SHORT REPORT
doi: 10.2306/scienceasia1513-1874.2022.065

ScienceAsia 48 (2022): **491–495**

Development and application of InDel markers based on sudangrass RAD-seq data

Lihua Wang[a,b], **Jieqin Li**[b], **Qiuwen Zhan**[b], **Yixin Shen**[a,*]

[a] College Of Agro-Grassland Science, Nanjing Agricultural University, NanJing 210095 China
[b] Anhui Science and Technology University, Fengyang 233100 China

*Corresponding author, e-mail: yxshen@njau.edu.cn

Received 19 Jul 2021, Accepted 12 Feb 2022
Available online 15 Apr 2022

ABSTRACT: Two sudangrass varieties, Sa and S722, were sequenced using restriction site-associated DNA sequencing (RAD-seq) on the Illumina Miniseq platform. After quality control, 920 542 and 892 626 clean reads were obtained for Sa and S722, respectively. Compared with the sorghum reference genome, 2341 and 2123 single nucleotide polymorphisms (SNPs) were obtained from Sa and S722, respectively. A total of 543 and 472 insertion-deletion (InDel) loci were obtained by sequence analysis in Sa and S722, respectively. From these InDel loci, 100 InDels were randomly selected to design InDel markers from Sa and S722. Polymorphism analyses were performed between sorghum Tx623B and Sa and between Tx623B and S722 using the InDel markers. The results showed that the polymorphism between Tx623B and Sa was 85%, and that between Tx623B and S722 was 87%. Diversity analysis was performed using 39 InDel markers for 42 sorghum and 6 sudangrass germplasms. Statistical analyses showed that the Shannon information index was 0.10–1.09 with an average value of 0.54. The polymorphism information content of the InDel markers ranged from 0.04 to 0.66 with an average of 0.35. Two genetic maps of chromosomes 1 and 2 of the Wancao No. 2 recombinant inbred line population were constructed using 16 InDel markers. Thus, it is an effective method to develop InDel markers for sorghum and sudangrass using sudangrass RAD-seq data.

KEYWORDS: RAD-seq, InDel marker, genetic map, sudangrass, sorghum

INTRODUCTION

InDel markers are developed based on known sequences and are suitable for the development of genome-wide molecular markers [1]. They have excellent stability, high polymorphism, and simple typing systems [2]. Moreover, they can be used both intraspecifically and interspecifically [3]. Compared with single nucleotide polymorphism (SNP) marker detection, InDel detection is simpler, more convenient, and cheaper as it is based on polymerase chain reaction (PCR) technology. InDel markers were found to be more accurate than SNP markers in identifying genetic relationships among samples based on a SNaPshot typing platform [4]. However, the development of InDel markers relies on the genomic information. For non-model species, it is difficult to develop InDel markers because of the limited genomic information.

Many species have been sequenced because of the development of next-generation sequencing (NGS) technology. However, it is still expensive and unnecessary to develop markers by whole genome sequencing, and it is not a cost-effective way to develop markers for these species without a reference genome. Therefore, the most economical approach is to develop markers using reduced representation-sequencing. Restriction-site associated DNA sequencing (RAD-seq) is a simplified genome sequencing method developed from NGS. This method can reduce the representation of complex genomes using enzymatic digestion. It can develop up to 10 times more molecular markers than the traditional molecular marker development techniques [5]. RAD-seq has been used in several plant species to discover SNP, InDel, and simple sequence repeat (SSR) markers for germplasm collection [6, 7], genetic analysis, and molecular characterization to determine the existence of a reference genome [8, 9]. This technique has a higher accuracy and data utilization rate and requires less time at a lower cost than the traditional marker development techniques [5].

Sorghum sudanense, commonly called sudangrass, is used as a forage for ruminants. It is widely planted in Russia, Eastern Europe, and South Asia. Sudangrass can be easily crossed with sorghum (*Sorghum bicolor* (L.) Moench) and shows vigorous heterosis. Compared to sorghum and sudangrass, the hybrid of sorghum and sudangrass is a forage with higher yield, drought tolerance, and lodging resistance. Researchers have used this method to develop a new type of forage, Sorghum-Sudangrass grass [10–12]. Wancao No. 2, which is a hybrid of sorghum Tx623A and sudangrass S722, is a widely cultivated variety in China and has a higher forage yield and drought tolerance compared to its parents [11]. To date, the genome of sudangrass has not been sequenced. Sorghum, sudangrass, and sorghum-sudangrass hybrids were clustered into the same group and belonged to the same species of sorghum [13]. They cannot be distinguished completely using molecular markers [14]. It is a good strategy to employ reference-based approaches to a closely related genome in RAD-seq studies and transcriptome sequenceing [15, 16]. Yang et al [17] also

called SNPs in sugarcane using the sorghum genome as a reference genome. Owing to the close similarity between sorghum and sudangrass [13], it is feasible to develop molecular markers using the sorghum genome as a reference genome.

In this study, we developed InDel markers based on RAD-seq data of sudangrass and sorghum BTx623 (version 3.1.1) reference genome (https://phytozome-next.jgi.doe.gov/info/Sbicolor_v3_1_1) [18]. We also validated the development of InDel markers by constructing a genetic map of the recombinant inbred line (RIL) population from sorghum and sudangrass (Wancao No. 2). Our research provides an effective and economical way to develop genome-wide InDel markers for non-model species.

