Published May 6, 2016

Development of a High-Density Linkage Map and Tagging Leaf Spot Resistance in Pearl Millet Using Genotyping-by-Sequencing Markers

Somashekhar M. Punnuri,* Jason G. Wallace, Joseph E. Knoll, Katie E. Hyma, Sharon E. Mitchell, Edward S. Buckler, Rajeev K. Varshney, and Bharat P. Singh

Abstract

Pearl millet [*Pennisetum glaucum* (L.) R. Br; also *Cenchrus americanus* (L.) Morrone] is an important crop throughout the world but better genomic resources for this species are needed to facilitate crop improvement. Genome mapping studies are a prerequisite for tagging agronomically important traits. Genotyping-by-sequencing (GBS) markers can be used to build high-density linkage maps, even in species lacking a reference genome. A recombinant inbred line (RIL) mapping population was developed from a cross between the lines 'Tift 99D_2B_1' and 'Tift 454'. DNA from 186 RILs, the parents, and the F_1 was used for 96-plex *Ape*KI GBS library development, which was further used for sequencing. The sequencing results showed that the average number of good reads per individual was 2.2 million, the pass filter rate was 88%, and the CV was 43%. High-quality GBS markers were developed with stringent filtering on sequence data from 179 RILs. The reference genetic map developed using 150 RILs contained 16,650 single-nucleotide polymorphisms (SNPs) and 333,567 sequence tags spread across all seven chromosomes. The overall average density of SNP markers was 23.23 SNP/cM in the final map and 1.66 unique linkage bins per cM covering a total genetic distance of 716.7 cM. The linkage map was further validated for its utility by using it in mapping quantitative trait loci (QTLs) for flowering time and resistance to *Pyricularia* leaf spot [*Pyricularia grisea* (Cke.) Sacc.]. This map is the densest yet reported for this crop and will be a valuable resource for the pearl millet community.

Core Ideas

- Pearl millet [*Pennisetum glaucum* (L.) R. Br; also *Cenchrus americanus* (L.) Morrone] is an important forage and grain crop in many parts of the world but genomic resources for this species are needed to facilitate crop improvement.
- The reference genetic map developed using 150 recombinant inbred lines contained 16,650 single-nucleotide polymorphisms and 333,567 sequence tags spread across all seven chromosomes.
- This map is the densest yet reported for this crop and will be a valuable resource for the pearl millet community.
- Genome mapping studies are a prerequisite for tagging agronomically important traits.
- Genotyping-by-sequencing markers can be used to build high-density linkage maps, even in species lacking a reference genome.

S.M. Punnuri and B.P. Singh, Agricultural Research Station, Fort Valley State Univ., 1005 State University Drive, Fort Valley, GA 31030; J.G. Wallace, Dep. Crop and Soil Sciences, the Univ. of Georgia, Athens, GA 30602 and Inst. for Genomic Diversity, Cornell Univ., Ithaca, NY 14853; J.E. Knoll, USDA-ARS, Crop Genetics and Breeding Research Unit, Tifton, GA 31793; K.E. Hyma and S.E. Mitchell, Genomic Diversity Facility, Inst. of Biotechnology, Cornell Univ., Ithaca, NY 14853; E.S. Buckler, USDA-ARS, Inst. for Genomic Diversity, Dep. Plant Breeding & Genetics, Cornell Univ., Ithaca, NY 14853; R.K. Varshney, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, 502324 Telangana, India. SMP and JGW contributed equally to this work. Received 14 July 2015. Accepted 29 Jan. 2016. *Corresponding author (punnuris@fvsu.edu).

Abbreviations: DArT, Diversity Arrays Technology; GBS, genotyping-by-sequencing; H^2, broad-sense heritability; LD, linkage disequilibrium; LG, linkage group; LOD, logarithm of odds; NGS, next-generation sequencing; QTL, quantitative trait locus; RFLP, restriction fragment length polymorphism; RIL, recombinant inbred line; SNP, single-nucleotide polymorphism; SSR, simple sequence repeat

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2015.10.0106

Pearl millet, widely known for its tolerance to heat, drought and soil toxicity, is grown for both grain and forage in many parts of the world, particularly in warm, dry regions (Burton and Powel, 1968; Chemisquy et al., 2010). Pearl millet has higher water-use efficiency and nitrogen-use efficiency than many other cereals (Muchow, 1988; Maman et al., 2006; Vadez et al., 2012) and shows useful genetic variation for tolerance to high temperatures during seedling establishment (Peacock et al., 1993; Howarth et al., 1994) and during reproductive growth stages (Gupta et al., 2015) and can thrive on acidic, sandy, or infertile soils where few other crops can grow (Andrews and Kumar, 1992). For these reasons, pearl millet is an essential staple food grain and/or fodder crop in many developing countries.

The market for pearl millet grain is also increasing in the United States because of consumers preferring gluten-free food and demand for millet flour by many ethnic groups (Dahlberg et al., 2004; Gulia et al., 2007). In addition, alternative sources to maize- (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.]-based livestock feed are sought to lower production costs for the poultry industry in the southeastern United States (Durham, 2003; Farrell, 2005; Cunningham and Fairchild, 2012). Whole pearl millet grain has been shown to be a satisfactory feed ingredient for broiler chickens and for egg production while reducing feed costs (Collins et al., 1997; Davis et al., 2003; Garcia and Dale, 2006). Compared to sorghum [*Sorghum bicolor* (L.) Moench], pearl millet grain offers lower starch, superior protein quality and content, a higher protein efficiency ratio, and greater metabolizable energy levels for poultry diets (Sullivan et al., 1990; Bramel-Cox et al., 1992; Andrews et al., 1993; Nambiar et al., 2011). Over 70% of the approximately 10 Mha of pearl millet grown annually in India is sown to F_1 hybrids (Yadav and Rai, 2011; Yadav et al., 2011a) and the development of pearl millet grain hybrids in the United States has shown some progress. For example, the USDA-ARS at Tifton, GA, in collaboration with the University of Georgia, released 'TifGrain 102' as a commercial grain hybrid (Durham, 2003; Lee et al., 2004). TifGrain 102 offers several advantages compared to other row crops, especially its ability to grow on sandy, acidic soils with minimum inputs and its resistance to root knot nematode (*Meloidogyne incognita* Kofoid & White), rust (*Puccinia substriata* Ellis & Barth. var. *indica* Ramachar & Cummins), and *Pyricularia* leaf spot [*Pyricularia grisea* (Cke.) Sacc. (teleomorph: *Magnaporthe grisea* (T.T. Herbert) M.E. Barr](Hanna and Wells, 1989; Wilson et al., 1989; Timper et al., 2002; Gupta et al., 2012). Because of its high forage quality, pearl millet is also grown as an annual fodder crop in the southeastern United States (Burton and Powel, 1968; Chemisquy et al., 2010).

Pearl millet is diploid with seven pairs of homologous chromosomes and an estimated genome size of 2350 Mb (or 2C = 4.71 pg based on flow cytometry), much of which consists of repetitive sequences (Bennett and Smith, 1976; Wimpee and Rawson, 1979; Martel et al., 1997; Jauhar and Hanna, 1998; Thomas et al., 2000). Some DNA markers have been developed and used over the past two decades in pearl millet for genetic research or for applied breeding and selection (Hash and Bramel-Cox 2000; Bidinger and Hash, 2004; Gale et al., 2005). Nonetheless, pearl millet crop improvement suffers from a relative lack of genetic and genomic resources compared to most other cereals. Characterization and utilization of pearl millet diversity can be aided by expanding the (currently few) genomic resources available in this crop.

Genetic markers are the building blocks for constructing linkage maps. Linkage maps further support numerous applications in plant breeding. Genetic maps of several pearl millet populations have been made using different marker sets over the past 20 y (Liu et al., 1994; Devos et al., 2000; Qi et al., 2004; Pedraza-Garcia et al., 2010; Supriya et al., 2011; Sehgal et al., 2012). Recently, a simple-sequence repeat (SSR) consensus map with 174 loci was developed using four RIL mapping populations (Rajaram et al., 2013). Despite these efforts, pearl millet linkage maps frequently have large gaps at the distal ends, which is probably caused by (i) a lack of either sufficient markers or polymorphisms in these regions, (ii) extremely high rates of genetic recombination in these regions requiring large numbers of physically closely linked markers to permit linkage detection, and/or (iii) the nature of the markers and parents used in these studies (Devos et al., 2000; Vadez et al., 2012). A suggestion that these gaps are caused by some combination of the latter two explanations is provided by Supriya et al. (2011), who demonstrated greatly improved genome coverage with Diversity Arrays Technology (DArT) markers compared with that provided by available SSR markers. Most of the remaining previous maps have generally relied on SSRs, restriction fragment length polymorphisms (RFLPs), or related markers; however, in many crops, linkage maps based on SNPs are now becoming common because of the low cost of high-throughput sequencing methods (Ganal et al., 2009; Kumar et al., 2012). Because of their abundance in the genome, SNPs can be used to build much denser linkage maps than other types of markers. Such SNP-based genetic maps are highly informative as they not only reveal the complexity of genome architecture (structure and organization) but also trace the genetic basis of QTLs underlying a trait with better resolution (Krawczak, 1999; Mammadov et al., 2012). Next-generation sequencing (NGS) technologies have facilitated the rapid detection of genome-wide SNP markers. Genotyping-by-sequencing is one such powerful approach to develop genome-wide SNP datasets (Elshire et al., 2011). This technique uses restriction enzymes to selectively digest genomic DNA; next, 'barcoded' DNA adapters are ligated to the fragments to multiplex many samples in a single sequencing lane (Elshire et al., 2011). The choice of restriction enzyme(s) and multiplexing makes GBS a versatile system and the ability to multiplex enables low-cost, high-throughput-marker discovery (Poland and Rife, 2012). Importantly,

it also works in less exploited crops, including those for which no reference genome sequence is available publicly, such as pearl millet. The potential utility of GBS markers in developing high-density molecular maps for several cereal crops, including maize, barley (*Hordeum vulgare* L.) and oat (*Avena sativa* L.) has been extensively reviewed and shown to be useful (He et al., 2014). Recently, a pearl millet linkage map was also developed with 2809 high-quality SNP markers using a modified GBS protocol (Moumouni et al., 2015).

