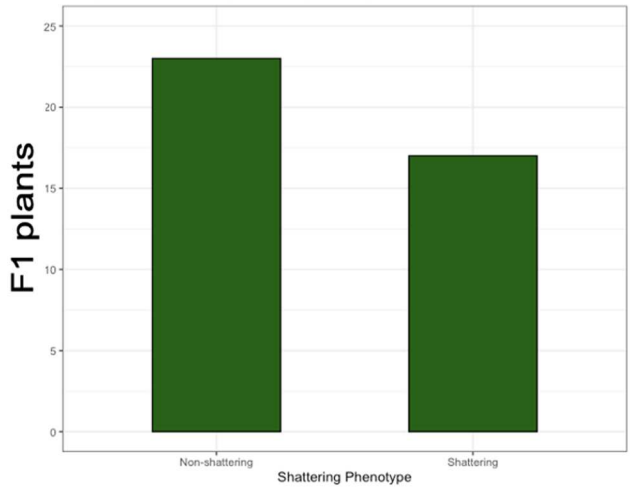

Figure 12 Among 40 F1 putative plants, 24 plants were recorded as shattering while 16 plants were recorded as non-shattering, which are believed to be selfed with unsuccessful fertilization from wild pollen.

Figure 13 (Left) presence of shoot hairs only in nodal region of plants, (Right) presence of dense shoot hairs all over the plants.

Figure 14 Presence of brace/aerial roots in one of the putative F1 plants

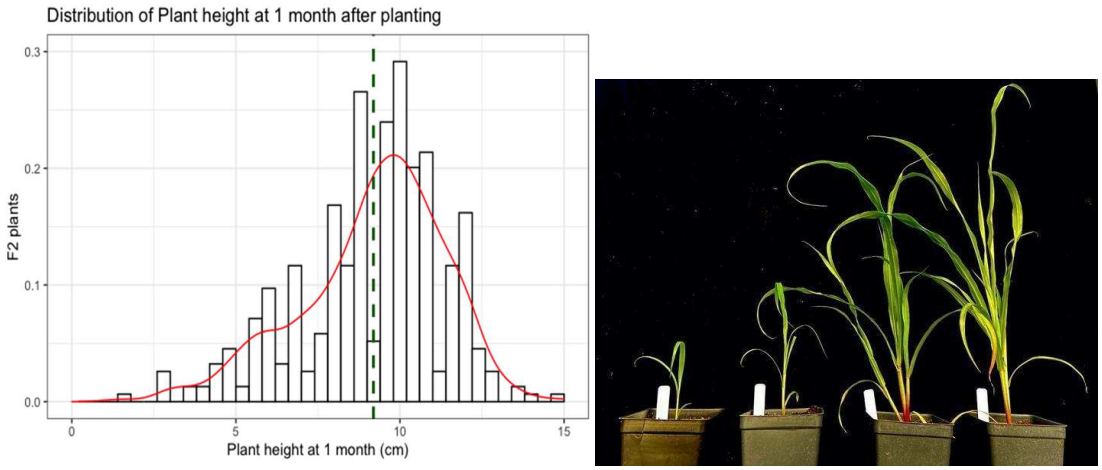

Figure 15 Frequency distribution of plant height of F2 mapping population at 1 month after planting (left) and Plants showing variation in plant height at 1 month after planting (right).

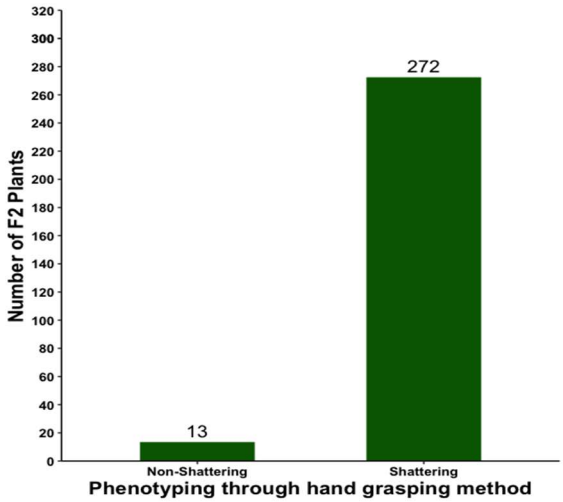

Figure 16 Phenotyping through hand grasping method where 13 F2 lines were characterized as non-shattering and 272 F2 lines were characterized as shattering.

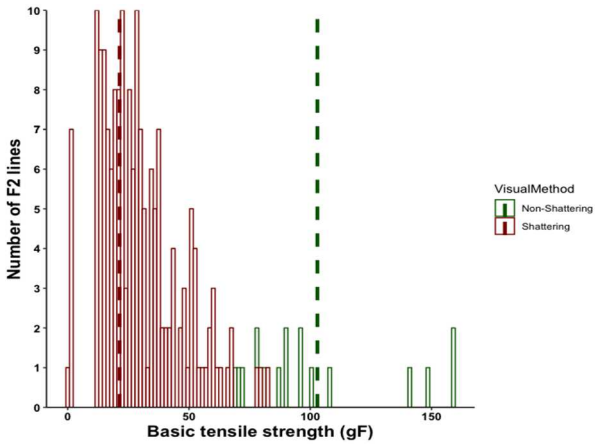

Figure 17 Histogram showing basic tensile strength of the F2 lines which were characterized as shattering and non-shattering. Red color represents the F2 lines characterized as shattering and green color represents the F2 lines characterized as non-shattering. While dashed line represents the mean od respective groups.

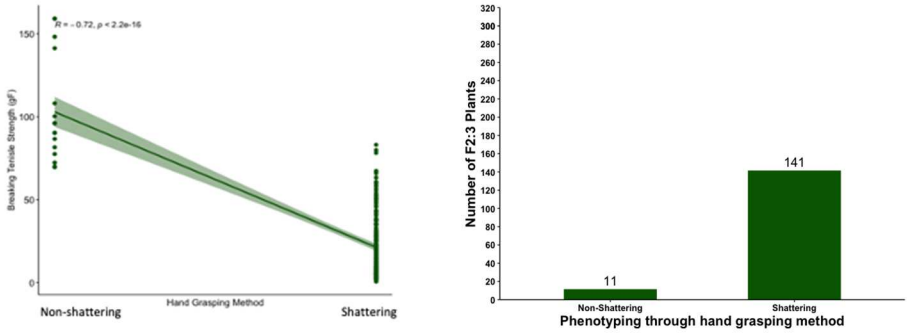

Figure 18 (Left) Significant higher positive between phenotyping done through hand grasping method and measurement of breaking tensile data where higher BTS data refers to the plants characterized as the shattering through hand grasping. (Right) Simple bar graph showing the phenotype of shattering and non-shattering qualitative data from $F_{2:3}$ population.