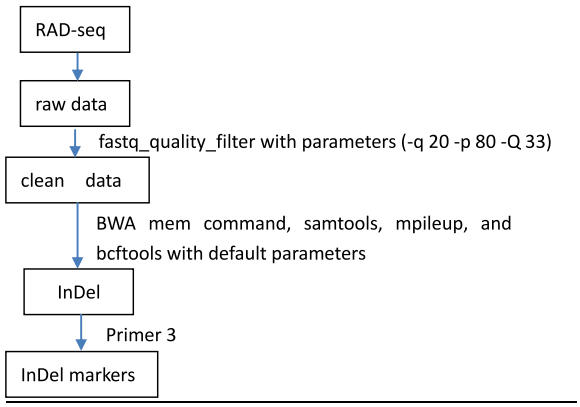

Fig. 1 The workflow of development of InDel markers based on RAD-seq.

MATERIALS AND METHODS

Plant materials

Two sudangrass varieties, Sa and S722, were used for RAD-seq. The RIL population, including 102 lines, was constructed with sorghum Tx623A and sudangrass S722 as parents to validate the InDel markers from RAD-seq.

To ensure high genetic homogeneity between sorghum and sudangrass, 48 germplasms, including 42 of sorghum and 6 of sudangrass, were used for the analysis of InDel polymorphic markers (Table S1).

DNA extraction and RAD-seq

Genomic DNA was extracted from the fresh leaves of sudangrass using a DNAsecure Plant Kit (TIANGEN Biotech Co., LTD, Beiing, China). DNA quality was determined using a BioDrop Touch spectrophotometer (Biochrom Ltd., Cambridge, UK). The two samples were normalized to 50 ng/μl and digested with the enzymes, PstI (CTGCAG) and MspI (CCGG), first at 37 °C for 2 h and then at 65 °C for 20 min. The digested samples were ligated with adapters and pooled for PCR ampliïňĄcation. The genotyping-by-sequencing (GBS) library was sequenced using the Miniseq system (Illumina Inc., San Diego, CA, USA).

RAD-seq data process and InDel primer design

The workflow is illustrated in Fig. 1. The sequencing reads of Sa and S722 were extracted from the raw data of RAD-seq and filtered using fastx_barcode_splitter and fastq_quality_filter with parameters (-q 20 -p 80 -Q 33) of fastx_toolkit-.0.13.2 (http://hannonlab.cshl.edu/fastx_toolkit). High-quality sequencing data were aligned using the BWA-MEM algorithm. Then, samtools mpileup and bcftools were used to call InDels from the alignment files of the samples [9, 19]. InDel primers were designed based on these called InDels of Sa and S722 and sorghum reference genome version 3.1.1 using Primer Premier 3.0 [20]. The parameters were as follows: product size of 100–300 bp; an annealing temperature of 50–60 °C; the criterion of >4 bp difference in the bases of the InDel locus.

Verification of polymorphism of InDel markers and construction of RIL genetic map

InDel primers were screened with Tx623B and Sa as well as with Tx623B and S722. The PCR reaction volume was 10 μl, including 1 μl template DNA (50 ng/μl), 5 μl 2X PCR mixture (10X PCR buffer with Mg^{2+}, 2.5 mmol/l dNTPs, and 0.5 U Taq DNA polymerase), 1 μl primers (2 μmol), and 3μl ddH_2O. The PCR procedure was as follows: pre-denaturation at 94 °C for 5 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 30 s, and extension at 72 °C for 40 s, and a final extension at 72 °C for 7 min. The PCR products were detected by silver staining on an 8% non-denaturing polyacrylamide gel.

Data analysis

Shannon's information index (H') and polymorphism information content (PIC) values were calculated using PopGene 1.32: $H' = -\sum P_i \ln P_i$ [21] and $\text{PIC} = 1 - P_i^2$, where P_i is the allele frequency [22].

A genetic map of the RIL population was constructed using IciMapping [23]. The map was drawn using Map Chart v2.3 [24].

RESULTS AND DISCUSSION

Analysis of RAD-seq

The RAD-seq results showed that 920 542 and 892 626 clean reads were obtained from Sa and S722, respectively. The mapping rates for Sa and S722 with the reference genome of sorghum version 3.1.1 were 96.37% and 95.15%, respectively. The average Q30 for Sa and S722 were 92.44% and 92.31%, respectively. Q30 was high, indicating that the sequence data were reliable. The GC content for Sa and S722 were 53.75% and 53.82%, respectively. After quality control and comparison with the sorghum reference genome, 2341

and 2123 SNPs were obtained from Sa and S722, respectively. Therefore, the RAD-seq data can be used for subsequent research.

Primer design and polymorphism of InDel markers

Based on the criterion of > 4 bp difference in the bases of the InDels, 543 and 472 InDel loci were obtained from Sa and S722, respectively. Then, 100 loci in both Sa and S722 were selected randomly from these InDel loci for primer design (Table S2). Polymorphism verification was performed between Tx623B and Sa and between Tx623B and S722. The results showed that the polymorphism between Tx623B and Sa was 85% with 5% markers having no amplified products. The polymorphism between Tx623B and S722 was 87%. Polymorphism rate of 50 InDels was found to be 76% in a previous study, which was based on RAD-seq between the two sequencing materials [25]. Therefore, it is a reliable method for developing InDel markers based on RAD-seq.

To further identify the universality of these markers, 39 InDel markers were used to characterize and evaluate 42 sorghum and 6 sudangrass germplasms (Table S1). Statistical analysis showed that the length of the amplified fragments was 150–300 bp and Shannon's information index was 0.10–1.09 with an average value of 0.54 (Table S3). Shannon's index of 24 markers (61%) was higher than 0.50, indicating great genetic diversity of the tested materials. The PIC ranged from 0.04 to 0.66 with an average of 0.35 (Table S3). Previous studies showed that $0.25 < \text{PIC} < 0.50$ indicates that the markers are moderately polymorphic [26]. Therefore, these markers can be used for the genetic analysis of sorghum and sudangrass.

