

Article

Not peer-reviewed version

---

# MMdb: Genomic Resource for the Improvement of the Family Musaceae

---

[Manosh Kumar Biswas](#) <sup>\*</sup>, [Dhiman Biswas](#), [Ou Sheng](#), [Ganjun Yi](#), [Guiming Deng](#)

Posted Date: 28 December 2023

doi: [10.20944/preprints202312.2134.v1](https://doi.org/10.20944/preprints202312.2134.v1)

Keywords: Microsatellite; SNP; ILP; database; *Musa* sp.



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

## Article

# MMdb: Genomic Resource for the Improvement of the Family MUSACEAE

Manosh Kumar Biswas <sup>1,3,\*</sup>, Dhiman Biswas <sup>2</sup>, Ou Sheng <sup>3</sup>, Ganjun Yi <sup>3</sup> and Guiming Deng <sup>3</sup>

<sup>1</sup> Department of Genetics and Genome Biology, University of Leicester, Leicester LE1 7RH, UK

<sup>2</sup> Department of Computer Science and Engineering, Maulana Abul Kalam Azad University of Technology, Kolkata, West Bengal, India

<sup>3</sup> Institute of Fruit Tree Research, Guangdong Academy of Agricultural Sciences, Tianhe District, Guangzhou 510640, China

\* Correspondence: manosh24@yahoo.com

**ABSTRACT:** The availability of genomic sequences and bioinformatics tools has facilitated genome-scale marker discovery. Microsatellite (SSR), Single Nucleotide Polymorphism (SNP), and Intron Length Polymorphism (ILP) markers are widely used in crop improvement and population genetics studies, developed using extensive bioinformatics pipelines. Here we introduce the Musa Molecular Marker Database (MMdb) as a valuable online resource for enhancing Musa species. It is a diverse array of molecular markers resources, including SSR, SNP, and ILP. This database contains 2115474 SSR, 63588 SNP, and 91547 ILP markers developed from twelve Musa species and three of its relative species. This online resource is freely accessible, following a 'three-tier architecture,' and organizes marker information in MySQL tables. It has user-friendly interface, that is written in JavaScript, PHP, and HTML code. Users can employ flexible search parameters, including marker location in the chromosome, transferability, polymorphism, and functional annotation, among others. This unique feature sets MMdb apart from existing marker databases, offering a novel approach that caters to the needs of the Musa research community. Despite being an *in silico* method, searching for markers based on various attributes holds promise for Musa research. These markers serve various purposes, including germplasm characterization, gene discovery, population structure analysis, and QTL mapping.

**Keywords:** microsatellite; SNP; ILP; database; Musa sp.

## 1. Introduction

Molecular markers are widely used in modern agriculture for crop improvement programs. They have a wide range of applications in plant breeding, including germplasm characterization, parent selection, genetic diversity [1] [2], population structure [3], trait tagging, etc. The easy accessibility of the genome sequence and molecular marker mining tools has revalorized genome-scale marker discovery and subsequent utility. Among the different types of molecular markers, SSR, SNP, and ILP markers are frequently used in plant breeding due to their genomic abundance, easy assay techniques, and reproducibility. The developments of genome scale SSR, SNP and ILP markers are straightforward, and it has become routine work for molecular breeders and bioinformaticians. Genome-wide molecular markers, especially SSR markers have been developed for many of the sequenced plant genomes, but the molecular marker databases have been created for only a few of them. For example, the Cotton Microsatellite Database [4] (CMD <http://www.cottonssr.org>) was created to present 5,484 microsatellites markers that were developed from nine major cotton microsatellite projects. Kazusa Marker DataBase [5] (KMD, <http://marker.kazusa.or.jp/>) was developed to provide linkage-map, physical-maps, and information on ~68000 SSR and ~1400000 SNP primers for 14 agronomical important crop species. This database provides forward and reverse primer sequences, as well as a unique primer id linked with additional informant about the primer. The NABIC marker database was developed base on 7250 markers from published articles on different crop plants [6]. Each marker in this database contains information about the marker name, gene definition, general marker information and expressed sequence tag number. The Pigeon Pea

Microsatellite Data Base (PIPEMicroDB) catalogs 123387 short tandem repeat information from the pigeon pea genome [7]. The Foxtail millet Marker Database (FmMDB) is an online searchable and downloadable database developed from Foxtail millet genome sequences. This database catalogs 21315 genomic SSRs, 447 genic SSRs and 96 ILP markers information [8]. Chickpea Microsatellite Database (CicArMiSatDB <http://cicarmisatdb.icrisat.org>) is a relational database created by mining Chickpea genome and it provides information on SSRs along with their features including genomic coordination's, primer-pairs, annealing temperature, and SSR repeat motif etc [9]. The PIP (Potential Intron Polymorphism) marker database developed based on 57658 ILP markers from 59 plant species. PIP database doesn't include any ILP markers from *Musa* and its sister species [10]. The 'SSRome' database represents 45.1 million microsatellite markers across all taxa including plant, metazoa, archaea, bacteria, virus, fungi and protozoa. Users can explore microsatellites from 6533 organisms including nuclear, mitochondrial and chloroplast genomes [11]. All the makers in this database are classified as genic or non- genic. The pineapple genomics database (PGD) is a core online platform that integrates genomic, transcriptomic and genetic marker (SSR, SNP and ILP) data for pineapple [12]. Lily genomic database represents SSR and SNP marker for the lily species [13].

