Mol Breeding (2015) 35:5
DOI 10.1007/s11032-015-0212-x

Construction of a genetic map for pearl millet, *Pennisetum glaucum* (L.) R. Br., using a genotyping-by-sequencing (GBS) approach

K. H. Moumouni · B. A. Kountche · M. Jean · C. T. Hash · Y. Vigouroux · B. I. G. Haussmann · F. Belzile

Received: 9 June 2014 / Accepted: 11 December 2014 / Published online: 15 January 2015

Abstract Pearl millet is the main component of traditional farming systems and a staple grain in the diet of sub-Saharan Africa and India. To facilitate breeding work in this crop, a genetic map consisting of single nucleotide polymorphism (SNP) markers was constructed using an F_2 population of 93 progenies, from a wild × cultivated pearl millet cross. We used a modified genotyping-by-sequencing (GBS) protocol involving two restriction enzymes (*Pst*I–*Msp*I) and PCR amplification with primers including three selective bases to generate 3,321 SNPs. Of these, 2,809 high-quality SNPs exhibited a minor allele frequency ≥0.3. In total, 314 non-redundant haplotypes and 85 F_2 individuals were used to construct a genetic map spanning a total distance of 640 cM. These SNPs were evenly distributed over seven linkage groups ranging considerably in size (62–123 cM). The average density for this map was 0.51 SNP/cM, and the average interval between SNP markers was 2.1 (±0.6) cM. Finally, to establish bridges between the linkage groups of this and previous maps, 19 SSR markers were examined for polymorphism between the parents of this population. We could only tentatively suggest a correspondence between four of our linkage groups and those of previous maps. Overall, GBS enabled us to quickly produce a genetic map with a density and uniformity of markers greater than previously published maps. The availability of such a map will be useful for the identification of genomic regions associated with *Striga* resistance and other important agronomic traits.

Keywords Pearl millet · Genetic map · SNP · Genotyping-by-sequencing

Electronic supplementary material The online version of this article (doi:10.1007/s11032-015-0212-x) contains supplementary material, which is available to authorized users.

K. H. Moumouni · M. Jean · F. Belzile (✉)
Département de Phytologie, Faculté des Sciences de l'Agriculture et de l'Alimentation and Institut de Biologie Intégrative et des Systèmes, Université Laval, Quebec, QC G1V 0A6, Canada
e-mail: francois.belzile@fsaa.ulaval.ca

B. A. Kountche · C. T. Hash · B. I. G. Haussmann
International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), BP 12404, Niamey, Niger

B. A. Kountche · Y. Vigouroux
Institut de Recherche pour le Développement (IRD), 34394 Montpellier Cedex 5, France

B. A. Kountche
Faculté d'Agronomie, Université Abdou Moumouni, BP 10960, Niamey, Niger

B. I. G. Haussmann
Institute of Plant Breeding, Seed Science and Population Genetics (350a), University of Hohenheim, 70599 Stuttgart, Germany

Introduction

Pearl millet *Pennisetum glaucum* (L.) R. Br., [syn. *Cenchrus americanus* (L.) Morrone] is one of the most important cereal crops in the arid and semi-arid regions of India and sub-Saharan Africa (Chemisquy et al. 2010; Hash et al. 2000; Senthilvel et al. 2008; Sehgal et al. 2012). Ranked sixth cereal crop in the world, pearl millet is grown as a source of nutrient-rich food grain for humans as well as a feed/forage/fodder crop for livestock (Jauhar et al. 2006). In these regions, pearl millet is the staple food of more than 90 million people, so it plays a crucial role in food security (Gowda and Rai 2006). Despite its low grain yield (600 kg ha^{-1}) due to abiotic stresses (drought, low soil fertility) and biotic stresses (*Striga*, millet head miner, downy mildew and other diseases), this crop provides nutritious grains that are highly recommended for certain diets (Rai et al. 2012).

Pearl millet is a diploid species with seven pairs of chromosomes and is native to Africa (Manning et al. 2011). The genome size is about 2.4 pg, comparable to that of maize. The plant is sexual, hermaphrodite and cross-pollinated. These characteristics contribute to a high level of heterozygosity.

To minimize yield losses and to facilitate the development of improved varieties, research efforts are being conducted in several areas including resistance to *Striga* and downy mildew (Kountche et al. 2013), drought and salinity tolerance (Bidinger et al. 2007, Sehgal et al. 2012; Sharma et al. 2014), cytoplasmic male sterility, use of dwarfing genes and the use of heterosis or hybrid vigor (Tostain and Marchais 1993; Presterl and Weltzien 2003; Dhuppe et al. 2005). In addition, molecular tools are increasingly available for pearl millet and have proven their efficiency in other species including sorghum, rice, maize and cowpea (Haussmann et al. 2004; Omanya et al. 2004; Gurney et al. 2006; Amusan et al. 2008; Yoshida and Shirasu 2009).