Genetic map construction using InDel markers

Genetic map construction is the basis of quantitative trait loci (QTL) mapping. To further validate these InDel markers, 16 InDel markers located on chromosomes 1 and 2 were selected to construct genetic maps of the RIL population of Wancao No. 2. The results showed that the total length of the chromosome 1 was 85.2 cM with the average distance between the markers being 12.17 cM (Fig. 2). The length of chromosome 2 was 76.8 cM, and the average distance between the markers was 8.5 cM. Compared to the sorghum genome version 3.1.1 [18], only two markers on chromosome 1 (RAD1-7 and RAD1-14) were not consistent with the order of these loci in the sorghum genome. However, on chromosome 2, all markers were consistent with the order of these InDel loci in the genome. This implies that these InDel markers can be used for the genetic analysis of hybrids of sorghum and sudangrass.

RAD-seq is based on second-generation sequencing technology and has been widely applied to develop markers for genetic analysis [27–30]. Compared with whole-genome resequencing, RAD-seq is cheaper and more feasible for developing markers for non-model species [31]. A large number of markers can be obtained in a single sequencing round using RAD-seq. In this study, we performed RAD-seq in sudangrass species, which do not have a reference genome. We used the genome of its relative species, sorghum, as a reference because of the similarity between the two species [13]. We used this method to develop InDel markers for sudangrass. We proved that these markers can be useful for the genetic analysis of sorghum, sudangrass, and the hybrids of sorghum and sudangrass. Hence, we have provided a feasible method using RAD-seq to develop markers for species whose genomes have not been sequenced.

Considering sudangrass and sorghum as the same species is still controversial. Snowden considered su-

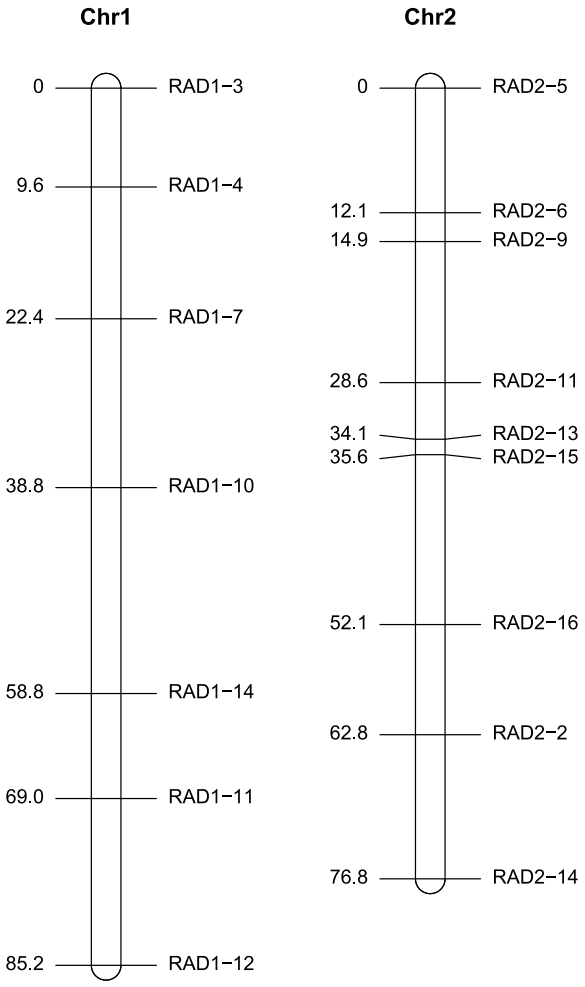

Fig. 2 Genetic maps of chromosomes 1 (Chr1) and 2 (Chr2) of the RIL population derived from the cross between sorghum Tx623A and sudangrass S722 and constructed using the developed InDel markers.

dangrass (*Sorghum sudanense*) a different species from sorghum based on anthotaxy and phenotype [32], but De Wet and Huckabay [33] suggested that sudangrass should be considered a subspecies, *drummondii* (steud), of *S. bicolor* (L.) Moench. Previously, the transcriptome of sudangrass was characterized, and high genomic similarity was observed between sudangrass and sorghum by RNA-seq analysis [34]. In this study, we sequenced a part of the genome of the two sudangrass varieties. The mapping rates of Sa and S722 with the sorghum reference genome were 96.37% and 95.15%, respectively. These results indicated that sudangrass is more suitable for consideration as a subspecies of sorghum than as a different species.

CONCLUSION

We designed InDel markers of sorghum-based on RAD-seq data in this study. The InDel markers were used for diversity analysis of 48 germplasms using PopGene 1.32. The results showed that the InDel markers had an excellent polymorphism and strong discrimination ability in sorghum and sudangrass. We also constructed genetic maps of the RIL population of Wancao No.2 using the InDel markers. Sixteen InDel markers were located on chromosomes 1 and 2. It shows that InDel markers based on sudangrass RAD-seq data can be used in sorghum, sudangrass, and sorghum-sudangrass hybrid. So, RAD-seq is an effective way to develop markers in sorghum, sudangrass, and sorghum- sudangrass hybrid.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at http://dx.doi.org/10.2306/scienceasia1513-1874.2022.065.