Until today, many online databases have been developed to store genomic data in an accessible platform. Among these, some focus on different types of molecular markers, some store only genomic sequences data, while others integrate both marker and sequence data. Most of the molecular marker databases deals with SSR markers as described earlier section of this article. Although some of the existing marker databases integrate molecular markers from *Musa* sp. among them SSRome [11], PMDBase [14], Banana Genome Hub [15] and BanSatDb [16] marker database are notable. SSRome and PMDBase only integrate SSR marker data from the *Musa acuminata* genome; BGH and TropGENE store only *Musa* SNP markers. BanSatDb database was created for *Musa* SSR markers and it allows for the design SSR markers from the three genomes of *Musa* species: *M. acuminata*, *M. balbisiana*, and *M. itinerans*. BanSatDb does not permit to search of primers from the *M. schizocarpa* genome. All the existing marker databases have many limitations and their major deficiency are (i) lacking the classification of SSR loci as classI(>20bp) or classII(<20bp), (ii) missing polymorphism and transferability information, which are key attributes for selecting high therapeutic markers in *in silico* methods, (iii) lack of associated gene function information, and (iv) missing SSR motif information such as motif richness (as AT, GC or AT/GC balance), motif located in CDS or UTR etc. Over all, existing marker databases generally present a limited set of information for each marker, such as forward and reverse primer sequences, SSR types, primer annealing temperature, and PCR product size. Moreover, search parameters are often inflexible and narrow. To address these limitations, we developed a comprehensive marker database called MMdb that integrates a large number of SSR, SNP, and ILP markers, each with ~20 unique features. This database covers 12 genomes of the *Musa* sp. and provides a more robust platform for marker selection and analysis.

## 2. Materials and Methods

### 2.1. Data Collection and Processing

The whole genome sequences of 12 banana species (Listed in Table S1) were downloaded from Musa Genome Hub (<https://banana-genome-hub.southgreen.fr/species-list>). A total of 25261 Banana GSS sequences were obtained from the NCBI database. EST sequences of *Musa* spp. were obtained from the EST Tool Kit and NCBI databases. All EST sequences were merged into a single FASTA file and then the "est\_trimmer.pl" tool ([http://pgrc.ipk-gatersleben.de/misa/download/est\\_trimmer.pl](http://pgrc.ipk-gatersleben.de/misa/download/est_trimmer.pl)) was used to remove low-quality sequences, poly A/T and low-complexity regions at the 5' and 3' ends. The resulting sequences were then assembled to remove redundant sequences using CD-Hit with default parameters and the non-redundant sequences were used for subsequent analysis.

### 2.2. Marker Development

**SSR markers:** The microsatellite (SSR) markers were developed using a pipeline called 3GMAT (<https://github.com/mkbcit/3GMAT>), with a minimum of 12nt long SSR loci. SSR loci were

characterized based on (i) the length of repeat motifs (Class I >20 bp, Class II ≤20 bp) and (ii) the nucleotide composition of repeat motifs (AT-rich, GC-rich and AT/GC balance). The E-PCR strategy was applied to assess the *in-silico* transferability and polymorphism analysis of the developed SSR markers.

**SNP markers:** The EST sequences were mapped on the *Musa acuminata* genome using Bowtie2 with default parameters, and then samtools was used to extract SNP and indels. A Perl script was used to extract flanking sequences of 400bp in length, including 200bp upstream and 200bp downstream of the SNP loci. SNP primer pairs were designed using Primer3 with default parameters. The Functional annotation and gene association of the developed SNP markers were estimated using Blast2Go analysis.

**ILP markers:** Intron information was extracted from the gff3 file of the annotated *Musa* genomes. Then, a Perl script was used to retrieve intron flanking regions along with the 100 bp on each side of the targeted introns. Subsequently, intron lengths ranging from 100 to 3000 bp were extracted for primer design. Primer3 software with default parameters was used to design the primers. The E-PCR method was used for the *in-silico* transferability assessment of the ILP markers.

### 2.3. User Interface and Database Construction

MMdb is a web-based interactive, searchable, downloadable and relational database server, developed using MySQL 5.0 ([www.mysql.com](http://www.mysql.com)), and it has tree-tiers: a client-tier, middle-tier, and database tier. All the developed Markers with their attributes have been stored in a MySQL database, which is accessed through PHP and Apache. User friendly web interface designed using HTML5, Bootstrap4, CSS, JavaScript, and jQuery.

