Available online at [www.sciencedirect.com](http://www.sciencedirect.com)

ScienceDirect



# BanSatDB, a whole-genome-based database of putative and experimentally validated microsatellite markers of three *Musa* species

Vasu Arora<sup>a</sup>, Neera Kapoor<sup>b</sup>, Samar Fatma<sup>a</sup>, Sarika Jaiswal<sup>a</sup>, Mir Asif Iquebal<sup>a</sup>, Anil Ra<sup>a</sup>, Dinesh Kumar<sup>a,\*</sup>

<sup>a</sup>Center for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

<sup>b</sup>School of Sciences, Indira Gandhi National Open University, New Delhi 110068, India

## ARTICLE INFO

### Article history:

Received 31 May 2017

Received in revised form 17

November 2017

Accepted 20 February 2018

Available online 7 March 2018

### Keywords:

Musaceae

Musa

Banana

Microsatellite

Short tandem repeat (STR)

Primers

## ABSTRACT

The genus *Musa* is one of three genera in the family *Musaceae*, which includes bananas and plantains, which are monocotyledonous plants. Bananas have valuable nutritional content of vitamin C, B6, minerals, and dietary fiber and are a rich food energy source, given that carbohydrates account for 22%–32% of fruit weight. Molecular markers are valuable for crop improvement and population genetics studies. The availability of whole-genome sequence and *in silico* approaches has revolutionized bulk marker discovery. We describe an online web genomic resource, BanSatDB (<http://webtom.cabgrid.res.in/bansatdb/>) having the highest number (>341,000) of putative STR markers from *Musa* genera so far, represented by three species: *M. acuminata* (110,000), *M. balbisiana* (107,000), and *M. itinerans* (124,000) from 11 chromosomes of each species. BanSatDB has also been populated with 580 validated STR markers from the published literature. It is based on a three-tier architecture using MySQL, PHP and Apache. The markers can be retrieved by use of multiple search parameters including chromosome number(s), microsatellite types (simple or compound), repeat nucleotides (1–6), copy number, microsatellite length, pattern of repeat motif, and chromosome location. These markers can be used for Distinctness, Uniformity and Stability (DUS) tests of variety identification and for marker assisted selection (MAS) in variety improvement and management. These STRs have also proved to be helpful in classification of *Musa* germplasm to distinguish individual accessions and in the development of a standardized procedure for genotyping. These markers can also be used in gene discovery and QTL mapping. The database represents a source of markers for developing and implementing new approaches for molecular breeding, which are required to enhance banana productivity.

© 2018 “Crop Science Society of China and Institute of Crop Science, CAAS”. Production and hosting by Elsevier B.V. on behalf of KeAi Communications Co., Ltd.

\* Corresponding author.

E-mail address: [dinesh.kumar@icar.gov.in](mailto:dinesh.kumar@icar.gov.in) (D. Kumar).

Peer review under responsibility of Crop Science Society of China and Institute of Crop Science, CAAS.

## 1. Introduction

The *Musa* genus belongs to one of three genera of the family *Musaceae* of monocotyledonous plants. The *Musa* genus contains >70 species, represented by bananas and plantains. Banana is the fourth most important food in the world after rice, wheat, and maize in terms of production [1]. Banana crops are grown in >135 countries, accounting for >130 Mt annual world production, one fourth of which is produced in India. More than 1000 varieties of bananas are known worldwide, categorized into 50 groups. Limited variety differentiation of banana by SSR and AFLP markers has been reported [2,3]. Banana productivity can be increased enormously by genetic improvement [4]. An attempt at variety differentiation by molecular markers has been made [2,3]. Intellectual property (IP) of banana varieties having commercial valuable epitopic genes (edible vaccine) are required [5]. In such a situation, STR-based profiling can be a rapid and economical protection of IP-containing banana varieties/lines. Markers are also required for genetic improvement including testing of ploidy level. With the availability of whole genome sequence, marker discovery can be done *in silico*. Such genomic resources can be made available online for convenient use by the global community.

Since the whole genome sequence of all three major species, *M. balbisiana* (550 Mb) [6], *M. acuminata* (523 Mb) [7], and *M. itinerans* (615 Mb) [8], have become available, their genomes can be used for mining of microsatellite repeats to be used as DNA markers. Such markers are useful in developing a standardized protocol to classify *Musa* germplasm and to distinguish individual accessions by genotyping. They can be used in studies of genetic diversity, germplasm characterization, and selection [9,10]. The wild relative of banana, *M. itinerans*, can be a valuable source of genomic resources for disease resistance and can be used in breeding programs [8].

STR markers flanking to the targeted genes can be used in an introgression program if there is an allele size difference between donor and recipient parental population. In segregating populations in variety improvement programs, such markers offer the advantage of rapidity and cost effectiveness in genotyping [11]. With the advent of next generation sequencing (NGS) technology, *in silico* STR discovery has drastically reduced the cost and time of genotyping [12]. Limited numbers of markers have been developed to classify *Musa* germplasm, distinguish individual accessions, estimate genetic diversity estimation, develop linkage maps, and perform MAS in *Musa* [9,10]. Flanking regions of STR are usually highly conserved and thus suitable for designing locus-specific primers [3,13]. An *in silico* approach can be used for whole-genome-based mining along with automated primer design at a desired STR locus of a specific chromosome carrying a gene of interest. Though limited numbers of validated markers have been reported, they have not been compiled in the form of a web genomic resource allowing with option to mine STR from all available genomes of *Musa* species.