Acknowledgements: This study was supported by the National Natural Science Foundation of China (31971993), the Anhui Provincial Natural Science Fund (2008085MC73), the Anhui Provincial Key R&D Programs (202004b11020003).

REFERENCES

1. Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, et al (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. *BMC Genom* **11**, 1248–1251.
2. Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) Arabidopsis map-based cloning in the post-genome era. *Plant Physiol* **129**, 440–450.
3. Väli U, Brandström M, Johansson M, Ellegren H (2008) Insertion-deletion polymorphisms (Indels) as genetic markers innatural populations. *BMC Genet* **9**, 715–720.
4. Santos C, Fondevila M, Ballard D, Banemann R, Bento AM, Børsting C, Branicki W, Brisighelli F, et al (2015) Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: results of a collaborative EDNAP exercise. *Forensic Sci Int Genet* **19**, 56–67.
5. Wang Y, Hu Y, Zhang T (2014) Current status and perspective of RAD-seq in genomic research. *Hereditas* **36**, 41–49.
6. Marrano A, Birolo G, Prazzoli ML, Lorenzi S, Valle G, Grando MS (2017) SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (*V. vinifera* L.). *PLoS One* **12**, e0170655.
7. Willyard A, Bower A, Hipkins V, Snelling J, DeWoody J (2020) Genetic diversity and population structure in *Chrysolepis chrysophylla* (golden chinquapin; Fagaceae): SSRs vs SNPs. *Can J For Res* **4**, ID 0009.
8. Dong W, He F, Wei S, Qiu Z, Chen Q (2021) Identification and characterization of SSR markers in taro [*Colocasia esculenta* (L.) Schott] by RAD sequencing. *Genet Resour Crop Evol* **7**, 2897–2905.
9. Siddra I, Imran UH, Bukhtawer N (2020) *In silico* identification of expressed sequence tags based simple sequence repeats (EST-SSRs) markers in *Trifolium* species. *ScienceAsia* **46**, 6–10.
10. Qian Z-Q (1990) Forage breeding by using sorghum male-sterile lines. *Chin Grassl* **6**, 61–63.
11. Zhan Q-W, Qian Z-Q (2004) Heterosis utilization of hybrid between sorghum [*Sorghum bicolor* (L.) Moench] and sudangrass [*Sorghum sudanense* (Piper) Stapf]. *Zuo Wu Xue Bao* **30**, 73–77.
12. Brad V, Bryan K (2008) Forage and biomass feedstock production from hybrid forage sorghum and sorghum-sudangrass hybrids. *Grassl Sci* **54**, 189–196.
13. Zhan QW, Zhang TZ, Wang BH, Li JQ (2008) Diversity comparison and phylogenetic relationships of *S. bicolor* and *S. sudanense* as revealed by SSR markers. *Plant Sci* **174**, 9–16.
14. Li J, Wang L, Zhan Q, Li Y (2007) A study on genetic variation of two species of sorghum by RAPD markers. *Cao Ye Xue Bao* **16**,140–144.
15. Shafer ABA, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Jochen BWW (2017) Bioinformatic processing of rad-seq data dramatically impacts downstream population genetic inference. *Methods Ecol Evol* **8**, 907–917.
16. Peng H, Gao J, Zhang C, Ou L, Luo M(2020) Novel SSR and SNP markers in *Viola yedoensis* Makino resistant to cadmium stress. *ScienceAsia* **46**, 280–287.
17. Yang X, Song J, You Q, Paudel DR, Zhang J, Wang J (2017) Mining sequence variations in representative polyploid sugarcane germplasm accessions. *BMC Genom* **18**, ID 594.
18. McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, et al (2018) The sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J* **93**, 338–354.
19. Li J, Tang W, Zhang Y-W, Chen K-N, Wang C, Liu Y, Zhan Q, Wang C, et al (2018) Genome-wide association studies for five forage quality-related traits in Sorghum (*Sorghum bicolor* L.). *Front Plant Sci* **9**, ID 1146.
20. Zhao L-N, Liu N, Wang X-S, Cui S-H, Lu Y (2021) Determination of quinolones-resistance genes in raw milk-derived *Escherichia coli* real-time fluorescence PCR. *J Food Saf Qual* **14**, 5629–5635.
21. Guo GJ, Sun Q, Liu JB, Pan B-g, Diao WP, Ge W, Gao CZH, Wang SB (2015) Development and application of pepper InDel markers based on genome re-sequencing. *Jiangsu*