## 3. Result

### 3.1. Markers

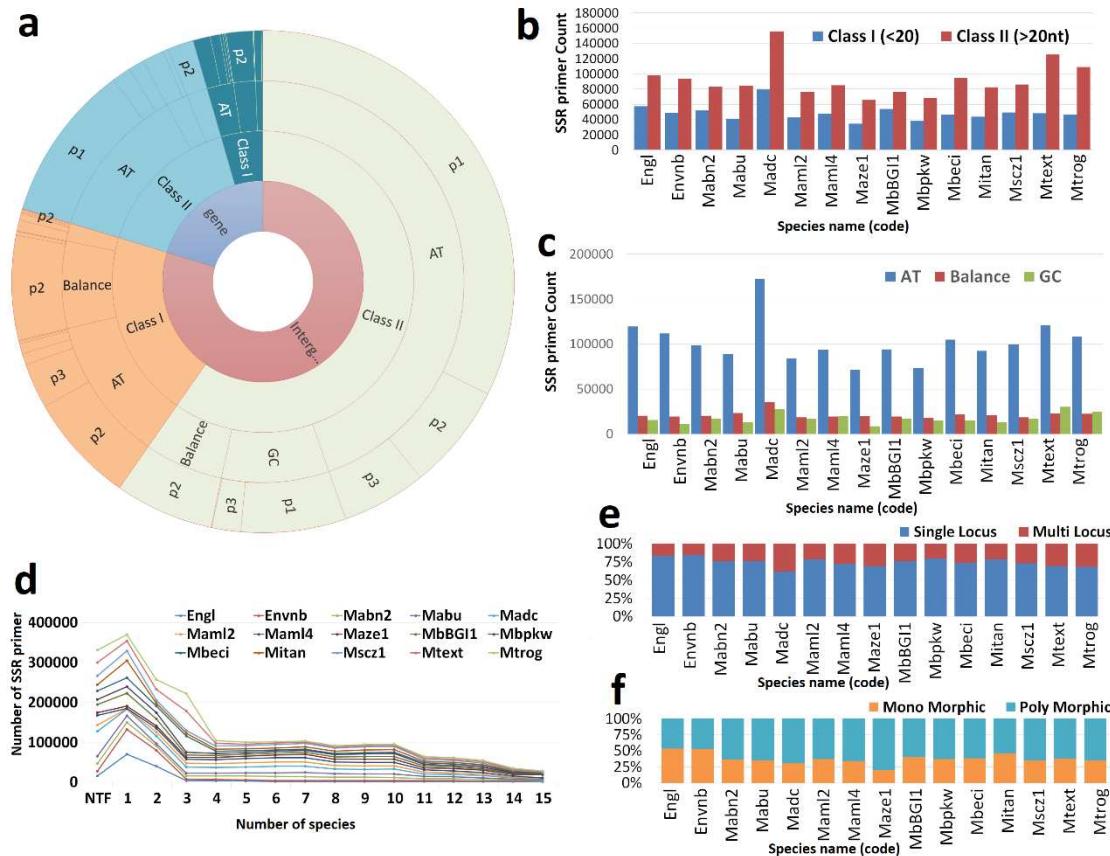
We conducted a genome-wide analysis to identify, distribute, and classify simple sequence repeats (SSRs) using 15 whole genome assemblies: 13 from *Musa* sp and 2 from *Ensete* sp, as detailed in Table S1. In total, we identified 2,761,190 SSRs, encompassing various types of desirable repeat motifs ranging from mono- to hexanucleotide repeats in the 15 datasets (see Table S1). The SSR densities exhibited variations across the tested genomes, with a range of 1,425 to 5,718 base pairs per one SSR. An average SSR was found at every 3,085 base pairs in the genome. Class II SSRs (motif length less than 20 base pairs) were more prevalent than Class I SSRs (motif length greater than 19 base pairs) in all *Musa* species. Among the tested genomes, AT-rich SSR motifs were dominant in comparison to other types of SSR motifs, including GC-rich and AT/GC-balanced motifs. All the identified SSRs were utilized for SSR marker development, resulting in 2,115,474 SSR loci successfully developed as SSR markers, representing 77% of the total SSRs. The remaining SSRs failed to develop markers, possibly due to the absence of perfect genome fragments, low-quality base representation, or insufficient flanking regions to generate primer sequences (Table 1.).

**Table 1.** Summary of SSR marker development and characterizations.

List of the genome	Genome code	Total Number of SSR	Primer design (count)	%
<i>Musa acuminata</i> banksii	Mabn2	183911	135187	74
<i>Musa acuminata</i> Dwarf_Cavendish	Madc	267698	235211	88
<i>Musa acuminata</i> burmannica	Mabu	141919	125946	89
<i>Musa acuminata</i> malaccensis V2	Maml2	147255	118834	81
<i>Musa acuminata</i> malaccensis V4	Maml4	185328	132721	72
<i>Musa acuminata</i> zebrina	Maze1	111705	100182	90
<i>Musa balbisiana</i> BGI11	MbBGI1	320858	130377	41
<i>Musa balbisiana</i> pkw	Mbpkw	131403	106302	81