The four existing Web genomic resources for the *Musa* genus include Banana Genome Hub (BGH) ([\[genome-hub.southgreen.fr/home1\]\(http://genome-hub.southgreen.fr/home1\)\) \[14\], Musa Marker Database \(MMDb\) \(<http://www.agrogene.ac.cn:8088/mumdb/index.html>\) \[15\], TropGENE Database \(TropGeneDB\) \(<http://tropgenedb.cirad.fr/tropgene/JSP/index.jsp>\) \[16\] and Musa Germplasm Information System \(MGIS\) \(<https://www.cropdiversity.org/mgis/gigwa>\) \[17\]. Although BGH and MGIS have SNP markers, they lack STR markers. The latest version of TropGeneDB \(2016\) contains 47,607 STRs mined from NGS sequences along with a set of 25 STRs from a study of Witherup et al. \[18\]. Among these STR records, only 625 include chromosome number or linkage group. TropGeneDB does not permit a user to select a target chromosome with a specific STR location for use of the STR as a marker closely linked to a gene targeted in a molecular breeding program. Being a static database, there is no option for mining of STRs along with primer designing for genotyping. MMDb has the limitation of containing <175 thousand STRs from only two genomes of \*Musa\* species: \*M. acuminata\* and \*M. balbisiana\*, and none from \*M. itinerans\*. Although all these three \*Musa\* species are previously reported by Rotchanapreeda et al. \[19\] but it is confined to limited 24 universal STR primer pairs, based on STR discovery by genomic library in one species \(\*M. balbisiana\*\).](http://banana-</a></p>
</div>
<div data-bbox=)

The present study was aimed at genome-wide mining of microsatellite repeats and development of a user-friendly database containing microsatellites from *M. balbisiana*, *M. acuminata*, and *M. itinerans* along with all previously validated markers. A further aim was to identify primers of desirable loci on specific chromosomes for rapid genotyping.

## 2. Materials and methods

### 2.1. Data collection

For short tandem repeat (STR) mining, nucleotide sequences of 11 chromosomes of *M. acuminata* DH Pahang v2 (473 Mb), the draft sequence of *M. balbisiana* (403 Mb), and the draft sequence assembly of *M. itinerans* (463 Mb) were downloaded from BGH. MicroSatellite identification tool (MISA) (<http://pgrc.ipk-gatersleben.de/misa/>) [20] and custom Perl scripts were used to mine microsatellite markers from each of the three genome. In order to overcome computational limitations of larger size, scripts were written to fragment genomic sequences, perform marker mining, and analyze output. Experimentally validated markers were mined from the literature, curated manually, and deposited in a database (Fig. 1).

### 2.2. Database development

BanSatDB is a Web-based relational database of microsatellites mined from the banana genome. It is an online relational database with a three-tier architecture consisting of client, middle, and database tiers (Fig. 2). It was developed with MySQL and includes tables for predicted STR and experimentally validated markers. The database is linked to a user-friendly Web interface designed with the open-

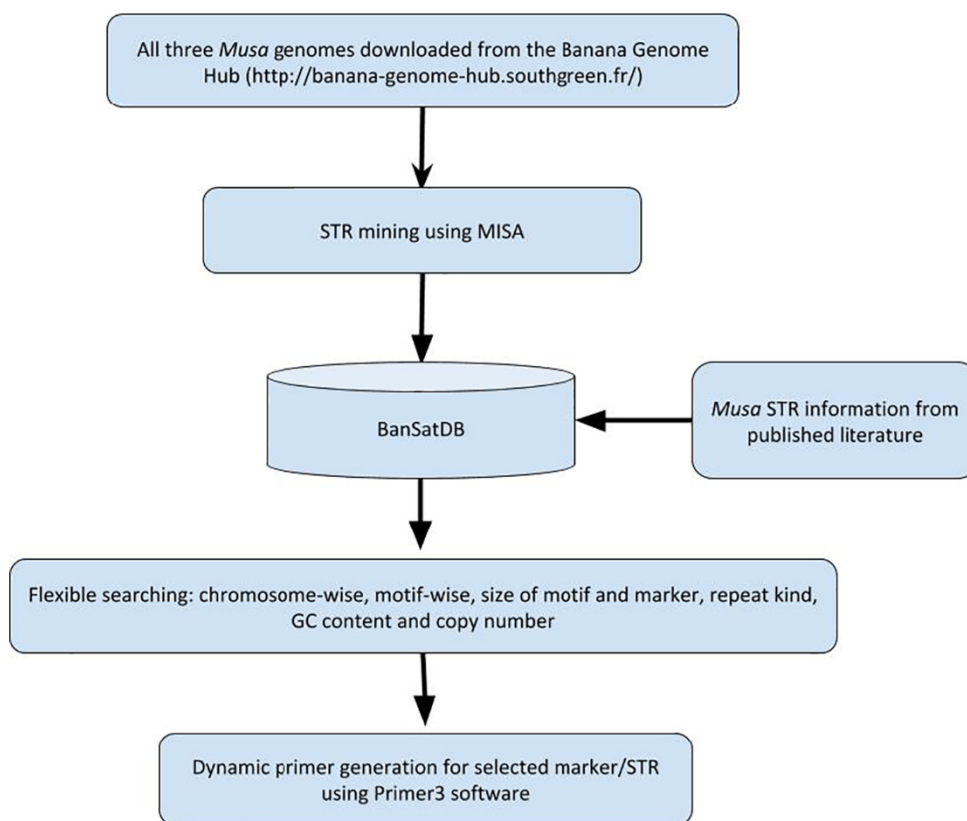


Fig. 1 – Workflow of BanSatDB.

source scripting language PHP. The Primer3 [21] standalone tool is integrated for designing forward and reverse primers for microsatellite sequences.