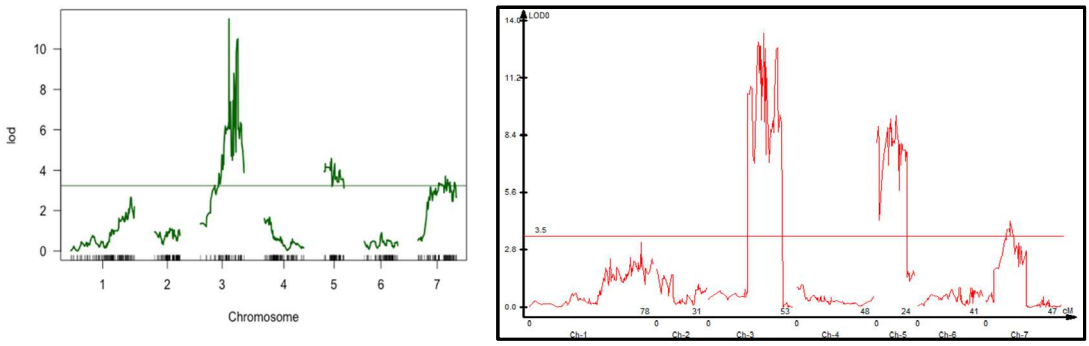

Figure 19 **QTL plot of quantitative data**: F2 generation quantitative breaking tensile strength (BTS) data QTLs significant on the chromosome 3 and 5 identified using simple interval mapping method. (Right) Significant Quantitative BTS data QTL of F2 generation on chromosome 3 and 5 identified using composite interval mapping.

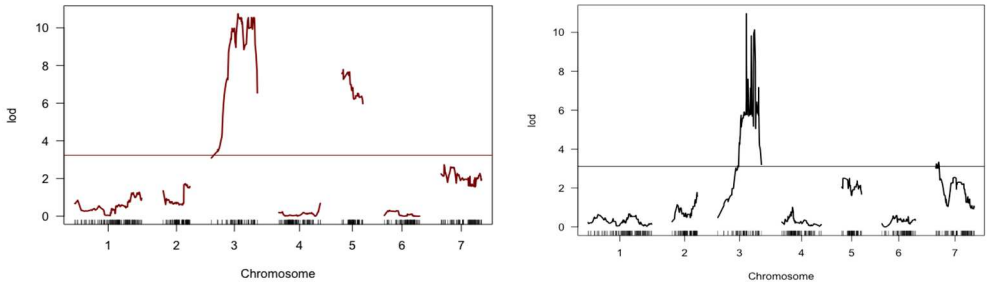

Figure 20 **QTL plot of quanlitative data from F_2 and $F_{2:3}$ generation**: Left (Red) F_2 generation qualitative data QTLs significant on the chromosome 3 and 5 identified using simple interval mapping method using binary model. Right (black) $F_{2:3}$ generation qualitative data QTLs significant on the chromosome 3 and 7 identified using simple interval mapping method using binary model.

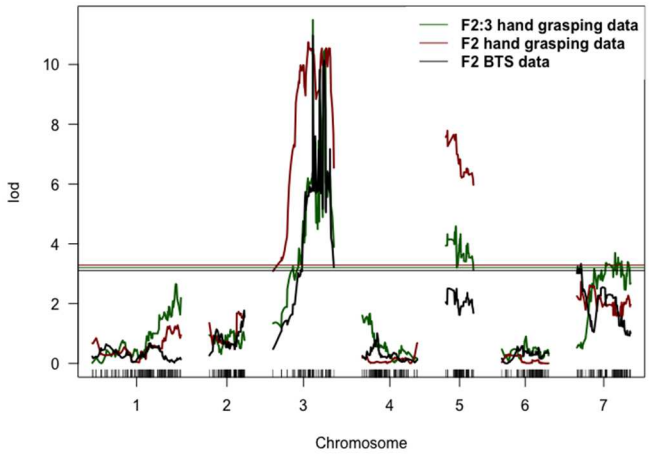

Figure 21 Major QTL on chromosome 3 identified for three sets of data co-localized with each other. Minor QTL identified on chromosome 5 co-localize with each other that has been identified from quantitative and qualitative data from F_2 generation.

