Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Genomic Data Reveal the Genetic Variation among Natural Mangifera Casturi Kosterm. Hybrids, an Underutilized Fruit Tree Under 'Extinct in Wild' Status from Kalimantan Selatan, Indonesia

Deden Derajat Matra (dedenmatra@apps.ipb.ac.id)
Bogor Agricultural University
Muh Agust Nur Fathoni
Bogor Agricultural University
Muhamad Majiidu
Bogor Agricultural University
Hanif Wicaksono
Tunas Meratus Conservation Organization
Agung Sriyono
Banua Botanical Garden
- Gunawan
Lambung Mangkurat University
Hilda Susanti
Lambung Mangkurat University
- Fitmawati
Riau University
Rismita Sari
Indonesian Institute of Sciences
Iskandar Zulkarnaen Siregar
Bogor Agricultural University
Roedhy Poerwanto
Bogor Agricultural University

Research Article

Keywords: Hybridization, Chloroplast, Polymorphisms, Phylogenetic, Tropical Fruit

Posted Date: March 3rd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-258288/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Abstract

Background: *Mangifera casturi* Kosterm. is an endemic local mango fruit from Kalimantan Selatan, Indonesia. The limited genetic information available on this fruit has severely limited the scope of research into its genetic variation and phylogeny. This study aimed to collect genomic information from *M. casturi* using next-generation sequencing technology and to develop microsatellite markers and perform Sanger sequencing for DNA barcoding analysis.

Results: The clean reads of the Kasturi accession of *M. casturi* were assembled *de novo* using a Ray assembler, producing 259,872 scaffolds with an N50 value of 1,445 bp. Fourteen polymorphic microsatellite markers were developed from 11,040 sequences containing microsatellite motifs. In total, 55 alleles were produced, and the mean number of alleles per locus was 3.93. Results from the microsatellite marker analysis revealed broad genetic variation in *M. casturi*. Phylogenetic analysis was performed using internal transcribed spacers (ITS), matK, rbcL, and trnH-psbA. The phylogenetic tree of chloroplast markers showed that Kasturi, Mawar, Pelipisan, Pinari, and Hambawang belong to one group, with *M. indica* as the female ancestor. In comparison, the phylogenetic tree of ITS markers indicated several *Mangifera* species as multiple ancestors of *M. casturi*.

Conclusions: This study strongly suggested that *M. casturi* originated from the cross-hybridization of multiple ancestors. Further, crossing the F1 hybrids of *M. indica* and *M. quadrifida* with other *Mangifera* spp. was hypothesized to produce the observed high genetic variation. The genetic information for this fruit is also a resource for the breeding and improvement as well as for conservation studies of this species.

Background

Mangifera casturi Kosterm. (Anacardiaceae), or Kalimantan mango, is an endemic fruit from Kalimantan Selatan, Indonesia and is classified as extinct in the wild according to the IUCN Red List [1]. This local mango belongs to the *Mangifera* genus within the Anacardiaceae [2] and is classified as a common ancestor of *Mangifera* in Indonesia [3]. The fruit is known by various local names, such as kasturi, mawar, pelipisan, and pinari [2]. Based on phylogenetic analysis using single nucleotide polymorphisms (SNPs), *M. casturi* is proposed to be a natural hybrid of *M. indica* and *M. quadrifida* [4]. This local mango is a prospective genetic resource that can be used to improve mango varieties in the future as *M. casturi* bears small fruits, attractive purple colors, and a distinctive aroma [5]. It is also known to contain useful metabolite compounds, including lupeol, an antioxidant and anticancer agent [6]. However, genomic information on this local mango is still limited. In nucleotide repositories such as the NCBI, only one nucleotide accession has been deposited (MF678493.1) and one SRA study (SRP183190) has been reported.

Recently, sequencing technology has developed significantly, from Sanger sequencing to next-generation sequencing technology such as whole genome sequencing (WGS), which can provide comprehensive

information on species [7]. From these data, it is easier to obtain genetic information such as microsatellite markers, which have advantages over other markers such as RAPD and AFLP, also used in other *Mangifera* species [8, 9]. Microsatellite markers can determine distinct variations on several levels as they are codominant and thus widely used in populations and genetics [10]. Microsatellite markers have been reported to determine genetic variation in *M. indica* [11].

Previously, phylogenetic studies using DNA barcoding methods, such as rbcL, matK [12], and trnH-PsbA [13], as well as internal transcribed spacers (ITS) and the second internal transcribed spacers (ITS2) from nuclear ribosomal DNA [14], have been widely used for phylogenetic analysis at various taxonomic levels. These markers can also be determined at the genus or family level because of their inheritance from a maternal ancestor. However, ITS can determine the barcoding of both parentals. This study aimed to collect genomic information from *M. casturi* using next-generation sequencing technology to develop microsatellite markers and Sanger sequencing for DNA barcoding. Currently, there is no clear information about the genetic variation in *M. casturi* and its ancestor, which originates from natural hybrids of *M. indica* and *M. quadrifid*a. These markers were also used to determine the genetic variations of *M. casturi* hybrids.

