Folia Hort. 37(1) (2025): 139–155

DOI: 10.2478/fhort-2025-0009

FOLIA HORTICULTURAE

Published by the Polish Society for Horticultural Science since 1989

ORIGINAL ARTICLE | Open access | https://sciendo.com/journal/FHORT

Diversity analysis and cross-species amplification of custard apple (*Annona squamosa*) using simple sequence repeats developed via next-generation sequencing technology

Abhilasha Krishnamurthy[1], *Thandavarayan Sakthivel*[2], *Kundapura V. Ravishankar*[1,*]

[1] Division of Basic Sciences, ICAR-Indian Institute of Horticulture Research, Hesaraghatta Lake Post, Bangalore 560089, India
[2] Division of Fruit Crops, ICAR-Indian Institute of Horticulture Research, Hesaraghatta Lake Post, Bangalore 560089. India

ABSTRACT

In this study, we isolated microsatellite markers for *Annona squamosa* through partial genome sequencing using next-generation sequencing (NGS) technology. From the NGS data, a total of 179,080 microsatellites were identified. Seventy randomly selected microsatellite markers were standardised for PCR amplification of *A. squamosa* DNA, and diversity was assessed in 40 genotypes. The number of alleles per locus ranged from 15 (ASIIHR38) to 46 (ASIIHR19, ASIIHR31 and ASIIHR35), with a mean value of 26.82. The polymorphic information content (PIC) values ranged from 0.748 (ASIIHR21) to 0.965 (ASIIHR40), with an average of 0.90. Dendrogram analysis classified the accessions into two major groups, which were further subdivided into multiple subgroups based on genetic relatedness. Most of the markers exhibited high levels of cross-species transferability, successfully amplifying in related species such as *Annona cherimola* (94.2%), *Annona reticulata* (95.7%), *Annona glabra* (97.1%), *Annona muricata* (90%) and *Annona atemoya* (80%). The results of this study demonstrate that species-specific genomic simple sequence repeats (SSRs) are more accurate for assessing genetic diversity. The SSR markers developed in this study will be valuable for diversity analysis, species identification, linkage map construction, DNA fingerprinting and crop improvement.

Keywords: *Annona squamosa* species, cluster analysis, genetic diversity, SSR marker

INTRODUCTION

India is considered as a secondary centre of origin for custard apple (*Annona squamosa* L.), which belongs to the family Annonaceae. The chromosome number of *A. squamosa* is 2n = 2x = 14 (Anuragi et al., 2016). Annonaceae, a family in the class Magnolideae, comprises 129 genera and more than 2000 species. Of these, *Annona cherimola* Mill. (cherimoya), *Annona glabra* L. (pond apple), *Annona muricata* L. (soursop), *Annona reticulata* L. (custard apple), *Annona atemoya* (a hybrid of *A. squamosa* and *A. cherimola*) and *A. squamosa* L. (sweetsop or sugar apple) are of major commercial importance (Vinay et al., 2017).

A. squamosa L. is commonly known as 'custard apple' or 'sugar apple'. While most species of *Annona* are thought to have originated in South America and the Antilles, wild soursop (*A. muricata*) is believed to have originated in Africa (Pinto et al., 2005). Now, these important species are found in nearly all continents, with soursop and sugar apple being widely distributed, particularly in the tropical regions.

*Corresponding author.
e-mail: kv_ravishankar@yahoo.co.in (Kundapura V. Ravishankar).

The pulp of *Annona* fruits is rich in minerals and vitamins (Gyamfi et al., 2011) and is a potential source of dietary fibre (up to 50% w/w dry basis). The seeds, especially those of *A. squamosa*, contain a significant amount of oil, which can be used for industrial purposes (Mariod et al., 2010). Custard apple is a versatile plant with multiple uses, and it is hardy and deciduous by nature. However, its cultivation is still in the early stages of domestication (Van Zonneveld et al., 2012).

The genetic resources and plant diversity of *Annona* species are being eroded due to agricultural modernisation, urbanisation and land-use changes. Therefore, the genetic resources of edible custard apples, mostly found *in situ* and in natural populations, need to be conserved. Characterising genetic diversity is essential for the efficient conservation and utilisation of genetic resources. Despite this, few efforts have been made to identify the diverse germplasm of custard apples and assess the diversity using molecular markers.

There are limited reports on the use of molecular markers to assess genetic diversity in *Annona* species. Some of these markers include random amplified polymorphic DNA (RAPD) markers (Ronning et al., 1995; Bharad et al., 2009), amplified fragment length polymorphism (AFLP) markers (Rahaman et al., 1998; Zhichang et al., 2011) and simple sequence repeat (SSR) markers derived from related species like *A. cherimola* (Escribano et al., 2004; Kwapata et al., 2007; Pereira et al., 2008; Escribano et al., 2008; Van Zonneveld et al., 2012; Anuragi et al., 2016; Thanachseyan et al., 2017). However, no species-specific SSR markers have been developed for *A. squamosa*, which would provide more precise information on the diversity of this economically important species. In light of this, in this study, we have attempted to develop SSR markers and assess the genetic diversity of *A. squamosa* in the Indian subcontinent and to examine transferability across related species.

MATERIALS AND METHODS

Plant material and DNA isolation

For molecular diversity analysis, a total of 40 *A. squamosa* cultivars, along with five related species – *A. cherimola*, *A. reticulata*, *A. glabra*, *A. muricata* and *A. atemoya* – were selected from the field gene bank of Indian Council of Agricultural Research-Indian Institute of Horticultural Research (ICAR-IIHR), Bengaluru, India (Supplementary Table 1). The 40 cultivars were collected from different regions of India and are maintained in the field germplasm bank. Total genomic DNA was isolated from young, tender leaves using a modified CTAB method (Ravishankar et al., 2000). The quantity and quality of DNA were assessed using a UV spectrophotometer (NABI, Microdigital) at 260 nm and visualised via agarose gel electrophoresis (0.8%) under a UV transilluminator. DNA samples were then diluted to 20 ng · μL^{-1} with sterile water and stored at 4°C for further analysis.

Table 1. Distribution of microsatellite types in *A. squamosa* genome.

Motif types	Number of SSRs	Frequency (%)
Mono-nucleotide	100658	56.20
Di-nucleotide	46493	25.96
Tri-nucleotide	15590	8.70
Tetra-nucleotide	2266	1.26
Penta-nucleotide	495	0.27
Hexa-nucleotide	202	0.11
Compound nucleotide	13376	7.4
Total	179080	

SSR, simple sequence repeat.

Genome sequencing and microsatellite identification

Total genomic DNA from the *A. squamosa* cultivar Balanagar was used for partial genome sequencing and microsatellite marker identification. Genome sequencing was performed using the Illumina HiSeq 2500 platform at M/Sandor Specialty Diagnostics Pvt. Ltd., Hyderabad, following the manufacturer's instructions. The reads were assembled into scaffolds using the *de novo* assembly tool SOAPdenovo2 (version 2.04; Altschul et al., 1990). Microsatellites (SSRs) were identified from the assembled genome fragments using MISA software (Beier et al., 2017), and primers for the identified microsatellites were designed using PRIMER3 software (Rozen and Skaletsky, 2000). Sequence data have been submitted to NCBI (Bioproject PRJNA682654).

PCR amplification and fragment size analysis

A total of 100 SSR primers were randomly selected for PCR standardisation. Pooled DNA was initially used to standardise PCR conditions at various annealing temperatures. Primers were tailed with M13 sequences, and PCR was conducted using M13-tailed primers labelled with fluorophores (Schuelke, 2000). M13 tails were added to both the forward (GTAAAACGACGGCCAGT) and reverse (GTTTCTT) primers at the 5′ end. We tested the 40 *A. squamosa* genotypes and the 5 related species (*A. cherimola*, *A. reticulata*, *A. glabra*, *A. muricata* and *A. atemoya*).

