RESEARCH ARTICLE

SABRAO Journal of Breeding and Genetics
55 (5) 1573-1586, 2023
http://doi.org/10.54910/sabrao2023.55.5.11
http://sabraojournal.org/
pISSN 1029-7073; eISSN 2224-8978

IN SILICO MINING AND VALIDATION OF SIMPLE SEQUENCE REPEAT (SSR) MARKERS DERIVED FROM THE *FRAGARIA X ANANASSA* GENOME ASSEMBLY

V.H. ELEC, C.A.E. CADORNA, and J.D. REY*

Plant Molecular Phylogenetics Laboratory, Institute of Biology, University of the Philippines, Quezon City, Philippines
*Corresponding author's email: jdrey@up.edu.ph
Email addresses of co-authors: vhelec@up.edu.ph, cecadorna@up.edu.ph

SUMMARY

Strawberry (*Fragaria x ananassa Duchesne ex Rozier*) is an economically important fruit crop grown commercially worldwide due to its known nutritional value. The demand for strawberries in the Philippines has been increasing but varietal development is slow due to the lack of genetic information on locally available germplasm. In addition, molecular marker resources for strawberries are limited. Here, we developed Simple Sequence Repeats (SSR) markers to characterize strawberries cultivated in tropical environments, particularly in the Philippines. *In silico* mining generated a total of 219,239 SSR markers, from which 160,025 unique markers were identified. Polymorphism scoring and the variation on the expected alleles resulting to the synthesis of the top 160 markers were validated using representative strawberry cultivars from the core collection. All 70 markers produced successful amplicons from all the genotypes, but only 67 markers displayed polymorphism. Using UPGMA, genetic similarities among the cultivars were calculated generating a dendrogram which categorized the strawberry cultivars into 3 groups. The grouping clearly showed the separation of the diploid *Fragaria vesca* from the octoploid *F. x ananassa* cultivars, considered as group 1. Furthermore, Groups 2 and 3 contained cultivars from different geographical origins (US, Japan, and North America). The eight novel markers developed and validated here can be added to the genetic marker resource for cultivar verification, assessment of genetic diversity, and most importantly, marker-assisted breeding.

Keywords: *In silico* mining, *Fragaria* genome, Philippine strawberry cultivars, simple sequence repeats

Key Findings: *In silico* mining from *F. x ananassa* genome sequence generated a total of 160,025 unique markers from which a subset of 70 markers was validated and approximately 95.7% (67 markers) are polymorphic. These markers were able to discriminate representative strawberry cultivars from the core collection of the BSU breeding program. It can also be utilized to further characterize and fingerprint these sets of germplasm as a prerequisite for a successful breeding program.s.

Communicating Editor: Dr. Kamile Ulukapi

Citation: Elec VH, Cadorna CAE, Rey JD (2023). *In silico* mining and validation of simple sequence repeat (SSR) markers derived from the *Fragaria x ananassa* genome assembly. *SABRAO J. Breed. Genet.* 55(5): 1573-1586. http://doi.org/10.54910/sabrao2023.55.5.11.

Manuscript received: April 5, 2023; Accepted: August 14, 2023.

INTRODUCTION

Strawberry (*Fragaria x ananassa Duchesne ex Rozier*) is an economically important fruit crop grown commercially worldwide due to its high nutritional value and unique aromatic flavor (Gonzalez *et al*., 2020). Over the past three decades, the worldwide strawberry production has seen an increase of more than 230%. In the year 2017, a total of 77 countries produced a combined 9.2 million tons of fresh fruit (FAOSTAT, 2021). Similarly, the demand for strawberries in the Philippines has been increasing for the past years. Importation increased from up to 2,800 metric tons to meet the gap requirement of strawberry production of only 28 metric tons (FAOSTAT, 2021). The major producing region for the strawberry in the country is the Cordillera Administrative region (CAR) which accounted for 99.9% of the total production in 2020. The only strawberry breeding program in the Philippines is also located in Benguet State University (BSU), La Trinidad Benguet. Like other breeding programs, it aimed at producing strawberry varieties with high yield potential, fruit quality and resistance to pests and diseases (Padua, 2011). Although progress has been made, the breeders only utilize morphological traits in characterizing their germplasm (Padua, 2011). Furthermore, the genetic profile of the strawberry germplasm in the breeding program has not been reported yet.