The objectives of this study were to construct a high-density linkage map using GBS-derived markers to provide a platform for downstream studies and to develop genomic resources for the greater pearl millet research community. The two parents used in this experiment (Tift $99D_2B_1$ and Tift 454) are also the parents of the commercial grain hybrid TifGrain 102 (Hanna et al., 2005a,2005b). Flowering time was chosen as a phenotypic trait for QTL analysis to demonstrate the utility of this map. Also, Tift $99D_2B_1$ carries genes for resistance to *Pyricularia* leaf spot (Hanna and Wells, 1989) and hence this mapping population was used to evaluate resistance to this disease as well.

Materials and Methods

Pearl Millet Mapping Population

The parental lines used in this study are Tift $99D_2B_1$ and Tift 454, where Tift $99D_2B_1$ was used as the female parent. Both Tift $99D_2B_1$ and Tift 454 are dwarf, early-maturing grain types that share genes from Tift $23D_2$. Tift $99D_2B_1$ has rust and *Pyricularia* leaf spot resistance alleles and Tift 454 has nematode resistance and pollen fertility restorer capability. This population was developed by Dr. Jeffrey P. Wilson (USDA-ARS (retired), Tifton, GA) and was provided to Fort Valley State University as part of the collaborative pearl millet project funded by USDA-National Institute of Food and Agriculture, Grant # GEOX-2008–02595 to Dr. Bharat Singh (retired). The population used for sequencing was a set of 184 RILs at the F_7 generation.

Plant DNA Preparation for Sequencing

Plant leaf tissue was collected from 1.5-mo-old seedlings raised in the greenhouse. The tissue was lyophilized for 8 h and then genomic DNA was isolated with a DNeasy 96 Plant Kit (6) (Qiagen Inc., Valencia, CA). The DNA was quantified to contain 10 ng μL^{-1} per sample and 50 μL of each sample from 184 lines was sent in 96-deep well plates to the Genomic Diversity Facility at Cornell University in Ithaca, NY, for GBS marker development. Each plate included DNA samples from both parents and TifGrain 102 in random wells as well as a random blank well containing only water.

Genotyping-by-Sequencing

Library preparation and sequencing were performed by the Genomic Diversity Facility at Cornell University, Ithaca, NY. Genomic complexity reduction was performed with the *Ape*KI restriction enzyme (recognition site G/CWCG) and samples were sequenced in 96-plex on an Illumina HiSeq 2000 (Illumina Inc., San Diego, CA). One hundred and eighty-four RILs were sequenced; five samples yielded less than 5000 reads each and were excluded from further analysis. Single-nucleotide polymorphisms were called from the remaining 179 lines.

Single-Nucleotide Polymorphism Calls

Raw FASTQ files were processed to SNP calls using the GBS pipeline in TASSEL (version 4.3.6) (Glaubitz et al., 2014). Reads were aligned against ~19,000 contigs of pearl millet genome sequence provided by the Pearl Millet Genome Sequencing Consortium (Varshney et al., unpublished data, 2015) using Bowtie2 (Langmead and Salzberg, 2012). To see the effect of using a reference genome for sequence alignment, we also generated a map that did not use the reference genome to align tags. This pipeline was identical to that used for the reference-based map, except that tags were aligned to each other using the UNEAK (Lu et al., 2013) filter in TASSEL version 5.2.1.15 (commands --UTagCountToTagPairPlugin and --UTagPairToTOPMPlugin).

Initial Map Generation and Ordering

Map creation was done in three iterative steps. All scripts and parameters used in this process are included in Supplemental File S1. First, high-quality SNP calls were selected by filtering for those with at least 70% coverage across RILs and with allele frequencies between 0.25 and 0.75. Sites showing >12% heterozygosity were removed as probable paralog misalignments. All RILs showing >50% missing data or >10% heterozygosity were then removed. Potential outcrosses were also identified by using the fraction of rare alleles (minor allele frequency ≤ 0.05) in each RIL to define a normal distribution. All RILs whose value had <1% probability after Benjamini–Hochberg correction (Benjamini and Hochberg, 1995) were excluded. Filtering resulted in a dataset of 146 RILs and 17,400 SNPs.

To order SNPs, heterozygous calls were first set to "missing" and the genotypes were transformed to numerical equivalents using TASSEL (Bradbury et al., 2007). The SNPs were then clustered using the hclust() function in R (R Core Team, 2014). The cluster trees were split at various levels and linkage disequilibrium (LD) among clusters were manually inspected for the smallest level that clearly separated all seven linkage groups (LGs). Each LG was separated and markers were imputed on the basis of nearest-neighbor analysis; only perfectly cosegregating SNPs were used to impute each other. Redundant markers were then removed and 100 bootstraps of each LG were made by randomly resampling the RILs. Each bootstrap was ordered independently using MSTmap (Wu et al., 2008) and the results were merged, keeping the 95% most stable markers. Marker position was fine-tuned with the ripple() function in R/qtl

(Broman et al., 2003). Map distances were also estimated with R/qtl using the Kosambi mapping function.

Using the first iteration map as a base, a second iteration map was built by testing all original SNPs' linkage to one of the first iteration LGs; only those with an R^2 value $\geq$0.6 were taken as being anchored. These SNPs were then filtered for those with calls in at least 60 RILs, minor allele frequencies $\geq$0.25, and heterozygosity $\leq$0.05. Each LG was then bootstrapped and reordered with MSTmap (Wu et al., 2008) as above, then cleaned with PLUMAGE (Spindel et al., 2014) and rippled with R/qtl (Broman et al., 2003).

The marker order from this second iteration was used to impute marker genotypes using FSFHap (Swarts et al., 2014), which uses a hidden Markov model to impute genotypes in bi-parental populations. The imputed genotypes were again bootstrapped, ordered, and cleaned as above. To get the final map, we put the original genotypes into the order identified by the imputed map, cleaned them with PLUMAGE (Spindel et al., 2014), and estimated map distances with R/qtl (Broman et al., 2003).

Comparison to the Consensus Map

Map LGs were numbered and oriented on the basis of their correlation to the consensus map of Rajaram et al. (2013). Three hundred and five SSR primer sequences from the consensus map were aligned to the contigs used in SNP-calling using Bowtie2 (Langmead & Salzberg, 2012). The position of the SSR was taken at the contig's location in the consensus map. Its corresponding location in the current map was calculated as the consensus location of the SNPs originating from each contig. Linkage groups were numbered and oriented on the basis of their best correlation to the consensus map.

Anchoring Sequencing Tags

Sequence tags were anchored based on the dominant-marker method of Elshire et al. (2011), where each tag's distribution across RILs was compared to the SNPs from the final map using a binomial test of segregation. The SNPs whose best *p*-value was below 0.0001 were considered to be anchored; all others were discarded. In this way, 333,567 tags (out of 9.33 million in total) were anchored to the genetic map.

Test for Segregation Distortion

In a recombinant inbred population, the expected segregation ratio for any given marker should be 50% from each parent. Each of the 16,650 markers was tested for segregation distortion using a χ^2 test with 1 degree of freedom at $\alpha = 0.05$ using Microsoft Excel (Microsoft Corp., Redmond, WA). The critical χ^2 value was adjusted for multiple testing using the false discovery rate procedure of Benjamini and Hochberg (1995).

Field Layout for RILs

One hundred and seventy-nine RILs, two parental lines, and TifGrain 102 were sown in single-row plots that were 1.5 m long and 0.7 m apart, with a 1-m alley between plots at Fort Valley Agricultural Research Station farm (32°31′N, 83°53′W) on 16 July 2013. The experimental design was a randomized complete block with three replications. Grain sorghum was planted as a border around the experiment plot.

Measurement of Phenotypic Traits

Multiple phenotypic traits were scored among the 179 RILs for one season; two traits are reported here as test cases for the linkage map and the others will be reported in a separate publication. The number of days from sowing to 50% flowering was recorded for each plot. The 50% flowering date was decided when at least half of the plants in each plot had started flowering and half of the panicles on individual plants had exserted stigmas. *Pyricularia* leaf spot infestation had occurred under natural conditions because of the rainy, humid weather during our experiment. Ten plants in each plot were visually scored and given an average rating for that plot. The disease manifestation was very clear and conspicuous on all RILs and on their parents. The disease scoring was per ICRISAT using a 1–9 scale (Thakur et al., 2011), where 1 indicates no disease and 9 indicates complete death of the plant from disease. As disease progress depends on growth stage, some of the late-maturing lines showed a different disease response from other lines. Therefore, disease scores for plants that vary widely in maturation rates were adjusted based on their maturity and disease progress curve (Wilson and Hanna, 1992).

Broad-sense heritability (H^2) for these two traits was calculated using the Type III means squares from PROC GLM in SAS version 9.3 (SAS Institute, Cary, NC) using the formula:

$$H^2 = MSG\,(MSG + MSR + MSE)^{-1},$$

where *MSG* is the mean square for genotype, *MSR* is the mean square for replication, and *MSE* is the mean square error. The denominator is thus the total phenotypic variance.

QTL Analysis

Before performing QTL mapping, the raw flowering time scores were transformed using Box–Cox transformation as coded in the MASS package for R (Venables and Ripley, 2002) because the data were not normally distributed. The optimal value for λ was determined by testing all values between −2.0 and +2.0 in steps of 0.008 (500 steps in total); $\lambda = -1.335$ had the highest log-likelihood value and so was used for transformation.

Mapping of QTLs was then performed using single-marker regression as coded in the R/qtl package for R (Broman et al., 2003). The phenotypes used were the raw disease scores and Box–Cox transformed flowering time scores and the genotypes were the final linkage map. We also smoothed the logarithm of odds (LOD) scores in 5-cM sliding windows, taking the maximum LOD within each window to identify peaks of association more clearly.