J Agr Sci **6**, 1400–1406.
22. Zhou RR, Zhou JH, Nan TG, Jiang C, Duan HY, Zhao YP, Huang LQ, Yuan Y (2019) Analysis of genomic SSRs in *Pueraria lobata* and *P. thomsonii* and establishment of DNA identity card for different germplasms of *P. thomsonii* of Jiangxi province. *Zhongguo Zhong Yao Za Zhi* **17**, 3615–3621.
23. Li HH, Ye GY, Wang JK (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* **175**, 361–374.
24. Jin D, Wang DZ, Wang HX, Li RZ, Chen SL, Yang WL, Zhang AM, Liu D, et al (2019) Fine mapping and candidate gene analysis of awn inhibiting gene B2 in common wheat. *Zuo Wu Xue Bao* **6**, 807–817.
25. Li HY, Li RH, Xia YS, Yuan QH, Zhang ZC, Zhao WC, Guo PG (2018) Development of polymorphic SSR markers in tobacco based on RAD sequencing. *Chin Tob Sci* **1**, 1–9.
26. Li QS, Chen JB, Gu HP, Yuan XX, Chen X, Cui J (2019) Genetic diversity and fingerprint analysis of mungbean varieties from China based on InDel markers. *J Plant Genet Resour* **1**, 122–128.
27. Barchi L, Lanteri S, Portis E, Acquadro A, Valè G, Toppino L, Rotino GL (2011) Identification of SNP and SSR markers in eggplant using RAD tag sequencing. *BMC Genom* **12**, ID 304.
28. Gupta SK, Baek J, Carrasquilla N, Penmetsa RV (2015) Genome-wide polymorphism detection in peanut using next-generation restriction-site-associated DNA (RAD) sequencing. *Mol Breed* **35**, ID 145.
29. Zhu JC, Guo YS, Su K, Liu ZD, Ren ZH, Li K, Guo XW (2018) Construction of a highly saturated genetic map for Vitis by next-generation restriction site-associated DNA sequencing. *BMC Plant Biol* **18**, ID 347.
30. Urasaki N, Goeku S, Kaneshima R, Takamine T, Tarora K, Takeuchi M, Moromizato C, Yonamine K, et al (2015) Leaf margin phenotype-specific restriction-site-associated DNA-derived markers for pineapple (*Ananas comosus* L.). *Breed Sci* **3**, 276–284.
31. Davey JW, Cezard T, Fuentes-Utrilla P, Eland C, Gharbi K, Blaxter ML (2013) Special features of rad sequencing data: implications for genotyping. *Mol Ecol* **11**, 3151–3164.
32. Snowden JD (1936) *The Cultivated Races of Sorghum*, Adlard & Son, Ltd. London, UK.
33. De Wet JMJ, Huckabay JP (1967) The origin of *Sorghum bicolor*. II. Distribution and domestication. *Evol* **21**, 787–802.
34. Li J, Wang L, Zhan Q, Liu Y, Yang X (2016) Transcriptome characterization and functional marker development in *Sorghum sudanense*. *PLoS One* **5**, e0154947.

Appendix A. Supplementary data

Table S1 The source of the 48 germplasms for InDel marker diversity analysis.

No.	Name	Source country
1	IS 1041	India
2	IS 4060	India
3	IS 4360	India
4	IS 4515	India
5	IS 4581	India
6	IS 4613	India
7	IS 7310	Nigeria
8	IS 7679	Nigeria
9	IS 13549	Mexico
10	IS 13782	South Africa
11	IS 14861	Cameroon
12	IS 15170	Cameroon
13	IS 15466	Cameroon
14	IS 15478	Cameroon
15	IS 20625	USA
16	IS 20632	USA
17	IS 20679	USA
18	IS 20697	USA
19	IS 20713	USA
20	IS 22986	Sudan
21	IS 23216	Zambia
22	IS 23514	Ethiopia
23	IS 25301	Ethiopia
24	IS 25548	Rwanda
25	IS 27786	Morocco
26	IS 27887	South Africa
27	IS 27912	South Africa
28	IS 28141	Yemen, Republic of
29	IS 29100	Yemen, Republic of
30	IS 29187	Swaziland
31	IS 29233	Swaziland
32	IS 29239	Swaziland
33	IS 29358	Lesotho
34	IS 29441	Lesotho
35	IS 29606	South Africa
36	IS 29627	South Africa
37	IS 29654	China
38	IS 29689	Zimbabwe
39	IS 30079	Zimbabwe
40	IS 30092	Zimbabwe
41	IS 30231	Zimbabwe
42	IS 31446	Uganda
43	Sa	China
44	XinsuNo2	China
45	Gw3105	USA
46	S722	China
47	Gw-01684	USA
48	Africa-Su	India

Table S2 The 100 InDel primers developed by RAD-seq.