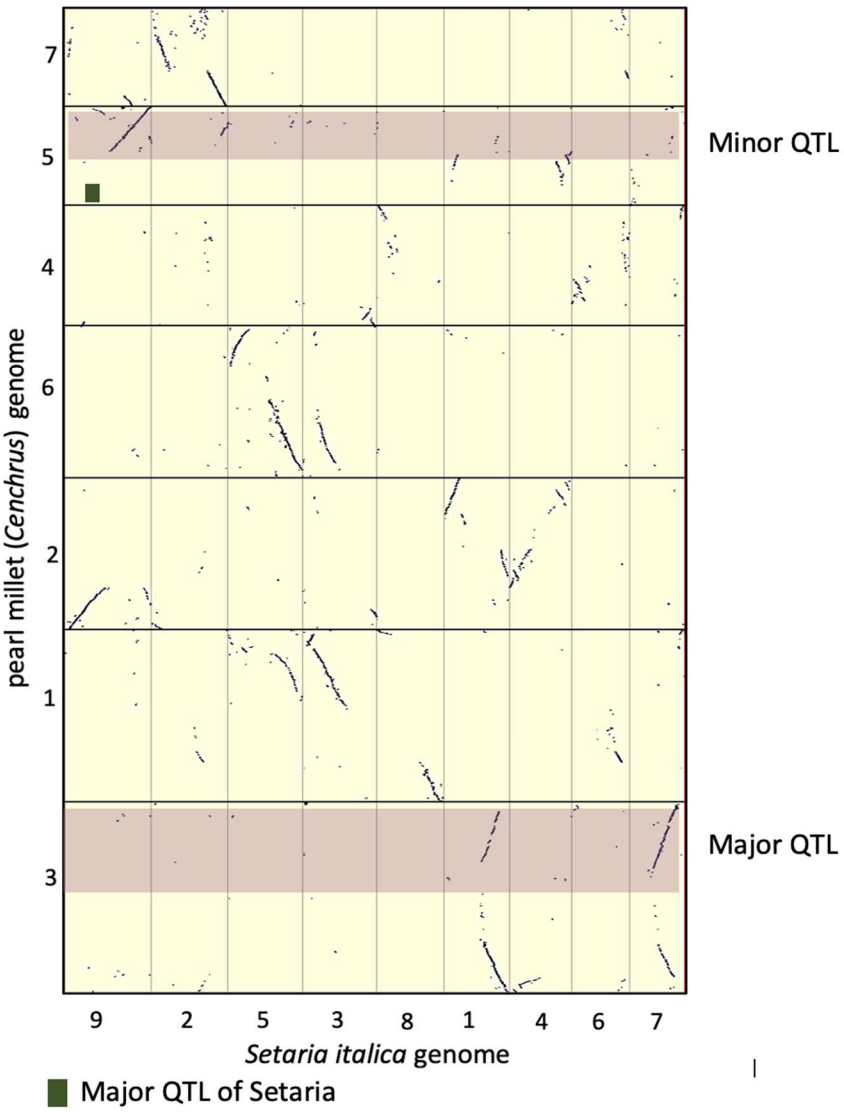

Figure 22 Whole genome dot plot of pearl millet vs *Setaria*. Diagonal blue lines in each cell indicated region of synteny between two genomes. Faded pink highlight indicates the chromosome with QTL identified in our study. Green rectangular box represents the major QTL identified in *Setaria* for shattering (Doust et al., 2014). Only QTL location from *Setaria* is retained in the dot plot if it lies on the QTL located chromosome of pearl millet.

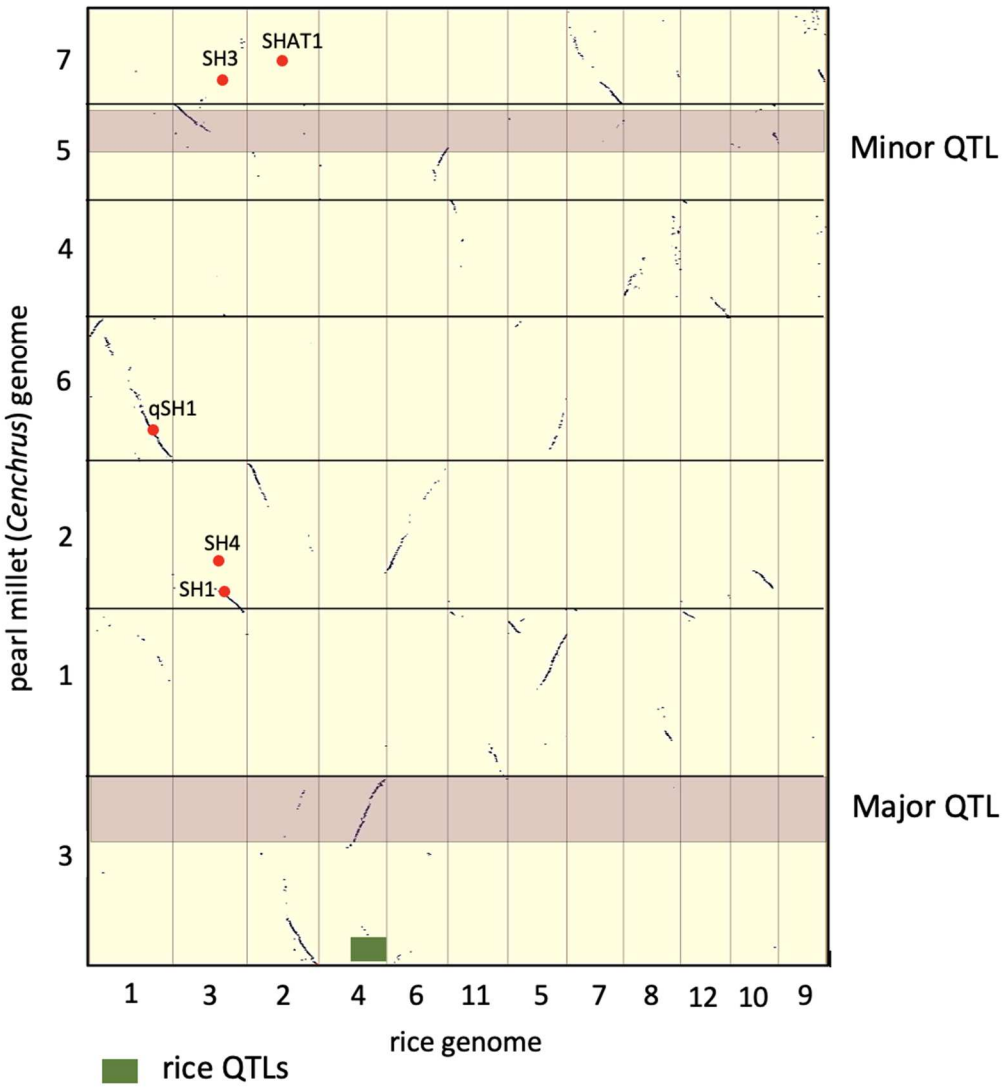

Figure 23 Whole genome dot plot of pearl millet vs rice. Diagonal blue lines in each cell indicate region of synteny between two genomes. Diagonal blue lines in each cell indicated region of synteny between two genomes. Faded pink highlight indicates the chromosome with QTL identified in our study. Green rectangular box represents the major QTL identified in rice for shattering (Doust et al., 2014). Only QTL location from rice is retained in the dot plot if it lies on the QTL located chromosome of pearl millet. Red dots represent the location of previously identified shattering genes.