DNA amplification was performed in a 20 µL reaction volume, containing 2.5 µL of template DNA (20 ng · μL^{-1}), 1 µL of each forward and reverse primers (5 pM), 0.4 µL of Taq DNA polymerase (3 U · μL^{-1}), 2.0 µL of Taq buffer A (with 15 mM $MgCl_2$) (10X), 2 µL of dNTPs (1 mM), 1 µL of M13 primers with fluorescent labels (5 pM) (fluorophores such as FAM, ATTO 550, ATTO 565 and YAKAMA YELLOW were synthesised using M/S Eurofins, Bengaluru) and 9.6 µL of sterile water. PCR amplification was carried out using the T100™ Thermal Cycler (BIO-RAD, California, USA).

Table 2. Genetic analysis of SSR primers data using *A. squamosa* genotypes.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR01 F	ACATGCCCTCAATCATCTCC	$(TA)_6$	55	23	0.50	0.93	0.91	0.0103
ASIIHR01 R	AGTTAAAAATTAATGAATGAGCCAGG							
ASIIHR02 F	CTGAATTCTAACAGATGTGCTGGG	$(TA)_8$	60	20	0.60	0.91	0.90	0.0150
ASIIHR02 R	GCTTTGGACGTCTACGCAT							
ASIIHR03 F	TCAGACTTCAACTCAAGTCGATCC	$(AT)_8$	55	25	0.62	0.93	0.91	0.0114
ASIIHR03 R	TAATACCCAAGTCATGAAGACCAA							
ASIIHR04 F	ACTCGTACTGCTATAAAAGTGGGT	$(AT)_7$	55	37	0.65	0.97	0.95	0.0032
ASIIHR04 R	ATGAGAGCTGCCCACGAC							
ASIIHR05 F	AAAACTGTCGGCTTCCATGT	$(TA)_6$	55	20	0.80	0.84	0.82	0.0364
ASIIHR05 R	AGAACCAACACAGCTTCCATT							
ASIIHR06 F	CTTCTGTTGTCATCACTCGCA	$(TA)_6$	55	22	0.52	0.90	0.88	0.0186
ASIIHR06 R	CACGATAGAAATCGAAGAACAGA							
ASIIHR07 F	AAATATCAAGTTTAAAGCAGTATTTGC	$(TA)_6$	55	30	0.97	0.95	0.93	0.0064
ASIIHR07 R	TACAAGCGGGACATAGAAGC							
ASIIHR08 F	GATAGGAAGACAAACAGTTAGTTTAGG	$(AG)_6$	55	28	0.92	0.91	0.90	0.0143
ASIIHR08 R	CGGCAGCTTTCTTCTTGTTC							
ASIIHR09 F	CCAAGAAATCTCACGTTCGC	$(AG)_{13}$	60	30	0.82	0.95	0.93	0.0065
ASIIHR09 R	TCTTTCTTTAGCAAGTAGTGGTTTTT							
ASIIHR10 F	TGTGGGTTTATATTGACCATCATT	$(GA)_8$	60	31	0.97	0.93	0.94	0.0057
ASIIHR10 R	GTGGAGCTAGGGTTCTTCGTC							
ASIIHR11 F	ACGCTTTTCTTCTCCGGC	$(GA)_9$	55	26	0.52	0.95	0.91	0.0110
ASIIHR11 R	CTCGCCTGCTCCTCTCAC							
ASIIHR12 F	GCTTTGAGAGAAAATGAGAGACAA	$(AG)_{11}$	55	30	0.92	0.91	0.93	0.0065
ASIIHR12 R	CTACCTTTCCGGCGAATCC							
ASIIHR13 F	GATATTCAAAGAGCACGAGAGGA	$(GA)_6$	55	33	0.70	0.88	0.89	0.0147
ASIIHR13 R	AAAAGTTAGTCGGCAAATCCC							
ASIIHR14 F	GTGAGAGAGAGAGAAGGAAGGC	$(GA)_{10}$	55	19	0.70	0.93	0.85	0.0284
ASIIHR14 R	GCGAATCTCTCTTCCGACAG							

(*Continued*)

Table 2. Continued.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR15 F	TTTTCTCTTTTCTTCGTTCTTGC	$(GA)_{12}$	60	25	0.82	0.92	0.91	0.0107
ASIIHR15 R	GAGGGGGCTGGTGACCTAT							
ASIIHR16 F	CCTAATCGGAAAGGTGCAAA	$(AC)_7$	55	23	0.70	0.85	0.91	0.0126
ASIIHR16 R	TGGCTTTATTGGATGTGTTTGA							
ASIIHR17 F	GCTAAGACGGGGCCAACC	$(AC)_6$	60	16	0.75	0.87	0.83	0.0388
ASIIHR17 R	TGCCTTATTTCTTTAAACAGGGTC							
ASIIHR18 F	CTCTCTCTTGTGCTTCTCCCA	$(AC)_6$	55	18	0.82	0.97	0.84	0.0338
ASIIHR18 R	CTTTCTCTCTCTCCCCCTCC							
ASIIHR19 F	TGACGAGATCGAATTAAGTACCC	$(CA)_8$	60	46	0.92	0.94	0.96	0.0022
ASIIHR19 R	TGATGCCTATAAAAGGGGCA							
ASIIHR20 F	ACCAGCAAATCCTGGGAAG	$(AC)_6$	55	29	1.00	0.78	0.93	0.0076
ASIIHR20 R	TGAATCTGCAACTCAAAACTGA							
ASIIHR21 F	TTGATGCAATTCTTCAGTTTGA	$(AC)_8$	55	16	0.92	0.92	0.74	0.0758
ASIIHR21 R	TACAAGGGTTGGGAATTTGG							
ASIIHR22 F	CATACATTTTGCCCACGACC	$(GT)_8$	60	25	0.77	0.95	0.90	0.0132
ASIIHR22 R	CACAGATACACACACACGAACAA							
ASIIHR23 F	AAAAAGTCCATTCTTTTTCTCCA	$(GT)_6$	55	28	0.92	0.94	0.93	0.0066
ASIIHR23 R	TCATTTCTTCATCCCATTGC							
ASIIHR24 F	CACATCACCCATATAAAAAGCG	$(TG)_6$	60	35	0.82	0.93	0.93	0.0065
ASIIHR24 R	GTTGGGACAACTCTTCACCC							
ASIIHR25 F	CAGCGATGGTTGCTTAATTTG	$(GT)_6$	55	25	0.77	0.93	0.92	0.0099
ASIIHR25 R	AGTAGGTGGAGAGACCCACG							
ASIIHR26 F	AGCAAAAGTGGTCATCCGAA	$(TG)_{13}$	60	30	0.80	0.96	0.91	0.0102
ASIIHR26 R	TTTTAGATCGCTAAGAAGATCACA							
ASIIHR27 F	TCGCTATTTCAAAATTAAGTAAAAGAA	$(TG)_8$	55	37	0.82	0.96	0.95	0.0044
ASIIHR27 R	TGTTTTAGTTGAGCAAGGCTAGG							
ASIIHR28 F	TCTTGTTTTTGCCAGTTCCC	$(GT)_8$	60	34	0.87	0.87	0.95	0.0038
ASIIHR28 R	GCTTGAACTCAAAGCATGTTG							
ASIIHR29 F	TGTTACTGTTGGGCATGGAA	$(GT)_6$	55	23	0.67	0.95	0.85	0.0263

(*Continued*)