Results

Genotyping-by-Sequencing Analysis and SNP Calling

One hundred and eighty-four RILs were sequenced, which generated a total of 438.6 million reads that, with the exception of five failed samples, are spread mostly evenly across the samples (Supplemental Fig. S1). As the two parental lines and their commercial hybrid were sequenced twice, we had high-depth coverage of 5,964,312 reads for Tift $99D_2B_1$, 3,077,835 reads for Tift 454, and 4,704,803 reads for TifGrain 102. The mean read depth across all successful samples was 2.2 ± 0.95 million (CV = 0.43), the pass filter rate was 88%, and the median was 2.16 million reads per sample. The total number of good reads among 179 RILs was 387,339,046; the individual with the fewest reads had 390,047 and the individual with the highest reads had 4,854,147. Raw reads were then converted to SNP calls using the TASSEL-GBS pipeline (Glaubitz et al., 2014; see the Methods section for the parameters used). Since pearl millet does not yet have a published reference genome, we aligned the reads against a collection of ~19,000 scaffolds and contigs kindly provided by the Pearl Millet Genome Sequencing Consortium (Varshney et al., unpublished data 2015). During SNP calling, 88.8% of the total reads were mapped to scaffolds and contigs from the pearl millet genomic sequence.

Relationship between Founder Lines and the RILs

The sequencing data from the RILs shows a close relationship to the parental lines used in this study (Supplemental Fig. S2). The RILs cluster around the theoretical value of 50% relatedness to each parent (0.5, 0.5). As expected, a few individuals show up to 80% or higher relatedness to one parent or the other, as a result of stochastic gamete sampling during meiosis.

Identification of Polymorphic Markers

Calling of SNPs resulted in >500,000 raw SNPs, many of which were false positives caused by sequencing errors. Filtering for SNPs with calls in at least 60% of lines and with minor allele frequencies above 0.25 resulted in ~24,000 high-quality polymorphic SNPs. To filter out false SNPs from paralogous sequences aligning together, we also removed sites that showed >12% heterozygosity (the 12% cutoff was determined empirically by looking at the distribution of heterozygous sites). We then also removed any RILs with >50% missing data or >10% heterozygosity. This resulted in using 17,400 sites across 146 RIL individuals, where missing data was in the range of 0.5 to 43.4% per individual (median 8.7%) and 12.7% missing across the entire dataset.

Construction of the Genetic Map

We built the linkage map in a series of iterative steps. First a subset of very high-quality "core" SNPs was taken and used to define LGs and an initial ordering. We obtained a core set of 1192 unique markers with stringent filtering (see Materials and Methods) covering seven LGs. Once the core map was assembled, lower-quality SNPs were anchored to LGs and the ordering was repeated. This second ordering was then used to impute all the markers using FSFHap (Swarts et al., 2014), which uses a hidden Markov model to impute individuals in biparental populations. These iterative steps added another 15,458 markers to the map across 150 RILs. A heat map of LD shows clear clusters between seven different LGs corresponding to the seven pearl millet chromosomes (Supplemental Fig. S3). These three iterative steps resulted in the final genetic map of 16,650 SNPs in 1191 unique recombination bins (Fig. 1). Our LGs were then renumbered and reoriented to match those of Rajaram et al. (2013), which were based on mapping to the amplicons used to generate their map. For comparison, we also created a map without using the genomic contigs to anchor the sequencing reads. Instead, GBS reads were aligned against each other with the UNEAK filter (Lu et al., 2013; see Materials and Methods); all other steps were identical. The final genome-free map included 4900 markers. This is still a significant number of markers, and if no genomic data were available, they would still form a useful map. However, the >3× higher number of markers from the original map demonstrate the value of having genomic sequences to align against, even if these sequences are not assembled into a reference genome.

Expanding the Genetic Map with Sequencing Tags

After obtaining the final map, we then anchored sequencing tags (the same 64-bp reads used in the GBS pipeline) to it. We used the dominant-marker method of Elshire et al. (2011), which anchored 333,567 (out of 9.33 million) tags onto the genetic map.

To gauge the accuracy of mapping, we looked at the overlap between sequencing reads and the SNPs they generated. There is only partial overlap between the set of tags that give rise to SNPs in the map and the tags that were mapped on their own (Fig. 2). This is mostly because (i) a tag can still be anchored even if any SNPs it gives rise to are filtered out, (ii) some tags are caused by presence–absence variation and so will not give rise to SNPs themselves but can still be anchored to nearby SNPs, and (iii) sequencing errors can make a tag appear unique, so even if a good SNP can be called in one part of a tag, an error elsewhere in it makes the tag too rare for the binomial segregation test to work.

Of the tags that do overlap, ~87% of them anchor to within 10 cM of their associated SNP, many of them to the exact same recombination bin. This implies that our mapping accuracy is high and the positions of the reads should be very close to their true position.

In total, 16,650 SNPs and 333,567 additional tags were distributed on all seven LGs (Table 1). Overall, 20.10% of data were missing for 16,650 loci across 150 individuals. The missing values across these loci ranged from 4 to 55 per locus (2.6–36.6%). In the final map

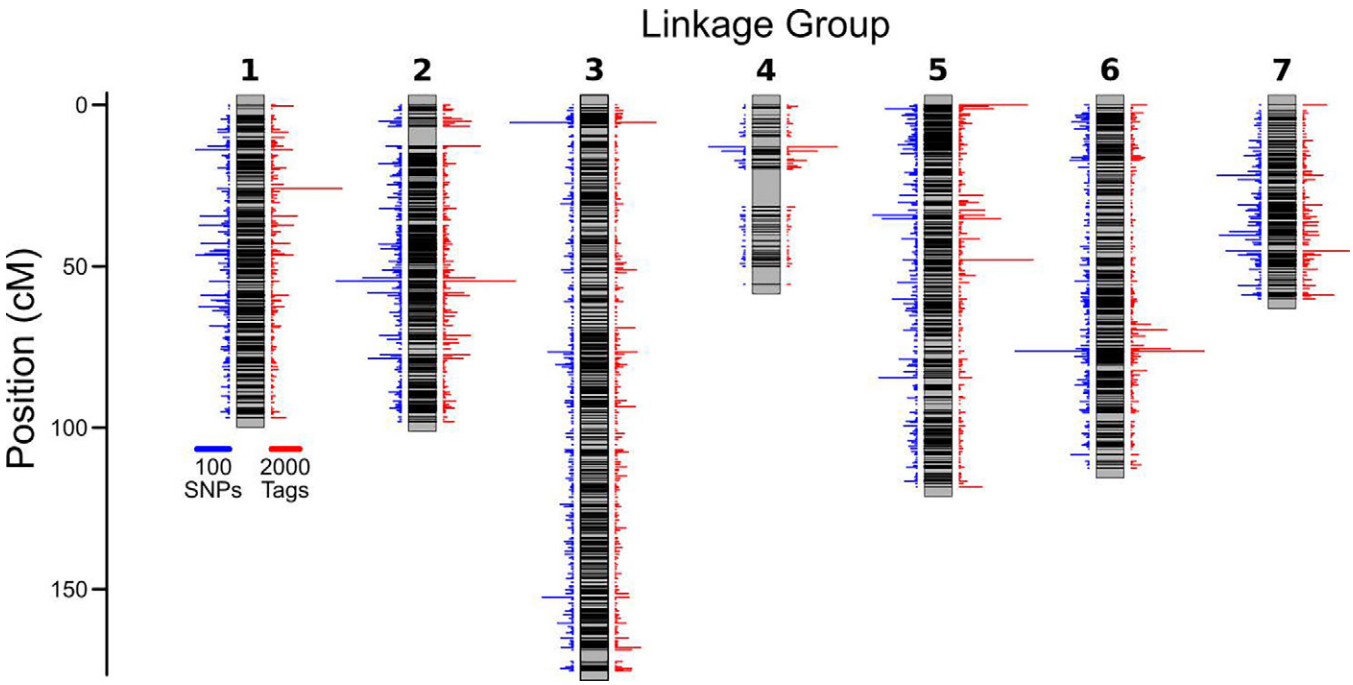

Fig. 1. Linkage map of pearl millet developed using genotyping-by-sequencing (GBS) markers. Gray bars represent each linkage group, with black bands showing the unique map locations on each linkage group. (Linkage groups were extended past the first and last markers for visual clarity.) Blue bars to the left of each group are proportional to the number of single-nucleotide polymorphisms (SNPs) at each location; red bars to the right show the number of sequencing tags mapped to each location.

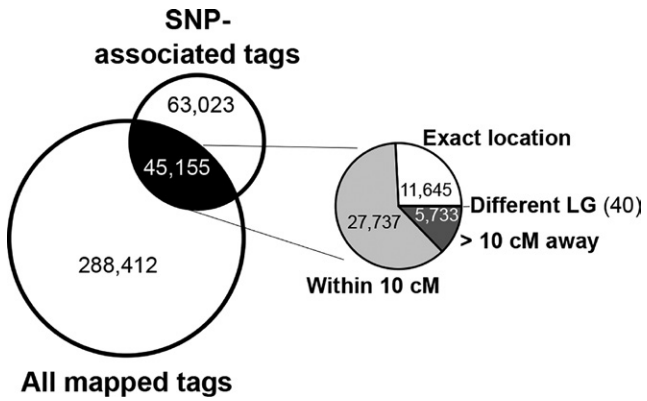

Fig. 2. Single-nucleotide polymorphism–tag concordance in pearl millet. The overlap between sequencing reads and the SNPs they generated is shown. Of the tags that do overlap, ~87% of them anchor to within 10 cM of their associated SNP, many of them to the exact same recombination bin.

Table 1. Linkage group (LG) statistics with marker and tag density in pearl millet.