### 2.3. BanSatDB interface

The BanSatDB web server comprises of Home, About, Search, Analysis, Tutorial, and Team. The “Home” page describes about database and its potential use in molecular breeding. The “About” page describes the banana plant, its sequencing information, and the application of microsatellite markers. The “Analysis” page comprises genome-wide analysis of data that is shown graphically and in tabular form to present

different attributes of STR loci. “Search” is the main page, with two submenus: “Predicted STR” and “Experimentally Validated STR”. The “Predicted STR” search page provides user with two search criteria: chromosome and whole genome. A microsatellite search is further categorized by motif type (mono, di, tri, tetra, penta, or hexa), repeat type (by typing the repeat motif) and repeat kind (simple or compound). For more customized criteria like location of chromosome, range of GC % and copy number, there is an option of “Advance search” which is also having option for primer designing for genotyping. “Experimentally Validated STR” allows a user to retrieve polymorphic or monomorphic loci. It also includes a more customized search option by motif type (mono, di, tri, tetra,

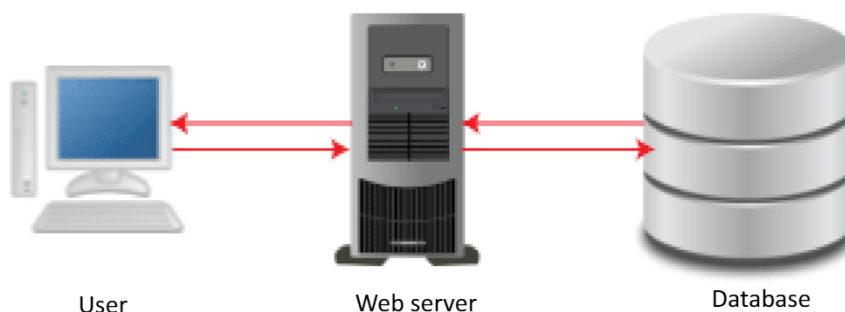


Fig. 2 – Three-tier architecture of BanSatDB.



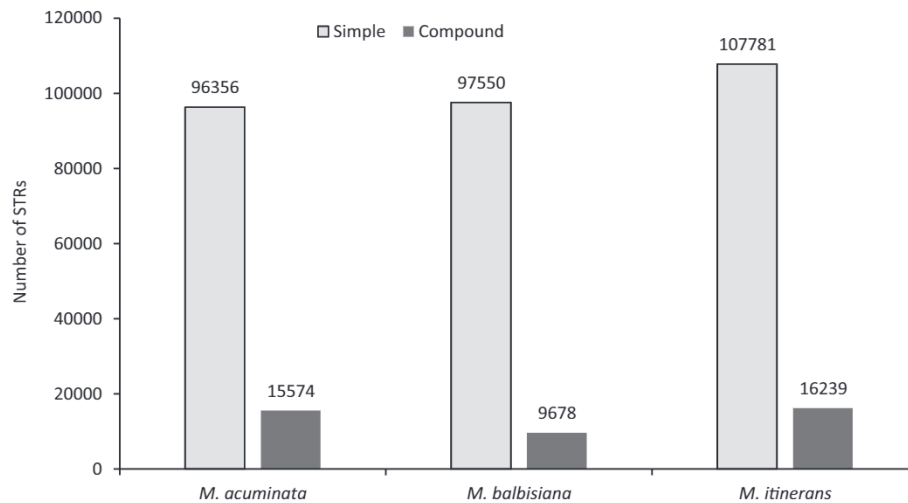


Fig. 3 – Distribution of microsatellite types in *Musa* genomes.

**Table 1 – Motif-type distribution of microsatellites in *Musa* genomes.**

Motif type	<i>M. acuminata</i>	<i>M. balbisiana</i>	<i>M. itinerans</i>
Mono	47,318 (49.11%)	39,164 (40.15%)	49,665 (46.08%)
Di	32,577 (33.81%)	40,932 (41.96%)	38,918 (36.11%)
Tri	14,649 (15.20%)	15,238 (15.62%)	17,097 (15.86%)
Tetra	1313 (1.36%)	1714 (1.76%)	1582 (1.47%)
Penta	307 (0.32%)	317 (0.32%)	312 (0.29%)
Hexa	192 (0.20%)	185 (0.19%)	207 (0.19%)

penta, and hexa), repeat type, and repeat kind (simple or compound).

#### 2.4. Transferability studies across species

Custom Perl scripts were written to study the transferability of STR markers across the three species of genus *Musa*. MISA

Tool and Primer3 were used to generate markers and primers, respectively, for all three genomes.