Table 2. Continued.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR29 R	GGTTGGGTTGAAAATTTAAGCA							
ASIIHR30 F	CCTTCCACCCTTGGATCTTA	$(CT)_6$	55	33	0.82	0.97	0.94	0.0058
ASIIHR30 R	GATCAATGTGGATAAAACTTCGC							
ASIIHR31 F	CTTTTCTTCTCCATTTTCCCG	$(CT)_8$	55	44	0.95	0.95	0.95	0.0031
ASIIHR31 R	CGTTGGAGATTCCAAGAGAAA							
ASIIHR32 F	AGGTGGATCGCTTAAGATGAA	$(TC)_6$	55	29	0.80	0.92	0.93	0.0069
ASIIHR32 R	TCTGAGCAGGTGTGTAGTCCT							
ASIIHR33 F	ACTGGCCGAGGAAAGGGT	$(TC)_{12}$	55	28	0.97	0.97	0.91	0.0117
ASIIHR33 R	GAGAGGGAGGAAAAAGTTAATCG							
ASIIHR34 F	CTCCCCGTTACCCAACTG	$(CT)_9$	55	35	0.72	0.96	0.95	0.0034
ASIIHR34 R	GGTTCCTTCCGTCCTCCTG							
ASIIHR35 F	TTTCATAGCTTTTATTGCTTTCTTAG	$(GA)_8$	55	40	0.85	0.96	0.95	0.0036
ASIIHR35 R	AGTCGGGTCGAATCCACA							
ASIIHR36 F	CTTCTGTCTTCCTCATTTTCTCG	$(AG)_8$	60	35	0.77	0.94	0.94	0.0047
ASIIHR36 R	TGGGAGAGAAGTGAGAGAGCTT							
ASIIHR37 F	GGCCACACTTGCTCAAAAAT	$(GC)_7$	55	19	0.92	0.90	0.92	0.0090
ASIIHR37 R	TTCTTATTGATGCGTAGCGG							
ASIIHR38 F	GGGAGGAAACTTGATCCCTT	$(GC)_6$	60	15	0.27	0.94	0.88	0.0196
ASIIHR38 R	AAAATTATGGTGCAGTGGCG							
ASIIHR39 F	GGCCACACTTGCTCAAAAAT	$(GC)_7$	55	28	0.27	0.97	0.92	0.0087
ASIIHR39 R	TTCTTATTGATGCGTAGCGG							
ASIIHR40 F	CCAATCCCTTTATCCAAGCA	$(GC)_7$	55	35	0.27	0.93	0.96	0.0027
ASIIHR40 R	GTGCAATTTGAAAAGCAGCA							
ASIIHR41 F	CATCTCCGCAACACCAGATA	$(GC)_8$	55	27	0.67	0.94	0.92	0.0092
ASIIHR41 R	GCCAGAAGAGGCAGTGATTC							
ASIIHR42 F	AGAGGAAAACTTACAAAAACATAGACG	$(ATG)_6$	55	26	0.30	0.93	0.92	0.0089
ASIIHR42 R	GCCCTAAGAGAGCAGTTTACCC							
ASIIHR43 F	GTATGTCATGGAGGATACAGGGA	$(GAA)_5$	55	23	0.25	0.90	0.92	0.0091

(*Continued*)

Table 2. Continued.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR43 R	CATCCTCATCATCTCCCACA							
ASIIHR44 F	ACTGCTGCTGAGATGTGCG	$(GAT)_7$	55	23	0.30	0.92	0.91	0.0112
ASIIHR44 R	CTGCTGCTGCTGTTGACG							
ASIIHR45 F	TTATTGTATAAAACACCCCAAAGAA	$(TTC)_{10}$	60	18	0.45	0.96	0.89	0.0183
ASIIHR45 R	TTCTTATCATTTTGCCCGTCT							
ASIIHR46 F	TTGGCAACCATCAGAATAAGA	$(AAT)_5$	60	17	0.35	0.94	0.90	0.0152
ASIIHR46 R	ACAACCCTGCTTCCATTCAA							
ASIIHR47 F	AAAAACCTTGGGCTTGTGC	$(TCT)_6$	55	32	0.37	0.95	0.95	0.0037
ASIIHR47 R	CCTTGCCCCTATTATTTTCC							
ASIIHR48 F	TTGGTGAAGCATTCAAAAATTC	$(ATG)_{10}$	55	32	0.52	0.93	0.93	0.0072
ASIIHR48 R	CCTTGCCCCTATTATTTTCC							
ASIIHR49 F	TCAAACGCCCGCATATTTA	$(AGC)_5$	55	26	0.40	0.94	0.93	0.0070
ASIIHR49 R	GCTGGAGAAAGACGGCAAG							
ASIIHR50 F	ACCTCAAAGCTAGGGGGTAAA	$(GAA)_5$	55	26	0.35	0.93	0.92	0.0103
ASIIHR50 R	CCGAAGTAGAGATACGCCTCTT							
ASIIHR51 F	AAAATGAGCATGAAGAAAAGAAAAA	$(ATT)_{10}$	55	22	0.30	0.94	0.93	0.0081
ASIIHR51 R	GATTGTAAGACAAATTGAGATGTAATG							
ASIIHR52 F	TCCCATTTTCTGATCGAGTTG	$(TCA)_5$	55	23	0.42	0.93	0.92	0.0101
ASIIHR52 R	TAACCCTCGCCGTGAATAG							
ASIIHR53 F	AGACTAATCTAAGTTTAAAGCAAGCAA	$(AAG)_5$	55	21	0.45	0.94	0.92	0.0090
ASIIHR53 R	TGATCTTTGTTGAAGCGTCTCT							
ASIIHR54 F	TTCAAGAATCATCTTTTAAGTCAACC	$(GTG)_5$	55	30	0.45	0.94	0.92	0.0079
ASIIHR54 R	TTTTCATGGAAACAACCAAATG							
ASIIHR55 F	TCACTTGGAATAATGTGGAACG	$(GTC)_{11}$	55	19	0.40	0.91	0.89	0.0181
ASIIHR55 R	CAGTCCCCAGCTCCAAAC							
ASIIHR56 F	CCCCAATCCCAATCCTTAGT	$(AGC)_5$	55	28	0.75	0.96	0.94	0.0047
ASIIHR56 R	GGAGTTCGTGTGCTTTACCG							
ASIIHR57 F	GACGTGCTGCTG	$(CGA)_5$	55	24	0.35	0.92	0.90	0.0134

(*Continued*)