Results

In this study, 11.01 gb of *M. casturi* DNA was produced with high-throughput sequencing using a Illumina HiSeq 4000 with 2 × 150 bp paired-ends. The raw data were registered in DDBJ with accession number DRA011022. After filtering, clean reads were obtained, and 10.95 gb of *de novo* genome assembly was performed using a Ray Assembler. The scaffold obtained was 259,872 bp with an N50 value of 1,445 bp and a maximum scaffold length of 144,601 bp (Table 1). The annotation process used BUSCO, assessed using details of the complete categories and single-copy BUSCOs (S), with a ratio of 42.3% (Table 2).

Table 1
Statistics of de novo assembly from *M. casturi* using Ray Assembler

Features	Number
Raw reads (Bases)	73 438 066 M (11 015 710 G)
Clean reads (bases)	73 102 518 M (10 954 176 G)
Number of Scaffold	259 872
N50 (bp)	1 445
Mean (bp)	947.68
Longest (bp)	144 601

Table 2
Summarized benchmarks in BUSCO notation from Ray Assembler of *M. casturi*

No	Categories	Number	Ratio (%)
1	Complete and single-copy BUSCOs (S)	608	42,2
2	Complete and duplicated BUSCOs (D)	36	2,5
3	Fragmented BUSCOs (F)	241	16,7
4	Missing BUSCOs (M)	555	38,5

Microsatellites were identified using the MISA program and 11,040 sequences containing microsatellite motifs, and 770 sequences with more than one microsatellite site were extracted (Table 3). The trinucleotide motif exhibited the highest proportion (52.77%), followed by the dinucleotide motif (33.3%). (Table 3). Fourteen candidate sequences were selected and identified (Table 4). All confirmed primers were amplified and then registered with the DDBJ accession numbers shown in Table 4.

Table 3
Number and motif of microsatellite region from Ray assembler of *M. casturi* scaffold

Characteristics	Number
- Total number of identified SSRs	11 040
- Number of SSR containing contig	10 160
- contig containing more than 1 SSR	770
- SSRs present in compound formation	272
Motif	
- Dinucleotide	3 680
- Trinucleotide	5 826
- Tetranucleotide	1 194
- Pentanucleotide	213
- Hexanucleotide	102
- Heptanucleotide	25

Table 4
Characteristics of 14 microsatellite loci of *M. casturi*

No	Locus	Nucleotides	Motif	Ranged size	NA	DDBJ accession
1	mc-122955	F : TGTTGATGGTAAGGATTTGGTGT	(GGATG)6	168–178	2	LC594546
		R : TCAGGTGAGTATGTATTGTGCA				
2	mc-148231	F : TCCCTCCCCTAAACCCTTCT	(ACCCTAA)5	188–209	4	LC594549
		R : GCTTCTCCTTGCCTCTAAATCCT				
3	mc-151578	F : GAGCCTTGTACTCGTTCAATGA	(CAAGCT)8	273–279	5	LC594547
		R : ACGAGCTTAAAATGAGTTTGACT				
4	mc-167596	F : AGCTGAACCTTGTTGCCCTT	(GA)27	192–224	3	LC594539
		R : TCTGCTTGTTGGAACTGAACA				
5	mc-176197	F : TGTATGCCCGAATTGTTCCAAC	(AC)19	237–250	3	LC594537
		R : GCTGGCTTTAATGGAAGTTGCA				
6	mc-211123	F : GGATGGTGGATGTCAGATTTTCG	(TGAAGT)6	323–339	5	LC594548
		R : CGAAGAGAACGGGTCCCTTG				
7	mc-21672	F : TGGTTGGTAAGAAGTAGGATTC	(ATAC)11	263–264	4	LC594543
		R : CACAATGCAAATCACTCCTC				
8	mc-230178	F : AGACAGCCATAATTTGCCCCA	(ATG)12	162–188	6	LC594541
		R : GCTGGAGGTTGATCAGGGTC				
9	mc-28107	F : GGTGTGCGTTCTGTTTTGACA	(TG)28	211–250	5	LC594540
		R : CAGCAGCATCAACACAAGCA				
10	mc-4673	F : TTTCCAAAGCCAAGACTCTC	(TAAACCC)5	231–245	3	LC594550

No	Locus	Nucleotides	Motif	Ranged size	NA	DDBJ accession
		R : AAAATTGTATTCATTAAGCCCCT				
11	mc-58089	F : TCTTGTCGTCGAATCAAACTCA	(AT)22	264–287	6	LC594538
		R : CTCGGTCTATCAATGGTGTAGGT				
12	mc-8693	F : CGAAGGGTTGAGGTTTGGGT	(CTTTT)7	159–183	4	LC594545
		R : AAAGAGTGAGAGGGTTGCGT				
13	mc-88075	F : CTCCAATCGAACAACCCAGC	(TTA)15	278–286	3	LC594542
		R : AGGGGTGCATATGGAGGATT				
14	mc-88387	F : CCATTTCGACGATGTTGGAAGT	(TATG)10	251–252	2	LC594544
		R : GCAACCCTTACCAACAAGCA				