### 3. Results and discussion

Since *Musa* genome is of 523 Mbp having 11 chromosomes in the haploid set thus, we expect an average chromosome size of around 50 Mb [7]. Cultivars of *Musa* have been developed by crossbreeding between the two species, *M. acuminata* and *M. balbisiana*, harboring the A and B genomes, respectively [22]. Both are triploid, but apparently may also occur as diploid [7]. Recently sequenced genome of *M. itinerans* species which is a close relative to banana progenitors thus can be used for testing of cross species conservation by *in silico* approach [8]. The genomes of these cultivars have a complex genetic structure due to whole genome duplication and paralogous clustering in its course of evolution [6,23]. Using MISA, a total

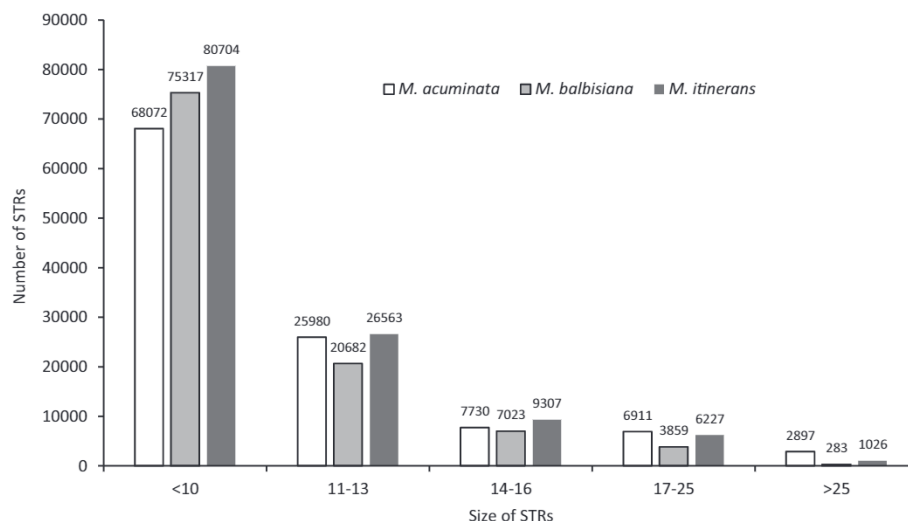


Fig. 4 – Frequencies of STRs based on their sizes in *Musa* genomes.

**Table 2 – Chromosome-wise abundance of STRs in the *M. acuminata* genome.**

Chromosome no.	Number of simple markers	Number of compound markers	Total markers	Percentage (%)	Size of chromosome (Mb)	Density (STR per Mb)
1	6816	1037	7853	7.02	29.07	270.14
2	6743	1085	7828	6.99	29.51	265.25
3	8917	1428	10,345	9.24	35.02	295.40
4	10,155	1731	11,886	10.62	37.11	320.33
5	9853	1587	11,440	10.22	41.85	273.34
6	9841	1667	11,508	10.28	37.59	306.12
7	8119	1282	9401	8.40	35.03	268.39
8	10,566	1720	12,286	10.98	44.89	273.70
9	9609	1519	11,128	9.94	41.31	269.40
10	8722	1417	10,139	9.06	37.67	269.12
11	7015	1101	8116	7.25	27.95	290.33

of 111,930, 107,228, and 124,020 microsatellite markers were identified in the *M. acuminata*, *M. balbisiana* genome, and *M. itinerans* genome sequences, respectively.

BanSatDB was populated with all of the microsatellites mined from the three genomes. Simple STR (86%) were more abundant than compound STR (14%) in all of the genomes (Fig. 3).

In *M. acuminata* and *M. itinerans* genome, among simple types, the “mono” repeat type (for example, (A)<sub>n</sub>) was most prevalent at 49.11% and 46.08%, respectively, followed by “di” repeats (for example, (CA)<sub>n</sub>), at 33.81 and 36.11%. In contrast, in *M. balbisiana*, the proportions of “mono” and “di” were 40.15% and 41.96%, respectively. Although the dinucleotide repeat type is observed abundantly in eukaryotes [24], mononucleotide repeats were the most abundant type in the *M. acuminata* and *M. itinerans* genomes. Since stringency of MISA parameters for detectability “mono” repeats were not compromised, thus higher abundance of mono-repeats might be due to the inherent limitation of the NGS technology used having sequencing error [25].

“Tri”- type repeats (for example, (CAG)<sub>n</sub>) constituted 15% of total markers in all three genomes, followed by “tetra” (1%–2%) and “penta”, and “hexa” types (<0.5%) (Table 1).

For >60% of the STR markers predicted in all three genomes, the length of the markers was <10 bp, followed by 11–13 bp with approximately 21%–22%, of total STR markers. Markers in the length range 14–25 bp accounted for 6%–7% of the total and those longer than 25 bp for 2%–3% (Fig. 4).

The chromosome-wise abundance of STRs and their motif frequencies in *M. acuminata* were determined. Total repeat content was proportional to chromosome length, as expected for the ubiquitously distributed STR markers [26]. Chromosome 8 harbored the most markers and chromosome 2 the fewest. Chromosome 4 showed the highest marker density (320.33 per Mb) and chromosome 2 the lowest (265.25 per Mb). The average densities were 237 and 266 markers per Mb for *M. acuminata* and *M. balbisiana*, respectively, exceeding the density in *Arabidopsis* (157 markers per Mb) [27]. Other crops with similar densities include cucumber (367 markers per Mb) [28], rice (370–490 markers per Mb) [29], poplar (485 markers per Mb) [30], and grape (487 markers per Mb) [31]. The STR markers are evenly distributed with homogeneity in terms of distance (Table 2).