A genetic map represents a key tool for genetic studies and should contain enough information to identify and position genes/QTLs that control traits of agronomic interest (Pedraza-Garcia et al. 2010). Liu et al. (1994) constructed the first DNA marker-based genetic map in pearl millet; it consisted of 181 RFLP markers and covered a total distance of 303 cM. An integrated genetic map was built by Qi et al. (2004) by using a combination of both RFLP (353) and SSR (65) markers from two F_2 populations. In the latter case, the total distance covered was 473 cM. For both of these maps, F_2 populations were used to analyze the segregation of markers. However, these maps suffered from a high degree of marker clustering such that many of the marker pairs were very closely linked whereas others were widely separated. Other genetic maps were followed, mostly based on SSR and DArT markers (Senthilvel et al. 2008; Supriya et al. 2011; Rajaram et al. 2013), and despite increasing marker coverage, these maps still suffer from a lack of uniformity in marker coverage. In these maps, many gaps greater than 20 or even 30 cM are still found.

Thus, the production of a dense map providing both extensive and uniform coverage remains a challenge in pearl millet. Furthermore, the development of RFLP and SSR markers requires a fair amount of upstream work to determine which sets of enzymes/probes (RFLPs) or primers (SSRs) are capable of revealing polymorphic loci segregating in any mapping population. Also, these markers are not suitable for analysis on a very large scale as, for the most part, these can only be analyzed one (or very few) at a time on each individual. As for DArT markers, even though a genotyping array with 7,680 clones has recently been developed for pearl millet (Supriya et al. 2011), these are dominant markers and some DArT loci originating from duplicated genomic regions can suffer from inconsistent map locations (Tinker 2013). New genotyping technologies capable of highly parallel analysis would represent a major step forward in this crop.

Recently, a high-throughput and low-cost genotyping method named genotyping-by-sequencing (GBS) has been developed and has proven its efficiency in other crops such as maize and barley (Elshire et al. 2011). The proposed approach is simple and suitable for rapidly generating high-density genetic maps. GBS can yield hundreds to thousands of SNPs (single nucleotide polymorphisms) without the need for any prior characterization of candidate loci. A few protocols have been described to date including the original protocol involving the enzyme *Ape*KI (Elshire et al. 2011), demonstrated in maize and barley, and an alternative protocol involving enzymes *Pst*I and *Msp*I, demonstrated in barley and wheat (Poland et al. 2012). The choice of enzymes is a key factor in determining the degree of complexity reduction that is achieved. More recently, Sonah et al. (2013) showed that it was

possible to further optimize the degree of complexity reduction and thus increase both the number and quality of markers obtained, by using PCR primers including selective bases during the preparation of GBS libraries. In addition, appropriate sequencing and genetic analysis tools are available to provide its efficiency to GBS method.

In the present study, we chose to explore the usefulness of such a GBS approach and demonstrate that it is possible to quickly produce a genetic map densely populated with SNP markers for pearl millet using such an approach. The availability of such a genetic map will certainly provide breeders and geneticists with a much-wanted tool to identify genomics regions associated with *Striga* resistance and other important agronomic traits.

Materials and methods

Plant materials

A mapping population, with *Striga* resistance as its primary target trait, consisting of 93 F_2 progenies derived from a cross between "116_11-(PS202-14)-121" (wild millet) and "SOSAT-IBL-197" (cultivated millet) was developed and used in this study. Both parents are from West Africa; 116_11-(PS202-14)-121 is an inbred line derived from six self-pollination cycles using the wild relative accession PS202 (*Pennisetum glaucum* subsp. *monodii*), which is reported to be resistant to *Striga hermonthica* (Del.) Benth. (Wilson et al. 2004). Similarly, at least six inbreeding cycles were conducted using the open-pollinated improved variety SOSAT-C88, which resulted in the development of the SOSAT-IBL-197 inbred line (Haussmann et al. unpublished data). This inbred is highly susceptible to *Striga.*

DNA extraction, library preparation and sequencing

Genomic DNA of individual F_2 plants and their parental lines was extracted from 50 mg fresh leaves, collected from single plants, using the MATAB protocol (a modified CTAB/β-mercaptoethanol method) (Mariac et al. 2006). DNA concentration was adjusted to 20 ng μl^{-1}, and *Pst*I–*Msp*I GBS libraries were prepared following the protocol described by Poland et al. (2012) with the exception that a further complexity reduction was achieved using PCR primers with three selective bases (CAC) as per Sonah et al. (2013). The resulting 95-plex library was sequenced on a single lane of an Illumina HiSeq 2000 at the McGill University and Génome Québec Innovation Centre (Montreal, Canada).