Table 2. Continued.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR57 R	GGATTCTTCACCAGGCAGTT							
ASIIHR58 F	AAAATGCATGCCTTGTTTGT	$(ATTC)_7$	60	23	0.55	0.91	0.90	0.0146
ASIIHR58 R	TAGTCCTTGAGCAACACATGC							
ASIIHR59 F	AAGGGCATGTTGTCTTCTCAA	$(TTGT)_6$	60	29	0.45	0.95	0.93	0.0070
ASIIHR59 R	TTTGCAAGTTTATTTCTGCTCAA							
ASIIHR60 F	TCCCGACCTTTCTTACGGAT	$(ATCTC)_6$	60	22	0.67	0.87	0.85	0.0276
ASIIHR60 R	TTTATCTCTATCTCTCCACCGGA							
ASIIHR61 F	AATATGTTAACCCGAAACTCAACC	$(TAGGGT)_5$	55	16	0.60	0.95	0.78	0.0571
ASIIHR61 R	AATAATTACATAATTGATGAGGGTCAA							
ASIIHR62 F	CTCTGTTTCTATCTCTCTCAAACTCA	$(TC)_9$tgtctctcatt$(TC)_6$	55	34	0.50	0.87	0.94	0.0058
ASIIHR62 R	GGAATGGGAGAGTATTTGAAGG							
ASIIHR63 F	TACCGGATCTCTCATTTTCG	$(TC)_8$cgccatttctatctct ctctccgtcattcct ctctctctttctctgttttccttt ggaaaaatcggca aacccaaat$(TC)_6$	55	24	0.62	0.81	0.90	0.0139
ASIIHR63 R	GCTGGGAGAGTGAGCTGAAA							
ASIIHR64 F	CATGTTGGACATGTGAGCCA	$(A)_{10}g(A)_{10}$	55	28	0.85	0.95	0.91	0.0101
ASIIHR64 R	ATGCCTATTTTAGGCTGGGT							
ASIIHR65 F	GGCGTCCAAAAATTGAGATT	$(AT)_7$gtatttgccccatggg ccccaaaaaaaataaa$(AT)_7$ ttttagactttcaa$(AT)_7$	55	31	0.67	0.92	0.92	0.0082
ASIIHR65 R	TTTGGGGAGTATCTACTCAGGC							
ASIIHR66 F	TCTACGCTACCCAGCAAATG	$(T)_{11}(TA)_6$*	55	28	0.67	0.93	0.90	0.0128
ASIIHR66 R	AATGACAATATGATTTGCCTTGA							
ASIIHR67 F	CCGACTTCAACCTTCTGAGC	$(A)_{11}$ttaaaacagattta tcaaaaatgttctta cgagaaagggaaaa taggagaaaaaggt agaatgagggttttc tttttgtgcgtgttttgga$(AG)_8$	55	27	0.60	0.94	0.91	0.0075

(*Continued*)

Table 2. Continued.

Locus	Forward sequence and reverse sequence 5′–3′	Repeat motif*	Tm (°C)	No. of alleles per locus (k)	Ho	He	PIC	PI
ASIIHR67 R	TCAATGGAATATTCACTTTTCTAGG							
ASIIHR68 F	TGAGCCATTTGGAGATTTTTG	$(AAT)_6$	55	26	0.55	0.95	0.92	0.0066
ASIIHR68 R	GTGGATACTCCCCGACTGG							
ASIIHR69F	TTCACATACTTTTGCCTGAGTAGA	$(AT)_6$gacctcttcgctcaatc caagcctcaatgtc$(A)_{10}$	55	21	0.67	0.92	0.90	0.0143
ASIIHR69 R	GCCATCTTGGGCTTTTTAGA							
ASIIHR70F	GAAGGGAGATGCAAACGTTAAG	$(A)_{11}(T)_{10}$	55	27	0.60	0.92	0.91	0.0118
ASIIHR70 R	AACTGCTTGCTTTATGCACTTT							

*Indicates the number of times a particular motif was repeated in the microsatellite locus. For example: $(TA)_6$ – means TA has been repeated six times in the sequence results analysed during SSR identification. He, expected heterozygosity; Ho, observed heterozygosity; PI, probability of identity; PIC, polymorphic information content; SSR, simple sequence repeat.

The temperature profile was 94°C for 2 min, followed by 35 cycles of 94°C for 30 s, annealing at 50°C/55°C/60°C (Table 3) for 30 s, 72°C for 1 min and a final extension at 72°C for 10 min. Amplified products were confirmed via 1.5% agarose gel electrophoresis, and the primers with strong, distinct amplification products were selected based on Tm values.

The PCR products from the four different fluorescent dyes were multiplexed and resolved using an automated ABI 3730 DNA analyser (Applied Biosystems, USA) at M/S Eurofins, Bengaluru. The resulting data were analysed using Peak Scanner software (Applied Biosystems) to determine the exact fragment size of the PCR products in base pairs (bp).

Statistical analysis

Using the fragment size data, polymorphic information content (PIC), probability of identity (PI), observed heterozygosity (Ho), expected heterozygosity (He) and the number of alleles per locus were calculated using CERVUS software (version 3.0; Kalinowski et al., 2007). SSR genotypic data were used to construct a dendrogram via unweighted pair group method with arithmetic mean (UPGMA), employing a neighbour-joining (NJ) tree and simple matching (SM) dissimilarity matrix using DARwin software. Bootstrap analysis was done with 10000 iterations (version 6.0.10; Perrier and Jacquemoud-Collet, 2003).

A Bayesian model-based clustering was performed using STRUCTURE software (version 2.3.1; Pritchard et al., 2000). The number of subgroups (K) was set between 2 and 10, with 10 iterations for each K value. The project parameters included a burn-in period of 10000, followed by 100000 Monte Carlo Markov Chain (MCMC) replications. Structure Harvester software was used to generate ΔK (Evanno et al., 2005) to estimate the number of populations. Additionally, analysis of molecular variance (AMOVA) was conducted to assess genetic variability among and within populations using GenAlEx software (version 6.5; Peakall and Smouse 2006).

RESULTS

Genome sequencing and microsatellite identification

Next-generation sequencing (NGS) technology, using the Illumina HiSeq 2500 platform, was employed to sequence the total genomic DNA isolated from the *A. squamosa* cultivar Balanagar. The sequencing run produced 3.9 million bases from 47476322 reads, after low-quality reads were filtered out. A total of 1388525 contigs were generated, and three assemblies were created based on different K-mer lengths of 40, 62 and 89. These assemblies were compared using QUAST software, and the K-mer length of 89 was selected as the best genome assembly based on the number of contigs, overall size, N50 value (954) and contig length

Table 3. Cross species amplification of 70 SSR loci derived from *A. squamosa.*

No.	Locus	*A. cherimola*	*A. reticulata*	*A. glabra*	*A. muricata*	*A. atemoya*
1	ASIIHR01	A	A	A	A	A
2	ASIIHR02	A	A	A	A	A
3	ASIIHR03	A	A	A	A	A
4	ASIIHR04	A	A	A	A	A
5	ASIIHR05	A	A	A	A	A
6	ASIIHR06	A	A	A	A	A
7	ASIIHR07	NA	A	NA	A	NA
8	ASIIHR08	A	A	A	A	A
9	ASIIHR09	A	A	A	A	A
10	ASIIHR10	A	A	A	A	A
11	ASIIHR11	A	A	A	A	A
12	ASIIHR12	A	A	A	A	A
13	ASIIHR13	A	A	A	A	A
14	ASIIHR14	A	A	A	A	A
15	ASIIHR15	A	A	A	A	A
16	ASIIHR16	A	A	A	A	A
17	ASIIHR17	A	A	A	A	A
18	ASIIHR18	A	A	A	A	A
19	ASIIHR19	A	A	A	A	A
20	ASIIHR20	A	A	A	A	A
21	ASIIHR21	A	A	A	A	A
22	ASIIHR22	A	A	A	NA	NA
23	ASIIHR23	A	A	A	A	A
24	ASIIHR24	A	A	A	A	A
25	ASIIHR25	A	A	A	A	A
26	ASIIHR26	A	A	A	A	A
27	ASIIHR27	NA	A	A	A	NA
28	ASIIHR28	A	A	A	A	A
29	ASIIHR29	A	A	A	A	A
30	ASIIHR30	A	A	A	A	A
31	ASIIHR31	A	A	A	A	A
32	ASIIHR32	A	A	A	A	A
33	ASIIHR33	A	A	A	A	A
34	ASIIHR34	A	A	A	A	A
35	ASIIHR35	A	A	A	A	NA
36	ASIIHR36	A	A	A	A	A
37	ASIIHR37	A	A	A	A	A
38	ASIIHR38	A	A	A	A	A
39	ASIIHR39	A	A	A	A	A
40	ASIIHR40	A	A	A	A	A
41	ASIIHR41	A	A	NA	A	NA
42	ASIIHR42	A	A	A	A	A
43	ASIIHR43	A	A	A	A	A
44	ASIIHR44	A	A	A	A	A
45	ASIIHR45	A	A	A	A	A
46	ASIIHR46	A	A	A	A	A

(*Continued*)

Table 3. Continued.