The proportion of GC content in STR in the range in terms of percentage were: 0–25 was greatest (61%–66%) followed by the range 25–50 (24%–30%), with the lowest proportion in the ranges

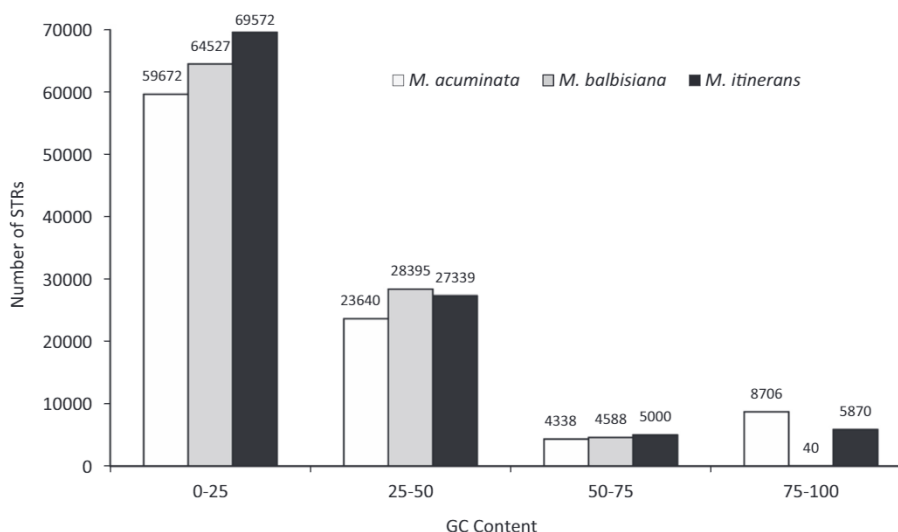
**Fig. 5 – GC content of STRs in *Musa* genomes.**



Fig. 6 – The flow of a database search in BanSatDb.

50–75 (4%–5%) in *M. acuminata* and *M. itinerans* genome and 75–100 (0.04%) in the *M. balbisiana* genome. Fig. 5 shows the distribution of microsatellite length in relation to GC proportion. No correlation was found between size of STR locus and GC content.

As the number of these predicted STR is large, geneticists and breeders may wish to select STR markers from specific genomic regions, suggesting the development of a database that enables the selection of STR markers. Many such STR databases have been developed, including for pigeon pea [32], chickpea [33], tomato [34], rice [35], maize, sorghum, soybean [36], and cotton [37]. User can mine the SSR markers on desired chromosome and search for experimentally verified SSRs along with primer generation in BanSatDb (Fig. 6).

### 3.1. Transferability of STR markers across species

Extent of conservation of STR loci were obtained successfully by computational approach depicting common loci across three genomes. This analysis revealed that out of a total of 207,983 STR markers obtained from all the three genomes, only 5979 (2.87%) were found to be conserved among all the three genomes. Two species comparison revealed that 13,649 and 17,020 STRs are

conserved between *M. acuminata* and *M. balbisiana* and *M. acuminata* and *M. itinerans* genomes, respectively (Fig. 7).

The results suggest that these transferable STR markers could be used for identification of markers associated with specific traits in other *Musa* species. EST-STR markers were found to be 100% transferable among the eight *Musa* genomes in a study [38]. Similar tests have been made among species of sugarcane [39], grapes [40], sunflower [41], and pine [42]. Cross-taxon transferability of STR markers has also been observed among *Musa* and related species [15].

### 3.2. BanSatDB utility

BanSatDB supports a microsatellite search in any of the three sequenced genomes: *M. acuminata*, *M. balbisiana*, and *M. itinerans*, using multiple parameters including microsatellite type (simple or compound), repeat type (mono- to hexanucleotide), copy number, length of marker, pattern of repeat motif, and chromosome location of a marker. Chromosome-wise microsatellites can be retrieved from any of the three *Musa* genomes species. Once the marker is located, all its information (chromosome, contig, or scaffold



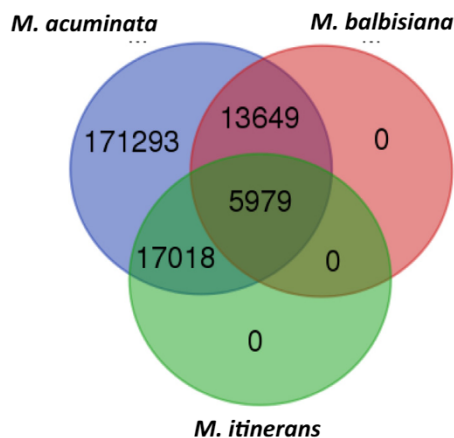


Fig. 7 – Transferability of STR across the three *Musa* genomes.

number, repeat kind, motif, repeat type, length of repeat, size of repeat, start of repeat, end of repeat, STR sequence, and GC content) are displayed on the result page. BanSatDB also supports this customized search on ranges of GC content, location (base pair length) and copy number.

User can view neighboring markers just by clicking on a chromosome, contig, or scaffold number. Users are also provided the option to replace degenerate bases with any alternative base (A, T, G, or C) and specify the length of the flanking sequence. The feature has been added to resolve the issue of degeneracy of bases present in genome assemblies, making primer design difficult without such an option.