Sequencing data analysis and SNP calling

The Universal Network-Enabled Analysis Kit (UNEAK) pipeline was used to call SNPs from the DNA sequence reads (Lu et al. 2013). UNEAK was used with minor allele frequency (MAF) set at 0.05. An in-house perl script was used to both filter and rescore genotypes that were the output of the UNEAK pipeline (listed in the HapMap.hmp.txt file). A stringent filter was first used to remove and replace by missing data (N) genotypes that had been called with fewer than 11 reads/SNP/individual. A genotype correction step was then carried out to remove "false" heterozygotes using one of the two criteria. If the total read count for a SNP in one individual was between 11 and 40, and the number of reads for the minor allele was ≤ 2, this individual was called homozygous for the major allele. Alternatively, if the total read count was >40 and the proportion of reads for the minor allele was <5 %, the individual was again scored as homozygous for the major allele.

SNP dataset filtering

The corrected HapMap.hmp.txt file resulting from these modifications was then uploaded into TASSEL 3.0_Standalone (Bradbury et al. 2007) for further filtering. SNP loci with more than 20 % missing data were removed, as were those with a minor allele frequency (MAF) <0.3. In addition, individuals with more than 19 % missing data were removed. The SNP markers that were heterozygous in one or both parents were also removed. Chi-square tests were performed on both the allelic (1:1) and genotypic (1:2:1) segregation ratios to assess the amount of segregation distortion. Finally, the data were recoded such that homozygotes for the alleles of SOSAT-IBL-197 were coded "A", homozygotes for the alleles of 116_11-(PS202-14)-121 were coded "B" and heterozygotes were coded "H" using an in-house python script.

Genetic map construction

SNPs with both MAF <0.3 and less than five missing data were grouped using MAPMAKER/EXP software Version 3.0b (Lander et al. 1987) at a LOD score threshold of 3.0 and a maximum distance of 30 cM. The JOIN HAPLOTYPES ON command was used to group markers sharing an identical segregation pattern in haplotype groups, and only the first SNP listed in each haplotype group was conserved to build linkage groups. Within each linkage group, SNP positions were determined using the ORDER command with an informativeness criterion set to a minimum distance of 4 cM and a minimum of 95 % informative individuals. PLACE and TRY commands were used to position SNPs that could not be mapped initially. The FRAMEWORK command produced the framework for each linkage group. Finally, the RIPPLE command was used to test the final SNP order in each linkage group. During all of these procedures, error detection mechanisms were on. The resulting genetic map was then drawn using the MapChart 2.2 software (Voorrips 2002). Distances between SNPs were determined in centimorgans (cM) using the Haldane function. The loci detected have been named with the prefix "TP" (for "tag pair").

Use of SSRs as bridge markers

A set of 19 SSR primer pairs reported by Qi et al. (2004) were assessed for polymorphism between the two parents. Polymorphic SSR markers were analyzed on the entire mapping population. PCR amplifications were performed in 20 µl reactions containing 1.5× PCR buffer, 50 ng genomic DNA, 0.125 µM of M13-tailed forward and reverse primers, 2 mM dNTPs, 0.002 µM IRD700-labeled M13 primer and 0.5 U *Taq* DNA polymerase. Amplifications were performed on a TProfessional Basic Thermocycler (Biometra, Göttingen, Germany) as follows: initial denaturation for 2 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C, 45 s at 72 °C and final extension for 10 min at 72 °C. Loading buffer (1 µl) was added to 5 µl of each PCR, denatured at 94 °C for 5 min and stored at 4 °C for 10 min. Next, 0.8 µl of each mixture was separated on a 6.5 % polyacrylamide gel in a TBE buffer on a LI-COR Global Edition IR^2 DNA Analyzer (LI-COR, Lincoln, USA) set to 1,500 V with a moderate scan speed at 50 °C during 1 h 25 min to 3 h (depending on PCR product size). For each SSR, genotypes were determined for each F_2 progeny and the resulting genotypic data were added to the SNP data to integrate these SSR markers to the genetic map produced using MAPMAKER as described above.

Results

Sequencing read quantity and quality

A single 95-plex GBS library was produced to generate SNP data on the parents and 93 F_2 progenies of the cross 116_11-(PS202-14)-121 × SOSAT-IBL-197. The GBS protocol achieved complexity reduction using both a two-enzyme restriction digest (*Pst*I–*Msp*I) and selective amplification using a PCR primer with three selective bases (CAC) to amplify only a subset of all the ligation products. After high-throughput sequencing, a total of 147M reads were obtained from a single lane and 145M "good" reads (99 % of all reads) met the quality standards of the UNEAK pipeline (Supplementary Figure 1). The number of reads per sample ranged between 0.56M and 2.45M, for an average of 1.53M. A total of 7M distinct 64-bp sequence tags were found in the entire set of reads (Table 1). Of these, 471,904 distinct 64-bp sequence tags, corresponding to 136.43M reads (94.3 % of the total number of reads), were present in sufficient number (≥5 reads in the population) to be used by the UNEAK pipeline for SNP calling (Supplementary Figure 1). On a sample basis, the number of tags ranged between 121K and 344K for an average of 220K tags (Table 1). These sequence tags were examined to identify pairs of tags that were identical or differed by at most a single base. A set of 18,346 "raw" SNPs with a MAF > 0.05 were initially called by the UNEAK pipeline and subsequently rescored