InDel primer	Reference genome position	Forward	Reverse	Product size (bp)
RAD1-1	428811	GCTTCCTTATGGCTGACTGC	TGTTGTCTACGTCGAGGTGC	182
RAD1-2	4495399	GAGCTAGCTAACGACGACGC	TGTCTTGCAGTTTCCTGACG	281
RAD1-3	4720323	GAACTCGATGCGAAAAGGAG	AGGTAATGACGTGGGTGAGC	281
RAD1-4	12267178	ATCAGATCAATGGCAGAGGG	ACATGCACATCAGTCGCAAG	228
RAD1-5	14250727	ACTGTGATCCCAACAGAGCC	CGACAGCAACTACGTCCTCA	266
RAD1-6	15349767	GAGAGCTAGCGAGGAAGGGT	CAGACTAGTCGCTGCGTGAG	156
RAD1-7	22026784	AATTTGAGCCCAAGAGAGGG	GAGGAAGGAAAATTCCAGGC	255
RAD1-8	57943004	AGCACAGACAGCTCCGGTAT	TTAGGATGGCGATGGAGAAG	298
RAD1-9	66574242	TGGATTCCATTCCATTCCAT	TCCACTTCAACACCAACCAA	215
RAD1-10	68960417	ATCCTAAGGCCTAAGCTCCG	TATGACGAGTGAGCAAACGC	216
RAD1-11	69534230	TATGATGGGTAAGCGCAACA	GCAGCACAAACAAGGGAAAT	291
RAD1-12	74485886	GGCAGATAGACGGGGTTGTA	GTCGTGAACCAGAGCTCCTC	245
RAD1-13	77773916	ATGGTTGCGACATGTTGTGT	ACTAGCCATATAACCGGCCC	278
RAD1-14	79376774	ATACTGCTTTGCCCACGTTC	ATCATACCGCCCTCCACATA	298
RAD2-1	4347547	ACTGGTGGGAGTCCGTGTAG	ACTAGAGGCAGCAAAGCGAG	221
RAD2-2	4586661	ATGGTCCCATGCATCAATTT	GCGATGAAGTCCAGGCTATC	278
RAD2-3	6049073	CCACATCGTGATCCACACTC	GAATAGCCAAGCTCTGACCG	237
RAD2-4	9066373	ACCGAATCTTCACGAACCAC	CCAAGAGCTCCACATCCCTA	261
RAD2-5	16643870	CTGAACCACCACACCAAGTG	CTTTGTATCTGGCATGGCCT	202
RAD2-6	56513212	ACTCGTACGTCCGTCCAAAC	TAGGACGATAACCACTGCCC	300
RAD2-7	61589978	GCCCAAAGAACATGGTGAGT	CGTAGATCTTGCACACGACG	182
RAD2-8	62098116	GATCGAGAAGGACCAATCCA	TAAACAGGGCCTAACAACGG	177
RAD2-9	62921095	CATCAATCAACGTACCACCG	TGCTGCATCACATTCCTTCT	186
RAD2-10	63964411	GCATGATGGATCTTGGCTTT	CAGCACACAGCCTACTCAGC	291
RAD2-11	66319653	ACTCTGCAGGTAAGGCAAGC	TACGTGCTCGTTCGTTGTTC	271
RAD2-12	67733940	TAAACATTAAACCCGGCCAA	TGGCATCTGCAGAATGAATC	230
RAD2-13	69051131	GTCTCGTCGTCGTCTCCTTC	GGCACGGAAACTAAATACGG	208
RAD2-14	70410298	GAGCAAGTAAACGACGACCG	GCTTTGTCCGCTTCATCTTC	200
RAD2-15	72597347	GCTCTGCAGGACAGGATAGG	CAGCTCCTTCGTTTGTAGCC	168
RAD2-16	73494865	TTAACAAGCCCAAACCCAAG	GGAAAATCATAGAATGGGCG	164
RAD3-1	568730	TGCATGCTGCAGTAGCTAGG	TATCTGAACAGGGCCCAAAG	233
RAD3-2	8116186	ATCTACCGAATCCAACTGCG	CAAGCTACTCGTTCCCAAGC	152
RAD3-3	9648432	CCTTCTTGTCGACGTTCTCC	AAAGTTGCTTTCGATCGCTG	241
RAD3-4	12324616	AGAGGAAAGCGAAGGAGAGG	CCAGGACAGGAGATGACGAT	233
RAD3-5	15625836	TTCGTCTCGTCAGTCGTCAC	TAAGCTAGGCAGAGTCCCCA	254
RAD3-6	54099661	CGGCCATTACTGGATCTGTT	TAGAGCAGGACGATGAGGGT	295
RAD3-7	55172360	CAACGTACCTTTCCTGGCAT	ATCAGAGGTACCCAACACCG	215
RAD3-8	63286176	GCCAGTTGATGGTGTTTCCT	TGCATGGAGATTAACTTGCG	292
RAD3-9	64260889	AAAGAGTTCTCAGCCCGGAT	CGTGTGGTAGAAGACGTCCC	299
RAD3-10	64552122	GAAAACATTTCGCCAAGGAA	GGCTGGGCATCTAAAGAACA	215
RAD3-11	71068746	AGTTAGGGTTGTTGCCGATG	ATCTGGTTGGTTTTTGCCTG	185
RAD3-12	71771984	GATCGATCCCTTCCTTCCTC	CGAAAGATTCTCCGCTCAAC	267
RAD3-13	72825949	TGAGCAAAGCAATTTCTCCC	TCTGCGCACTCTTTCATCAC	295
RAD3-14	73447306	ACTTGAACACCACTAGCCCC	CACCGCCACCATATCTATCC	206
RAD3-15	73831673	GGGTCAGCAACAATGGAAGT	GCGAGTTCATCCTCTTCGTC	196
RAD4-1	148889	CCAGCATCAGCAAAGATCAA	CCTACTGCTGCGACGGAT	269
RAD4-2	1663210	CTGCAGCTGCTCTCACTCAC	GCAAGCTCGATATGCACGTA	253
RAD4-3	2153046	ATAATGCCGAAGGACACGAC	CCTCGCTGCTGGTAGAGGTA	279
RAD4-4	4060069	TCTGCTTGGGTTCTTCACCT	GGCAACTGGGCTAAGACAAA	270
RAD4-5	7213923	GGAAGCGCAATTGGATAGAG	TGATCTGTTCCCTCCCTGTC	191
RAD4-6	48095866	TAGCTGCTATGGGGCAGAGT	CAAGAAGGCAGCTGAAAACC	253
RAD4-7	58774804	AAAGGCAACTCGGTAGCTCA	CCAATATGGACTACGGGCTC	250
RAD4-8	59554523	CGTCCCCGTCTGAGGACTAC	ATGAAAAAGGCCCAACACAG	272
RAD4-9	61653817	CCTAGCGGTGGTAGCTTTTG	GGGCTTGCCTTTCTCTCTCT	238
RAD4-10	66415290	ATCGATTCACGGCAAGAAAC	TCAAGCCATCTCCATTTTCC	247
RAD4-11	67204768	CGTCGTCACTGCTTAATTGC	AGTCGTAGTCGCTCCAGCAT	296

Table S2 The 100 InDel primers developed by RAD-seq (continued).