BanSatDB has been developed as a dynamic database that allows users to select STR loci having high probabilities of polymorphism, saving the time and cost of wet lab-based polymorphism discovery. Given that STR polymorphism depends on type, length, and structure of repeats (simple, compound, interrupted) [43], flexible provisions have been made for STR mining. It is well known that the longer the STR repeat array, the higher is the probability of finding a large allele number or high degree of polymorphism in a given gene pool. Simple repeats are faster-mutating than compound or interrupted repeats. Given that STR alleles are generated by slippage in DNA replication, a longer array of STR loci mutates faster. For dinucleotide repeats, the threshold value for such slippage events starts with STR loci containing >8 repeat unit [44]. For example, CA<sub>(20)</sub> will mutate faster than CA<sub>(10)</sub>. User may select a larger repeat array that is closely linked to or lies in the flanking regions of their gene of interest for genotyping. The option to specify amplicon size further enables multiplexing of genotyping with fluorescent dyes, supporting economical and accurate allele sizing during generation of STR allelic data [45]. This genomic resource can be valuable for Distinctness, Uniformity and Stability (DUS) tests of variety identification and for MAS in varietal improvement and germplasm management [34].

### 3.3. BanSatDB as a tool to mine experimentally validated STR-primer pairs

Validated STR markers along with their forward and reverse primers previously reported in the literature can be retrieved

using BanSatDB. This retrieval can be further customized by option to select polymorphic/monomorphic along with type, repeat and kind of motif. The result page shows tabulated information of the markers with motif details, polymorphism, forward and reverse primer pairs, and references and links to the publications in which the markers were reported.

## 4. Conclusions

Using three *Musa* genomes publicly available, we have developed an online web genomic resource, BanSatDB (<http://webtom.cabgrid.res.in/bansatdb/>) that catalogues highest number (>341,000) of putative STR markers reported so far. It also has compilation of 580 validated STR markers which are already published. This web-genomic resource has provision to select marker based on chromosome location, microsatellite types, repeat nucleotides, copy number, microsatellite length and pattern of repeat motif. These markers can be used for genotyping in DUS tests for varietal identification and also for MAS in germplasm improvement and management.

## Acknowledgments

The authors thank the Indian Council of Agricultural Research (ICAR), Ministry of Agriculture and Farmers' Welfare, Government of India for providing all facilities. The authors also thank the Director, ICAR-IASRI, New Delhi for the use of the ASHOKA computational facility where all analyses were performed using public-domain data. English editing help of Prof. James Nelson, Kansas State University, USA is acknowledged with sincere thanks.

## REFERENCES

- [1] E.A. Frison, J.V. Escalant, S. Sharrock, The global *Musa* genomic consortium: a boost for banana improvement, in: S. M. Jain, R. Swennen (Eds.), *Banana Improvement: Cellular, Molecular Biology, and Induced Mutations*, Science Publishers, Enfield, New Hampshire, USA 2004, pp. 341–349.