Eight samples were used to validate and determine allele size using QIAxcel®. The 14 primers produced 55 alleles in total, and the mean number of alleles per locus was 3.93 (Table 4). All loci were polymorphic (Table 5). The mc-230178 and mc-58089 loci produced six alleles, while mc-122955 and mc-88387 produced two alleles. Some loci showed the same alleles between Kasturi and Mawar, namely mc-176197, mc-21672, and mc-88075. In the mc-88387 locus, only the Kasturi sample was not amplified, and it was proposed that this locus was a null allele of Kasturi. Therefore, mc-88387 can be used to identify *M. casturi* in the population, as it is otherwise similar to other *Mangifera* species, such as Mawar. The UPGMA tree was produced using 14 loci (Fig. 2). *M. quadrifida* and Rawa-rawa were placed in the same clade. All accessions of *M. casturi* were in the same clade as *M. indica*, even as an out-group for this analysis (Hambangan or *M. foetida*). Mawar accessions were most closely related to *M. indica*. Kasturi and Pelipisan had the same clade. Some markers showed allele similarity between Kasturi and Pelipisan; these accessions, thus, had a closer genetic relationship to each other than to Mawar. However, Pinari also exhibited distinct genetic differences from the other accessions of *M. casturi*, even though Mawar was quite distant from another *M. casturi* accession.

Table 5
Allele size information per microsatellite locus

No.	Locus	Allele size (bp)
1	mc-122955	168,178
2	mc-148231	188,196,203,209
3	mc-151578	256,270,273,279,284
4	mc-167596	193,224,226
5	mc-176197	235,237,249
6	mc-211123	318,323,333,335,340
7	mc-21672	255,256,257,263
8	mc-230178	162,164,170,173,178,185
9	mc-28107	204,211,214,225,250
10	mc-4673	231,238,246
11	mc-58089	261,264,273,279,282,287
12	mc-8693	157,160,167,182
13	mc-88075	267,278,286
14	mc-88387	251,253

Phylogenetic analysis was performed using three widely used chloroplast markers (Fig. 3). The matK phylogenetic tree showed that Kasturi, Mawar, Pelipisan, Pinari, and Hambawang belonged to one group with *M. indica* and *M. sylvatica*. In comparison, the rbcL phylogenetic tree placed Mawar and Pelipisan into the same clade for almost all *M. indica* accessions. Meanwhile, Pinari and Hambawang were separated from this clade and joined by the *M. laurina, M. flava, M. cochinchinensis, M. odorata*, and *M. duperreana* clades. In contrast, the phylogenetic tree analysis results using trnH-psbA led to Kasturi, Pelipisan, and Hambawang being grouped with *M. indica*. Pinari was close to *M. odorata, M. griffithii, M. pajang, M. andamanica*, and *M. indica*.

The ITS phylogenetic tree produced three large groups, namely Indica 1, Indica 2, and a group containing Kasturi, Mawar, Pelipisan, and Pinari. Hambawang was included in the Indica 2 group. Pinari was placed in a sub-group with *M. oblongifolia, M. camptosperma, M. gedebe*, and *M. flava*. Kasturi, Mawar, and Pelipisan were included in the other sub-groups with *M. casturi* (MF678493.1), *M. griffthii, M. quadrifida, M. kemanga, M. torquenda*, and *M. sumatrana*.

Discussion

The *Mangifera* genus originates from southeast Asia and has polyembryony seeds, which originate from gametes or nucellar cell components [2]. Most *Mangifera* flowers are either hermaphrodites or males [32]. Self-crossing can occur in a variety of species. However, self-incompatibility in the *Mangifera* genus has been reported in several mango types [33]. This evidence suggests that *Mangifera* can be crossed among varieties and species [2, 34]. Many interspecies that have resulted from cross-hybridization in natural populations have been reported, including *M. odorata* (Kuini), a natural hybrid between *M. indica* and *M. foetida* [9].

Based on the indications from the 14 microsatellite loci, different allele sizes were obtained from four accessions of *M. casturi*. A high level of genetic variation was found to occur in *M. casturi* accessions and may have arisen from interspecific hybridization. The accessions of Kasturi, Mawar, and Pelipisan were more closely related compared to those of Pinari. Kasturi and Mawar and were very similar in terms of fruit size; however, the fruit shape of Kasturi was more oval than that of Mawar. In contrast, the Pelipisan variety fruit was more oval and oversized. Pinari showed the largest fruit size among the *M. casturi* accessions. Pinari was classified into the *M. casturi* group by the locals, based on a purplish skin color, which is similar to that of other *M. casturi* accessions.

Intraspecies genetic variation can occur because of multiple cross-hybridizations involving several species. For instance, in the natural hybrid of Kuini (*M. odorata*), cross-hybridization between *M. indica* and *M. foetida* was revealed by AFLP analysis and represented a simultaneous backcross between the F1 hybrid and *M. foetida* [4, 9]. Based on SNP analysis using double-digest restriction-site-associated DNA (ddRAD) data, *M. casturi* was revealed to be a natural hybrid between *M. indica* and *M. quadrifida*, whereas their F1 hybrid was a backcross with *M. indica*. Morphologically, *M. casturi* was very close to *M. quadrifida*, with purplish skin and a small fruit size [4].