No.	Locus	*A. cherimola*	*A. reticulata*	*A. glabra*	*A. muricata*	*A. atemoya*
47	ASIIHR47	A	A	A	A	A
48	ASIIHR48	A	A	A	A	NA
49	ASIIHR49	A	A	A	A	A
50	ASIIHR50	A	A	A	A	A
51	ASIIHR51	NA	A	A	A	A
52	ASIIHR52	A	A	A	NA	NA
53	ASIIHR53	NA	A	A	A	NA
54	ASIIHR54	A	A	A	A	A
55	ASIIHR55	A	A	A	A	A
56	ASIIHR56	A	A	A	A	A
57	ASIIHR57	A	A	A	A	A
58	ASIIHR58	A	NA	A	NA	NA
59	ASIIHR59	A	A	A	A	A
60	ASIIHR60	A	A	A	NA	NA
61	ASIIHR61	A	A	A	A	A
62	ASIIHR62	A	NA	A	NA	NA
63	ASIIHR63	A	A	A	A	A
64	ASIIHR64	A	A	A	A	A
65	ASIIHR65	A	A	A	NA	NA
66	ASIIHR66	A	A	A	A	A
67	ASIIHR67	A	A	A	A	A
68	ASIIHR68	A	NA	A	NA	NA
69	ASIIHR69	A	A	A	A	A
70	ASIIHR70	A	A	A	A	A
Transferability %		94.2	95.7	97.1	90.00	80.00

A, amplified, NA, not amplified; SSR, simple sequence repeat.

distribution. A total of 58527 scaffolds were assembled and screened for microsatellite identification using MISA software (Supplementary Table 2).

The results revealed that the scaffolds contained 179080 microsatellites in total. Among the microsatellite repeats, mono-nucleotide repeats were the most abundant, accounting for 56.2% of all repeat types, followed by di-nucleotide repeats (25.96%), tri-nucleotide repeats (8.7%), tetra-nucleotide repeats (1.26%), penta-nucleotide repeats (0.27%), hexa-nucleotide repeats (0.11%) and compound nucleotide repeats (7.4%) (Table 1).

Using Primer 3.0 software, 58527 primers were designed (excluding mono-nucleotide repeats). Of these, 100 microsatellite markers were randomly selected, and primers were synthesised. After initial screening for amplification of reproducible PCR products, 70 SSR primers were selected for future analysis. Seventy of these primers were used to amplify DNA from 40 *A. squamosa* genotypes and five related *Annona* species. The product sizes for the 70 microsatellites ranged from 200 bp to 450 bp. The analysis of data from these microsatellite markers showed a total of 1878 alleles, with an average of 26.82 alleles per locus. The number of alleles ranged from 15 (ASIIHR38) to 46 (ASIIHR19, ASIIHR31 and ASIIHR35) per locus. The three markers with the highest number of alleles (ASIIHR19, ASIIHR31 and ASIIHR35) had simple di-nucleotide repeat motifs. The PIC ranged from 0.78 (ASIIHR61) to 0.96 (ASIIHR40 and ASIIHR19), with a mean of 0.90.

Ho ranged from 0.250 to 1.000, with a mean of 0.647, while He ranged from 0.781 to 0.978, with a mean of 0.926. The PI showed a maximum value of 0.0758 for the ASIIHR21 locus (16 alleles) and a minimum value of 0.0022 for the ASIIHR19 locus (46 alleles), with an average value of 0.01283 across all SSR loci. The characteristics of the 70 polymorphic SSR markers are summarised in Table 2.

Genetic diversity and population structure analysis of A. squamosa genotypes

A dendrogram (Figure 1) was generated using the UPGMA method, based on a shared allele matrix, revealing two major clusters. Of the 40 cultivars, 23 (57.5%) were grouped in Cluster I, while the remaining 17 (42.5%) were grouped in Cluster II, based on their

genetic relatedness. Cultivars native to Andhra Pradesh were distributed across both clusters. A bar plot of ΔK, following the method described by Evanno et al. (2005), indicated that the optimal ΔK was at K = 2 (Figure 2).

An AMOVA revealed limited variation among populations, with 30% of the variation occurring among individuals and 70% within individuals of the populations. The Fst value from AMOVA was 0.065, indicating low genetic differentiation (Table 4).

DISCUSSION

In this study, we utilised the Illumina HiSeq 2500, an NGS platform, to sequence the genome and isolate microsatellites for *A. squamosa* (L.). NGS technologies have become essential tools in plant biology for whole-genome sequencing, marker development and gene identification. Our sequencing efforts yielded 3.9 million bp sequences, and the assembly of 47476322 raw reads resulted in 58527 scaffolds. Microsatellite analysis revealed that mono-nucleotide repeats were the most abundant (56.20%), followed by di-nucleotides (25.96%), tri-nucleotides (8.7%) and tetra-nucleotides (1.26%). Mono-, di-, tri- and tetra-nucleotide repeats accounted for the majority (92.12%) of microsatellites in custard apple. This pattern of mono-repeat predominance and followed by di-repeats is consistent with findings in other plant species (Ravishankar et al., 2015, 2017a, 2017b).

Among the 179080 identified SSRs, 46493 were di-nucleotide repeats, 15590 were tri-nucleotide repeats, 2266 were tetra-nucleotide repeats, and the remaining were penta- and hexa-nucleotide repeats (Table 2). From the 58527 designed primers, 70 SSRs were randomly selected and standardised for PCR conditions. These SSRs were employed for DNA amplification of 40 *Annona* cultivars and 5 related species (*A. cherimola*, *A. reticulata*, *A. glabra*, *A. muricata* and *A. atemoya*) (Table 3).

The SSR markers showed high PIC values, ranging from 0.78 to 0.96, which were notably higher than those reported in earlier studies using SSRs from *A. cherimola* in *A. squamosa* (Anuragi et al., 2016). For example, Anuragi et al. (2016) examined molecular diversity among 20 *A. squamosa* genotypes using 12 *A. cherimola* SSR markers, which had PIC values ranging from 0.169 to 0.694. Thus, species-specific SSR are more efficient in assessing the diversity. Similarly, Thanachseyan et al.

Table 4. AMOVA analysis of genetic variances within and among populations of 40 custard apple accessions.

Source of variation	df	SS	MS	Estimated variance	% Variation
Among populations	5	213.769	42.74	0.021	0
Among individuals	34	1447.244	42.56	9.870	30
Within individuals	40	913.00	22.825	22.825	70
Total					100

AMOVA, analysis of molecular variance; df, degrees of freedom; MS, mean sum of squares; SS, sum of squares.