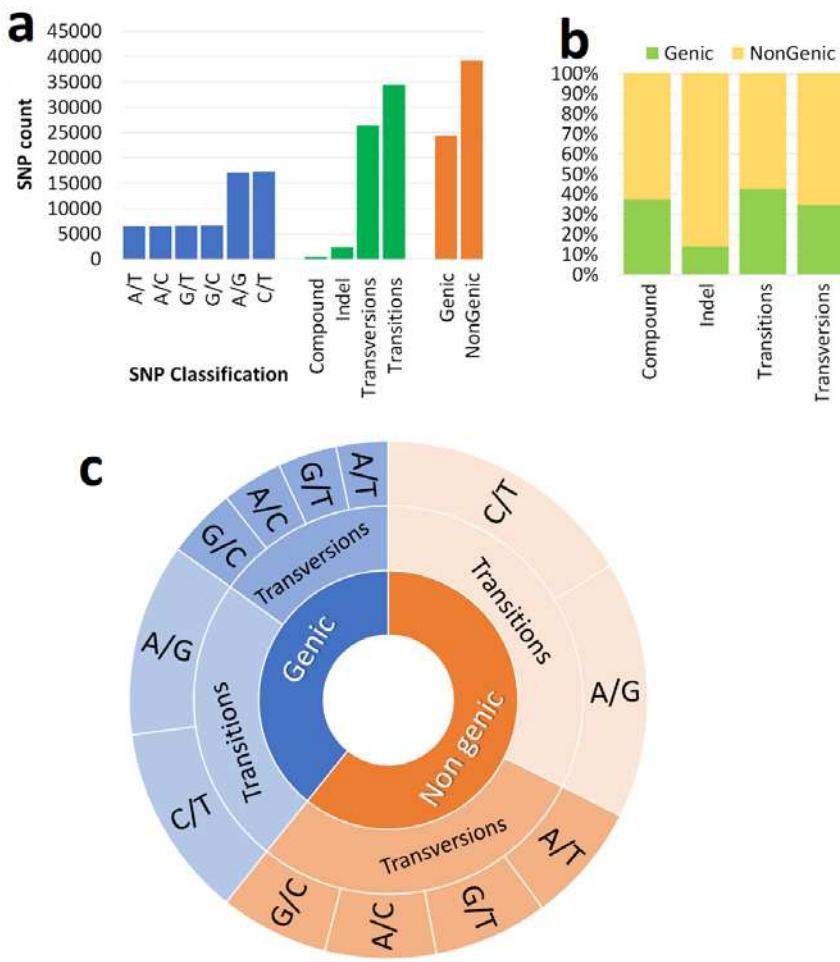
<i>Musa beccariei</i>	Mbeci	181767	141466	78
<i>Musa itinerans</i>	Mitan	151683	126107	83
<i>Musa schizocarpa</i>	Mscz1	178889	135556	76
<i>Musa textilis</i>	Mtext	215660	173840	81
<i>Musa troglodytarum</i>	Mtrog	197524	155701	79
<i>Ensete glaucum</i>	Engl	181817	155508	86
<i>Ensete ventricosum</i> Bedadeti	Envnb	163773	142536	87
<b>Total</b>		<b>2761190</b>	<b>2115474</b>	<b>77</b>

We systematically characterized all the SSR markers based on their transferability, genomic location, polymorphism, locus type, and the distribution of repeat type, repeat richness, and repeat length among genomic regions. The results are presented in Figure 1a. The findings revealed that 80% of the SSRs are located in intergenic regions. Furthermore, we distributed Class I and Class II SSRs between these two regions, genic and intergenic. However, we did not observe any significant preference for the distribution of Class I and Class II SSR types among these locations in both regions. Class II types were more frequent than Class I types SSRs in both regions (Figure 1a-c). Single locus markers were found to be prevalent among both *Musa* and *Ensete* species. Based on the e-polymerase chain reaction (e-PCR) results, a significant number of polymorphic markers were identified, indicating their potential utility for further applications (Figure 1d-f).



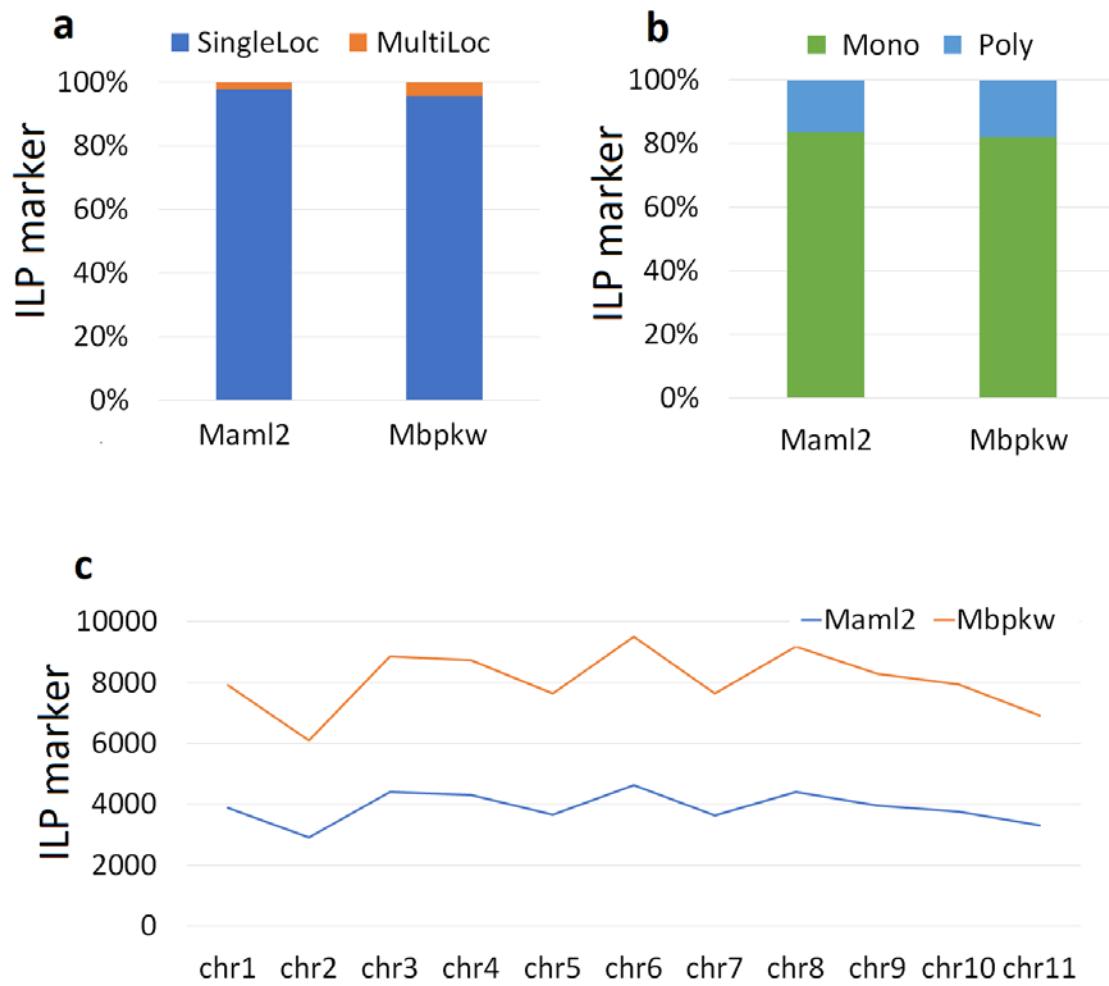
**Figure 1.** SSR Marker Development and Characterization: **(a)** Shows the distribution of Simple Sequence Repeats (SSRs) in different gene locations, divided into categories like Class I and Class II. We've further categorized these SSRs based on their richness (AT, Balanced, GC) and displayed the specific motifs (mono to hexa) associated with them; **(b)** Depicts the distribution of Class I and Class II SSR markers among *Musa* and *Ensete* species; **(c)** Demonstrates the distribution of motif richness within SSR markers among *Musa* and *Ensete* species; **(d)** Offers a comparative analysis of SSR marker transferability among different species; **(e)** Illustrates the distribution of single locus and multi locus SSR markers among *Musa* and *Ensete* species; **(f)** Presents the distribution of polymorphic and monomorphic SSR markers among *Musa* and *Ensete* species.