Table 1 Total, average, minimum and maximum number of reads and the corresponding tags analyzed by the UNEAK pipeline

	Number of reads	Sequence tags
Total	144,573,417	7,047,526
Average	1,530,772	220,803
Minimum	560,812	121,772
Maximum	2,446,012	344,004

using our in-house script to remove poorly supported genotype calls and false heterozygotes. Approximately 6 % of reads, 93 % of tags and 56 % of SNP loci were lost during the UNEAK pipeline work to call SNPs (Supplementary Figure 1).

SNP filtering

To ensure that the SNP data were of high-quality, stringent filtering was performed in multiple steps using TASSEL software (Fig. 1). A total of 3,321 SNPs had less than 20 % missing data and a MAF $\geq$ 0.3. In addition, eight individuals forming a distinct group with especially high levels of missing data (>19 %) were removed. These individuals exhibited between 521 and 1,427 missing data. Together, they accounted for more than 40 % of all the missing data, and each of these was characterized by a low coverage (<900K reads/individual). The resulting dataset showed an average of 35 reads per SNP, and the amount of missing data per individual and per SNP both averaged 3 % (Supplementary Table 1).

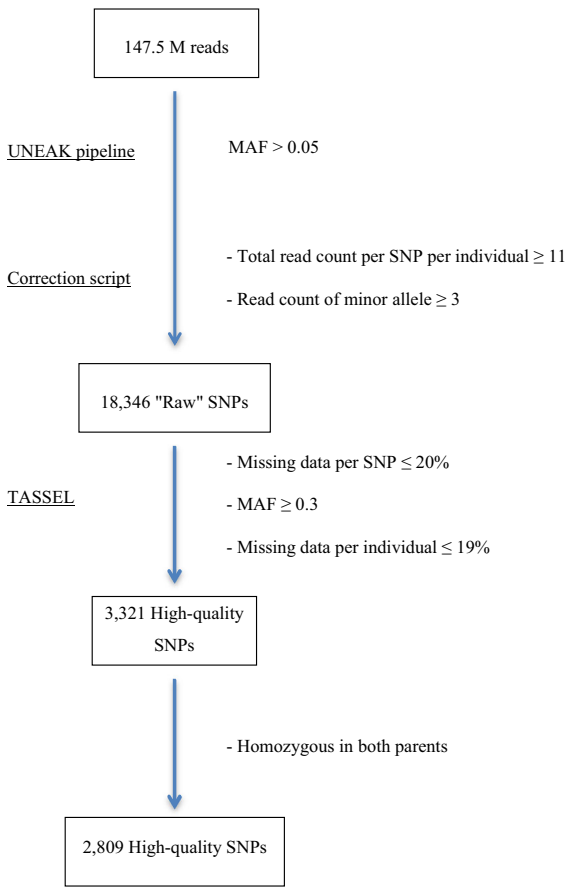

Fig. 1 Flow diagram of the different SNP calling and filtering steps

From this initial set of 3,321 markers, we retained 2,809 SNPs that were homozygous for contrasting SNP alleles in the inbred parental lines of the F_2 mapping population. Of these, 1,381 had no missing data at all, 2,156 had <5 % and 2,541 had <10 % missing data (Supplementary Figure 2). For the purpose of building a high-quality and high-density map, the 2,156 segregating SNPs with <5 % missing data and a MAF $\geq$ 0.3 were used.

Genetic map

Because of the limited size of the mapping population (n = 85), many of the 2,156 SNPs co-segregated in the F_2 population (or only differed due to missing data), suggesting that these loci were in close proximity and had not been resolved by recombination events. These formed a total of 314 non-redundant haplotypes for which a single representative SNP marker (hereafter referred to simply as a SNP) was used for map construction. Despite the use of a MAF $\geq$ 0.3, 11.8 % of the SNPs (37 out of 314) presented a significant segregation distortion (at $P < 0.01$). A total of six distortion regions (SDRs) containing between 1 and 19 SNPs were observed (Supplementary Table 2). Upon grouping (at LOD = 3.0), all SNP markers fell onto one of the seven linkage groups (LGs A–G; see Table 2; Fig. 2). Linkage groups contained between 32 (LG E) and 53 SNPs (LG A). The length of linkage groups ranged between 61.7 cM for LG G and 123.1 cM for LG A, for a total map length of 640.6 cM. The mean distance between SNPs was 2.1 ($\pm$0.6) cM for an average density of 0.51 SNPs/cM. Linkage group G had the smallest distance between markers (1.5 cM) and the highest marker density (0.70 SNPs/cM), while LG D had the largest mean distance between markers (2.7 cM) and the lowest marker density (0.38 SNPs/cM). The seven linkage groups exhibited good coverage and uniformity of the distribution of markers, with the largest distance between markers being 19.7 cM (Fig. 2). The resulting map had no interval greater than 20 cM in length, and only ten intervals were larger than 10 cM.