In the allopolyploid plant mangosteen (*Garcinia mangostana*), microsatellite markers show cross-hybridization with multiple ancestors, including *G. malaccensis*, *G. celebica*, and *G. porrecta* [22]. Our microsatellite analysis results showed that four accessions of *M. casturi*, Kasturi, Mawar, Pelipisan, and Pinari had allelic differences in all microsatellite loci. However, allele sharing between four accessions of *M. casturi* was shown in the mc-8693 locus with an allele size of 160/182. This evidence suggests that these accessions were derived from the same ancestor. In contrast, the allele differences in microsatellite loci indicated a possibility that the four accessions of *M. casturi* species underwent cross-hybridization with multiple ancestors.

DNA barcoding analysis using matK and rbcL suggested very high nucleotide similarity between the four accessions of *M. casturi*. This evidence indicated that the maternal ancestor of these accessions was the same and that *M. indica* is one of the maternal ancestors. Additional evidence was shown in the trnH-psbA phylogenetic tree, where Pinari showed a different maternal ancestor from the other accessions of *M. casturi*. It is possible that one of the *M. casturi* hybrids crossed with other *Mangifera* spp. as the mother ancestor. The ITS of the phylogenetic tree also revealed that three accessions of *M. casturi*, excluding Pinari, belonged to the same sub-group, which contrasts with a previous hypothesis that *M.*

casturi is a cross-hybrid between *M. indica* and *M. quadrifida*. Our results also support the hypothesis that F1 hybrids crossing with other Mangifera spp. produced variations in *M. casturi* in the natural population.

Conclusions

The results of this study demonstrated broad genetic variation in *M. casturi*. This represents an important source of genetic resources for breeding and the improvement of mango characteristics in the future. More intensive conservation efforts are needed, as *M. casturi* is currently classified as extinct in the wild, and its habitat is severely threatened. Kalimantan Selatan is well known to contain abundant coal that could be exploited extensively in the near future. This poses a serious threat to the existence of this local mango. Moreover, *M. casturi* has never been confirmed as a variety by authorities. The results of this study can help breeders and local governments to officially register one of their precious germplasms.

Methods

M. casturi accessions were collected from the Banjar, Kalimantan Selatan, in the southern region of Borneo Island (Fig. 1; Supplementary Table 1). To analyze whole genome sequencing, genomic DNA was isolated from *M. casturi* (Kasturi accession) using a DNeasy Power Plant kit (Qiagen) following the manufacturer's protocol. The quality and quantity of DNA were analyzed using a NanoPhotometer® NP80 Touch (IMPLEN) spectrophotometer. Genomic DNA samples were sent to Novogen-AIT Singapore with 150 paired-ends (PE) collected using an Illumina HiSeq4000 system. Raw reads were quality controlled using FASTQC [15], and clean reads were filtered using the Fastp program with default parameters [16]. Clean reads were assembled using a Ray assembler [17] under the Maser Platform facility [18]. BUSCO analysis [19], using the Maser Platform with default parameters, was performed to check the assembled contig quality.

Microsatellite marker development and validation

Microsatellite markers were extracted using the MISA program [20], the parameters being set to the following minimum repeat levels: six for two bases, and five for three, four, five, and six bases. The difference between microsatellite motifs was 100 bases. The primer was designed using the web version of Primer 3 [21].

Genomic DNA was isolated using the modified CTAB method, with a slight alteration [22]. The quality and quantity of DNA were assessed using a NanoPhotometer® NP80 Touch (IMPLEN). A microsatellite PCR kit (QIAGEN) was used to analyze the microsatellite markers. PCR master mix was prepared with a mixture of 3.2 µL RNase-free water, 5 µL 2x Type-it Multiplex PCR Master Mix, 0.4 Q solution, 0.2 µL of 10 µM forward primer, and 0.2 µL of 10 µM reverse primer. PCR was performed using a SimpliAMP Thermo Cycler (Applied Biosystems). The PCR conditions were as follows: initial conditions of PCR pre-denaturation at 95°C for 5 min, followed by 32 cycles of denaturation at 95°C for 30 s, annealing at 57°C

for 1 min 30 s, extension at 70°C for 30 s, and final extension at 60°C for 30 min. The amplicons were checked using 1% electrophoresis gel in TAE buffer for 20 min at 100 v. Before loading the sample into capillary electrophoresis (QIAxcel®, Qiagen), the sample was diluted twice and then run using a QIAxcel DNA High-Resolution Kit. Allele size data were confirmed and processed manually using QIAxcel ScreenGel Software (Qiagen).

The molecular data were processed using the Phylip program version 3.695 with the unweighted pair group method and arithmetic mean (UPGMA) method. The resulting dendrogram was edited using the program MEGA-X [23].