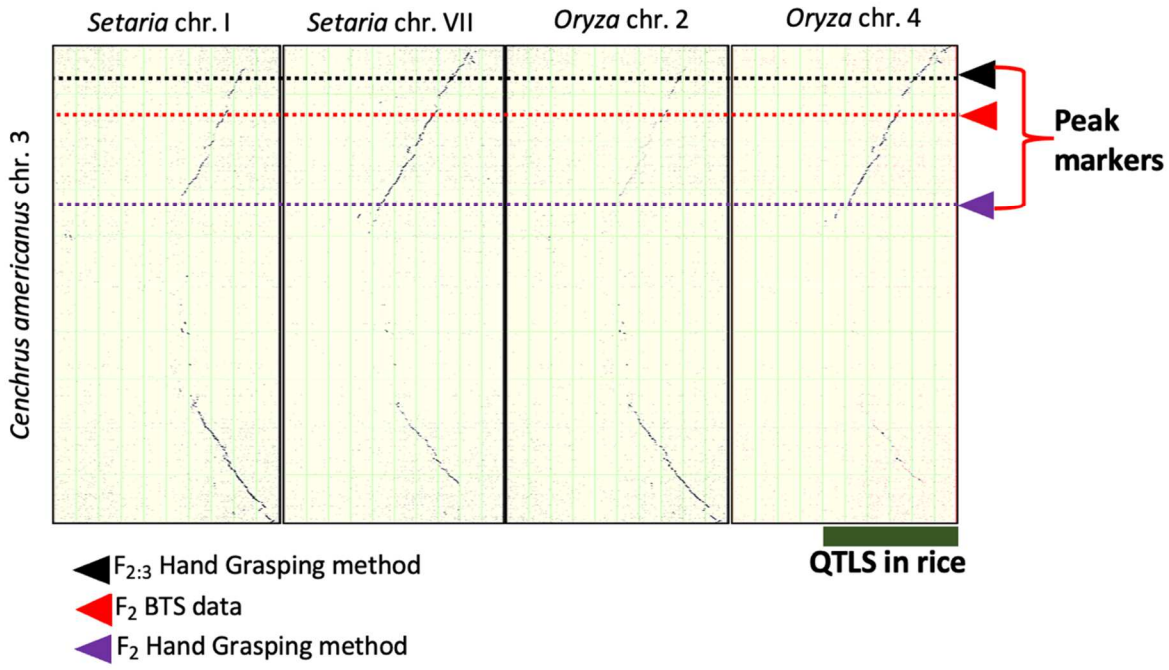

Figure 24 Analysis of pearl millet chromosome 3 and corresponding regions in *Setaria* and *Oryza* (rice). Each panel represents a chromosome-by-chromosome dot plot of pearl millet chr. 3 (vertical axis) vs., in turn, *Setaria* chromosome I, *Setaria* chromosome VII, *Oryza* chr. 2 and *Oryza* chr. 4 (horizontal axis). Horizontal dot lines represent the peak marker positions of QTL located on chromosome 3 where black dot line represents the position of peak marker identified in F2:3 generation for qualitative data from hand grasping method, red dot line represents the position of peak marker identified in F2 generation for quantitative data from breaking tensile strength method and purple dot line represents the position of peak marker identified in F2 generation for qualitative data from hand grasping method. On the bottom axis is the six QTL regions colocalized and drawn together in green filled box previously identified in rice (Doust, Mauro□Herrera, et al., 2014).

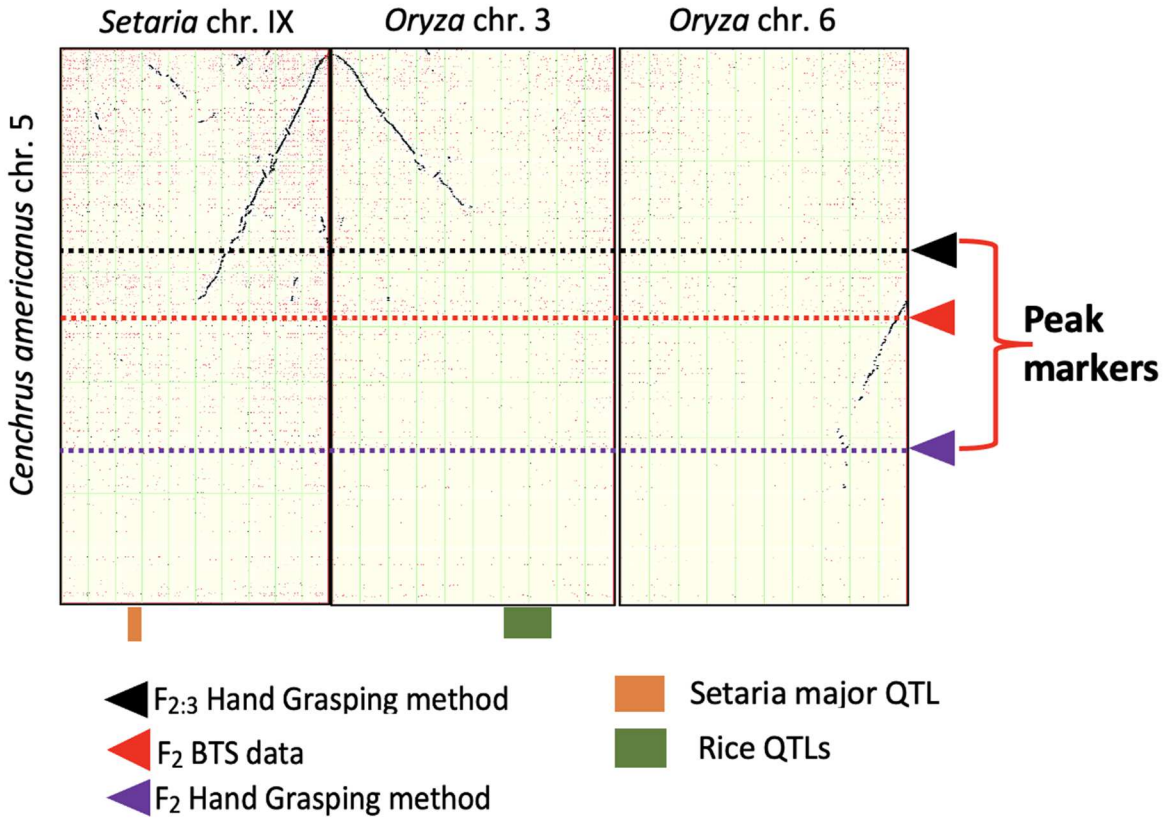

Figure 25 Analysis of pearl millet chromosome 5 and corresponding regions in *Setaria* and *Oryza* (rice). Each panel represents a chromosome-by-chromosome dot plot of pearl millet chr. 5 (vertical axis) vs., in turn, *Setaria* chromosome IX, *Oryza* chr. 3 and *Oryza* chr. 6 (horizontal axis). Horizontal dot lines represent the peak marker positions of QTL located on chromosome 3 where black dot line represents the position of peak marker identified in F2:3 generation for qualitative data from hand grasping method, red dot line represents the position of peak marker identified in F2 generation for quantitative data from breaking tensile strength method and purple dot line represents the position of peak marker identified in F2 generation for qualitative data from hand grasping method. On the bottom axis is the major QTL region previously identified in *Setaria* drawn in orange filled box (Doust, Lukens, et al., 2014). and four QTL region previously identified in rice

colocalized and drawn together in green filled box previously identified in rice (Doust, Mauro□Herrera, et al., 2014).