In this study, SNP markers were developed and characterized based on sequences obtained from the public domain. The results are presented in Figure 2 and Supplementary Table S2. A total of 63588 SNP markers were developed in this study, with 24375 of them located in the genic regions. Transition mutations ( $C \leftrightarrow T$ ;  $A \leftrightarrow G$ ) were found to be dominant in both regions, although their frequency was not significantly higher than that of transversion mutations. Indels are more abundant in the non-genic regions than in the genic regions.



**Figure 2.** SNP marker mining and characterization. (a) distribution of different class of SNP (b) distribution of SNP in geneic and non-genic regions (c) distribution of SNP types, Class among the geneic and non-genic regions.

A total of 91,547 ILP markers were developed from the genomes of *Maml2* and *Mbpkw*. The percentage of single and multi-locus markers significantly varied in both genomes (*Maml2* and *Mbpkw*). The results reveal that single-locus ILP markers constitute more than 98% of both genomes, while multi-locus markers are less common. The majority of the ILP markers show monomorphism, accounting for approximately 82% of the total (Figure 3a-c). The chromosomal level distributions of ILP markers in both genomes do not exhibit any significant differences.



**Figure 3.** ILP Marker Development and Characterization: (a) Comparative distribution of single-locus and multi-locus ILP markers in two banana species; (b) Distribution of polymorphic and monomorphic ILP markers in two banana species; (c) Chromosome-level distribution of ILP markers.

### 3.2. Interface and Search Criteria

The MMdb is a comprehensive and user-friendly resource that enables exploration, searching, downloading, and comparison of molecular markers such as SSR, SNP and ILP across 12 *Musa* species. It has an interactive web interface, which is accessible via a top navigation bar, enabling users to easily access various sections and tools in the database.

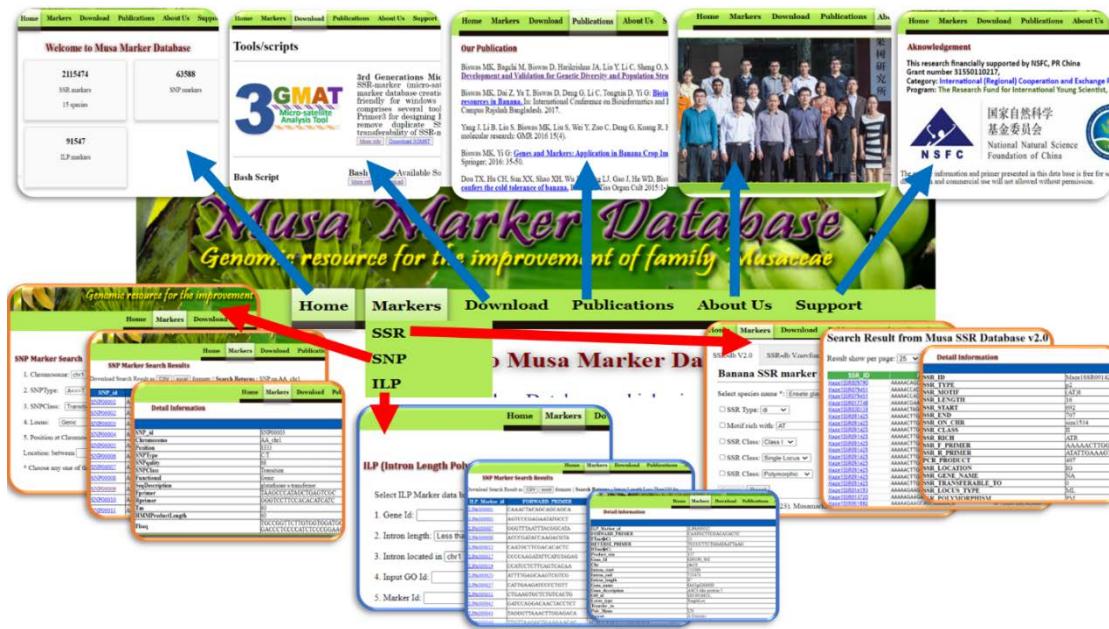
The “Home” page presents an overview of the MMdb and its potential use in banana breeding (Figure 4). The “Marker” menu lists three different marker search interfaces for SSR, SNP and ILP primers and their information’s. The “Download” page comprises a list and links to downloadable data, while the “Publication” page lists research articles published by our research group. The “About” page represent information of the group members involved in building this database, while the “Support” page describes the financial support information of this project.