DNA barcoding and ancestral phylogenetic analysis

For the DNA barcoding analysis, we used three chloroplast genes: matK [24], rbcL [25], and trnH-psbA [26], as well as one nuclear DNA region, the internal transcribed spacer (ITS) [27]. PCR barcoding was performed using KOD Plus (Toyobo) according to the manufacturer's protocol. The PCR products were cleaned using ExoSAP-IT™ PCR Product Cleanup Reagent (Applied Biosystems). Then, PCR sequencing was carried out with a BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), followed by purification using a BigDye XTerminator ™ Purification Kit (Applied Biosystems) according to the manufacturer' s protocol. The products were sequenced using a 3500 Series Genetic Analyzer (Applied Biosystems). Sequence data were analyzed using Sequencing Analysis Software v6.0 (Applied Biosystems), and the data were processed with ATGC-MAC version.7 (Genetyx Co.) and MEGA-X software [23].

Phylogenetic trees were inferred using the maximum likelihood method and constructed using MEGA X software [28]. The best DNA model was calculated using MEGA X for each marker [29, 30]. Phylogenetic trees were tested using 10,000 bootstrap replicates [31].

Declarations

Ethics approval and consent to participate

All experiments were performed in accordance with relevant guidelines and regulations. The experimental research has complied with Bogor Agricultural University research regulation (No: 11/SA-IPB/P/2016 on research and publication ethics), and the field study was in accordance with the national legislations of Indonesian Law Number 5/1990 on biological diversity conservation and Indonesia Law Number 11/2013 on the ratification of the Nagoya Protocol. M. casturi samples were collected and exported from Banjar, South Kalimantan to Bogor, West Java with the permission (No : 2020.2.1702.0.K12.000044) from Plant Quarantine Division of National Agency for Agricultural Quarantine in Banjarmasin, South Kalimantan following permit approvals from South Kalimantan Natural Resources Conservation Agency/BKSDA of the Ministry of Environment and Forestry of Republic of Indonesia (KLHK) as agency in charge of managing conservation areas including protected plant in the territory, particularly the nature reserve forests (wildlife, nature reserves) and national park. The M. casturi samples from Banjar, South

Kalimantan as herbaria (received by Agung Sriyono) were duplicated and are stored in Banua Botanical Garden, Province of South Kalimantan.

Consent for publication

Not applicable

Availability of data and materials

All sequence data from the next generation sequencing during the current study have been submitted to the DDBJ Read Archive (DRA) under the BioProject accession number PRJDB10715: http://trace.ddbj.nig.ac.jp/BPSearch/bioproject?acc=PRJDB10715. All sequence data from DNA barcoding analysis during the current study have been submitted to the DDBJ Nucleotide Sequence Submission System under the accession number of LC602976- LC602993.

Competing interests

The authors declare that they have no conflicts of interest.

Funding

The Ministry of Research and Technology/National Agency for Research and Innovation (RISTEK-BRIN), the Republic of Indonesia

Authors' contributions

study and initiated the work: DDM, RPO. Conceived and designed the analyses: DDM, MAF, MM. Performed sample collection: DDM, HW, AS, HS. Performed Microsatellite experiments: DDM, MAF, MM, IZS. Performed phylogenetic analyses: G, F, RS. Analyzed the data: DDM, MAF. Wrote the paper: DDM, MAF, RPO. all authors revised and approved the final version.

Acknowledgements

This research was supported by the Ministry of Research and Technology/National Agency for Research and Innovation (RISTEK-BRIN), the Republic of Indonesia, through the Fundamental Research Scheme of IPB University (PDUPT) grant 2020. The exploration of M. casturi in Kalimantan Selatan was also supported by The Nagao Natural Environment Foundation (NEF) through the NEF Research Grant Scheme 2020.