- [2] A. Lamare, S.R. Rao, Efficacy of RAPD, ISSR and DAMD markers in assessment of genetic variability and population structure of wild *Musa acuminata* colla, *Physiol. Mol. Biol. Plants* 21 (2015) 349–358, <https://doi.org/10.1007/s12298-015-0295-1>.
- [3] P. Christelová, M. Valárik, E. Hřibová, I. Van den Houwe, S. Channelière, N. Roux, J. Doležel, A platform for efficient genotyping in *Musa* using microsatellite markers, *AoB Plants* 2011 (2011), plr024.
- [4] M. de Vries, C.A. Ferwerda, J.D. Flach, Choice of food crops in relation to actual and potential production in the tropics, *Netherlands, J. Agric. Sci.* 15 (1967) 241–248.
- [5] K.L. Durell, Intellectual Property Protection for Plant Derived Vaccine Technology: Here They Come are we Ready or Not? *Lex Electronica*, Vol. 10, No. 3, Hiver/Winter, [http://www.lex-electronica.org/files/sites/103/10-3\\_durell.pdf](http://www.lex-electronica.org/files/sites/103/10-3_durell.pdf) 2006, Accessed date: 16 November 2017.
- [6] A. D'Hont, F. Denoeud, J.M. Aury, F.C. Baurens, F. Carreel, O. Garsmeur, B. Noel, S. Bocs, G. Droc, M. Rouard, C. Da Silva, K. Jabbari, C. Cardi, J. Poulain, M. Souquet, K. Labadie, C. Jourda, J. Lengellé, M. Rodier-Goud, A. Alberti, M. Bernard, M. Correa, S. Ayyampalayam, M.R. McKain, J. Leebens-Mack, D. Burgess, M. Freeling, D. Mbéguié-A-Mbéguié, M. Chabannes, T. Wicker, O. Panaud, J. Barbosa, E. Hribova, P. Heslop-Harrison, R. Habas, R. Rivallan, P. Francois, C. Poirion, A. Kilian, D. Burthia, C. Jenny, F. Bakry, S. Brown, V. Guignon, G. Kema, M. Dita, C. Waalwijk, S. Joseph, A. Dievert, O. Jaillon, J. Leclercq, X. Argout, E. Lyons, A. Almeida, M. Jeridi, J. Dolezel, N. Roux, A. M. Risterucci, J. Weissenbach, M. Ruiz, J.C. Glaszmann, F. Quétier, N. Yahiaoui, P. Wincker, The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants, *Nature* 488 (2012) 213–217.
- [7] M.W. Davey, R. Gudimella, J.A. Harikrishna, L.W. Sin, N. Khalid, J. Keulemans, A draft *Musa balbisiana* genome sequence for molecular genetics in polyploid, inter- and intra-specific *Musa* hybrids, *BMC Genomics* 14 (2013) 683.
- [8] W. Wu, Y. Yang, W. He, M. Rouard, W. Li, M. Xu, Whole genome sequencing of a banana wild relative *Musa itinerans* provides insights into lineage-specific diversification of the *Musa* genus, *Sci. Rep.* 6 (2016), 31586.
- [9] B.J. Till, J. Jankowicz-Cieslak, L. Sági, O.A. Huynh, H. Utsushi, R. Swennen, R. Terauchi, C. Mba, Discovery of nucleotide polymorphisms in the *Musa* gene pool by Ecotilling, *Theor. Appl. Genet.* 121 (2010) 1381–1389.
- [10] G.L. Jiang, Molecular markers and marker-assisted breeding in plants, in: S.B. Andersen (Ed.), *Plant Breeding from Laboratories to Fields*, InTech, 2013, <https://doi.org/10.5772/52583> (Available from) <https://www.intechopen.com/books/plant-breeding-from-laboratories-to-fields/molecular-markers-and-marker-assisted-breeding-in-plants>.
- [11] A.D. Twyford, R.A. Ennos, Next-generation hybridization and introgression, *Heredity* 108 (2012) 179–189.
- [12] J.E. Zalapa, H. Cuevas, H. Zhu, S. Steffan, D. Senalik, E. Zeldin, B. McCown, R. Harbut, P. Simon, Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences, *Am. J. Bot.* 99 (2012) 193–208.
- [13] H. Ellegren, Microsatellites: simple sequences with complex evolution, *Nat. Rev. Genet.* 5 (2004) 435–445.
- [14] G. Droc, D. Larivière, V. Guignon, N. Yahiaoui, D. This, O. Garsmeur, A. Dereeper, C. Hamelin, X. Argout, J.F. Dufayard, J. Lengelle, F.C. Baurens, A. Cenci, B. Pitollat, A. D'Hont, M. Ruiz, M. Rouard, S. Bocs, The banana genome hub, *Database* 2013 (2013), bat035.
- [15] M.K. Biswas, Y. Liu, C. Li, O. Sheng, C. Mayer, G. Yi, Genome-wide computational analysis of *Musa* microsatellites: classification, cross-taxon transferability, functional annotation, association with transposons & miRNAs, and genetic marker potential, *PLoS One* 10 (2015), e0131312.
- [16] C. Hamelin, G. Sempere, V. Jouffe, M. Ruiz, TropGeneDB, the multi-tropical crop information system updated and extended, *Nucleic Acids Res.* 41 (2013) D1172–D1175.
- [17] M. Ruas, V. Guignon, G. Sempere, J. Sardos, Y. Hueber, H. Duvergey, A. Andrieu, R. Chase, C. Jenny, T. Hazekamp, B. Irish, K. Jelali, J. Adeka, T. Ayala-Silva, C.P. Chao, J. Daniells, B. Dowiya, B. Effa effa, L. Gueco, L. Herradura, L. Ibobondji, E. Kempenaers, J. Kilangi, S. Muhangi, P. Ngo Xuan, J. Paofa, C. Pavis, D. Thiemele, C. Tossou, J. Sandoval, A. Sutanto, G. Vangu Paka, G. Yi, I. Van den Houwe, N. Roux, M. Rouard, MGIS: managing banana (*Musa* spp.) genetic resources information and high-throughput genotyping data, *Database* 2017 (2017) (bax046).
- [18] C. Witherup, D. Ragone, T. Wiesner-Hanks, B. Irish, B. Scheffler, S. Simpson, F. Zee, M.I. Zuberi, N.J.C. Zerega, Development of microsatellite loci in *Artocarpus altilis* (Moraceae) and cross-amplification in congeneric species, *Appl. Plant Sci.* 1 (2013), 1200423.
- [19] T. Rotchanapreeda, S. Wongniam, S.C. Swangpol, P.P. Chareonsap, N. Sukkaewmanee, J. Somana, Development of SSR markers from *Musa balbisiana* for genetic diversity analysis among Thai bananas, *Plant Syst. Evol.* 302 (2016) 739–761.
- [20] T. Thiel, W. Michalek, R. Varshney, A. Graner, Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.), *Theor. Appl. Genet.* 106 (2003) 411–422.
- [21] S. Rozen, H. Skaletsky, Primer3 on the WWW for general users and for biologist programmers, in: S. Misener, S.A. Krawetz (Eds.), *Methods in Molecular Biology, Bioinformatics Methods and Protocols*, Vol. 132, Humana Press Inc, Totowa, New Jersey, USA 2000, pp. 365–386.
- [22] N. Simmonds, *The Evolution of the Bananas*, Longmans, London, UK, 1962.
- [23] R. Ortiz, R. Swennen, From crossbreeding to biotechnology-facilitated improvement of banana and plantain, *Biotechnol. Adv.* 32 (2014) 158–169.
- [24] C. Shioiri, N. Takahata, Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry, *J. Mol. Evol.* 53 (2001) 364–376.
- [25] G. Haseneyer, T. Schmutzer, M. Seidel, R. Zhou, M. Mascher, C.C. Schön, S. Taudien, U. Scholz, K.F. Mayer, E. Bauer, From RNA-seq to large-scale genotyping-genomics resources for rye (*Secale cereale* L.), *BMC Plant Biol.* 11 (2011) 131.
- [26] E. Zietkiewicz, A. Rafalski, D. Labuda, Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification, *Genomics* 20 (1994) 176–183.
- [27] M.J. Lawson, L. Zhang, V. Tech, Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes, *Genome Biol.* 7 (2006) R14.
- [28] P.F. Cavagnaro, D.A. Senalik, L. Yang, P.W. Simon, T.T. Harkins, C.D. Kodira, S. Huang, Y. Weng, Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus* L.), *BMC Genomics* 11 (2010) 569.
- [29] I.R.G. Sequencing Project, The map-based sequence of the rice genome, *Nature* 436 (2005) 793–800.
- [30] S. Li, T. Yin, Map and analysis of microsatellites in the genome of *Populus*: the first sequenced perennial plant, *Sci. China Ser. C Life Sci.* 50 (2007) 690–699.
- [31] J. Chen, N. Wang, L.C. Fang, Z.C. Liang, S.H. Li, B.H. Wu, Construction of a high-density genetic map and QTLs mapping for sugars and acids in grape berries, *BMC Plant Biol.* 15 (2015) 28.
- [32] V. Arora Sarika, M.A. Iquebal, A. Rai, D. Kumar, PIPEMicroDB: microsatellite database and primer generation tool for pigeonpea genome, *Database* 2013 (2013), bas054.
- [33] D. Doddamani, M.A. Katta, A.W. Khan, G. Agarwal, T.M. Shah, R.K. Varshney, CicArMiSatDB: the chickpea microsatellite database, *BMC Bioinf.* 15 (2014) 212.