The SSR search interface features three tabbed menus: GW-SSR (Genome Wide SSR Marker), Novel Functional SSR Marker Search (NFS), and SSR-db v1.0. The “GW-SSR” tab, there are five different search criteria that users can employ either individually or in various combinations. On the “NFS” search page enables users to explore SSR markers from the Functional SSR marker sets developed by Biswas et al. (2020). This search interface comprises eight individual search criteria, as illustrated in Figure 4. The third tab, SSR-db V1.0, contains SSR markers published by Biswas et al. (2015) [17]. It offers two distinct search criteria: a basic search with 14 individual search parameters

and an advanced search with seven parameters. Users can select these parameters in various combinations, empowering them to obtain more specific and tailored results.

The SNP search page features five distinct search criteria, each operating independently based on the specific characteristics of the SNP. It returns specific types of SNP information, presenting search results in a tabulated format. Each entry in the table is linked to detailed information for the corresponding SNP data.

The ILP search page includes five independent search criteria, aiding users in selecting the best search results from the database. Each search yields results in a tabulated format, with each data entry linked to additional details for each ILP marker.



**Figure 4.** Overview of the MMdb.

### 3.3. Unique Feature of MMdb Compare with Others Existing Marker Database

There are currently three marker databases for *Musa*, including the SSRome, TropGENE Database, and BanSatDb database. A comprehensive overview of their comparative features are listed in Table S3, providing a detailed analysis and insights into the distinct attributes of each database. However, BGH and TropGENE only store *Musa* SNP markers, while BanSatDb presents SSR markers with very limited data and search parameters. For example, BanSatDb developed markers from only three *Musa* species (*M. acuminata*, *M. balbisiana*, and *M. itinerans*), and does not offer information on primer redundancy, transferability, or polymorphism to other *Musa* or non-*Musa* species. Additionally, there is no unique identifier for the designed primers, which can affect reproducibility in downstream marker applications by different researchers. In contrast, MMdb offers several unique features compared to other existing *Musa* marker databases. For instance, MMdb characterizes and classifies each marker based on various parameters, such as SSR motif size, type, location in the genome (genic, non-genic, CDS, UTR), nucleotide base composition, chromosome position, transferability to other *Musa* species, and *in silico* prediction of PCR polymorphism. Moreover, MMdb contains a much larger number of markers and up-to-date marker information for sequenced *Musa* species. Overall, MMdb provides a comprehensive and user-friendly resource for exploring, downloading, and comparing various molecular markers across multiple *Musa* species.

### 3.4. Utility and Future Directions

The MMdb is a valuable resource for banana researchers and breeders, as it allows users for easy access to a large collection of molecular markers, including SSRs, SNPs, and ILPs, along with detailed

characterization and classification information. This database can aid in marker-assisted selection, genetic diversity analysis, banana germplasm management, and genome mapping studies in *Musa* species. Additionally, the *in-silico* transferability and polymorphism feature save time and resources by allowing researchers to pick the best marker set without requiring wet lab experiments.

In the future, the database could be extended to include more *Musa* species and additional marker types. The incorporation of functional annotation data and gene expression information could also enhance its usefulness in molecular breeding applications. Additionally, regular updates and improvements to the database interface and search features could improve user experience and make it an even more valuable tool for the *Musa* research community.

#### 4. Discussion

Molecular markers play a crucial role in modern plant breeding by allowing breeders to identify and track desirable traits, evaluate genetic diversity [18] [19] [2] and develop more targeted breeding programs. These markers can speed up the breeding process by helping to select plants that are more likely to possess desirable traits, as well as aid in the understanding of the genetic basis of complex traits and the development of molecular breeding strategies to improve these traits. Molecular markers are particularly useful in crops with long breeding cycles or complex traits that are difficult to measure directly. Availability of whole genome sequences has enabled the development of various types of molecular markers and the creation of molecular marker databases to maximize their utility [20] [21]. In this study, a novel molecular marker database was developed from whole genome sequences, EST, and GSS sequences of 12 *Musa* species. SSR, SNP, and ILP markers were mined and characterized for optimal marker data quality, resulting in the creation of a searchable *Musa* marker database that is freely accessible from the following link ([www.genomicsres.org/mmdb/](http://www.genomicsres.org/mmdb/)).

In terms of SSR analysis, our study identified a substantial number of SSRs, amounting to 2,761,190, with variable densities in the tested genomes. The prevalence of Class II SSRs (shorter than 20bp), characterized by shorter motifs, was consistent with previous findings in plant genomes [21] [20] [22]. Furthermore, the dominance of AT-rich SSR motifs aligns with similar observations in *Musa* and other plant species [23] [24]. The high percentage of successfully developed SSR markers (77%) suggests the potential for efficient marker development in these genomes. This outcome is in line with previous research demonstrating the utility of SSR markers for genetic mapping and breeding applications [25] [26].