InDel primer	Reference genome position	Forward	Reverse	Product size (bp)
RAD5-1	734950	CAGCAAGCTGCAGAAACAAA	GCCATTGGCCCATCTATCTA	200
RAD5-2	2981419	TCGACTGAGAGGCAGCAATA	TGCCACTCTCCCTAGGCTAA	223
RAD5-3	3819355	GTGAAACTAACCTCGGCTGC	GGTCAGTGGACATCTTCCGT	185
RAD5-4	6781339	CAATTGCACGATCAATTTGC	CCAGCTTCATGCTGTAGTGC	250
RAD5-5	8329129	TCTCGCCTCACGTACTTCCT	TTCTTCCGGCACTTGGTAAC	237
RAD5-6	11058088	ATGGCCCCTAGATCGAGTTT	TAGTGGCAGAGGTGTTGCTG	212
RAD5-7	20920762	CAAGAGTTAACCGTGACGCA	ACGCCCTGTGTTTTTGTCAT	221
RAD5-8	53391992	GCCTCCTTAGTCGTCCCTCT	GCATTTGCTCTTTGCTGTGA	297
RAD5-9	62745820	TTGCCCACCAACCTTCTAAC	ATGGACATCCTCCCTCTCCT	159
RAD6-1	40967679	GCCCAAGACGAAGACAAAAG	GGTTGCCATTTCCGGTAGTA	181
RAD6-2	41371517	AGCATTACACACCCCCTCAC	GTGAGTACGACGACCACCAC	160
RAD6-3	42392965	AAAGGGGGTTTAAATTTGCG	TCACGTTCTTCTTGTGTGGC	268
RAD6-4	58591572	CAGGGTGGGCTGCAAGTAT	GCATCCATCTCCAACAGGAT	281
RAD6-5	60970169	CTGAAAATTAGCCTGCCTGC	GCTAGAGAGCGATCGAGGAA	234
RAD7-1	54525558	TGCTGGAAAGCACAAAGCTA	GTTTTTCTTGTCCTCGCTGG	158
RAD7-2	58919634	GCGTCCACGTACACCTACAC	AAAGCTTCCTTTCTCCCTCG	230
RAD7-3	64443302	CTGTTGTCACAGCTGGTGCT	AATTCGTCCGGCCTACTTCT	227
RAD7-4	64557481	CGCGGTCTCTAACAACCTTC	TGTGGATGACCCTTCTAGGC	251
RAD7-5	64691867	TGATCGACCAACTCCAATCA	TCCAAGCACCTATTTCTGGG	155
RAD8-1	6233934	ATGAGATGACAATACCCGCC	CTGACCTCCCACCACCATAG	216
RAD8-2	6719636	CGTGTAGATGATGACGTGGC	CCATGTTGAATATTTGGACGG	278
RAD8-3	51347169	CTCTTGGAGATGCTCAAGCC	CTAGCCGAGGAGGTGAGGTC	222
RAD8-4	53214539	CCTCTTCGTCGTTGTCATCA	CTACCTCTGGTGGGCTGACT	281
RAD8-5	58362705	TCCTCTTCCTCACGCTTGTT	TCCCTTTGTCGGCTAACATC	177
RAD8-6	62207694	CTAGTGCTCCCTGGAAGCTG	CCACCTGTCGATGCAAGTAA	295
RAD9-1	878319	GTTTTGCGAGATTTTGGAGG	GTTTGGATTGGGTTGTTGCT	296
RAD9-2	7179550	GCATGCGGAGAGCATACATA	CACTTTTGGTCTTGCTGGGT	285
RAD9-3	47254063	CATGCATCGATCAGAAATGG	CGGTTCAAGAACCTGATGGT	299
RAD9-4	47926089	TATCCGGAACTCCGTACCTG	GAGTAGTTTTGCTGGGTCCG	247
RAD9-5	56167222	CCTTTAAGTGCCAAAACGGA	GAAGTTTTCAAGTGCGAGCC	252
RAD9-6	56178918	AAAATGGTCTGAAACCGTCG	CTCCACGGATGAAGAAGAGC	195
RAD9-7	57370577	GACGTGGCACCACTAGGAAT	CGCCAGATCTCGATCATTTT	162
RAD9-8	57406929	TTGACCTGCAGAACGAAGTG	GACTGCACATGTTGTCGCTT	172
RAD10-1	1886341	CAATCCTTCTCGAGGCTGAC	TTGGCTCCTGTGTTGACTTG	284
RAD10-2	2074176	GGACATTTATCGCTGCCTGT	GACTTCACATTTGGGGCTGT	269
RAD10-3	5837638	TGGTGAGAGAATTCCCGAAG	GGAGGTGAGTCTTTTTCCCC	269
RAD10-4	6396010	GCCGGGTATATAGACAGCCA	ATGGGAAAAACCACTTGCAG	196
RAD10-5	6529525	GCATGAGGAGGATTTGAGGA	AAAAGGGAGGGTGGAGAAGA	154
RAD10-6	45567986	ATGTAGACCTCAAGCCACGC	CTAACGCACTCTCGTCCTCC	233
RAD10-7	45995834	AGCTGCAGTAGACATCGCCT	TTTGGCTCACCTACTGGGTC	184
RAD10-8	49744622	GTAGAATTCGCTTCGTGCGT	ATTCGAATCCTCATGGTTGC	192
RAD10-9	51350993	ACATGCACTGTTTGGCACAT	TTGTCCTGGGATAAAGTGGC	223
RAD10-10	55091920	CCTGCAGAACGATTCACAGA	GGTGTTCACCGTCCTCAAGT	212
RAD10-11	57163771	GCTCGGTTTCGAACAGTAGG	GTGGGCCCTTCTCTCTCTCT	162

Table S3 Shannon's information Index and PIC value of the polymorphic 39 InDel markers for 42 sorghum and 6 sudangrass germplasms.