	Gen#	Chr #	Marker interval	Marker distance	peak marker	R2	additive	LOD
BreakingTensile Strength	F_2	3	Pgl_GLEAN_207757021 -- Pgl_GLEAN_299415938	91658917	Pgl_GLEAN_260396484	21.57 %	0.4968	12.6
		5	Pgl_GLEAN_ 60776619 -- Pgl_GLEAN_147490303	86713684	Pgl_GLEAN_116928697	13.68 %	0.513	6.8
		7	Pgl_GLEAN_88605312 -- Pgl_GLEAN_123738771	35133459	Pg_GLEAN_111516474	5.83%	0.504	4.2
Hand Grasping Method Phenotype	F_2	3	Pgl_GLEAN_300052462 -- Pgl_GLEAN_20620825	279431637	Pgl_GLEAN_273684987	22.55 %	0.338	12.5
		5	Pgl_GLEAN_81419087-- Pgl_GLEAN_152287762	70868675	Pgl_GLEAN_43422339	18.92	0.334	7.5
Hand Grasping Method Phenotype	$F_{2:3}$	3	Pgl_GLEAN_20620825-- Pgl_GLEAN_298708044	278037219	Pgl_GLEAN_236314895	21.44	0.444	6.17
		7	Pgl_GLEAN_81419087-- Pgl_GLEAN_151681260	70262173	Pgl_GLEAN_108010786	7.11	0.433	2.83

Table 2 QTL for seed shattering phenotype through two different methods in F_2 populations.

Gene	Accession ID	Cenchrus transcript	Cenchrus genome position	Identity of first hit
Shat1	LOC_Os04g55560.2	Pgl_GLEAN_10036454	chr7:78667416:78671077	79.024%
SH1	Sevir.9G153200	Pgl_GLEAN_10005105	chr2:27006164:27010214	95.906%
SH4	Seita.3G020800.1	Pgl_GLEAN_10004492	chr2:24414529:24416080	42.473%
QSh1	XP_015641948	Pgl_GLEAN_10005878	chr6:33203838:33207146	80%
Sh3	XP_015647585	Pgl_GLEAN_10024334	chr7:30762633:30767385	91.6675

Table 3 Location of known and characterized shattering gene on the *Cenchrus* genome

IV. DISCUSSION:

In this study, histology and SEM analysis revealed the position of abscission zone in pearl millet domesticated and wild accessions along with differences in their function and degree of shattering. The position of the abscission zone is at the junction of primary branch and main branch; as seen also in early development using scanning electron microscopy (Doust and Kellogg 2002). Sectioning analysis revealed that both the position and the developmental process of abscission zone formation is unlike *Setaria,* which has its abscission zone located below the glumes with no distinct cell size differentiation both below, in and above the abscission layer (Hodge and Kellogg 2016). Anatomical studies are yet to be done in pearl millet abscission for further comparison and detailed analysis, but the position and morphological differences suggest that the abscission zone in pearl millet is not derived from anything similar in *Setaria*.

Following the identification of the abscission zone position and measurement of the different shattering ability of the two accessions, I identified two QTL, with both the BTS and hand-grasping method. The QTL on chromosome 3 explained most of the variance

and the QTL on chromosome 5 explains relatively little variance. As expected, the wild allele led to a decrease in breaking tensile strength data and an increase in shattering. Our results revealed that the SH1 ortholog, first identified in sorghum as a single locus responsible for shattering (Lin et al. 2012), is not located in the QTL loci identified for *Cenchrus*. Recent studies support our findings as they revealed that even the SH1 loci, which has been the prime gene suspected to be under parallel selection for shattering across grasses was enriched only in the abscission zone of *Setaria* but not in rice and *Brachypodium* (Yu et al. 2020).

A comparative genomics analysis was used to gain further insights into possible conservation of shattering loci between *Cenchrus* and *Setaria* and rice. In general, the analysis of syntenic blocks between pearl millet chromosome 3 and *Setaria* and rice suggests a 1:2 relationship while analysis of syntenic blocks of minor QTL chromosome 5 suggests a 1:1 relationship in *Setaria* but 1:2 relationship in rice. Similarly, the analysis of syntenic blocks between minor QTL on chromosome 5 of pearl millet with *Setaria* and rice suggests that there is 1:1 relationship with *Setaria* but 1:2 relationship with rice. *Cenchrus* chromosome 3 is collinear, in part, to syntenic blocks on *Setaria* chr. 1 and 7, suggesting that chromosome duplication in Setaria took place after the divergence of *Setaria* and *Cenchrus* ~8.3 Myr ago (Devos et al. 2017) (figure 23). However, further analysis needs to be done to confirm this possible duplication event. There are no QTLs for shattering identified in either syntenic block of *Setaria* that co-localize with major QTL identified in pearl millet. In *Setaria*, two QTLs were identified on chromosomes V and IX to control shattering using an F7 recombinant inbred lines mapping population derived from a cross between domesticated foxtail millet (*S. italica*) and its wild ancestor,

green millet (*S. viridis*) (Doust et al. 2014; Odonkor et al. 2018), but these do not co-localize with QTL regions in *Cenchrus*. This indicates non-conservation of major shattering loci among these closely related grass genera. This accords with the different position and mode of breakage of the abscission zone in pearl compared to *Setaria* (Hodge and Kellogg 2016; Yu et al. 2020).

Interestingly, the syntenic block of chromosome 3 in rice has six QTLs previously identified from independent studies, as well as the major shattering gene SH4 (Doust et al. 2014; Li et al. 2006). SH4 was cloned from population of cross between two rice parents which differed in the shattering phenotype, with the wild accession having a complete abscission zone and the domesticated accession having an abscission zone developed at the similar time but being incomplete and with a large vascular bundle at maturity. In rice, the substitution of an asparagine for lysine in the SH4 of cultivated rice led to reduced shattering. SH4 plays an important role in the establishment of the abscission layer from the early stage of flower development as well as a role in the activation of the abscission process at seed maturity by promoting the hydrolysis of the abscission cells during the abscission process. (Li et al. 2006). However, a BLASTP search of the SH4 gene resulted in the closest homolog being identified on *Cenchrus* chromosome 2, while a putative pearl millet SH4 homolog in chromosome 3 within the syntenic region had a similarity of only 36% similarity and did not recover the rice SH4 gene when reverse blasted.