LG	Length (cM)	Markers	Anchored tags	Marker density per cM	Tag density per cM
LG 1	96.9	2509	49,855	25.89	514.50
LG 2	98.1	2986	62,754	30.44	639.69
LG 3	175.3	3000	61,367	17.11	350.07
LG 4	55.5	513	15,789	9.24	284.49
LG 5	118.3	3085	58,902	26.08	497.90
LG 6	112.6	2449	46,685	21.75	414.61
LG 7	60.0	2108	38,215	35.13	636.92
Total	716.7	16650	333,567	23.23	465.42

containing 16,650 SNPS and 333,567 tags, the average densities of SNP markers and of additional tags across all chromosomes were 23.23 and 465.42 tags per cM, respectively, covering the genome length of 716.7 cM. The marker densities per LG were spread from a minimum of 9.24 cM^{-1} on LG 4 to a maximum of 35.13 cM^{-1} on LG 7. When only unique linkage bins were counted, marker densities are in the range of 0.81 bins cM^{-1} (LG 4) to 1.90 bins cM^{-1} (LG 2) with an average density of 1.66 bins cM^{-1} across the genome.

Among all the chromosomes, LG 4 had the fewest markers and tags, whereas LG 5 had the highest number of SNPs and LG 2 had the highest number of tags. The highest numbers of SNP markers were anchored and ordered on LG 5, which had 3085 markers in the final map.

Comparison to an Existing Pearl Millet Consensus Map

Rajaram et al. (2013) recently produced a consensus pearl millet map by combining SSR data from four different linkage populations. The current map was matched to the consensus using 305 SSR primer pairs from Rajaram et al. (2013). Of these, 191 aligned uniquely while being in the correct relative orientations and distances apart; 16 aligned concordantly but at multiple locations, one aligned discordantly (incorrect orientation), and 97 either did not align at all or had only partial alignments (meaning one primer was aligned but not both) (Supplemental Fig. S4).

The lengths of each chromosome in the current map ranged from 55.5 cM (LG 4) to 175.3 cM (LG 3), with an average length of 102.3 cM per chromosome. In the core map made from 1192 sites, the average intermarker distances between two adjacent markers ranged from 0.52 cM (LG 2) to 1.23 cM (LG 4). The inter-marker distance of 0.01 cM was least on LG 2 and LG 3, and the maximum distance of 11.71 cM was observed on LG 4, with an overall average marker distance of 0.67 cM across the entire core genetic map. There were three intervals [5.52 cM (LG 4), 6.14 cM (LG 2), and 11.71 cM (LG 4)] that were more than 5 cM between neighboring markers. The rest of the intervals were below 5 cM distances, which reflects that more than 99% of the map had small spacings between neighboring markers. Linkage Group 3 here appeared to be extended longer than LG 3 of the consensus map, whereas LG 7 fairly represented its counterpart. The rest of the chromosomes were shorter than the consensus map. We also compared our LG lengths with four LGs (LGA, LGB, LGC, and LGG) in the GBS-based SNP map by Moumouni et al. (2015), which revealed that our map was extended in LG 1 and LG 6, but it was shorter in LG 2, LG 4, and LG 7. These extensions are very common in telomeric regions, which also have been observed in DArT-based maps of pearl millet (Supriya et al., 2011). The maps reported in all the previous studies used Haldane mapping functions, whereas our map used the Kosambi mapping function distances, which could be one reason for discrepancies in map lengths.

The current map appears to have roughly equal coverage to the consensus map but with some caveats. Many individual markers and some groups of markers were localized to different locations in the two maps. Some of this may be a result of technical error, such as misalignment of the primer sequences or misassembly caused by sequencing errors. Some of the discrepancies are probably biological, however, and represent small- and large-scale structural variations between the populations used to make the two maps. Pearl millet has significant genetic diversity (Oumar et al., 2008), to the point that only a single SSR from the consensus map was mappable in all four of its input populations (Rajaram et al., 2013). In that context, finding significant variation with a fifth population (the one used in this study) is to be expected here as well.

Segregation Distortion

Of the 16,650 mapped SNP markers, 6652 (39.41%) showed significant segregation distortion after adjustment for multiple comparisons (Benjamini and Hochberg, 1995). Most of these distorted markers occurred in large linkage blocks. Linkage Group 3 showed the greatest amount of segregation distortion, with nearly the entire LG (98.67% of mapped markers) significantly biased in favor of Tift $99D_2B_1$. In contrast, LG 1 was also highly distorted (80.71% of mapped markers) but was biased in favor of the other parent, Tift 454. Linkage Group 2 also had several highly distorted blocks, biased toward the Tift $99D_2B_1$ parent, and LG 6 had one major linkage block biased toward Tift $99D_2B_1$. Linkage Group 4 showed the least segregation distortion, with only two markers (0.39%) distorted (Supplemental Fig. S5).

Mapping Leaf Spot Resistance and Days to 50% Flowering Traits

The linkage map developed in this experiment was used in regression analysis to identify QTLs for two phenotypic traits: leaf spot resistance and days to 50% flowering. The H^2 was quite high for these two traits. For days to flowering, $H^2 = 0.7578$. For Box–Cox transformed days to flowering, H^2 drops to 0.5110. For raw disease score, $H^2 = 0.7978$; for the adjusted disease score, it is 0.9163.

The two parents showed significant differences for these two traits in the field, whereas their F_1 hybrid, TifGrain 102, showed good leaf spot resistance, similar to Tift $99D_2B_1$, but flowered later, similar to Tift 454 (Supplemental Table S1). R/qtl results identified leaf spot resistance loci on LG 5 and LG 7 with significant threshold LOD values above 3.0 (Fig. 3, Table 2). These QTLs were found to be minor, with phenotypic variance of 4.83 to 5.05% and a favorable allelic effect (lower disease score) from Tift 454. Two more QTLs for leaf spot resistance with a favorable allelic effect from Tift $99D_2B_1$ were located on LG 2 and LG 3, having LOD values just above 2.0. A significant QTL for flowering time with a LOD value above 3.0 was located on the upper arm of LG 2, which explained 6.0% of the phenotypic variance, with the positive allelic effect (later flowering) coming from the parent Tift 454 (Fig. 3, Table 2). The rest of the QTLs for flowering time were detected below LOD 3.0 on LG 1, LG 5, and LG 7 with 0.49 to 4.75% phenotypic variance and positive additive effects coming from the other parent, Tift $99D_2B_1$.

Discussion

Importance of a High-Density Genetic Map and Its Comparison to Existing Maps

Next-generation sequencing technologies have revolutionized marker discovery and enabled high-throughput plant genotyping through several new marker platforms like GBS (Poland and Rife, 2012). Genotyping-by-sequencing is a cost-effective and efficient system for developing high-density markers, which are concurrently discovered and genotyped in larger mapping populations (He et al., 2014). These abundant markers, coupled with well-developed bioinformatics, facilitate the development of dense molecular linkage maps. In this experiment, we had high-depth coverage and abundant high-quality SNPs.

Ever since the first pearl millet genetic map was made from RFLPs in 1994 (Liu et al., 1994), there has been a continuous effort to improve such maps with greater marker density and uniformity. Many of these maps had large gaps in the distal regions of chromosomes, probably caused by very high recombination rates, so most improvement efforts targeted these