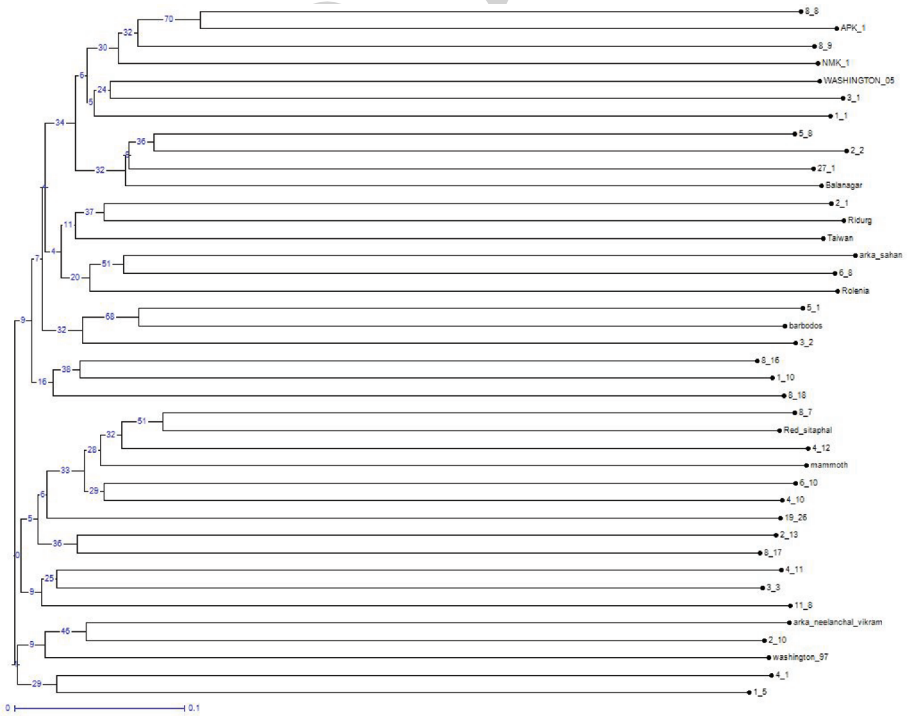

Figure 1. Dendrogram analysis of *A. squamosa* genotypes using SSR markers data. SSR, simple sequence repeat.

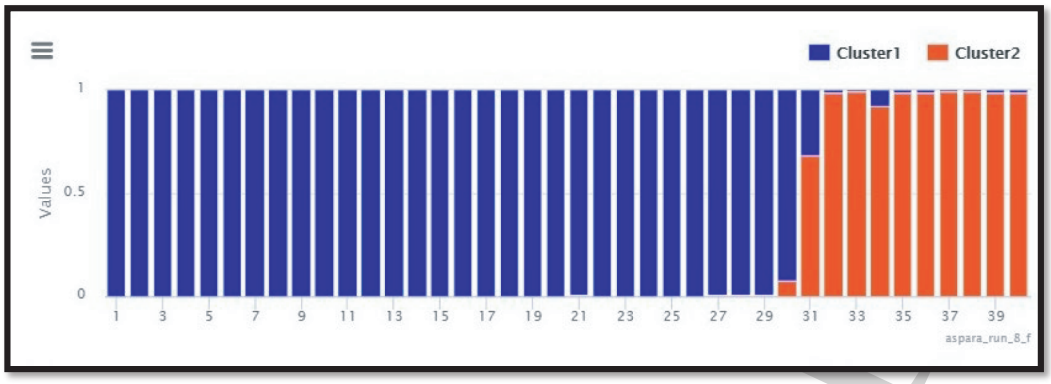

Figure 2a. Bar plot of the STRUCTURE analysis showing the partitioning of 40 *Annona* genotypes into K = 2 colored segments that designate the population's estimated membership fraction in the inferred subgroups.

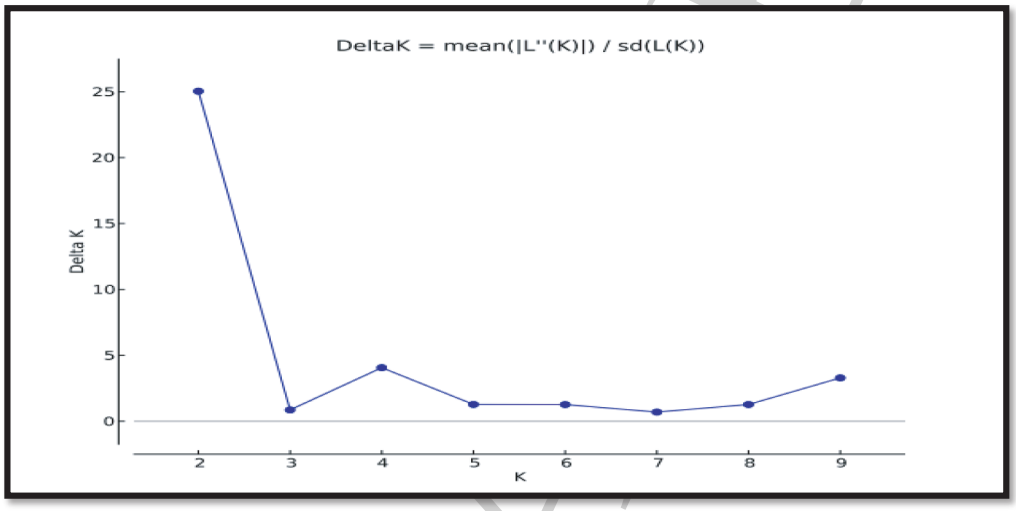

Figure 2b. Population structure analysis of custard apple genotypes showing; Delta K high probability of two (K = 2) subpopulation.

Figure 2. Structure analysis of *Annona* genotypes using SSR data.

(2017) studied 34 accessions of *A. muricata* using 8 *A. cherimola* SSR markers, reporting an average PIC value of 0.0131, which is much lower than our findings. Thus, it appears that analysis employing species-specific SSR markers gives a better indication of diversity and heterozygosity in the population.

In this study, all 70 SSR markers were highly polymorphic, with PIC values exceeding 0.5 (Table 3), indicating a high level of genetic diversity among the 40 *Annona* genotypes. This was further supported by the high Ho values, which ranged from 0.250 to 1.000, and He values, which ranged from 0.781 to 0.978 (Table 3). These results demonstrate the effectiveness of species-specific SSR markers for assessing genetic diversity in *A. squamosa*. Additionally, the Ho being higher than expected suggests the presence of at least two isolated populations among the genotypes studied.

Among the markers, three SSRs (ASIIHR19, ASIIHR31 and ASIIHR35) showed the highest number of alleles (46), all of which were derived from di-nucleotide perfect repeat motifs $(CA)_8$, $(CT)_8$ and $(GA)_8$, respectively. Previous studies (Liu et al., 2016 in kiwifruit; Biswas et al., 2014 in sweet orange; Chapman, 2019 in lablab) have also reported that microsatellite markers with di-nucleotide repeat motifs exhibit a significantly higher number of alleles compared with those with tri-, tetra- and penta-nucleotide repeats, likely due to the ease of mutation via DNA slippage during replication.

Furthermore, seven markers (ASIIHR04, ASIIHR19, ASIIHR31, ASIIHR34, ASIIHR35, ASIIHR40 and ASIIHR47) with low PI values ranging from 0.0022 to 0.0037 were found to be suitable for DNA fingerprinting of *A. squamosa* genotypes (Table 3). SSR markers with low PI values are highly useful for DNA fingerprinting (Ravishankar et al., 2017a, 2017b), making these markers ideal for the identification of custard apple varieties.

The dendrogram analysis classified the *A. squamosa* genotypes into two main groups, further subdivided into sub-clusters based on genetic relatedness (Figure 1). Low bootstrap values observed here may be due to only too few SSR markers employed to estimate meaningful

genetic differentiation or from a weak population structure due to recent divergence or gene flow. The later explanation is plausible given the recent introduction of *Annona* species to the Indian subcontinent. This clustering aligns with earlier studies, such as Anuragi et al. (2016), where 20 *A. squamosa* genotypes were grouped into 7 clusters using 12 *A. cherimola* SSR markers. The high genetic diversity observed in this study is likely due to the collection of germplasm from diverse locations and species-specific markers used, reflecting the presence of a broad genetic base in *A. squamosa*. A similar pattern of high diversity was reported for *A. cherimola* (Escribano et al., 2007) and *A. senegalensis* (Kwapata et al., 2007).