The cultivated strawberry *(F. x ananassa*) is an octoploid (2n=8x=56) generated by the natural hybridization between two octoploid strawberry species, *F. chiloensis* and *F. virginiana* (Hancock *et al*., 1990). An estimate of the haploid genome's size was around 813.4 Mb, with each sub-genome contributing around 170 Mb (Edger *et al*., 2019). The genome is also highly heterozygous, with a high degree of allelic variation between the sub-genomes and within each sub-genome. The group of Shulaev (2011) first sequenced the genome of the diploid *F. vesca*, followed by a more recent high-quality assembly of the octoploid *F.* × *ananassa* genome (Edger et al., 2019).

One of the most important and promising tools for strawberry research and commercial production of the strawberry industry is the use of molecular markers. The development and use of DNA markers has been reported as early as the 1980s (Haymes *et al*., 2000). In recent years, SSR markers have emerged as one of the most frequently utilized marker systems in plant molecular genetic research. Simple sequence repeats (SSRs), also known as microsatellites, are tandemly repeated short DNA sequences of 1-6 base pairs in length. They are widely distributed throughout the genomes of eukaryotic organisms and exhibit high levels of polymorphism, making them valuable tools for marker development, especially in crop improvement programs, where genetic diversity is a crucial factor in developing improved varieties (Varshney *et al*., 2013).

The traditional method of developing SSR markers involves the screening of genomic libraries for the presence of SSR sequences and subsequent validation through PCR amplification and sequencing (Varshney *et al*., 2016). Others involve the use of EST (expressed sequence tag) sequences, which represent transcribed regions of the genome and are generated by sequencing cDNA libraries (Choudhary *et al*., 2016). These approaches have been successful in identifying SSR markers in a wide range of plant species, including switchgrass (*Panicum virgatum*), cucumber (*Cucumis sativus*), and chickpea (*Cicer arietinum*) (Fu *et al*., 2016; Cheng *et al*., 2017; Nayak *et al*., 2017). However, these processes can be time-consuming and costly, as it requires the construction and screening of large genomic libraries (Kalia *et al*., 2011).

Recent advancements in sequencing technologies have enabled the generation of high-quality genome assemblies for a wide range of plant species, specifically for woodland strawberry (*F. vesca*) and cultivated strawberry *(F. x ananassa*) (Edger *et al*.,

2019). The availability of these genome sequences has facilitated the development of molecular markers by mining SSRs across the entire genome (Chen *et al*., 2018; Awad and Hassasin, 2022; Khaled *et al*., 2023). Advances in sequencing technologies and bioinformatics tools have facilitated the SSR markers development from genomic sequences, including whole-genome shotgun sequences, transcriptome data, and genomic resequencing data (Varshney *et al*., 2016). These approaches have allowed for the rapid and cost-effective identification and characterization of large numbers of SSR loci in a variety of crops and other organisms. Understanding the genetic characteristics of strawberry germplasm and employing cutting-edge and effective techniques could provide a lot of potential in improving the crops' productivity with the aid of molecular marker technologies and bioinformatic pipelines.

In this study, we aim to expand the limited genetic marker resource for strawberries cultivated in tropical environments through *in silico* analysis and SSR marker development based on *F. x ananassa* '*Camarosa*' genome sequence, which is the first genome assembly of a cultivated strawberry, and at the same time maintaining conserved regions of its diploid progenitors. The molecular markers generated in this study will accelerate basic research on the characterization of germplasm available in the Philippine strawberry breeding program and would provide value in the discovery and functional gene towards their utilization in trait improvement through marker-aided selection.