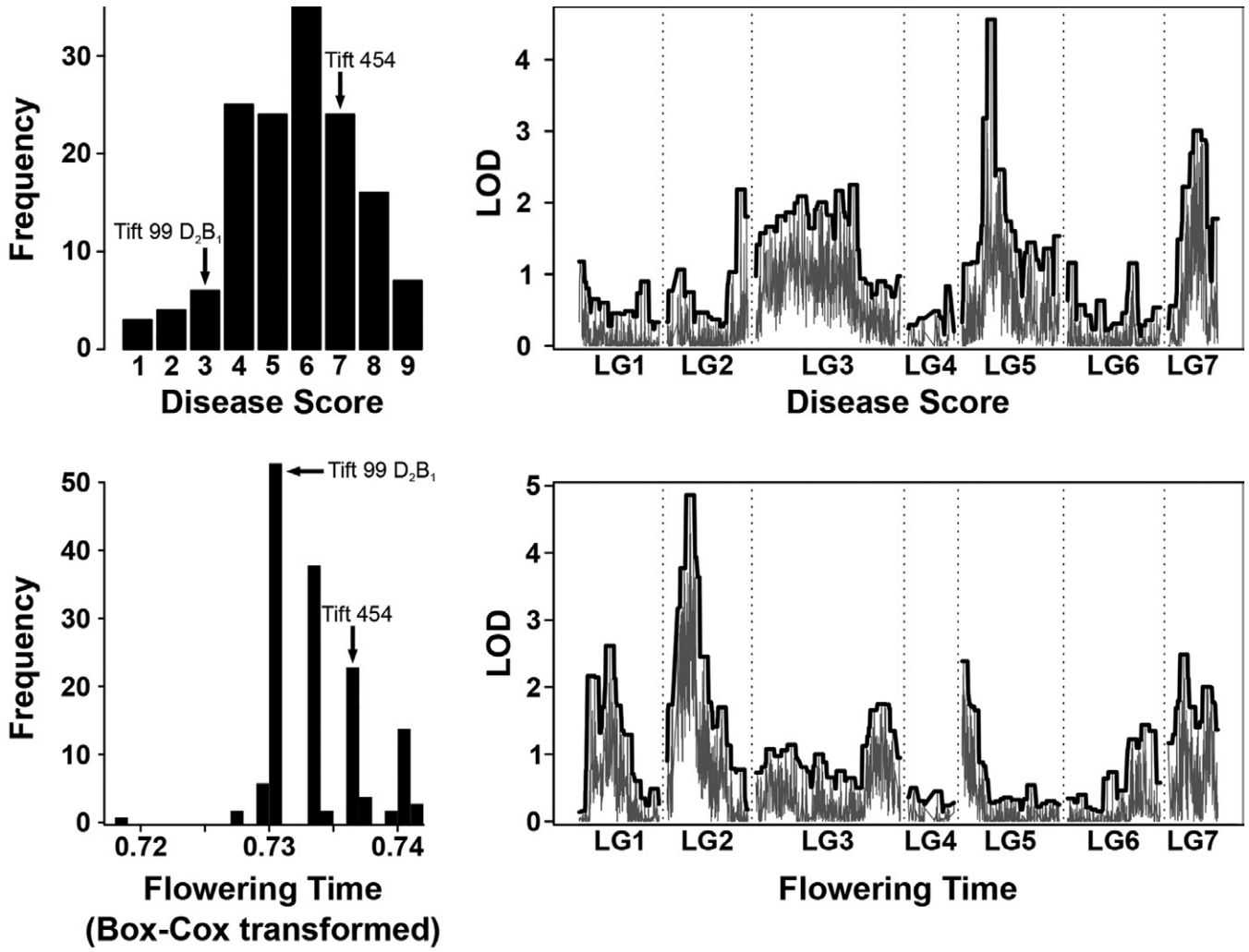

Fig. 3. Quantitative trait locus (QTL) mapping in pearl millet. Quantitative trait loci were identified for *Pyricularia* leaf spot (top) and flowering time (bottom). The distribution of phenotype scores is shown on the left and logarithm of odds (LOD) values from single-marker regression (using R/qtl; Broman et al., 2003) are shown on the right. The light gray traces show the raw LOD scores, which vary depending on the different levels of missing data for each marker. The solid black line shows a smoothed trace, taking the maximum value in 5-cM sliding windows.

regions. For example, expressed sequence tag and genomic SSRs were added by Senthilvel et al. (2008), DArT markers by Supriya et al. (2011), and gene-based SNP and conserved intron spanning primers markers by Sehgal et al. (2012). Despite these efforts, large gaps of more than 30 cM were still present in most of the distal regions of chromosomes. The most recent consensus map (Rajaram et al., 2013) used expressed sequence tag SSRs and also contained large gaps in the range of 18 to 27 cM on every chromosome. Using NGS, Moumouni et al. (2015) made a GBS map from 314 nonredundant SNPs. Although the map developed by Moumouni et al. (2015) was uniform in coverage with no interval greater than 20 cM in length and only 10 intervals larger than 10 cM, it still had a maximum gap of 19.7 cM on LG 2 that corresponds to 3.0% of the total map length. The linkage map in the current study has a maximum gap of 11.71 cM on LG 4, equating to 1.6% of total map length and representing a significant improvement in reduced gap size.

To our knowledge, this map represents the densest genetic map in pearl millet so far. It contains 16,650 SNPs and 333,567 sequence tags covering all seven LGs. Here, we report an average density of 1.66 linkage bins cM^{-1} and 23.23 SNP cM^{-1} in the final map, which significantly surpasses the 0.51 SNP cM^{-1} of the next-densest map (Moumouni et al., 2015). The linkage map constructed in this study is more dense, uniform, and highly saturated, which is reflected through smaller marker spacing (<5cM) than any previously published pearl millet genetic map. The mean distance between two neighboring markers is the least: 0.6 cM compared to 2.1 cM (Moumouni et al., 2015) and 3.7 cM (Supriya et al., 2011) published so far. The small marker spacings on every chromosome with several cosegregating redundant markers shows that with the exception of LG 4, this map is extensive and reasonably uniform in genome coverage. Therefore, our map complements the recent pearl millet linkage map developed by Moumouni et al. (2015), which contains 2809 GBS markers from 85 F_2 progenies. At 716.7 cM

Table 2. Quantitative trait loci for flowering time and *Pyricularia* leaf spot disease identified in a pearl millet recombinant inbred line population.

Flowering time						
LG§	Location	SNP interval	Peak SNP	LOD	Variance	Additive effect†
	cM				%	d
1	32.3	S1_1423–S1_3590	S1_2196	2.61	3.03	1.8
2	23.3	S2_1896–S2_2803	S2_2223	4.86	6.00	−2.0
5	0.0	S5_0012–S5_1669	S5_0451	2.38	4.75	1.5
7	14.4	S7_0244–S7_2067	S7_0774	2.48	0.49	1.3
Leaf spot disease						
LG	Location	SNP interval	Peak SNP	LOD	Variance	Effect‡
	cM				%	
2	85.0	S2_7773–S2_8331	S2_7983	2.18	1.78	−0.6
3	114.2	S3_0019–S3_4763	S3_4544	2.25	1.82	−0.5
5	30.5	S5_2145–S5_4145	S5_3817	4.56	4.83	0.9
7	30.5	S7_0738–S7_3864	S7_2251	3.01	5.05	0.9

† A negative sign indicates that the later flowering allele was derived from the Tift 454 parent, whereas a positive sign indicates that the allele from parent Tift $99D_2B_1$ delayed flowering.

‡ 1 indicates no disease symptoms; 9 indicates complete susceptibility. A negative sign indicates that the Tift $99D_2B_1$ allele increased resistance (lower score), whereas a positive sign indicates that the Tift 454 allele increased resistance.

§ LG, linkage group; SNP, single-nucleotide polymorphism; LOD, logarithm of odds.

in total length, our map is slightly longer than that of Moumouni et al. (2015) (640.6 cM), which used an F_2 population and thus is expected to be shorter. The high quality and quantity of markers found in this experiment were possible because of high-depth coverage for two parents in calling SNPs and the large number of RILs (150 individual progenies) available after stringent filtering.

Genetic map distances are relative distances based on recombination frequencies, unlike physical maps, which estimate actual distances in base pairs. The map distances and positions of individual markers can vary from one mapping population to the other depending on the parents used in the initial cross and type of mapping population used. Our map distances are represented through the Kosambi mapping function although previous studies used Haldane mapping function, which may explain some of the differences in map length. The comparison between our map and the previous consensus map has shown some agreement but also some discrepancies. For example, some markers are at different locations in the two maps (Supplemental Fig. S4). Our total map length is shorter than the total map lengths reported by Supriya et al. (2011), Sehgal et al. (2012), and Rajaram et al. (2013). Although some of these disagreements are probably caused by technical differences in the ways each map was prepared, many of the disagreements are probably a result of biological differences, including a few large linkage blocks that may represent actual translocations in one population relative to the other. Given the quality of LD within the current map (Supplemental Fig. S3), any major discrepancies are probably caused by structural variations originating from the germplasm used in the current study.

High-density maps developed through GBS not only support functional genomics through connecting phenotype to genotype but they also serve as reference maps in fundamental studies like genome sequencing to refine, order, and assemble scaffolds and contigs of pseudochromosomes (Poland and Rife, 2012; Ward et al., 2013). This map has been partly used in contig assembly of the pearl millet genome sequencing project led by ICRISAT. Furthermore, a well-ordered dense map allows a comparative genome structure analysis and informs about important evolutionary changes (Gale and Devos, 1998). This linkage map will also help other researchers working on mapping traits in pearl millet. For example, others can directly use the 64-bp tags used to develop SNPs in this study for the same purpose. The resulting datasets can be used to make genetic maps, mine alleles, and characterize diverse pearl millet accessions.

Imputation of SNP Data

The major drawback of sequencing-based genotyping technology is the large amount of missing data; GBS is no exception. Several approaches can be used to reduce these missing data, such as sequencing to high depth, filtering to save only high-quality data, or performing imputation of haplotypes (Poland and Rife, 2012). We used careful filtering to achieve a missing rate of 20.1% in our final (unimputed) genetic map, although one of the steps used to generate it included imputing another version down to only ~3% missing data. We focus our analyses on the unimputed map because imputation can introduce biases. Both the imputed map and unimputed map are available in Supplemental File S2.

The Parents and Their Ancestry

The parents of this mapping population, Tift $99D_2B_1$ and Tift 454, are dwarf, early-maturing grain types. Both parents carry the recessive dwarfing gene d_2, which lies on LG 4 (Parvathaneni et al., 2013). We discovered very few markers on LG 4 compared to other LGs. Since the two parents inherited genomic regions from Tift $23D_2B_1$, it is possible that this LG has few SNPs because of a region of common descent around the dwarfing gene d_2. The male-sterile A-line Tift $99D_2A_1$ and Tift 454 are the parents of the commercial hybrid known as TifGrain 102 (Hanna et al., 2005a, 2005b). Tift $99D_2B_1$ was selected for resistance to rust and is derived from Tift $89D_2$ and also shares some genomic regions with Tift $23D_2$ (Hanna and Wells, 1993; Hanna et al., 2005b). It also appears to have resistance to *Pyricularia* leaf spot. Tift 454 was derived from an interspecific cross between pearl millet Tift $23D_2A_1$ and a napiergrass [*Cenchrus purpureus* (Schum.) Morrone]–pearl millet hybrid and carries at least one A′ chromosome from the napiergrass parent (Hanna et al., 2005a). Tift 454 is resistant to nematodes [*Meloidogyne areniaria* (Neal) Chitwood and *Meloidogyne incognita* Kofoid & White] and has male-fertility restorer capability in A_1 cytoplasm.