Our study's exploration of SNP markers in genic and non-genic regions also adds to the growing body of knowledge regarding polymorphism in plant genomes. The higher occurrence of transition mutations (C↔T; A↔G) over transversions observed in both regions corresponds with studies in various plant species [27] [28]. Additionally, the prevalence of Indels in non-genic regions resonates with previous reports of the non-coding regions being more prone to structural variations [29]. These findings underscore the importance of considering both genic and non-genic markers for genetic diversity and trait association studies.

The results regarding ILP markers, particularly the dominance of single-locus ILP markers and their high monomorphism rate (82%), are consistent with studies in other plant species [30]. However, the fact that multi-locus ILP markers are less common raises questions about their potential applications, which may differ from those of single-locus markers. Further research may be needed to explore the specific utility of multi-locus ILP markers in these genomes. It is important to note that the chromosomal distribution of ILP markers showed no significant differences between the genomes of *Maml2* and *Mbpkw*, indicating that the genomic context does not strongly influence the distribution of these markers. This finding contrasts with some studies that have reported uneven distributions of markers across chromosomes [31] [30], suggesting the need for further investigation into the underlying factors governing marker distribution.

#### 5. Conclusions

In conclusion, the *Musa* Marker Database (MMdb) is a user-friendly and inclusive genomic resource for the banana research community, that offers its uses to explore, download and compare

molecular markers such as SSR, SNP, and ILP across 12 *Musa* species. The unique features of this database such as marker transferability, ePCR validation, and polymorphism information of each marker make it stand out from other existing marker databases. The current version of the MMdb contains a larger number of markers and up-to-date data for *Musa* species. This resource can be used to study genetic diversity, germplasm characterization, develop MAS strategies, perform GWAS, and compare markers across different *Musa* species.

**Supplementary Materials:** Table S1. Summary of data and SSR mining of *Musa* species. Table S2. Summary of the SNP Mining and characterization. Table S3. Comparison of MMdb with other related databases.

**Author Contributions:** The work presented here was carried out in collaboration among all authors. M.K.B. was involved in SSR, SNP, ILP analysis, database development, and drafted the manuscript; D.B. wrote and managed the server to host the database; O.S. and G.D. participate SSR data analysis and revised the MS; M.K.B. and G.Y. conceived and designed the experiments. All authors have read and approved the final manuscript.

**Funding:** This work was financially supported by the Natural Science Foundation of China (32261160375) and supported by the earmarked fund for CARS (CARS-31).

**Acknowledgments:** The authors express gratitude to the NSFC for awarding the International Young Scientist grant, enabling Dr. Manosh Biswas to initiate this research.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Biswas, M.K.; Darbar, J.N.; Borrell, J.S.; Bagchi, M.; Biswas, D.; Nuraga, G.W.; Demissew, S.; Wilkin, P.; Schwarzacher, T.; Heslop-Harrison, J.S. The landscape of microsatellites in the enset (*Ensete ventricosum*) genome and web-based marker resource development. *Scientific Reports* **2020**, *10*, 15312.
2. Biswas, M.K.; Xu, Q.; Deng, X. Utility of RAPD, ISSR, IRAP and REMAP markers for the genetic analysis of Citrus spp. *Scientia Horticulturae* **2010**, *124*, 254-261.
3. Biswas, M.K.; Bagchi, M.; Biswas, D.; Harikrishna, J.A.; Liu, Y.; Li, C.; Sheng, O.; Mayer, C.; Yi, G.; Deng, G. Genome-wide novel genic microsatellite marker resource development and validation for genetic diversity and population structure analysis of banana. *Genes* **2020**, *11*, 1479.
4. Blenda, A.; Scheffler, J.; Scheffler, B.; Palmer, M.; Lacape, J.; Yu, J.Z.; Jesudurai, C.; Jung, S.; Muthukumar, S.; Yellambalase, P. CMD: a cotton microsatellite database resource for *Gossypium* genomics. *BMC Genomics* **2006**, *7*, 1-10.
5. Shirasawa, K.; Isobe, S.; Tabata, S.; Hirakawa, H. Kazusa Marker DataBase: a database for genomics, genetics, and molecular breeding in plants. *Breed Sci* **2014**, *64*, 264-271.
6. Kim, C.; Seol, Y.; Lee, D.; Jeong, I.; Yoon, U.; Lee, G.; Hahn, J.; Park, D. NABIC marker database: A molecular markers information network of agricultural crops. *Bioinformation* **2013**, *9*, 887.
7. Sarika; Arora, V.; Iquebal, M.A.; Rai, A.; Kumar, D. PIPEMicroDB: microsatellite database and primer generation tool for pigeonpea genome. *Database* **2013**, *2013*, bas054.
8. Muthamilarasan, M.; Misra, G.; Prasad, M. FmMDB: a versatile database of foxtail millet markers for millets and bioenergy grasses research. *PLoS one* **2013**, *8*, e71418.
9. Doddamani, D.; Katta, M.A.; Khan, A.W.; Agarwal, G.; Shah, T.M.; Varshney, R.K. CicArMiSatDB: the chickpea microsatellite database. *BMC Bioinformatics* **2014**, *15*, 1-7.
10. Yang, L.; Jin, G.; Zhao, X.; Zheng, Y.; Xu, Z.; Wu, W. PIP: a database of potential intron polymorphism markers. *Bioinformatics* **2007**, *23*, 2174-2177.
11. Mokhtar, M.M.; Atia, M.A.M. SSRome: an integrated database and pipelines for exploring microsatellites in all organisms. *Nucleic Acids Res* **2019**, *47*, D244-D252.
12. Xu, H.; Yu, Q.; Shi, Y.; Hua, X.; Tang, H.; Yang, L.; Ming, R.; Zhang, J. PGD: pineapple genomics database. *Horticulture Research* **2018**, *5*.
13. Biswas, M.K.; Natarajan, S.; Biswas, D.; Howlader, J.; Park, J.-.; Nou, I.-. Lily Database: A Comprehensive Genomic Resource for the Liliaceae Family. *Horticulturae* **2024**, *23*, DOI <https://doi.org/10.3390/horticulturae10010023>.
14. Yu, J.; Dossa, K.; Wang, L.; Zhang, Y.; Wei, X.; Liao, B.; Zhang, X. PMDBase: a database for studying microsatellite DNA and marker development in plants. *Nucleic Acids Res* **2017**, *45*, D1046-D1053.
15. Droc, G.; Lariviere, D.; Guignon, V.; Yahiaoui, N.; This, D.; Garsmeur, O.; Dereeper, A.; Hamelin, C.; Argout, X.; Dufayard, J. The banana genome hub. *Database* **2013**, *2013*, bat035.