Table 2 Number of SNPs, map length, mean distance between SNPs and density for each linkage group

Linkage group (LG)	Number of SNPs	Length (cM)	Mean distance between SNPs (cM)	Density (SNP/cM)
A	53	123.1	2.4	0.44
B	52	116.4	2.3	0.45
C	44	88.2	2.1	0.50
D	40	106.6	2.7	0.38
E	32	66.3	2.1	0.48
F	50	78.3	1.6	0.64
G	43	61.7	1.5	0.70
Total	314	640.6	2.1	0.51

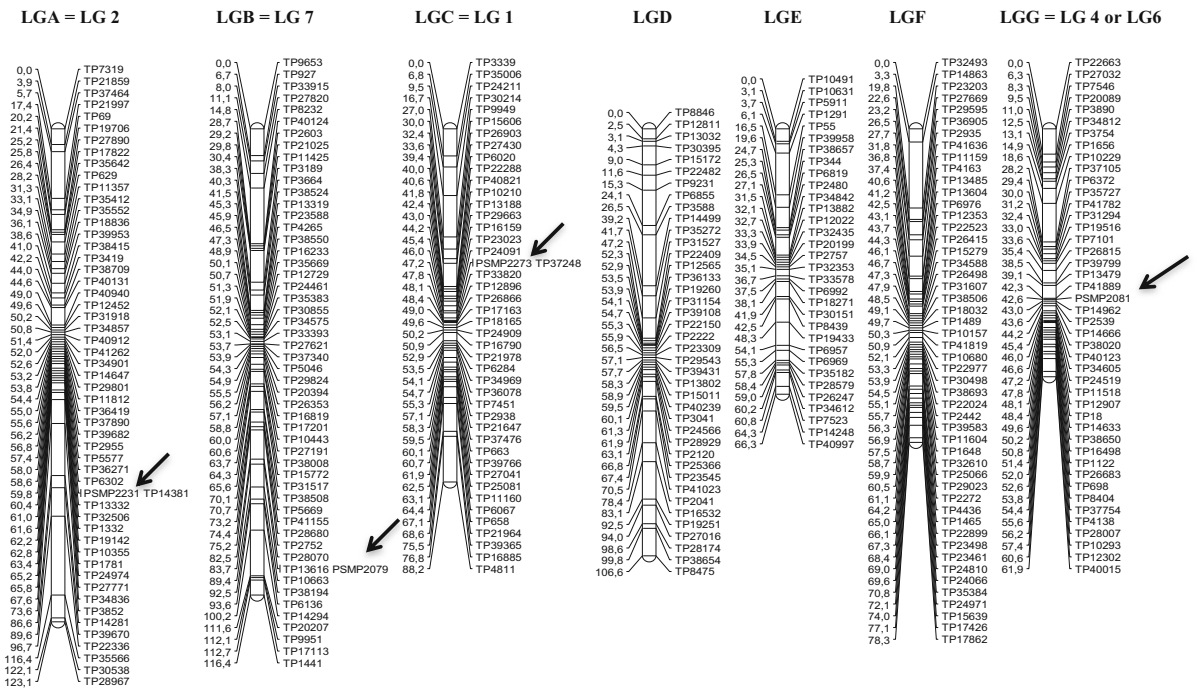

Fig. 2 SNP genetic map of pearl millet. Cumulative distances are indicated to the left (in cM). SNP markers are labeled TPXXXX, while the four SSR markers used for bridging maps are labeled PSMPXXXX and their position is shown using *arrows*

Relationship between linkage groups in different maps

To determine the relationship between the seven linkage groups of pearl millet described in this work and those described in previous maps, a set of 19 previously mapped SSRs were tested. Among these, only four (PSMP2231, PSMP2079, PSMP2273 and PSMP2081) proved polymorphic between the parents. These polymorphic SSR markers were then genotyped on the entire population, and the resulting data were used to integrate these SSR markers on the SNP map. This preliminary analysis suggests a possible correspondence between four of our linkage groups (LGs A, B, C and G) with four linkage groups (LG2, LG7, LG1 and LG4 or LG6) from the linkage maps of Qi et al. (2004) and Rajaram et al. (2013) (Table 3).