In the case of the syntenic block containing the minor QTL on chromosome 5 of pearl millet, comparison with rice revealed that no QTL was identified for shattering in rice chromosome 6 but four QTLs were identified from independent studies in rice

chromosome 3. However, QTLs located on the rice chromosome 3 did not colocalize within the syntenic block of pearl millet. This reveals no conservation of minor QTLs underlying shattering in rice as well as with closely related C4 grass *Setaria.* To further test the parallel selection hypothesis, BLASTP search of previously identified and characterized shattering genes was done against the *Cenchrus* genome to identify if any shattering genes underlie either major or minor QTLs identified in our study. However, results revealed no known genes underlying the QTLs identified; further strengthening the hypothesis that novel loci are responsible for shattering in pearl millet.

To further identify the novel loci in pearl millet, narrowing down of QTLs identified is a must. While QTLs identified from both methods and generations were co-localized in the same chromosomes, I suffered from a major problem of segregation distortion as well as lack of recombination across F2 progenies, resulting in wide QTL regions with low resolution. However, comparative genetics provides us some insights for candidate gene identification and rule out the hypothesis of parallel selection in pearl millet for SH1 and for other known shattering genes. The wide QTL intervals in our study are likely the result of issues in marker order as well as low rates of recombination and high segregation distortion.

Both QTL loci identified for shattering are different from those identified by Poncet et al. (2000, 2002), where QTL on LG6 was found to be associated with a functional abscission layer. There is some reason to be concerned that the few markers able to be used for this early map were not able to accurately construct a reliable map, as the same study identified a major QTL on linkage group 7 for plant height at maturity; yet fine mapping

of the major gene affecting height in pearl millet has shown it to be on LG4 of pearl millet (Parvathaneni et al., 2013).

The QTLs identified in our study associating with the shattering phenotype provide us direction for future refinement of the genomic region using $F_{2:3}$ generation seeds. Comparative mapping of their positions has confirmed that *Setaria* and *Cenchrus*, while closely related, do not share the same genetic or morphological shattering mechanism. Further exploration needs to be done to validate our results and test these hypotheses.

REFERENCES

Agte, V., Khot, S., Girigosavi, S., Paknikar, K., and Chiplonkar, S. (1999). Comparative performance of pearl millet-and sorghum-based diets vs. wheat-and rice-based diets for trace metal bioavailability. *Journal of trace elements in medicine and biology, 13*(4), 215-219.

Alexandratos, N. (2009). *World food and agriculture to 2030/50.* Paper presented at the Highlights and views from MID-2009: Paper for the Expert Meeting on "How to Feed the World in.

Andrews, D., and Kumar, K. A. (1992). Pearl millet for food, feed, and forage. In *Advances in agronomy* (Vol. 48, pp. 89-139): Elsevier.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., . . . Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS one, 3*(10), e3376.

Baltensperger, D. D. (2002). Progress with proso, pearl and other millets. *Trends in new crops and new uses*, 100-103.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *bioinformatics, 30*(15), 2114-2120.

Chandra, K., and Pandey, A. (2017). QTL mapping in crop improvement: a basic concept. *International Journal of Current Microbiology and Applied Sciences, 6*(12), 835-842.

CLARKE, J. M. (1981). Effect of delayed harvest on shattering losses in oats, barley and wheat. *Canadian Journal of Plant Science, 61*(1), 25-28.

Collard, B. C., Jahufer, M., Brouwer, J., and Pang, E. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica, 142*(1), 169-196.

Council, N. R. (2010). *Advancing the Science of Climate Change*. Washington, DC: The National Academies Press.

DeLuca, D. S., Levin, J. Z., Sivachenko, A., Fennell, T., Nazaire, M.-D., Williams, C., . . . Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *bioinformatics, 28*(11), 1530-1532.

Devos, K. M., Wu, X., and Qi, P. (2017). Genome structure and comparative genomics. In *Genetics and genomics of Setaria* (pp. 135-147): Springer.

Doust, A. N., Lukens, L., Olsen, K. M., Mauro-Herrera, M., Meyer, A., and Rogers, K. (2014). Beyond the single gene: How epistasis and gene-by-environment effects influence crop domestication. *Proceedings of the National Academy of Sciences, 111*(17), 6178-6183.

Doust, A. N., Mauro□Herrera, M., Francis, A. D., and Shand, L. C. (2014). Morphological diversity and genetic regulation of inflorescence abscission zones in grasses. *American journal of botany, 101*(10), 1759-1769.

Doust, A. N., Penly, A. M., Jacobs, S. W., and Kellogg, E. A. (2007). Congruence, Conflict, and Polyploidization Shown by Nuclear and Chloroplast Markers in the Monophyletic"Bristle Clade"(Paniceae, Panicoideae, Poaceae). *Systematic Botany, 32*(3), 531-544.

Dussert, Y., Remigereau, M. S., Fontaine, M., Snirc, A., Lakis, G., Stoeckel, S., . . . Robert, T. (2013). Polymorphism pattern at a miniature inverted□repeat transposable element locus downstream of the domestication gene Teosinte□ branched1 in wild and domesticated pearl millet. *Molecular ecology, 22*(2), 327-340.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one, 6*(5), e19379.

Fan, M., Miao, F., Jia, H., Li, G., Powers, C., Nagarajan, R., . . . Yan, L. (2021). O-linked N-acetylglucosamine transferase is involved in fine regulation of flowering time in winter wheat. *Nature communications, 12*(1), 1-12.

Gao, M., and Zhu, H. (2013). Fine mapping of a major quantitative trait locus that regulates pod shattering in soybean. *Molecular Breeding, 32*(2), 485-491.

Goswami, A., Sharma, K., and Sehgal, K. (1969). Nutritive value of proteins of pearl millet of high-yielding varieties and hybrids. *British Journal of Nutrition, 23*(4), 913-916.

Gualtieri, M., and Rapaccini, S. (1990). Sorghum grain in poultry feeding. *World's Poultry Science Journal, 46*(3), 246-254.