MATERIALS AND METHODS

Plant material

To validate the developed markers, this study used a subset of 10 cultivars from the available germplasm collection in Benguet State University (16°27'07"N 120°35'27"E) strawberry breeding program. Furthermore, this study collected disease-free fully expanded trifoliate of 10 leaf samples from each cultivar (Table 1). The obtained leaf samples underwent total genomic DNA extraction at the Plant Molecular Phylogenetics Laboratory (Institute of Biology, University of the Philippines, Diliman, Quezon City, Philippines).

Retrieval of genomic dataset for genome-wide SSR marker mining

This study downloaded the genome assembly of *F. x ananassa* 'Camarosa' (Edger et al., 2019) from the Genome Database of Rosaceae (GDR, https://www.rosaceae.org). Edger et al. (2019) used next-generation sequencing (NGS) technologies, which included Illumina (San Diego, CA), 10X Genomics (Pleasanton, CA), and Pacific Biosciences (PacBio; Menlo Park, CA) to generate the genome with a 615X coverage.

Table 1. Strawberry cultivars used in the study for screening the SSR markers.

Entry No.	Strawberry cultivars	Code	Origin	Ploidy level
1	Sweet Charlie (Argentina)	SCC	USA	octoploid
2	Sweet Charlie (Washington)	SCW	USA	octoploid
3	Benihoppe	BEN	Japan	octoploid
4	Honeoye	HON	USA	octoploid
5	Festival	FST	USA	octoploid
6	*Fragaria vesca*	FRV	North America	diploid
7	Whitney	WTY	USA	octoploid
8	Sweet Angel Bohol	SAB	Japan	octoploid
9	Earlibrite	EBR	USA	octoploid
10	Snow White	SNW	Japan	octoploid

In silico mining of SSR markers from genomic dataset and primer design

Using the Genome-wide Microsatellite Analyzing Toward Application (GMATA) software (Wang and Wang, 2016) with default parameters, this study performed *in silico* mining of simple sequence repeats (SSRs) in the downloaded *F. x ananassa* genome assembly v1.0 (Edger et al., 2019). The Primer3 program integrated into the GMATA software aided in generating primers for the SSRs identified in the assembly.

In silico polymorphism scoring and primer synthesis

Simulated marker mapping (e-mapping) in GMATA based on a forward e-PCR algorithm determined the polymorphism of the designed markers. The e-PCR parameters were on default values. All the amplified fragments from each marker in the genome produced size scores to calculate the polymorphism. Moreover, this study performed e-PCR analysis on various strawberry species and varieties available in GenBank. Macrogen Inc. (Seoul, South Korea), through its oligonucleotide synthesis service, synthesized a set of putative polymorphic SSR primers for experimental validation.

Total genomic DNA extraction

Total genomic DNA extraction utilized the DNeasy® Plant Mini Kit (Qiagen Inc., Germany) following the manufacturer's protocol. The DNA integrity assessment by electrophoresis used 1% agarose in 1x Tris-acetate-EDTA (TAE) running buffer at 100V for 45 min and stained using GelRed® (Biotium, USA). Then, DNA visualization utilized the Gel Doc EZ Documentation System (Biorad, USA). Finally, NanoDrop™ Lite (Thermofisher Scientific Inc., USA) spectrophotometer determined the relative purity and concentration of the isolated DNA.