Table 3 Relationship between LGs in this map and previously published maps established using four common SSR loci

SSR markers	Repeat unit	Linkage group (LG)		
		This work	Qi et al. (2004)	Rajaram et al. (2013)
PSMP2231	$(TG)_{12}$ $GG(TA)_4$	A	2	2
PSMP2079	–	B	7	–
PSMP2273	$(GA)_{12}$	C	1	1
PSMP2081	$(AC)_{15}$	G	4	4 and 6

Discussion

Genotyping-by-sequencing efficiency

The simultaneous identification and genotyping of SNPs were made possible because of the recent significant advances in sequencing. In this study, a complexity reduction approach relying both on a rare-cutting enzyme (*Pst*I) and primers with three selective bases produced a set of over 3,300 high-quality SNP markers segregating between two contrasting pearl millet lines. Using an identical level of multiplexing (95 individuals per Illumina HiSeq lane) and the UNEAK pipeline, Lu et al. (2013) obtained 3,000 segregating SNPs (with a minimal call rate >90 % and Mendelian allelic ratios) in a full-sib mapping population of switchgrass. As these authors used an enzyme that cuts relatively frequently to produce their libraries (*Ape*KI), they obtained a much larger number of SNPs (400,107), but these had a low median coverage (0.54X) and a low call rate (40 %). Thus, in their initial mapping work, less than 1 % (3,000 SNPs) of this large set of SNPs had a sufficient coverage and call rate to be used for building a map. In this work, almost one-fifth of our SNP markers (3,321 of 18,346) were of sufficient quality to be used for mapping.

We suggest that for genetic mapping in biparental populations, where a thousand mapped markers would be quite sufficient in most circumstances, it is more efficient to aim for a greater complexity reduction by using a less frequently cutting enzyme, amplification with selective primers or a combination of both. By properly adjusting the degree of achieved complexity reduction, Sonah et al. (2013) have shown that it is possible to increase the number of lines that can be genotyped in a single library and sequencing lane at practically no additional cost.

In comparison with a DArT assay, the only other highly parallel genotyping tool available in pearl millet, the GBS approach proved significantly more efficient while not requiring any prior marker discovery work in the form of array development. On an array of 7,680 spotted clones, only 574 (7.5 %) proved polymorphic among a panel of 24 pearl millet inbreds and only 389 were polymorphic in a biparental mapping population. A similarly small amount of genomic DNA yielded almost 10 times as many (3,321) polymorphic and codominant SNP markers.

Segregation distortion

A significant degree of segregation distortion was observed for 11.8 % of the genetic loci defined by the 314 non-redundant SNP markers used to build this map. Such distortion has been reported in all previous genetic maps in pearl millet. Supriya et al. (2011) reported that 35 % of their DArT markers presented a significant deviation, whereas Rajaram et al. (2013) encountered distortion in 31–38 % of their SSR markers in four mapping populations. Many other examples of segregation distortion have been reported in other crop species including barley (Devaux et al. 1995), rice (Xu et al. 1997), maize (Lu et al. 2002) and wheat (Quarrie et al. 2005).

High-density map

Although 2,156 SNP markers met all of our filtering criteria in this F_2 population, a large proportion of these produced identical segregation patterns, resulting in only 314 non-redundant mapped loci. This number of mapped loci is not directly comparable to previous maps (Liu et al. 1994; Qi et al. 2004; Supriya et al. 2011; Rajaram et al. 2013) in which redundant markers were not removed, but rather formed clusters of loci mapping at the same position on these maps. What is certain, however, is that a larger mapping population would have allowed more (possibly all) of the 2,156 high-quality SNP markers to be assigned a distinct position on a genetic map without any need to

develop a larger genotyping array for DArTs or analyzing more SSR loci. Given the dense and uniform coverage already achieved, however, it is not clear that this would lead to a significant improvement in map coverage. In the context of performing association analyses, however, the significant increase in the number of available markers would be extremely useful.

The genetic map reported here contains 314 SNPs distributed on seven linkage groups spanning a total of 640 cM. This is greater than the total reported for previous maps built using F_2 populations. The map of Liu et al. (1994) contained 181 loci and covered 303 cM, whereas the map built by Qi et al. (2004) contained 242 loci and spanned 473 cM. The fact that all of our SNP markers could be placed on a linkage group and that there was extensive redundancy in the observed segregation patterns both suggest that map coverage is quite extensive. An attempt was made to establish relationships between the linkage groups defined in the present map (based on SNPs) and those reported in previous maps (using SSRs). Given the low number of SSR markers that proved polymorphic between our parental lines, such relationships could only be tentatively established for four linkage groups. The upcoming release of the pearl millet genome, however, will remove any ambiguity in this regard and allow a very clear and extensive description of these relationships.

This total map distance, however, is almost exactly half of the total map distance reported by Supriya et al. (2011) that was obtained using RILs and about 50 % less than the consensus map produced by merging data from four RIL populations (Rajaram et al. 2013). The maps constructed using RILs did not take into account the fact that such populations typically exhibit twice the number of recombination events in any given genetic interval compared to F_2s and so are not directly comparable. To answer this, it would be necessary to examine the genetic distance between the same genetic loci in these various maps.