References

1. Rhodes L, Maxted N. Mangifera casturi. The IUCN Red List of Threatened Species. 2016.
2. Kostermans AJGH, Bompard JM. The Mangoes: Their Botany, Nomenclature, Horticulture, and Utilization. San Diego: Academic Press; 1993.
3. Fitmawati F, Hayati I, Mahatma R, Suzanti F. Phylogenetic Study of Mangifera from Sumatra, Indonesia using Nuclear and Chloroplast DNA Sequences. Sabrao J Breed Genet. 2018;50:295–312.
4. Warschefsky E. The Evolution and Domestication Genetics of the Mango Genus, Mangifera (Anacardiaceae). Doctoral dissertation. Florida International University; 2018.
5. Iyer CPA, Schnell RJ. Breeding and Genetics. In: The Mango, 2nd Edition. Botany, Production and Uses. Oxfordshire: CABI; 2009. p. 67–96.
6. Suhartono E, Viani E, Rahmadhan M, Gultom I, Rakhman M, Indrawardhana D. Total flavonoid and antioxidant activity of some selected medicinal plants in South Kalimantan of Indonesian. APCBEE Procedia. 2012;4:235–9.
7. Ekblom R, Galindo J. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity (Edinb). 2011;107:1–15. doi:10.1038/hdy.2010.152.
8. Eiadthong W, Yonemori K, Kanzaki S, Sugiura A, Utsunomiya N, Subhadrabandhu S. Amplified fragment length polymorphism analysis for studying genetic relationships among Mangifera species in Thailand. J Am Soc Hortic Sci. 2000;125.
9. Teo LL, Kiew R, Set O, Lee SK, Gan YY. Hybrid status of kuwini, Mangifera odorata Griff. (Anacardiaceae) verified by amplified fragment length polymorphism. Mol Ecol. 2002;11:1465–9. doi:10.1046/j.1365-294X.2002.01550.x.
10. Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review. Mol Ecol. 2002;11:1–16. doi:10.1046/j.0962-1083.2001.01418.x.
11. Viruel MA, Escribano P, Barbieri M, Ferri M, Hormaza JI. Fingerprinting, embryo type and geographic differentiation in mango (Mangifera indica L., Anacardiaceae) with microsatellites. Mol Breed. 2005;15:383–93. doi:10.1007/s11032-004-7982-x.
12. Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, van der Bank M, et al. A DNA barcode for land plants. Proc Natl Acad Sci. 2009;106:12794 LP – 12797. doi:10.1073/pnas.0905845106.
13. Pang X, Liu C, Shi L, Liu R, Liang D, Li H, et al. Utility of the trnH–psbA intergenic spacer region and Its combinations as plant DNA barcodes: a meta-analysis. PLoS One. 2012;7:e48833. doi:10.1371/journal.pone.0048833.
14. Li D-Z, Gao L-M, Li H-T, Wang H, Ge X-J, Liu J-Q, et al. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proc Natl Acad Sci. 2011;:201104551. doi:10.1073/pnas.1104551108.
15. Andrews S. FastQC a quality-control tool for high-throughput sequence data. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

16. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90. doi:10.1093/bioinformatics/bty560.
17. Boisvert S, Laviolette F, Corbeil J. Ray: Simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol. 2010;17:1519–33. doi:10.1089/cmb.2009.0238.
18. Kinjo S, Monma N, Misu S, Kitamura N, Imoto J, Yoshitake K, et al. Maser: one-stop platform for NGS big data from analysis to visualization. Database. 2018;2018. doi:10.1093/database/bay027.
19. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. doi:10.1093/bioinformatics/btv351.
20. Thiel T, Michalek W, Varshney R, Graner A. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). Theor Appl Genet. 2003;106:411–22. doi:10.1007/s00122-002-1031-0.
21. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3—new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115–e115. doi:10.1093/nar/gks596.
22. Matra DD, Poerwanto R, Santosa E, Sobir, Higashio H, Anzai H, et al. Analysis of allelic diversity and genetic relationships among cultivated mangosteen (Garcinia mangostana L.) in Java, Indonesia using microsatellite markers and morphological characters. Trop Plant Biol. 2016;9:29–41. doi:10.1007/s12042-016-9161-8.
23. Stecher G, Tamura K, Kumar S. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. Mol Biol Evol. 2020;37:1237–9. doi:10.1093/molbev/msz312.
24. Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. Am J Bot. 2002;89:132–44. doi:10.3732/ajb.89.1.132.
25. Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci U S A. 2005;102:8369 LP – 8374. doi:10.1073/pnas.0503123102.
26. Sang T, Crawford D, Stuessy T. Sang T, Crawford DJ, Stuessy TF.. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (Paeoniaceae). Am J Bot 84: 1120-1136. Am J Bot. 1997;84:1120.
27. Cheng T, Xu C, Lei L, Li C, Zhang Y, Zhou S. Barcoding the kingdom Plantae: new PCR primers for ITS regions of plants with improved universality and specificity. Mol Ecol Resour. 2016;16:138–49. doi:10.1111/1755-0998.12438.
28. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018;35:1547–9.
29. Jukes TH, Cantor CR. Evolution of Protein Molecules. In: Munro HN, editor. Mammalian Protein Metabolism. New York. USA: Academic Press; 1969. p. 21–132.
30. Tamura K. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases. Mol Biol Evol. 1992;9:678–87.

31. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783–91.
32. Ledesma N, Campbell RJ, Poor HW, Figueroa JJ, Zona S. Floral morphology of seven Mangifera species. Acta Hortic. 2017;1183:1–10.
33. Dutta SK, Srivastav M, Rymbai H, Chaudhary R, Singh AK, Dubey AK, et al. Pollen–pistil interaction studies in mango (Mangifera indica L.) cultivars. Sci Hortic (Amsterdam). 2013;160:213–21. doi:10.1016/j.scienta.2013.05.012.
34. Mukherjee SK, Litz RE. Introduction: Botany and Importance. In: Litz RE, editor. The Mango: Botany, production and uses. Wallingford: CAB International; 2009. p. 1–18.