Gulia, S., Wilson, J., Carter, J., and Singh, B. (2007). Progress in grain pearl millet research and market development. *Issues in new crops and new uses*, 196-203.

Harlan, J. (1975). Crops and Man, Madison, Wisconsin, American society of Agronomy and Crop Science Society of America. *WI, USA*.

Healey, A., Furtado, A., Cooper, T., and Henry, R. J. (2014). Protocol: a simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. *Plant Methods, 10*(1), 21. doi:10.1186/1746-4811-10-21

Hodge, J. G., and Kellogg, E. A. (2016). Abscission zone development in Setaria viridis and its domesticated relative, Setaria italica. *American journal of botany, 103*(6), 998-1005.

Ishikawa, R., Nishimura, A., Htun, T. M., Nishioka, R., Oka, Y., Tsujimura, Y., . . . Ishii, T. (2017). Estimation of loci involved in non-shattering of seeds in early rice domestication. *Genetica, 145*(2), 201-207.

Ishikawa, R., Thanh, P. T., Nimura, N., Htun, T. M., Yamasaki, M., and Ishii, T. (2010). Allelic interaction at seed-shattering loci in the genetic backgrounds of wild and cultivated rice species. *Genes and genetic systems, 85*(4), 265-271.

Jagodic, M., and Stridh, P. (2014). Positional Gene Cloning in Experimental Populations. In *Multiple Sclerosis* (pp. 3-24): Springer.

Ji, H.-S., Chu, S.-H., Jiang, W., Cho, Y.-I., Hahn, J.-H., Eun, M.-Y., . . . Koh, H.-J. (2006). Characterization and mapping of a shattering mutant in rice that corresponds to a block of domestication genes. *Genetics, 173*(2), 995-1005.

Jukanti, A., Gowda, C. L., Rai, K., Manga, V., and Bhatt, R. (2016). Crops that feed the world 11. Pearl Millet (Pennisetum glaucum L.): an important source of food security, nutrition and health in the arid and semi-arid tropics. *Food Security, 8*(2), 307-329.

Kong, W., Kim, C., Zhang, D., Guo, H., Tan, X., Jin, H., . . . Sezen, U. (2018). Genotyping by sequencing of 393 Sorghum bicolor BTx623× IS3620C recombinant inbred lines improves sensitivity and resolution of QTL detection. *G3: Genes, Genomes, Genetics, 8*(8), 2563-2572.

Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., and Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics, 1*(2), 174-181.

Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. *science, 311*(5769), 1936-1939.

Lin, Z., Li, X., Shannon, L. M., Yeh, C.-T., Wang, M. L., Bai, G., . . . Clemente, T. E. (2012). Parallel domestication of the Shattering1 genes in cereals. *Nature genetics, 44*(6), 720-724.

Liu, C., Witcombe, J. R., Pittaway, T., Nash, M., Hash, C., Busso, C., and Gale, M. (1994). An RFLP-based genetic map of pearl millet (Pennisetum glaucum). *Theoretical and Applied Genetics, 89*(4), 481-487.

Lyons, E., Pedersen, B., Kane, J., Alam, M., Ming, R., Tang, H., . . . Lisch, D. (2008). Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. *Plant physiology, 148*(4), 1772-1781.

Lyons, E., Pedersen, B., Kane, J., and Freeling, M. (2008). The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Tropical Plant Biology, 1*(3), 181-190.

Moumouni, K., Kountche, B., Jean, M., Hash, C., Vigouroux, Y., Haussmann, B., and Belzile, F. (2015). Construction of a genetic map for pearl millet, Pennisetum glaucum (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. *Molecular Breeding, 35*(1), 1-10.

Mutschler, M., Doerge, R., Liu, S., Kuai, J., Liedl, B., and Shapiro, J. (1996). QTL analysis of pest resistance in the wild tomato Lycopersicon pennellii: QTLs controlling acylsugar level and composition. *Theoretical and Applied Genetics, 92*(6), 709-718.

Myers, R. (2002). Alternative crop guide: pearl millet. *Jefferson Institute. Washington, DC http://www. jeffersoninst. org/(accessed 01 Aug. 2014)*.

Newman, Y., Jennings, E., Vendramini, J., and Blount, A. (2010). Pearl millet (Pennisetum glaucum): Overview and management. *SSAGR-337, one of a series of the Agronomy Department, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida*.

Odonkor, S., Choi, S., Chakraborty, D., Martinez-Bello, L., Wang, X., Bahri, B. A., . . . Devos, K. M. (2018). QTL mapping combined with comparative analyses identified candidate genes for reduced shattering in Setaria italica. *Frontiers in plant science, 9*, 918.

Panaud, O. (2006). Foxtail millet. In *Cereals and millets* (pp. 325-332): Springer.

Parvathaneni, R. K., Jakkula, V., Padi, F. K., Faure, S., Nagarajappa, N., Pontaroli, A. C., . . . Devos, K. M. (2013). Fine-mapping and identification of a candidate gene underlying the d2 dwarfing phenotype in pearl millet, Cenchrus americanus (L.) Morrone. *G3: Genes, Genomes, Genetics, 3*(3), 563-572.

Paterson, A. H., Lin, Y.-R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., . . . Irvine, J. E. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. *science, 269*(5231), 1714-1718.

Peters, J. L., Cnudde, F., and Gerats, T. (2003). Forward genetics and map-based cloning approaches. *Trends in plant science, 8*(10), 484-491.

Poncet, V., Lamy, F., Devos, K., Gale, M., Sarr, A., and Robert, T. (2000). Genetic control of domestication traits in pearl millet (Pennisetum glaucum L., Poaceae). *Theoretical and Applied Genetics, 100*(1), 147-159.

Punnuri, S. M., Wallace, J. G., Knoll, J. E., Hyma, K. E., Mitchell, S. E., Buckler, E. S., . . . Singh, B. P. (2016). Development of a high□density linkage map and tagging leaf spot resistance in pearl millet using genotyping□by□sequencing markers. *The Plant Genome, 9*(2), plantgenome2015.2010.0106.

Qi, P., Gimode, D., Saha, D., Schröder, S., Chakraborty, D., Wang, X., . . . Devos, K. M. (2018). UGbS-Flex, a novel bioinformatics pipeline for imputation-free SNP discovery in polyploids without a reference genome: finger millet as a case study. *BMC plant biology, 18*(1), 117.