InDel primer	Forward primer	Reverse primer	Shannon's information index	PIC value
RAD1-5	ACTGTGATCCCAACAGAGCC	CGACAGCAACTACGTCCTCA	0.69	0.50
RAD1-7	AATTTGAGCCCAAGAGAGGG	GAGGAAGGAAAATTCCAGGC	0.77	0.46
RAD1-10	ATCCTAAGGCCTAAGCTCCG	TATGACGAGTGAGCAAACGC	0.81	0.46
RAD1-12	GGCAGATAGACGGGGTTGTA	GTCGTGAACCAGAGCTCCTC	0.58	0.39
RAD1-13	ATGGTTGCGACATGTTGTGT	ACTAGCCATATAACCGGCCC	0.67	0.48
RAD1-14	ATACTGCTTTGCCCACGTTC	ATCATACCGCCCTCCACATA	0.33	0.19
RAD2-3	CCACATCGTGATCCACACTC	GAATAGCCAAGCTCTGACCG	0.65	0.46
RAD2-5	CTGAACCACCACACCAAGTG	CTTTGTATCTGGCATGGCCT	0.29	0.15
RAD2-6	ACTCGTACGTCCGTCCAAAC	TAGGACGATAACCACTGCCC	0.69	0.50
RAD2-7	GCCCAAAGAACATGGTGAGT	CGTAGATCTTGCACACGACG	0.62	0.43
RAD2-10	GCATGATGGATCTTGGCTTT	CAGCACACAGCCTACTCAGC	0.46	0.28
RAD2-13	GTCTCGTCGTCGTCTCCTTC	GGCACGGAAACTAAATACGG	0.10	0.04
RAD2-14	GAGCAAGTAAACGACGACCG	GCTTTGTCCGCTTCATCTTC	1.05	0.63
RAD2-16	TTAACAAGCCCAAACCCAAG	GGAAAATCATAGAATGGGCG	0.23	0.12
RAD3-2	ATCTACCGAATCCAACTGCG	CAAGCTACTCGTTCCCAAGC	0.66	0.46
RAD3-4	AGAGGAAAGCGAAGGAGAGG	CCAGGACAGGAGATGACGAT	0.29	0.15
RAD3-5	TTCGTCTCGTCAGTCGTCAC	TAAGCTAGGCAGAGTCCCCA	1.09	0.66
RAD3-8	GCCAGTTGATGGTGTTTCCT	TGCATGGAGATTAACTTGCG	0.45	0.28
RAD3-12	GATCGATCCCTTCCTTCCTC	CGAAAGATTCTCCGCTCAAC	0.69	0.50
RAD4-4	TCTGCTTGGGTTCTTCACCT	GGCAACTGGGCTAAGACAAA	0.68	0.49
RAD4-5	GGAAGCGCAATTGGATAGAG	TGATCTGTTCCCTCCCTGTC	0.69	0.49
RAD4-7	AAAGGCAACTCGGTAGCTCA	CCAATATGGACTACGGGCTC	0.42	0.25
RAD4-10	ATCGATTCACGGCAAGAAAC	TCAAGCCATCTCCATTTTCC	0.61	0.42
RAD5-7	CAAGAGTTAACCGTGACGCA	ACGCCCTGTGTTTTTGTCAT	0.23	0.12
RAD5-8	GCCTCCTTAGTCGTCCCTCT	GCATTTGCTCTTTGCTGTGA	0.50	0.26
RAD5-9	TTGCCCACCAACCTTCTAAC	ATGGACATCCTCCCTCTCCT	0.14	0.06
RAD6-3	AAAGGGGGTTTAAATTTGCG	TCACGTTCTTCTTGTGTGGC	0.17	0.08
RAD6-4	CAGGGTGGGCTGCAAGTAT	GCATCCATCTCCAACAGGAT	0.44	0.27
RAD6-5	CTGAAAATTAGCCTGCCTGC	GCTAGAGAGCGATCGAGGAA	0.54	0.36
RAD7-1	TGCTGGAAAGCACAAAGCTA	GTTTTTCTTGTCCTCGCTGG	0.58	0.39
RAD7-2	GCGTCCACGTACACCTACAC	AAAGCTTCCTTTCTCCCTCG	0.33	0.19
RAD7-3	CTGTTGTCACAGCTGGTGCT	AATTCGTCCGGCCTACTTCT	0.42	0.25
RAD7-4	CGCGGTCTCTAACAACCTTC	TGTGGATGACCCTTCTAGGC	0.68	0.49
RAD7-5	TGATCGACCAACTCCAATCA	TCCAAGCACCTATTTCTGGG	0.51	0.33
RAD8-1	ATGAGATGACAATACCCGCC	CTGACCTCCCACCACCATAG	0.32	0.17
RAD8-5	TCCTCTTCCTCACGCTTGTT	TCCCTTTGTCGGCTAACATC	0.66	0.47
RAD9-1	GTTTTGCGAGATTTTGGAGG	GTTTGGATTGGGTTGTTGCT	0.69	0.50
RAD10-10	CCTGCAGAACGATTCACAGA	GGTGTTCACCGTCCTCAAGT	0.72	0.45
RAD10-11	GCTCGGTTTCGAACAGTAGG	GTGGGCCCTTCTCTCTCTCT	0.66	0.47