- [34] M.A. Iquebal, V. Arora, N. Verma, A. Rai, D. Kumar, First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification, *BMC Plant Biol.* 13 (2013) 197.
- [35] B. Jayashree, R. Punna, P. Prasad, K. Bantte, C.T. Hash, S. Chandra, D.A. Hoisington, R.K. Varshney, A database of simple sequence repeats from cereal and legume expressed sequence tags mined *in silico*: survey and evaluation, *In Silico Biol.* 6 (2006) 607–620.
- [36] Z. Zhao, C. Guo, S. Sutharzan, P. Li, C.S. Echt, J. Zhang, C. Liang, Genome-wide analysis of tandem repeats in plants and green algae, *G3 Genes Genomes Genet.* 4 (2014) 67–78.
- [37] A. Blenda, J. Scheffler, B. Scheffler, M. Palmer, M. Lacape, J.Z. Yu, C. Jesudurai, S. Jung, S. Muthukumar, P. Yellambalase, S. Ficklin, M. Staton, R. Eshelman, M. Ulloa, S. Saha, B. Burr, S. Liu, T. Zhang, D. Fang, A. Pepper, D. Main, CMD: a cotton microsatellite database resource for *Gossypium* genomics, *BMC Genomics* 7 (2006) 132.
- [38] S. Backiyarani, S. Uma, P. Varatharaj, M.S. Saraswathi, Mining of EST-SSR markers of *Musa* and their transferability studies among the members of order the Zingiberales, *Appl. Biochem. Biotechnol.* 169 (2013) 228–238.
- [39] G.M. Cordeiro, R. Casu, C.L. McIntyre, J.M. Manners, R.J. Henry, Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *erianthus* and *sorghum*, *Plant Sci.* 160 (2001) 1115–1123.
- [40] V. Decroocq, M.G. Favé, L. Hagen, L. Bordenave, S. Decroocq, Development and transferability of apricot and grape EST microsatellite markers across taxa, *Theor. Appl. Genet.* 106 (2003) 912–922.
- [41] J.R. Ellis, J.M. Burke, EST-SSRs as a resource for population genetic analyses, *Heredity* 99 (2007) 125–132.
- [42] D. Chagné, P. Chaumeil, A. Ramboer, C. Collada, A. Guevara, M.T. Cervera, G.G. Vendramin, V. Garcia, J.M. Frigerio, C. Echt, T. Richardson, C. Plomion, Cross-species transferability and mapping of genomic and cDNA SSRs in pines, *Theor. Appl. Genet.* 109 (2004) 1204–1214.
- [43] Y.D. Kelkar, N. Strubczewski, S.E. Hile, F. Chiaromonte, K.A. Eckert, K.D. Makova, What is a microsatellite: a computational and experimental definition based upon repeat mutational behavior at A/T and GT/AC repeats, *Genome Biol. Evol.* 2 (2010) 620–635.
- [44] Y.D. Kelkar, K.A. Eckert, F. Chiaromonte, K.D. Makova, A matter of life or death: how microsatellites emerge in and vanish from the human genome, *Genome Res.* 21 (2011) 2038–2048.
- [45] M.J. Hayden, T.M. Nguyen, A. Waterman, K.J. Chalmers, Multiplex-ready PCR: a new method for multiplexed SSR and SNP genotyping, *BMC Genomics* 9 (2008) 80.