Figures

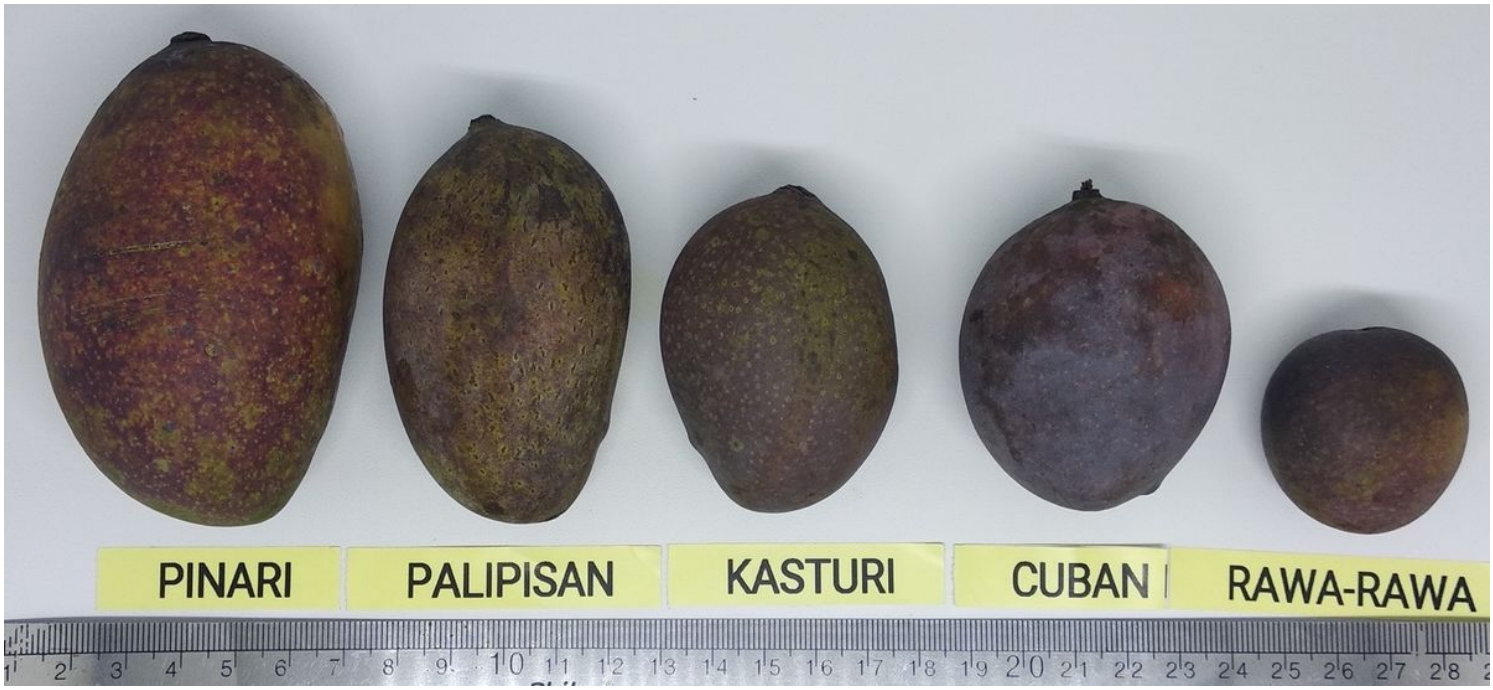

Figure 1

Appearance Fruits of M. casturi (Pinari, Palipisan (Pelipisan), Kasturi (Mawar)) and Rawa-rawa This photo taken by DDM

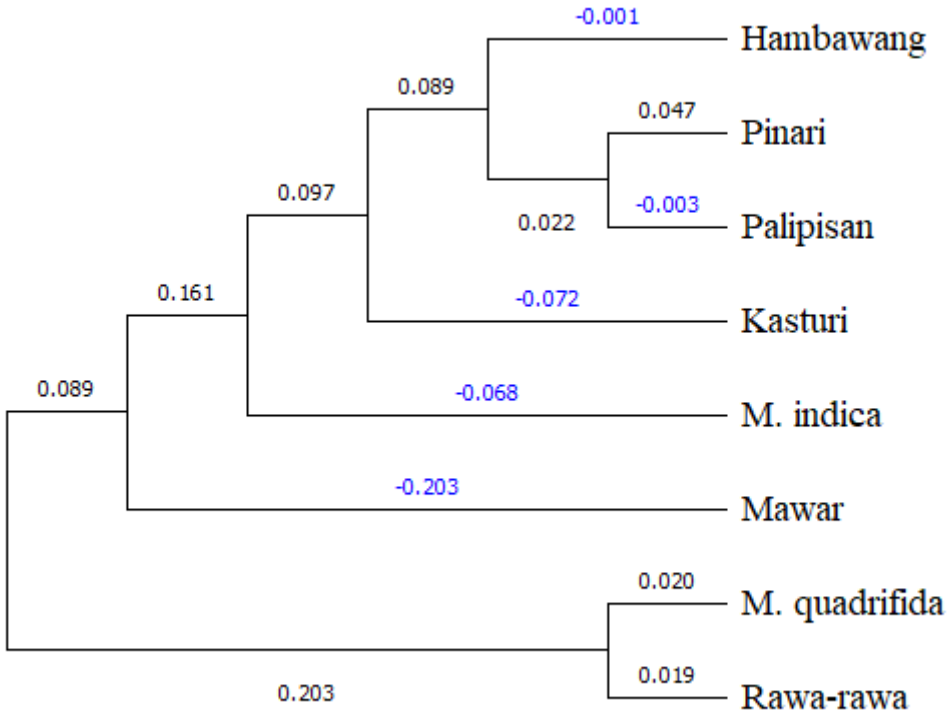

Figure 2

Dendrogram for UPGMA cluster analysis in four M. casturi accessions and four closely related species using 14 microsatellite loci