16. Arora, V.; Kapoor, N.; Fatma, S.; Jaiswal, S.; Iquebal, M.A.; Rai, A.; Kumar, D. BanSatDB, a whole-genome-based database of putative and experimentally validated microsatellite markers of three *Musa* species. *The Crop Journal* **2018**, *6*, 642-650.
17. Biswas, M.K.; Liu, Y.; Li, C.; Sheng, O.; Mayer, C.; Yi, G. Genome-wide computational analysis of *Musa* microsatellites: classification, cross-taxon transferability, functional annotation, association with transposons & miRNAs, and genetic marker potential. *PLoS One* **2015**, *10*, e0131312.
18. Salgotra, R.K.; Chauhan, B.S. Genetic diversity, conservation, and utilization of plant genetic resources. *Genes* **2023**, *14*, 174.
19. Dida, G. Molecular Markers in Breeding of Crops: Recent Progress and Advancements. *J.Microbiol.Biotechnol* **2022**, *7*.
20. Savadi, S.; Muralidhara, B.M.; Venkataravanappa, V.; Adiga, J.D. Genome-wide survey and characterization of microsatellites in cashew and design of a web-based microsatellite database: CMDB. *Frontiers in Plant Science* **2023**, *14*.
21. Biswas, M.K.; Xu, Q.; Mayer, C.; Deng, X. Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). *PloS one* **2014**, *9*, e104182.
22. Singh, J.; Sharma, A.; Sharma, V.; Gaikwad, P.N.; Sidhu, G.S.; Kaur, G.; Kaur, N.; Jindal, T.; Chhuneja, P.; Rattanpal, H.S. Comprehensive genome-wide identification and transferability of chromosome-specific highly variable microsatellite markers from citrus species. *Scientific Reports* **2023**, *13*, 10919.
23. Zhang, J.; Liu, T.; Rui, F. Development of EST-SSR markers derived from transcriptome of *Saccharina japonica* and their application in genetic diversity analysis. *J Appl Phycol* **2018**, *30*, 2101-2109.
24. Liu, S.; An, Y.; Li, F.; Li, S.; Liu, L.; Zhou, Q.; Zhao, S.; Wei, C. Genome-wide identification of simple sequence repeats and development of polymorphic SSR markers for genetic studies in tea plant (*Camellia sinensis*). *Mol Breed* **2018**, *38*, 1-13.
25. Varshney, R.K.; Pandey, M.K.; Bohra, A.; Singh, V.K.; Thudi, M.; Saxena, R.K. Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor Appl Genet* **2019**, *132*, 797-816.
26. Gautam, A.; Tantwai, K.; Nema, S.; Tripathi, N.; Tiwari, S. Cross Transferability of SSR Markers from Finger Millet, Pearl Millet and Rice to Indian Little Millet and their Genetic Diversity Analysis.
27. Liu, J.; Luo, W.; Qin, N.; Ding, P.; Zhang, H.; Yang, C.; Mu, Y.; Tang, H.; Liu, Y.; Li, W. A 55 K SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. *Theor Appl Genet* **2018**, *131*, 2439-2450.
28. Kaur, B.; Mavi, G.S.; Gill, M.S.; Saini, D.K. Utilization of KASP technology for wheat improvement. *Cereal Research Communications* **2020**, *48*, 409-421.
29. Li, B.; Zhang, N.; Wang, Y.; George, A.W.; Reverter, A.; Li, Y. Genomic prediction of breeding values using a subset of SNPs identified by three machine learning methods. *Frontiers in genetics* **2018**, *9*, 237.
30. Wang, X.; Zhao, X.; Zhu, J.; Wu, W. Genome-wide investigation of intron length polymorphisms and their potential as molecular markers in rice (*Oryza sativa* L.). *DNA Research* **2005**, *12*, 417-427.
31. Zhang, Z.; Min, X.; Wang, Z.; Wang, Y.; Liu, Z.; Liu, W. Genome-wide development and utilization of novel intron-length polymorphic (ILP) markers in *Medicago sativa*. *Mol Breed* **2017**, *37*, 1-8.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.