Interestingly, the map coverage achieved in the present study is quite uniform with an average distance of 2.1 (±0.6) cM between neighboring SNPs. This interval is smaller than the value of 3.6 (±1.5) cM obtained by Supriya et al. (2011), although here again the latter number may simply be inflated due to the use of RILs. Importantly, no large gaps (>20 cM) were found in our map contrary to genetic maps obtained by Senthilvel et al. (2008) and Supriya et al. (2011). In the map of Senthilvel et al. (2008), four individual intervals between adjacent SSRs were estimated to exceed 30 cM (the greatest spanning 62 cM). In the map built by Supriya et al. (2011), one individual interval between DArTs exceeded 30 cM (35 cM). This is in stark contrast to the map produced here, in which only ten intervals (3.3 %) were larger than 10 cM. In other words, greater than 96 % of all intervals were less than 10 cM.

The results of this study indicate that GBS can rapidly and efficiently provide high-quality, codominant SNP markers that can be used to construct densely populated genetic maps even in the absence of a reference genome. The availability of SNP markers and high-density genetic maps will not only facilitate gene and QTL mapping in biparental populations, but also make it possible to perform association analyses on panels of unrelated lines. As such panels typically exhibit much decreased linkage disequilibrium between adjacent markers (relative to F_2s or RILs), many thousands of markers will likely be required for such work. Such dense marker coverage can only be achieved using SNP genotyping and GBS provides a cost-effective means to reach this goal.

Acknowledgments The support of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) and the Institut de Recherche pour le Développement (IRD) are acknowledged. We would like also to thank the Institut de Biologie Intégrative et des Systèmes (IBIS) at Université Laval for their involvement in the process of GBS. This publication was finalized as part of the CGIAR Research Program on Dryland Cereals.

References

Amusan IO, Rich PJ, Menkir A, Housley T, Ejeta G (2008) Resistance to *Striga hermonthica* in a maize inbred line derived from *Zea diploperennis*. New Phytol 178:157–166

Bidinger FR, Nepolean T, Hash CT, Yadav RS, Howarth CJ (2007) Identification of QTLs for grain yield of pearl millet [*Pennisetum glaucum* (L.) R. Br.] in environments with variable moisture during grain filling. Crop Sci 47:969–980

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Chemisquy MA, Giussani LM, Scataglini MA, Kellogg EA, Morrone O (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 (Oxford, UK) 106:107–130

Devaux P, Kilian A, Kleinhofs A (1995) Comparative mapping of the barley genome with male and female recombination-derived, doubled haploid populations. Mol Gen Genet 249:600–608

Dhuppe MV, Chavan AA, More AW, More SS, Choudhary LM (2005) Heterosis in pearl millet. J Soils Crops 15(2):343–351

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE 6(5):e19379. doi:10.1371/journal.pone.0019379

Gowda CLL, Rai KN (2006) Evolution of hybrid parents research. In: Gowda CLL, Rai KN, Freddy BVS, Sana KB (eds) Hybrid parents research at ICRISAT. ICRISAT Center, Patancheru, pp 1–10

Gurney AL, Slate J, Press MC, Scholes JD (2006) A novel form of resistance in rice to the angiosperm parasite *Striga hermonthica*. New Phytol 169:199–208

Hash CT, Yadav RS, Cavan GP, Howarth CJ, Liu H, Qi X, Sharma A, Kolesnikova-Allen MA, Bidinger FR, Witcombe JR (2000) Marker-assisted backcrossing to improve terminal drought tolerance in pearl millet. In: Ribaut J-M, Poland D (eds) Molecular approaches for the genetic improvement of cereals for stable production in water-limited environments, a strategic planning workshop held at CIMMYT, El Batan, Mexico, 21–25 June 1999, International Maize and Wheat Improvement Center (CIMMYT), Mexico, pp 114–119

Haussmann BIG, Hess DE, Omanya GO, Folkertsma RT, Reddy BVS, Kayentao M, Welz HG, Geiger HH (2004) Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. Theor Appl Genet 109:1005–1016

Jauhar PP, Rai KN, Ozias-Akins P, Chen Z, Hanna WW (2006) Genetic improvement of pearl millet for grain and forage production: cytogenetic manipulation and heterosis breeding. In: Singh RJ, Jauhar PP (eds) Genetic resources, chromosome engineering, and crop improvement, 2nd edn. CRC Press, Taylor & Francis Group, Boca Raton, pp 281–307

Kountche BA, Hash CT, Dodo H, Oumarou L, Sanogo MD, Amadou T, Vigouroux Y, This D, Nijkamp R, Haussmann BIG (2013) Development of a pearl millet *Striga*-resistant genepool: response to five cycles of recurrent selection under *Striga*-infested field conditions in West Africa. Field Crop Res 154:82–90