A Bayesian model-based analysis further supported the clustering results, grouping the 40 *A. squamosa* genotypes into two major clusters (Figure 2A). The ΔK value derived from Evanno's algorithm predicted K = 2, consistent with the dendrogram results. Cluster I comprised genotypes from Andhra Pradesh, Tamil Nadu, Maharashtra, Karnataka, the USA and Taiwan, while Cluster II comprised genotypes from Andhra Pradesh, Telangana, Odisha and the USA.

Cross-species amplification of SSRs was highly successful, with a transferability rate of 80.0%–97.1% across five *Annona* species (Table 3). This high transferability is likely due to the conserved nature of flanking sequences around microsatellites in *Annona* species. Previous studies, such as those by Anuragi et al. (2016) and Kwapata et al. (2007), reported similar results, with SSR markers showing high cross-species transferability among different *Annona* taxa.

Although SSR development is expensive, once established, these markers are cost-effective and time-efficient for genetic analysis. In *A. squamosa*, SSRs have not been extensively used, but their high polymorphism suggests that they are valuable for assessing genetic diversity. Most of the genetic variation in this study was due to differences within populations, likely because *A. squamosa* is a recently introduced cultivated species. The AMOVA results, with an Fst value of 0.065, indicate significant genetic differentiation among the 40 custard apple genotypes, reflecting the coexistence of different genotypes within the same region (Ravishankar et al., 2015).

CONCLUSIONS

NGS proved to be an efficient method for identification and development of microsatellite markers. Compared with previous studies, this study revealed relatively higher heterozygosity and PIC values, emphasising the usefulness of species-specific SSR markers. The high PIC, polymorphism and Ho values indicate that the SSR markers developed in this study are effective for genetic diversity analysis in *A. squamosa*. These markers also demonstrated high transferability to closely related species, including *A. cherimola*, *A. reticulata*, *A. glabra*, *A. muricata* and *A. atemoya*. The microsatellite markers generated in this study will be valuable for genetic diversity studies, mapping, development of linkage map and other crop improvement programmes, like gene discovery, in *A. squamosa*.

ACKNOWLEDGEMENTS

We would like to thank the Indian Council of Agricultural Research-Indian Institute of Horticultural Research (ICAR-IIHR), Hesaraghatta Lake Post, Bangalore for providing facilities.

FUNDING

We would like to thank the Indian Council of Agricultural Research-Indian Institute of Horticultural Research (ICAR-IIHR), Hesaraghatta Lake Post, Bangalore for funding.

CONFLICT OF INTEREST

The authors report that there are no competing interests to declare. AI tool Kimi has been used to correct English language only. Sequence data has been submitted to NCBI (Bioproject PRJNA682654).

REFERENCES

Anuragi, H., Dhaduk, H. L., Kumar, S., Dhruve, J. J., Parekh, M. J., and Sakure, A. A. (2016). Molecular diversity of *Annona* species and proximate fruit composition of selected genotypes. *3 Biotech, 6*, 204, https://doi.org/10.1007/s13205-016-0520-9.

Altschul S. F., Gish W., Miller W., Myers E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Beier, S., Thiel, T., Münch, T., Scholz, U., and Mascher, M. (2017). MISA-web: a web server for microsatellite prediction. *Bioinformatics (Oxford, England), 33*(16), 2583–2585, https://doi.org/10.1093/bioinformatics/btx198.

Bharad, S. G., Kulwal, P. L., and Bagal, S. A. (2009). Genetic diversity study in *A. squamosal* by morphological, biochemical and RAPD markers. *Acta Horticulturae, 839*, 615–623, https://doi.org/10.17660/ActaHortic.2009.839.84.

Biswas, M. K., Xu, Q., Mayer, C., and Deng, X. (2014). Genome wide characterization of short tandem repeat markers in sweet orange (*Citrus sinensis*). *PloS ONE, 9*(8), p.e104182, https://doi.org/10.1371/journal.pone.0104182.

Chapman, M. A. (2019). Optimizing depth and type of high- throughput sequencing data for microsatellite discovery. *Applications in Plant Sciences*, *7*(11), p.e11298, https://doi.org/10.1002/aps3.11298.

Escribano, P., Viruel, M. A., and Hormaza, J. I. (2004). Characterization and cross species

amplification of microsatellite markers in cherimoya (*Annona cherimola* Mill. Annonaceae). *Molecular Ecology Notes, 4*, 746–748, https://doi.org/10.1111/j.1471-8286.2004.00809.x.

Escribano, P., Viruel, M. A., and Hormaza, J. I. (2007). Molecular analysis of genetic diversity and geographic origin within an ex situ germplasm collection of cherimoya by using SSRs. *Journal of the American Society for Horticultural Science, 132*, 357–367, https://doi.org/10.21273/JASHS.132.3.357.

Escribano, P., Viruel, M. A., and Hormaza, J. I. (2008). PERMANENT GENETIC RESOURCES: Development of 52 new polymorphic SSR markers from cherimoya (*Annona cherimola* Mill.): transferability to related taxa and selection of a reduced set for DNA fingerprinting and diversity studies. *Molecular Ecology Resources, 8*, 317–321, https://doi.org/10.1111/j.1471-8286.2007.01941.x.

Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611–2620.

Gyamfi, K., Sarfo, D., Nyarko, B., Akaho, E., Serfor-Armah, Y., and Ampomah-Amoako, E. (2011). Assessment of elemental content in the fruit of graviola plant, *Annona muricata*, from some selected communities in Ghana by instrumental neutron activation analysis. *Elixir Food Science, 41*, 5671–5675. https://www.elixirpublishers.com/articles/1685693363_201112015.pdf.

Kalinowski, S. T., Taper, M. L., and Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology, 16*, 1099–1106, https://doi.org/10.1111/j.1365-294X.2007.03089.x.

Kwapata, K., Mwase, W. F., Bokosi, J. M., Kwapata, M. B., and Munyenyembe, P. (2007). Genetic diversity of *Annona senegalensis* Pers. populations as revealed by simple sequence repeats (SSRs). *African Journal of Biotechnology, 6*, 1239–1247. https://www.internationalscholarsjournals.com/articles/geneticdiversity-of-annona-senegalensis-pers-populations-as-revealed-by-simplesequence-repeats-ssrs.pdf.

Liu, C., Zhang, Q., Yao, X., Zhong, C., Yan, C. and Huang, H. (2016). Characterization of genomewide simple sequence repeats and application in interspecific genetic map integration in kiwifruit. *Tree Genetics & Genomes*, *12*, 1–9.

Mariod, A. A., Elkheir, S., Ahmed, Y. M., and Matthaus, B. (2010). *A. squamosa* and *C. nilotica* seeds, the effect of the extraction method on the oil composition. *Journal of the American Oil Chemists' Society, 87*, 763–769, https://doi.org/10.1007/s11746-010-1548-3.

Peakall, R., and Smouse, P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes, 6*, 288–295, https://doi.org/10.1111/j.1471-8286.2005.01155.x.

Pereira, M. F., Bandeira, L. F., Blanco, J. V., Ciampi, A. Y., and Coelho, S. G. (2008). Development of microsatellite markers in *Annona crassiflora* Mart., a Brazilian Cerrado fruit tree species. *Molecular Ecology Resources, 8*, 1329–1331, https://doi.org/10.1111/j.1755-0998.2008.02275.x.

Perrier, X., and Jacquemoud-Collet, J. P. (2003). *DARwin software*. Retrieved from https://darwin.cirad.fr/.