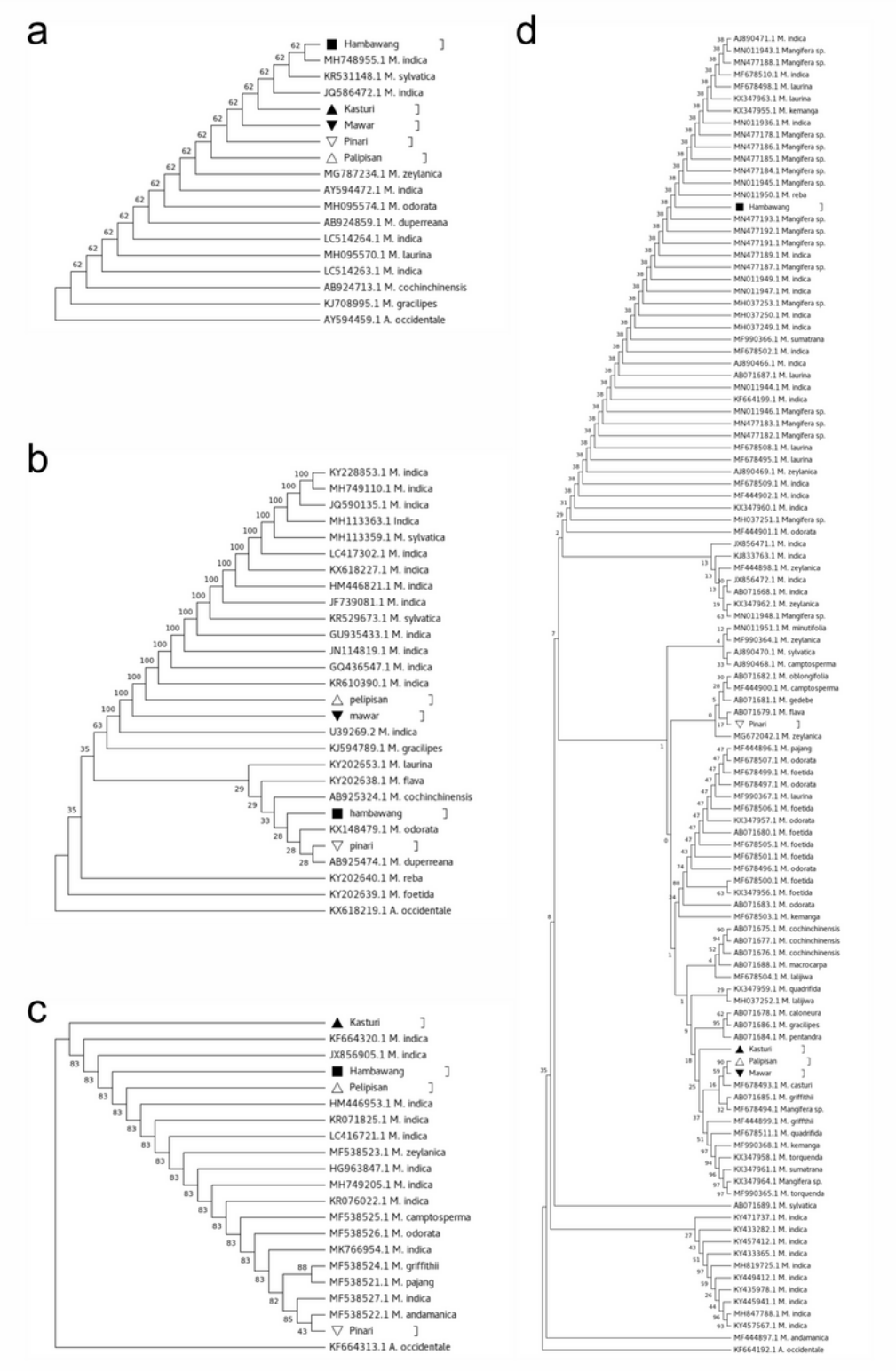

Figure 3

Phylogenetic analysis of (a) matK, (b) rbcL, (c) trnH-PsbA, and (d) internal transcribed spacers (ITS) by Maximum Likelihood method The evolutionary history was inferred by using the Maximum Likelihood method and Tamura 3-parameter model for matK, trnH-PsbA, ITS and Jukes-Cantor model for rbcL. The bootstrap consensus tree inferred from 10 000 replicates. The tree is rooted with the outgroup, Anacardium occidentale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryTable1.xlsx