PCR analysis

PCR amplification carried through with 15 uL reaction volume containing 1 uL genomic DNA, 0.5 uL of the forward primer (10 uM), 0.5 uL of the reverse primer (10 uM), 7.5 uL 2X MyTaq™ HS Red Mix (Bioline, USA), and 5.5 uL nuclease-free water. PCR conditions used are as follows: initial denaturation at 95℃ for a period of 1 min, followed by 35 cycles of denaturation at 95℃ at 15 sec, annealing at 60℃ at 15 sec, and extension at 72℃ at 10 sec, and a final extension step at 72℃ at 5 min. Amplifications proceeded in T100™ Thermal Cycler (Bio-Rad, USA). Electrophoresis resolved the PCR products using 10% non-denaturing polyacrylamide gel in 0.5X Tris-borate EDTA buffer at 100V for 90 min in the C.B.S. Scientific Triple Wide Mini Vertical System™ (C.B.S. Scientific Company San Diego, California, USA) and visualized using GelRed® Nucleic Acid Gel Stain (Biotium, USA). Gel imaging used Gel Doc EZ Documentation System (Bio-Rad, USA), with gels scored manually for the presence or absence of bands.

Data analysis

Genetic diversity analysis of 10 strawberry cultivars used 63 selected markers. SSR markers were scored as 1 (presence of band) or 0 (absence of band) for each cultivar. The iMEC program (Amiryousefi *et al.*, 2018) computed the polymorphism indices of the binary dataset for all strawberry cultivars assembled in a single matrix. Similarly, data analysis used the unweighted pair group method with arithmetic mean (UPGMA; htttp://genomes.urv.cat/UPGMA) algorithm utilizing the Jaccard coefficient to calculate the genetic similarities and between cultivars. Finally, the FigTree v1 4.4 (http://tree.bio.ed.ac.uk/software/figtree) software generated a dendrogram using the output data in Newick format.

RESULTS

Frequency and distribution of SSR markers in the genome of *F. x ananassa*

This study used the strawberry genome assembly consisting of 815 megabases (Mb) of sequences to mine SSR markers. *In silico* mining identified a total of 230,750 SSRs with motifs ranging from 2-9 bp. SSRs with dinucleotide repeats were the most abundant (83.95%), followed by trinucleotide repeats with 14%. Repeats motifs having penta-, hexa-, hepta-, octo-, and nona-nucleotide repeats comprise only approximately 1% as presented in Figure 1. Among the dinucleotide motifs, AT and TA make up 40% of the total SSRs with 2-bp. SSRs generated range from 2bp- 10bp in length. The most predominant are SSRs with 10-bp in length which makes up 29.2% whereas those that are 30-bp long were the least frequent (2%) as shown in Figure 2.

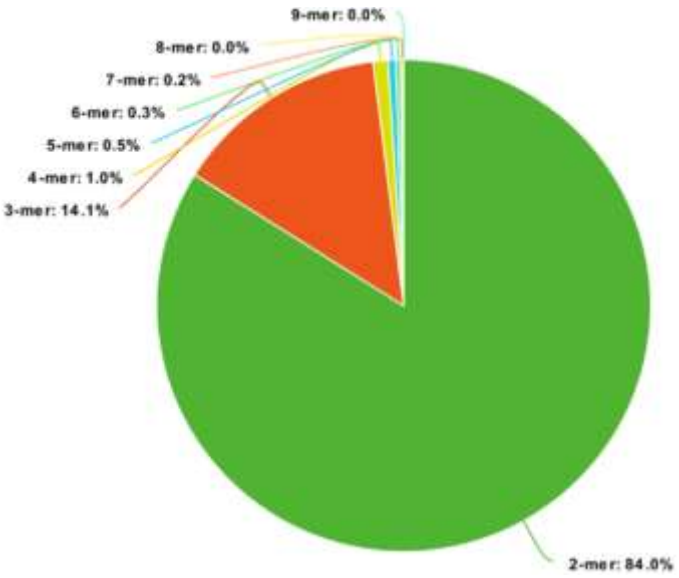

Figure 1. Abundance of repeat motifs based on *in silico* analysis of *Fragaria* × *ananassa* genome sequence.