Pinto, A. C., Cordeiro, M. C. R., Andrade, S. R. M., Ferreira, F. R., Filgueiras, H. A., Alves, R. E., and Kinpara, D. I. (2005). *Annona species*. UK: International Centre for Underutilized Crops, University of Southampton.

Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype Data. *Genetics, 155*, 945–959.

Rahman, M., Shimada, T., Yamamoto, T., Yonemoto, J. Y., and Yoshida, M. (1998). Genetic diversity of cherimoya cultivars revealed by Amplified Fragment Length Polymorphism (AFLP) analysis. *Breeding Science*, *48*, 5–10.

Ravishankar, K. V., Anand, L., and Dinesh, M. R. (2000). Assessment of genetic relatedness among mango cultivars of India using RAPD markers. *The Journal of Horticultural Science and Biotechnology, 75*, 198–201, https://doi.org/10.1080/14620316.2000.11511223.

Ravishankar, K. V., Dinesh, M. R., Nischita, P., and Sandya, B. S. (2015). Development and characterization of microsatellite markers in mango (*Mangifera indica*) using next-generation sequencing technology and their transferability across species. *Molecular Breeding, 35*, 93, https://doi.org/10.1007/s11032-015-0289-2.

Ravishankar, K. V., Gayathri, M., Pitchaimuthu, M., and Santhosh-Kumar, G. (2017a). Identification of novel microsatellite markers in okra (*Abelmoschus esculentus* (L.) Moench) through next-generation sequencing and their utilization in analysis of genetic relatedness studies and cross-species transferability. *Journal of Genetics, 97*, 39–47, https://doi.org/10.1007/s12041-018-0893-0.

Ravishankar, K. V., Vasudeva, R., Hemanth, B., Sandya, B. S., Sthapit, B. R., Parthasarathy, V. A., and Ramanatha Rao, V. (2017b). Isolation and characterization of microsatellite markers in *Garcinia gummi-gutta* by next-generation sequencing and cross-species amplification. *Journal of Genetics, 96*, 213–218, https://doi.org/10.1007/s12041-017-0756-0.

Ronning, C. M., Schnell, R. J., and Gazit, S. (1995). Using randomly amplified polymorphic DNA (RAPD) markers to identify *Annona* cultivars. *Journal of the American Society for Horticultural*

Science, 120, 726–729, https://doi.org/10.21273/JASHS.120.5.726.

Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for General Users and for Biologist Programmers. In S. Misener and S. A. Krawetz (Eds). *Bioinformatics Methods and Protocols. Methods in Molecular Biology*™, vol 132. Totowa, NJ, USA: Humana Press, https://doi.org/10.1385/1-59259-192-2:365.

Schuelke, M. (2000). An economic method for the fluorescent labelling of PCR fragments. *Nature Biotechnology, 18*, 233–234, https://doi.org/10.1038/72708.

Thanachseyan, P., Samarasinghe, W. L. G., and Thirukkumaran, G. (2017). Diversity assessment of *Annona muricata* L. through molecular characterization using SSR markers. *International Journal of New Technology and Research, 3*, 57–62. https://www.ijntr.org/download_data/IJNTR03090025.pdf.

Van Zonneveld, M., Scheldeman, X., Escribano, P., Viruel, M. A., Damme, P. V., Garcia, W., Tapia, C., Romero, J., Siguen, M., and Hormaza, J. I. (2012). Mapping genetic diversity of cherimoya (*Annona cherimola* Mill.): application of spatial analysis for conservation and use of plant genetic resources. *PloS ONE, 7*(1), e29845, https://doi.org/10.1371/journal.pone.0029845.

Vinay, G. M., Sakthivel, T., and Priyanka, H. L. (2017). Recent advances in Annona breeding: A review. *International Journal of Pure & Applied Bioscience*, *2*, 1168–1181.

Zhichang, Z., Guibing, H., Ruo, O., Yunchum, L., Yeyuan, C. and Shirong, L. (2011). Studies of the genetic diversity of seven sweetsop (*Annona squamosa* L.) cultivars by amplified fragment length polymorphism analysis. *African Journal of Biotechnology*, *10*(35), 6711–6715.

Received: February 6, 2025; accepted: May 20, 2025

SUPPLEMENTARY MATERIALS

Supplementary Table 1. List of 40 custard apple (*A. squamosa*) genotypes/cultivars belonging to *A. squamosa* and different species of *Annona* used in the study.

Sl. No.	Cultivars/accessions No.	Place of collection	State
1.	Balanagar	Sangareddy	Telangana
2.	Raidurg	Ananthapur	Andhra Pradesh
3.	Taiwan	Taiwan	Taiwan
4.	Arka_Sahan (IC No. 0632061)	IIHR	Karnataka
5.	NMK_1	Garmale	Maharashtra
6.	APK_1	Aruppukottai	Tamil Nadu
7.	WASHINGTON_05	Florida	USA
8.	Barbados	Florida	USA
9.	Washington_97	Florida	USA
10.	19/26 (IC no. 0632061)	IIHR, Bangalore	Karnataka
11.	Arka Neelanchal Vikram	Bhubaneswar	Odisha
12.	Mammoth	Florida	USA
13.	Red_sitaphal	Sangareddy	Telangana
14.	6_8	Nalakadoddi	Andhra Pradesh
15.	2_1	Vengalampalli	Andhra Pradesh
16.	3_1	Vengalampalli	Andhra Pradesh
17.	1_1	Vengalampalli	Andhra Pradesh
18.	8_8	Pythota	Andhra Pradesh
19.	8_9	Pythota	Andhra Pradesh
20.	2_2	Vengalampalli	Andhra Pradesh
21.	27_1	Jambugumpala	Andhra Pradesh
22.	5_8	Molakalmuru	Karnataka
23.	3_2	Vengalampalli	Andhra Pradesh
24.	5_1	Molakalmuru	Karnataka
25.	8_18	Pythota	Andhra Pradesh
26.	1_10	Vengalampalli	Andhra Pradesh
27.	8_16	Pythota	Andhra Pradesh
28.	11_8	Yercaud	Tamil Nadu
29.	3_3	Vengalampalli	Andhra Pradesh
30.	4_11	Molakalmuru	Karnataka
31.	8_17	Pythota	Andhra Pradesh
32.	1_5	Vengalampalli	Andhra Pradesh
33.	2_13	Vengalampalli	Andhra Pradesh
34.	4_1	Molakalmuru	Karnataka
35.	2_10	Vengalampalli	Andhra Pradesh
36.	4_10	Molakalmuru	Karnataka
37.	6_10	Nalakadoddi	Andhra Pradesh
38.	4_12	Molakalmuru	Karnataka
39.	8_7	Pythota	Andhra Pradesh
40.	Rolenia	Moodubidire	Karnataka
41.	*A. cherimola*	Sangareddy	Telangana
42.	*A. reticulata*	Sangareddy	Telangana
43.	*A. glabra*	Sangareddy	Telangana
44.	*A. muricata*	Sangareddy	Telangana
45.	*A. atemoya*	Sangareddy	Telangana

Supplementary Table 2. Summary of sample Balanagar assembled genome.

Plant WGS	Balanagar
Contigs generated	1388525
Maximum contig length	21462
Minimum contig length	100
Average contig length	466.126
Total contigs length	647227553 (647.2 MB)
Total number of non-ATGC characters	54429
Percentage of non-ATGC characters	8.40956E-05
Contigs >100 b:	1379816
Contigs >500 b:	350135
Contigs >1 Kb:	170254
Contigs >10 Kb:	19
Contigs >100 Kb:	0
n50 value:	954
n90_value:	151
kmer length	89
Total number of scaffolds	58528
Total number of identified SSRs	1790080

SSR, simple sequence repeat.