Qi, X., Pittaway, T., Lindup, S., Liu, H., Waterman, E., Padi, F., . . . Devos, K. (2004). An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, Pennisetum glaucum. *Theoretical and Applied Genetics, 109*(7), 1485-1493.

Rajaram, V., Nepolean, T., Senthilvel, S., Varshney, R. K., Vadez, V., Srivastava, R. K., . . . Kumari, B. R. (2013). Pearl millet [Pennisetum glaucum (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. *BMC genomics, 14*(1), 1-16.

Reshma, R. S., and Das, D. (2021). Molecular markers and its application in animal breeding. In *Advances in Animal Genomics* (pp. 123-140): Elsevier.

Sanchez, A., Subudhi, P., Rosenow, D., and Nguyen, H. (2002). Mapping QTLs associated with drought resistance in sorghum (Sorghum bicolor L. Moench). *Plant Molecular Biology, 48*(5), 713-726.

Sawaya, W., Khalil, J., and Safi, W. (1984). Nutritional quality of pearl millet flour and bread. *Plant Foods for Human Nutrition, 34*(2), 117-125.

Silva, L. D. C. E., Wang, S., and Zeng, Z.-B. (2012). Composite interval mapping and multiple interval mapping: procedures and guidelines for using Windows QTL Cartographer. In *Quantitative trait loci (QTL)* (pp. 75-119): Springer.

Supriya, A., Senthilvel, S., Nepolean, T., Eshwar, K., Rajaram, V., Shaw, R., . . . Narasu, M. (2011). Development of a molecular linkage map of pearl millet integrating DArT and SSR markers. *Theoretical and Applied Genetics, 123*(2), 239-250.

Tanksley, S. D., Young, N. D., Paterson, A. H., and Bonierbale, M. (1989). RFLP mapping in plant breeding: new tools for an old science. *Bio/technology, 7*(3), 257-264.

Terrill, T., Gelaye, S., Amoah, E., Miller, S., Kouakou, B., Gates, R., and Hanna, W. (1998). Protein and energy value of pearl millet grain for mature goats. *Journal of animal science, 76*(7), 1964-1969.

Tomkowiak, A., Bocianowski, J., Wolko, Ł., Adamczyk, J., Mikołajczyk, S., and Kowalczewski, P. Ł. (2019). Identification of Markers Associated with Yield Traits and Morphological Features in Maize (Zea mays L.). *Plants, 8*(9), 330.

Tsujimura, Y., Sugiyama, S., Otsuka, K., Htun, T. M., Numaguchi, K., Castillo, C., . . . Ishikawa, R. (2019). Detection of a novel locus involved in non-seed-shattering behaviour of Japonica rice cultivar, Oryzasativa 'Nipponbare'. *Theoretical and Applied Genetics, 132*(9), 2615-2623.

Varshney, R. K., Shi, C., Thudi, M., Mariac, C., Wallace, J., Qi, P., . . . Rathore, A. (2017). Pearl millet genome sequence provides a resource to improve agronomic traits in arid environments. *Nature biotechnology, 35*(10), 969-976.

Wingett, S. W., and Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research, 7*.

Wu, Y., Bhat, P. R., Close, T. J., and Lonardi, S. (2008). Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS genetics, 4*(10), e1000212.

Yadav, O., and Rai, K. (2013). Genetic improvement of pearl millet in India. *Agricultural Research, 2*(4), 275-292.

Yadav, S., Sandhu, N., Singh, V. K., Catolos, M., and Kumar, A. (2019). Genotyping-by-sequencing based QTL mapping for rice grain yield under reproductive stage drought stress tolerance. *Scientific reports, 9*(1), 1-12.

Yang, C., Wang, Z., Yang, X., and LU, B. R. (2014). Segregation distortion affected by transgenes in early generations of rice crop□weed hybrid progeny: Implications for assessing potential evolutionary impacts from transgene flow into wild relatives. *Journal of Systematics and Evolution, 52*(4), 466-476.

Yano, M., Harushima, Y., Nagamura, Y., Kurata, N., Minobe, Y., and Sasaki, T. (1997). Identification of quantitative trait loci controlling heading date in rice using a high-density linkage map. *Theoretical and Applied Genetics, 95*(7), 1025-1032.

Yu, Y., Hu, H., Doust, A. N., and Kellogg, E. A. (2020). Divergent gene expression networks underlie morphological diversity of abscission zones in grasses. *New Phytologist, 225*(4), 1799-1815.

Zegada-Lizarazu, W., and Iijima, M. (2005). Deep root water uptake ability and water use efficiency of pearl millet in comparison to other millet species. *Plant production science, 8*(4), 454-460.

Zeng, Q., and An, S. (2021). Identifying the biogeographic patterns of rare and abundant bacterial communities using different primer sets on the loess plateau. *Microorganisms, 9*(1), 139.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *bioinformatics, 30*(15), 2114-2120.

Broman, K. W., Wu, H., Sen, Ś., & Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *bioinformatics, 19*(7), 889-890.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Molecular ecology, 22*(11), 3124-3140.

Devos, K., Pittaway, T., Reynolds, A., & Gale, M. (2000). Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. *Theoretical and Applied Genetics, 100*(2), 190-198.

Doust, A. N., Mauro-Herrera, M., Hodge, J. G., & Stromski, J. (2017). The C4 model grass Setaria is a short day plant with secondary long day genetic regulation. *Frontiers in plant science, 8*, 1062.

Parvathaneni, R. K., Jakkula, V., Padi, F. K., Faure, S., Nagarajappa, N., Pontaroli, A. C., . . . Devos, K. M. (2013). Fine-mapping and identification of a candidate gene underlying the d2 dwarfing phenotype in pearl millet, Cenchrus americanus (L.) Morrone. *G3: Genes, Genomes, Genetics, 3*(3), 563-572.

VITA

Nikee Shrestha

Candidate for the Degree of

Master of Science

Thesis: Understanding the genetic basis of shattering in pearl millet (*Cenchrus americanus*)

Major Field: Plant Biology

Biographical:

Education:

Completed the requirements for the Master of Science in Plant Biology at Oklahoma State University, Stillwater, Oklahoma in April 2022.

Completed the requirements for the Bachelor of Science in Agriculture at Tribhuvan University, Kathmandu, Nepal in 2018

Experience:
Research assistant
Teaching assistant

Professional Memberships:
Botanical Society of America
Maize Genetics Corporation