Regions of significant segregation distortion have been reported in previous genetic mapping studies in pearl millet (Qi et al., 2004; Rajaram et al., 2013; Moumouni et al., 2015), so it is not surprising that they were detected in this population as well. However, we found two regions of segregation distortion in this population that each spans nearly an entire LG (LG 1 and LG 3) (Supplemental Fig. S5). Such large regions of segregation distortion have not been reported in previous studies in pearl millet. Linkage Group 1 and 3 also had the highest number of discrepancies in comparison to the map of Rajaram et al. (2013) (Supplemental Fig. S4). According to Hanna et al. (2005a), the parental line Tift 454 ($2n = 2x = 14$) carries at least one pair of chromosomes from the A′ genome of napiergrass in place of a homologous chromosome pair from the A genome of pearl millet. The evidence here, namely nearly complete segregation distortion of two entire LGs along with a large number of map discrepancies, suggests that Tift 454 may in fact carry two napiergrass chromosomes. Linkage Groups 1 and 3 appear to represent two A–A′ chromosome pairs. Though the A and A′ genomes are reported to be homologous (Hanna, 1990), it is possible that the rate of recombination between the napiergrass and pearl millet chromosomes is lower than the rate of recombination between chromosomes originating from the same species. Evidence reported by Techio et al. (2006) suggests that the A and A′ chromosomes are likely to be homeologous rather than homologous. In addition, meiotic irregularities have also been reported in triploid (Techio et al., 2006) and hexaploid (Paiva et al., 2012) pearl millet–napiergrass hybrids. Interestingly, most of LG 1 is biased in favor of the Tift 454 parent, suggesting that the A′ chromosome transmits more frequently, whereas LG 3 is biased in favor of Tift $99D_2B_1$, suggesting reduced frequency of transmitting this A′ chromosome. Though the RILs were selected randomly, the bias toward one parent or the other may also be an artifact of unintentional selection based on characteristics such as pollen viability or seed set under the selfing bag.

Utility of the Map in Tagging Disease Resistance Loci and Flowering Traits

The high-density GBS-based linkage map was validated by mapping QTLs for flowering time and *Pyricularia* leaf spot resistance. The leaf spot resistance loci identified in this study indicate that this trait is controlled by several loci from different LGs. In a previous study, a random amplified polymorphic DNA marker was identified as being associated with *Pyricularia* leaf spot resistance but was not assigned to any LG (Morgan et al., 1998). Research from ICRISAT, India, has mapped a leaf spot resistance QTL to LG 4 in a RIL population based on 'ICMB841-P3' × '863B-P2' (Dr. R K Srivastava, personal communication, 2015), which was also associated with stover quality traits and was introgressed into the hybrid seed parent 'ICMA/B 95222' (Nepolean et al., 2006). ICMA/B 95222 is the seed parent of hybrid 'HHB 146' released from Chaudhary Charan Singh Haryana Agricultural University, Hisar (Dwivedi et al., 2012). The present study also identified a significant flowering time QTL on LG2, the same LG where the *PHYC* gene was significantly associated with flowering time (Saïdou et al., 2009) and several other flowering and drought tolerance QTLs were reported (Yadav et al., 2002, 2004, 2011b; Bidinger et al., 2007; Sehgal et al., 2012). Primer sequences from these studies were used to compare their location on our map (Supplemental Table S2). Based on their marker position in our map, our flowering time QTL locations do not correspond to the locations reported in these previous studies. However, it will be interesting to explore the potential candidate genes once the complete pearl millet genome sequence is available. The amount of phenotypic variation explained by these QTLs was low for these two traits despite the fact that H^2 was quite high [$H^2 = 0.511$ for days to flower (transformed) and $H^2 = 0.916$ for adjusted disease score] and were comparable to other studies (Yadav et al., 2002, 2004; Nepolean et al., 2006; Dwivedi et al., 2012; Sehgal et al., 2015). The heritabilities for flowering trait were reported to be in the range of 47 to 94% in the previous studies (Yadav et al., 2004; Sathya et al., 2014; Sehgal et al., 2015). One explanation is that numerous QTLs, each with a very small effect, contribute to these traits (Yadav et al., 2003). Additionally, the QTL detection method used here (single-marker regression in R/qtl software) may underestimate individual QTL effects (Lander and Botstein, 1989; Zeng, 1994). The relative lack of markers on LG 4 [because of the apparent descent of much of LG 4 in both parents of the RIL populations from a common ancestor (*Pyricularia* leaf spot-susceptible Tift $23D_2B_1$)], where a leaf spot resistance QTL was previously identified, could also explain why we did not identify this QTL. When these traits were mapped using a genetic map made without genomic sequences, many of the QTLs were still identifiable but appeared to have lost some significance, probably because they lacked the SNPs that were in tightest linkage with the causal locus (Supplemental Fig. S6). This also reflects that having genome sequence information will enhance QTL mapping. The QTL results reported here are based on a single season of data, so they will need to be validated by additional studies in more environments. Even so, the examples presented here demonstrate the utility of this genetic map for identifying QTLs.

This study used a RIL population, which allowed for a replicated field screen for disease response and flowering time. Such replication increases the accuracy of phenotyping, despite having only one season of data, and is not possible with F_2 populations. Additionally, seeds of the RIL population can be distributed to other researchers to map other traits of interest without the need to reconstruct the genetic map.

Conclusions

Pearl millet is considered a minor crop in the United States and Europe, so development of genetic and genomic resources in this crop has lagged behind other

cereals. It is, however, an essential staple crop in many parts of the world, particularly developing countries in hot semiarid and arid regions where little else will grow. Thus improvement of this crop is critically important for food security in these areas and may become critical to currently more favorable areas if global climate change continues unabated. Tools like molecular markers can facilitate rapid advances in crop improvement but the development of such resources was a formidable task in pearl millet until the advent of NGS-based markers like GBS. In this experiment, GBS markers were successfully used to make a high-density map containing 16,650 SNPs and 333,567 additional sequence tags, which is the densest map yet created in pearl millet. High-density linkage maps provide better map resolution and abundant genomic resources. A recombinant inbred mapping population created from an elite germplasm was used to construct this map so that useful and repeatable variation can be studied using this resource. These genome-wide markers can be used for applications such as marker-assisted selection, genomic selection, diversity studies, and comparative genomic analyses. The results will also help to identify and tag several traits related to disease and nematode resistance in pearl millet. In addition, understanding the genes underlying important traits in pearl millet, such as drought tolerance and nitrogen use efficiency, could help to improve these traits in other crops.

Supplemental Information Available

Supplemental material is available with this article.

Supplemental Table S1: Leaf spot scores and days to 50% flowering for parental lines, their F_1 hybrid (TifGrain 102), and the RIL population.

Supplemental Table S2: Marker positions on the current map based on basic local alignment search tool (BLAST) hits.

Supplemental Figure S1: Read depth per sample. The number of sequencing reads matched to each individual is shown in order of increasing read depth. Gray bars represent RILs that were sequenced once each; black bars are the two parents (Tift $99D_2B_1$ and Tift 454) and their F_1 hybrid (TifGrain 102), which were sequenced twice (once on each plate). Five samples were removed because they had <5000 mapped reads each.

Supplemental Figure S2: Relatedness of RILs to parents. RILs (pale blue circles) are plotted according to their degree of relatedness relative to both parents. Darker colors indicate where points have stacked on top of each other.

Supplemental Figure S3: Linkage disequilibrium heatmap. Linkage disequilibrium (r^2) heat map shown across the final genetic map for all pairwise SNP comparisons. Single-nucleotide polymorphisms are arrayed in map order on both the *x* and *y* axes and each point shows the pairwise linkage disequilibrium between a set of SNPs. The size of each block is proportional to the number of SNPs in each LG; the small number of SNPs in LG 4 is probably caused by a large chromosomal segment that is identical in both parents that is likely to have been inherited from their common ancestor, Tift $23D_2B_1$.

Supplemental Figure S4: Comparison to existing pearl millet consensus map. Simple sequence repeat primer sequences from an existing SSR consensus pearl millet map (Rajaram et al., 2013) were aligned against the contigs used to call SNPs in the current map. The linkage map from this study (left-hand side, dark gray) is compared with the SSR consensus map (right-hand side, light gray). Black bars indicate markers that could be identified in both maps, with colored lines connecting each marker position to its corresponding position in the other map. Solid lines indicate markers that map to matching linkage groups (LGs); dashed lines indicate markers that map to different LGs; and line color indicates the LG in the current SNP-based map. Although many markers show good correlation, many also show inconsistent ordering. Large blocks of inconsistent markers may represent large translocations, such as between the consensus LG 1 and our LG 4 and between the consensus LG 6 and our LG 1.

Supplemental Figure S5: Map of segregation distortion in the pearl millet RIL population. Markers shaded in red are biased in favor of Tift $99D_2B_1$; markers shaded blue are biased in favor of Tift 454. Markers with a χ^2 value greater than the critical value are significantly distorted.

Supplemental Figure S6: Effect of genomic sequence on mapping quality. Quantitative trait locus maps for flowering time and leaf spot disease compared between the full linkage map and the map made without aligning sequences to the pearl millet genomic data.

Supplemental File S1 (Text files): All scripts and parameters used in the current experiment.

Supplemental File S2 (Excel files): Genotypic data for 16,550 loci used for final map creation and phenotypic data for leaf spot disease and flowering traits in 179 RILs.

Acknowledgments

We thank Eli Rodgers-Melnick for part of the R code for rippling LGs, the Pearl Millet Genome Sequencing Consortium for use of prepublication contigs and the Genomic Diversity Facility (Cornell University) for helpful advice on GBS analysis. We thank Ms. Chrisdon B. Bonner for helping us to improve the quality of the manuscript. We are also grateful for funding support received from the capacity building project by USDA-National Institute of Food and Agriculture Grant # GEOX-2008-02595, NSF Grant IOS-1238014, the University of Georgia, and the USDA-ARS. The authors declare that they have no competing interests related to the contents of this manuscript. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

References

Andrews, D.J., and K.A. Kumar. 1992. Pearl millet for food, feed, and forage. Adv. Agron. 48:89–139. doi:10.1016/S0065-2113(08)60936-0

Andrews, D.J., J.F. Rajewski, and K.A. Kumar. 1993. Pearl millet: New feed grain crop. In: J. Janick and J.E. Simon, editors, New crops. John Wiley & Sons, New York. p. 198–208.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc., B 57(1):289–300.