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newberg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Liu CJ, Witcombe JR, Pittaway TS, Nash M, Busso CS, Hash CT, Gale MD (1994) An RFLP-based genetic map of pearl millet (*Pennisetum glaucum*). Theor Appl Genet 89:481–487

Lu H, Romero-Severson J, Bernardo R (2002) Chromosomal regions associated with segregation distortion in maize. Theor Appl Genet 105:622–628

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet 9(1):e1003215. doi:10.1371/journal.pgen.1003215

Manning K, Pelling R, Higham T, Schwenniger JL, Fuller DQ (2011) 4500-Year old domesticated pearl millet (*Pennisetum glaucum*) from the Tilemsi Valley, Mali: new insights into an alternative cereal domestication pathway. J Archaeol Sci 38(2):312–322

Mariac C, Luong V, Kapran I, Mamadou A, Sagnard F, Deu M, Chantereau J, Gerard B, Ndjeunga J, Bezançon G, Pham J, Vigouroux Y (2006) Diversity of wild and cultivated pearl millet accessions [*Pennisetum glaucum* (L.) R. Br.] in Niger assessed by microsatellite markers. Theor Appl Genet 114:49–58

Omanya GO, Haussmann BIG, Hess DE, Reddy BVS, Kayentao M, Welz HG, Geiger HH (2004) Utility of indirect and direct selection traits for improving striga resistance in two sorghum recombinant inbred populations. Field Crop Res 89:237–252

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

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7(2):e32253. doi:10.1371/journal.pone.0032253

Presterl T, Weltzien E (2003) Exploiting heterosis in pearl millet for population breeding in arid environments. Crop Sci 43:767–776. doi:10.2135/cropsci2003.7670

Qi X, Pittaway TS, Lindup S, Liu H, Waterman E, Padi FK, Hash CT, Zhu J, Gale MD, Devos KM (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

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen EJ, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti M-C, Hollington PA, Aragués R, Royo A, Dodig D (2005) A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring 9 SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110:865–880

Rai KN, Govindaraj M, Rao AS (2012) Genetic enhancement of grain iron and zinc content in pearl millet. Qual Assur Saf Crops Foods 4(3):119–125. doi:10.1111/j.1757-837X.2012.00135.x

Rajaram V, Nepolean T, Senthilvel S, Varshney RK, Vadez V, Srivastava RK, Shah TM, Supriya A, Kumar S, Kumari BR, Bhanuprakash A, Narasu ML, Riera-Lizarazu O, Hash CT (2013) Pearl millet [*Pennisetum glaucum* (L.) R. Br.] consensus linkage map constructed using four RIL mapping populations and newly developed EST-SSRs. BMC Genom 14:159

Sehgal D, Rajaram V, Armstead IP, Vadez V, Yadav YP, Hash CT, Yadav RS (2012) Integration of gene-based markers in a 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, Jayashree B, Mahalakshmi V, Kumar PS, Nakka S, Nepolean T, Hash CT (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

Sharma PC, Singh D, Sehgal D, Singh G, Hash CT, Yadav RS (2014) Further evidence that a terminal drought tolerance QTL of pearl millet is associated with reduced salt uptake. Environ Exp Bot. doi:10.1016/j.envexpbot.2014.01.013

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G et al (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE 8(1):e54603. doi:10.1371/journal.pone.0054603

Supriya A, Senthilvel S, Nepolean T, Eshwar K, Rajaram V, Shaw R, Hash CT, Kilian A, Yadav RC, Narasu ML (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

Tinker N (2013) Genotype-by-sequence in oat. Agriculture and Agri-Food Canada. Retrieved from http://www.9ioc.net/02-Tinker_Nick.pdf

Tostain S, Marchais L (1993) Évaluation de la diversité génétique des mils (*Pennisetum glaucum* (L.) R. BR.) au moyen de marqueurs enzymatiques et relations entre formes sauvages et cultivées. Laboratoire Ressources Génétiques et Amélioration des Plantes Tropicales. ORSTOM, Montpellier Cedex 1

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93(1):77–78. doi:10.1093/jhered/93.1.77

Wilson JP, Hess DE, Hanna WW, Kumar KA, Gupta SC (2004) *Pennisetum glaucum* subsp. *monodii* accessions with *Striga* resistance in West Africa. Crop Prot 23:865–870

Xu Y, Zhu L, Xiao J, Huang N, McCouch SR (1997) Chromosomal regions associated with segregation distortion of molecular markers in F2, backcross, doubled haploid and recombinant inbred populations of rice (*Oryza sativa* L.). Mol Gen Genet 253:535–545

Yoshida S, Shirasu K (2009) Multiple layers of incompatibility to the parasitic witchweed, *Striga hermonthica*. New Phytol 183:180–189