Bennett, M.D., and J.B. Smith. 1976. Nuclear DNA amounts in angiosperms. Phil. Trans. Roy. Soc. Lond. Ser. B. 274:227–274. doi:10.1098/rstb.1976.0044
Bidinger, F.R., and C.T. Hash. 2004. Pearl millet. In: H.T. Nguyen and A. Blum, editors, Physiology and biotechnology integration for plant breeding. Marcel Dekker, New York. p. 221–233.
Bidinger, F.R., T. Nepolean, C.T. Hash, R.S. Yadav, and C.J. Howarth. 2007. Quantitative trait loci for grain yield in pearl millet under variable post flowering moisture conditions. Crop Sci. 47:969–980. doi:10.2135/cropsci2006.07.0465
Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/bioinformatics/btm308
Bramel-Cox, P.J., K. Anand Kumar, J.H. Hancock, and D.J. Andrews. 1992. Sorghum and millets for forage and feed. In: D.A.V. Dendy, editor, Sorghum and millets, chemistry and technology. American Association of Cereal Chemists, St. Paul, MN. p. 325–364.
Broman, K.W., H. Wu, S. Sen, and G.A. Churchill. 2003. R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. doi:10.1093/bioinformatics/btg112
Burton, G.W., and J.B. Powel. 1968. Pearl millet breeding and cytogenetics. Adv. Agron. 20:49–89. doi:10.1016/S0065-2113(08)60854-8
Chemisquy, M.A., L.M. Giussani, M.A. Scataglini, E.A. Kellogg, and O. Morrone. 2010. Phylogenetic studies favour the unification of *Pennisetum, Cenchrus* and *Odontelytrum* (Poaceae): A combined nuclear, plastid and morphological analysis, and nomenclatural combinations in *Cenchrus*. Ann. Bot. (Lond.) 106:107–130. doi:10.1093/aob/mcq090
Collins, V.P., A.H. Cantor, A.J. Pescatore, M.L. Straw, and M.J. Ford. 1997. Pearl millet in layer diets enhances egg yolk n-3 fatty acids. Poult. Sci. 76:326–330. doi:10.1093/ps/76.2.326
Cunningham, D.L., and B.D. Fairchild. 2012. Broiler production systems in Georgia costs and returns analysis. University of Georgia Cooperative Extension B 1240. http://www.caes.uga.edu/departments/agecon/extension/pubs/documents/B1240_3.PDF (accessed 11 Mar. 2016).
Dahlberg, J.A., J.P. Wilson, and T. Snyder. 2004. Sorghum and pearl millet: Health foods and industrial products in developed countries. In*:* Alternative Uses of Sorghum and Pearl Millet in Asia. CFC Technical Paper No. 34. Proceedings of an Expert Meeting, Patancheru, Andhra Pradesh, India. 1–4 July 2003. ICRISAT, India, p. 42–49.
Davis, A.J., N.M. Dale, and F.J. Ferreira. 2003. Pearl millet as an alternative feed ingredient in broiler diets. J. Appl. Poult. Res. 12:137–144. doi:10.1093/japr/12.2.137
Devos, K.M., T.S. Pittaway, A. Reynolds, and M.D. Gale. 2000. Comparative mapping reveals a complex relationship between the pearl millet genome and those of foxtail millet and rice. Theor. Appl. Genet. 100:190–198. doi:10.1007/s001220050026
Dwivedi, S.L., H. Upadhyaya, S. Senthilvel, C. Hash, K. Fukunaga, X. Diao, et al. 2012. Millets: Genetic and genomic resources. Plant Breed. Rev. 35:247–375. doi:10.1002/9781118100509.ch5.
Durham, S. 2003. New strain of pearl millet. Agric. Res. Mag. 51:19.
Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379. doi:10.1371/journal.pone.0019379
Farrell, D.J. 2005. Matching poultry production with available feed resources: Issues and constraints. World Poultry Sci. J. 61:298–307. doi:10.1079/WPS200456
Gale, M.D., and K.M. Devos. 1998. Plant comparative genetics after 10 years. Science 282:656–659. doi:10.1126/science.282.5389.656
Gale, M.D., K.M. Devos, J.H. Zhu, S. Allouis, M.S. Couchman, H. Liu, et al. 2005. New molecular marker technologies for pearl millet improvement. Int. Sorghum Millets Newslett. 42:16–22.
Ganal, M.W., T. Altmann, and M.S. Röder. 2009. SNP identification in crop plants. Curr. Opin. Plant Biol. 12:211–217. doi:10.1016/j.pbi.2008.12.009
Garcia, A.R., and N.M. Dale. 2006. Feeding of unground pearl millet to laying hens. J. Appl. Poult. Res. 15:574–578. doi:10.1093/japr/15.4.574
Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, et al. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9(2):e90346. doi:10.1371/journal.pone.0090346
Gulia, S.K., J.P. Wilson, J. Carter, and B.P. Singh. 2007. Progress in grain pearl millet research and market development. In: J. Janik and A. Whipkey, editors, Issues in new crops and new uses. ASHS Press, Alexandria, VA. p. 196–203.
Gupta, A.K., K.N. Rai, P. Singh, V.L. Ameta, K. Suresh, A.K. Jayalekha, et al. 2015. Seed set under high temperatures during flowering period in pearl millet (*Pennisetum glaucum* L.). Field Crops Res. 171:41–53. doi:10.1016/j.fcr.2014.11.005
Gupta, S.K., R. Sharma, K.N. Rai, and R.P. Thakur. 2012. Inheritance of foliar blast resistance in pearl millet (*Pennisetum glaucum* L. (R.) Br.). Plant Breed. 131:217–219. doi:10.1111/j.1439-0523.2011.01929.x
Hanna, W.W. 1990. Transfer of germplasm from the secondary to the primary gene pool in *Pennisetum*. Theor. Appl. Genet. 80:200–204. doi:10.1007/BF00224387
Hanna, W.W., and H.D. Wells. 1989. Inheritance of *Pyricularia* leaf spot resistance in pearl millet. J. Hered. 80:145–147.
Hanna, W.W., and H.D. Wells. 1993. Registration of parental line Tift $89D_2$, rust resistant pearl millet. Crop Sci. 33:361–362. doi:10.2135/cropsci1993.0011183X003300020048x
Hanna, W., J. Wilson, and P. Timper. 2005a. Registration of pearl millet parental line Tift 454. Crop Sci. 45:2670. doi:10.2135/cropsci2005.0171
Hanna, W., J. Wilson, and P. Timper. 2005b. Registration of pearl millet parental lines Tift $99D_2A_1/B_1$. Crop Sci. 45:2671. doi:10.2135/cropsci2005.0172
Hash, C.T., and P.J. Bramel-Cox. 2000. Marker applications in pearl millet. In: B.I.G. Haussmann, H.H. Geiger, D.E. Hess, C.T. Hash, and P. Bramel-Cox (eds.) Training manual for a seminar held at IITA, Ibadan, Nigeria, 16–17 Aug. 1999. ICRISAT, Patancheru, India. p. 112–127
He, J., X. Zhao, A. Laroche, Z.X. Lu, H. Liu, and Z. Li. 2014. Genotyping-by-sequencing (GBS), an ultimate marker assisted selection (MAS) tool to accelerate plant breeding. Front. Plant Sci. 5:484. doi:10.3389/fpls.2014.00484
Howarth, C.J., G.P. Cavan, K.P. Skot, R.W.H. Layton, C.T. Hash, and J.R. Witcombe. 1994. Mapping QTLs for heat tolerance in pearl millet. In: J.R. Witcombe and R.R. Duncan, editors, Use of molecular markers in sorghum and pearl millet breeding for developing countries. Overseas Development Administration, London, UK. p. 80–85.
Jauhar, P.P., and W.W. Hanna. 1998. Cytogenetics and genetics of pearl millet. Adv. Agron. 64:1–26. doi:10.1016/S0065-2113(08)60501-5
Krawczak, M. 1999. Informativity assessment for biallelic single nucleotide polymorphisms. Electrophoresis. 20:1676–1681.
Kumar, S., T.W. Banks, and S. Cloutier. 2012. SNP discovery through next-generation sequencing and its applications. Int. J. Plant Genomics 2012: 831460. doi:10.1155/2012/831460.
Lander, E.S., and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.
Langmead, B., and S. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9:357–359. doi:10.1038/nmeth.1923
Lee, D., W. Hanna, G.D. Buntin, W. Dozier, P. Timper, and J.P. Wilson. 2004. Pearl millet for grain. University of Georgia. http://extension.uga.edu/publications/files/pdf/B%201216_3.PDF (accessed 11 Mar. 2016).
Liu, C.J., J.R. Witcombe, T.S. Pittaway, M. Nash, C.T. Hash, C.S. Busso, and M.D. Gale. 1994. An RFLP-based genetic-map of pearl millet (*Pennisetum glaucum*). Theor. Appl. Genet. 89:481–487. doi:10.1007/BF00225384.
Lu, F., A.E. Lipka, R.J. Elshire, J.C. Glaubitz, J.H. Cherney, M.D. Casler, et al. 2013. Switchgrass genomic diversity, ploidy and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9:E1003215. doi:10.1371/journal.pgen.1003215
Maman, N., S.C. Mason, and D.J. Lyon. 2006. Nitrogen rate influence on pearl millet yield, nitrogen uptake, and nitrogen use efficiency in Nebraska. Commun. Soil Sci. Plant Anal. 37(1–2):127–141. doi:10.1080/00103620500406112
Mammadov, J., R. Aggarwal, R. Buyyarapu, and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. Int. J. Plant Genomics 2012: 728398: doi:10.1155/2012/728398.
Martel, E., N.D. De, S. Siljak-Yakovlev, S. Brown, and A. Sarr. 1997. Genome size variation and basic chromosome number in pearl millet and fourteen related *Pennisetu*m species. Heredity 88:139–143. doi:10.1093/oxfordjournals.jhered.a023072
Morgan, R.N., J.P. Wilson, W.W. Hanna, and P. Ozais-Akins. 1998. Molecular markers for rust and pyricularia leaf spot disease resistance in pearl millet. Theor. Appl. Genet. 96:413–420. doi:10.1007/s001220050757
Moumouni, K.H., B.A. Kountche, M. Jean, C.T. Hash, Y. Vigouroux, B.I.G. Haussmann, et al. 2015. Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach. Mol. Breed. 35:5. doi:10.1007/s11032-015-0212-x
Muchow, R.C. 1988. Effect of nitrogen supply on the comparative productivity of maize and sorghum in a semi-arid tropical environment leaf growth and leaf nitrogen. Field Crops Res. 18(1):131–143. doi:10.1016/0378-4290(88)90056-1.

Nambiar, V.S., J.J. Dhaduk, N. Sareen, T. Shahu, H. Shah, and R. Desai. 2011. Potential functional implications of pearl millet (*Pennisetum glaucum*) in health and disease. J. Appl. Pharm. Sci. 01(10):62–67.

Nepolean, T., M. Blümmel, A.G. Bhasker Raj, V. Rajaram, S. Senthilvel, and C.T. Hash. 2006. QTLs controlling yield and stover quality traits in pearl millet. Int. Sorghum Millets Newslett. 47:149–152.

Oumar, I., C. Mariac, J.-L. Pham, and Y. Vigouroux. 2008. Phylogeny and origin of pearl millet (*Pennisetum glaucum* [L.] R. Br.) as revealed by microsatellite loci. Theor. Appl. Genet. 117:489–497. doi:10.1007/s00122-008-0793-4

Paiva, E.A.A., F.O. Bustamante, S. Barbosa, A.V. Pereira, and L.C. Davide. 2012. Meiotic behavior in early and recent duplicated hexaploid hybrids of napier grass (*Pennisetum purpureum*) and pearl millet (*Pennisetum glaucum*). Caryologia 65(2):114–120. doi:10.1080/00087114.2012.709805

Parvathaneni, R.K., V. Jakkula, F.K. Padi, S. Faure, N. Nagarajappa, A.C. Pontaroli, et al. 2013. Fine-mapping and identification of a candidate gene underlying the d_2 dwarfing phenotype in pearl millet, *Cenchrus americanus* (L.) Morrone. G3. 3:563–572. doi:10.1534/g3.113.005587.

Peacock, J.M., P. Soman, R. Jayachandran, A.V. Rani, C.J. Howarth, and A. Thomas. 1993. Effect of high soil surface temperature on seedling survival in pearl millet. Exp. Agric. 29:215–225. doi:10.1017/S0014479700020664

Pedraza-Garcia, F., J.E. Specht, and I. Dweikat. 2010. A new PCR-based linkage map in pearl millet. Crop Sci. 50:1754–1760. doi:10.2135/cropsci2009.10.0560

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-sequencing for plant breeding and genetics. Plant Gen. 5:92. doi:10.3835/plantgenome2012.05.0005

Qi, X., T.S. Pittaway, S. Lindup, H. Liu, E. Waterman, F.K. Padi, et al. 2004. An integrated genetic map and a new set of simple sequence repeat markers for pearl millet, *Pennisetum glaucum*. Theor. Appl. Genet. 109:1485–1493. doi:10.1007/s00122-004-1765-y

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org (accessed 11 Mar. 2016).

Rajaram, V., T. Nepolean, S. Senthilvel, R.K. Varshney, V. Vadez, R.K. Srivastava, et al. 2013. Pearl millet [*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. BMC Genomics 14:159. doi:10.1186/1471-2164-14-159

Saïdou, A.A., C. Mariac, V. Luong, J.L. Pham, G. Bezançon, and Y. Vigouroux. 2009. Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. Genetics 182:899–910. doi:10.1534/genetics.109.102756

Sathya, M., P. Sumathi, N. Senthil, S. Vellaikumar, and A. John Joel. 2014. Genetic studies for yield and its component traits in RIL population of pearl millet (*Pennisetum glaucum* [L.] R. Br.). Electronic J. Plant Breeding 5(2):322–326.

Sehgal, D., L. Skot, R. Singh, R.K. Srivastava, S.P. Das, J. Taunk, et al. 2015. Exploring potential of pearl millet germplasm association panel for association mapping of drought tolerance traits. PLoS ONE 10:e0122165. doi:10.1371/journal.pone.0122165

Sehgal, D., V. Rajaram, V. Vadez, C.T. Hash, and R.S. Yadav. 2012. Integration of gene-based markers in pearl millet genetic map for identification of candidate genes underlying drought tolerance quantitative trait loci. BMC Plant Biol. 12:9. doi:10.1186/1471-2229-12-9

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

Spindel, J., M. Wright, C. Chen, J. Cobb, J. Gage, S. Harrington, et al. 2014. Bridging the genotyping gap: Using genotyping by sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. 126(11):2699–2716. doi:10.1007/s00122-013-2166-x

Sullivan, T.W., J.H. Douglas, D.J. Andrews, P.L. Bond, J.D. Hancock, P.J. Bramel-Cox, et al. 1990. Nutritional value of pearl millet for food and feed. In: G. Ejeta, E.T. Mertz, L. Rooney, R. Schaffert, and J. Yohe (eds) Proceedings of the International Conference on Sorghum Nutritional Quality, 26 Feb.–1 March 1990, Purdue University, West Lafayette, IN. Purdue University, West Lafayette, IN. p. 83–94

Supriya, A.,S. Senthilvel, T. Nepolean, K. Eshwar, V. Rajaram, R. Shaw, et al. 2011. Development of a molecular linkage map of pearl millet integrating DArT and SSR markers. Theor. Appl. Genet. 123:239–250. doi:10.1007/s00122-011-1580-1

Swarts, K., L. Huihui, J.A. Romero Navarro, A. Dong, M.C. Romay, S. Hearne, et al. 2014. Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. Plant Gen. 7. doi:10.3835/plantgenome2014.05.0023

Techio, V.H., L.C. Davide, and A.V. Pereira. 2006. Meiosis in elephant grass (*Pennisetum purpureum*), pearl millet (*Pennisetum glaucum*) (Poaceae, Poales) and their interspecific hybrids. Genet. Mol. Biol. 29(2):353–362. doi:10.1590/S1415-47572006000200025

Thakur, R.P., R. Sharma, and V.P. Rao. 2011. Screening techniques for pearl millet diseases. Information Bulletin No. 89. ICRISAT, Patancheru, Andhra Pradesh, India.

Thomas, G., T. Bhavna, and N.C. Subrahmanyam. 2000. Highly repetitive DNA sequences of pearl millet: Modulation among *Pennisetum* species and cereals. J. Plant Biochem. Biotechnol. 9:17–22. doi:10.1007/BF03263077

Timper, P., J.P. Wilson, A.W. Johnson, and W.W. Hanna. 2002. Evaluation of pearl millet grain hybrids for resistance to *Meloidogyne* spp. and leaf blight caused by *Pyricularia grisea*. Plant Dis. 86:909–914. doi:10.1094/PDIS.2002.86.8.909

Vadez, V., T. Hash, F.R. Bidinger, and J. Kholova. 2012. Phenotyping pearl millet for adaptation to drought. Front. Physiol. 3:1–12. doi:10.3389/fphys.2012.00386

Venables, W.N., and B.D. Ripley. 2002. Modern applied statistics with S. 4th ed. Springer, New York.

Ward, J.A., J. Bhangoo, F. Fernández-Fernández, P. Moore, J.D. Swanson, R. Viola, et al. 2013. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14:2. doi:10.1186/1471-2164-14-2

Wilson, J.P., and W.W. Hanna. 1992. Effects of gene and cytoplasm substitutions in pearl millet on leaf blight epidemics and infection by *Pyricularia grisea*. Phytopathology 82:839–842. doi:10.1094/Phyto-82-839

Wilson, J.P., G.W. Burton, H.D. Wells, J.D. Zongo, and I.O. Dicko. 1989. Leaf spot, rust and smut resistance in pearl millet landraces from central Burkina Faso. Plant Dis. 73:345–349. doi:10.1094/PD-73-0345

Wimpee, C.F., and J.R.Y. Rawson. 1979. Characterisation of the nuclear genome of pearl millet. Biochim. Biophys. Acta 562:192–206. doi:10.1016/0005-2787(79)90165-5

Wu, Y., P.R. Bhat, T.J. Close, and S. Lonardi. 2008. Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 4(10): E1000212. doi:10.1371/journal.pgen.1000212

Yadav, R.S.,C.T. Hash, F.R. Bidinger, G.P. Cavan, and C.J. Howarth. 2002. Quantitative trait loci associated with traits determining grain and stover yield in pearl millet under terminal drought stress conditions. Theor. Appl. Genet. 104:67–83. doi:10.1007/s001220200008

Yadav, O.P., and K.N. Rai. 2011. Hybridization of Indian landraces and African elite composites of pearl millet results in biomass and stover yield improvement under arid zone conditions. Crop Sci. 51:1980–1987. doi:10.2135/cropsci2010.12.0731

Yadav, O.P., K.N. Rai, I.S. Khairwal, B.S. Rajpurohit, and R.S. Mahala. 2011a. Breeding pearl millet for arid zone of north-western India: Constraints, opportunities and approaches. All India Coordinated Pearl Millet Improvement Project, Jodhpur, India.

Yadav, R.S., F.R. Bidinger, C.T. Hash, Y.P. Yadav, O.P. Yadav, S.K. Bhatnagar, et al. 2003. Mapping and characterization of QTL × E interactions for traits determining grain and stover yield in pearl millet. Theor. Appl. Genet. 106:512–520. doi:10.1007/s00122-002-1081-3.

Yadav, R.S., C.T. Hash, F.R. Bidinger, K.M. Devos, and C.J. Howarth. 2004. Genomic regions associated with grain yield and aspects of post-flowering drought tolerance in pearl millet across stress environments and tester background. Euphytica 136:265–277. doi:10.1023/B:EUPH.0000032711.34599.3a

Yadav, R.S., D. Sehgal, and V. Vadez. 2011b. Using genetic mapping and genomics approaches in understanding and improving drought tolerance in pearl millet. J. Exp. Bot. 62:397–408. doi:10.1093/jxb/erq265

Zeng, Z.-B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.