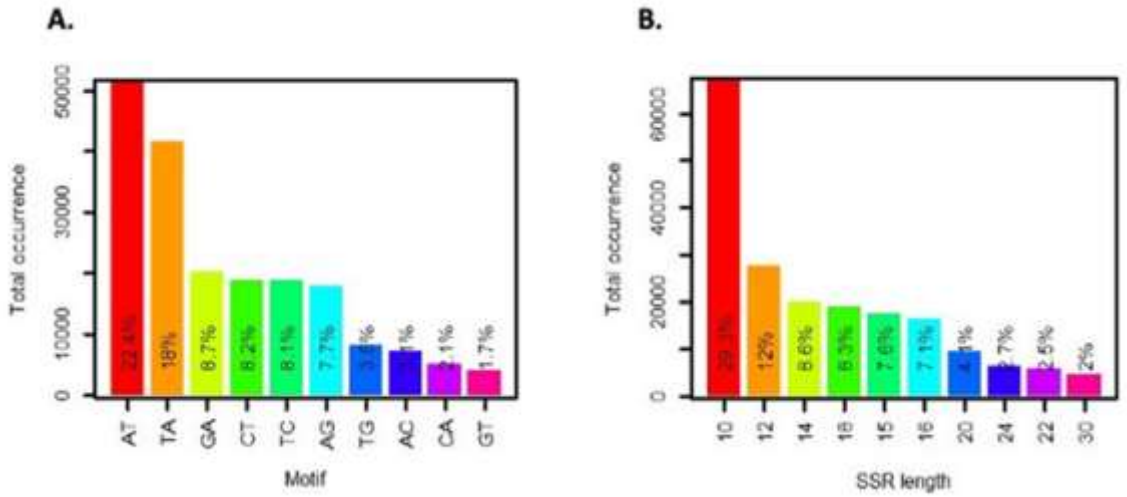

Figure 2. Frequency distribution of the different (A) SSR motifs and (B) SSR lengths throughout the *F.* × *ananassa* 'Camarosa' genome assembly v1.0 based on GMATA data.

Table 2. Pseudo molecule-wise distribution and density of the simple sequence repeat (SSR) loci in the *Fragaria* × *ananassa* genome.

Pseudo molecule ID	Size of Pseudo molecules	No. of SSRs	% SSRs	Frequency (SSRs/Mb)
Fvb1-1	27594200	7588	3.29	274.99
Fvb1-2	28910674	7827	3.39	270.73
Fvb1-3	27436561	7369	3.19	268.58
Fvb1-4	22887349	6909	2.99	301.87
Fvb2-1	26582685	7687	3.33	289.17
Fvb2-2	24782128	7763	3.36	313.25
Fvb2-3	24073015	7242	3.14	300.83
Fvb2-4	26692599	7618	3.30	285.40
Fvb3-1	32005440	9244	4.01	288.83
Fvb3-2	31459976	8878	3.85	282.20
Fvb3-3	29626823	8301	3.60	280.19
Fvb3-4	27809139	8321	3.61	299.22
Fvb4-1	20034018	5607	2.43	279.87
Fvb4-2	25974422	7090	3.07	272.96
Fvb4-3	31955388	9578	4.15	299.73
Fvb4-4	26295489	7355	3.19	279.71
Fvb5-1	29826953	8835	3.83	296.21
Fvb5-2	24981319	7243	3.14	289.94
Fvb5-3	27452983	7928	3.44	288.78
Fvb5-4	25211045	7373	3.20	292.45
Fvb6-1	36657112	10827	4.69	295.36
Fvb6-2	36124132	10508	4.55	290.89
Fvb6-3	43627644	12092	5.24	277.16
Fvb6-4	34274104	9895	4.29	288.70
Fvb7-1	32186896	9130	3.96	283.66
Fvb7-2	32354134	9051	3.92	279.75
Fvb7-3	25137763	6842	2.97	272.18
Fvb7-4	23534715	6649	2.88	282.52
Total	805488706	230750	100	
Average	28767453	8241	3.57	286.47

SSRs markers generated *in silico* revealed their wide distribution throughout the pseudo-molecules (Table 2). There is an average of 8,241 SSRs per pseudo molecule with an average frequency of 286 SSRs/Mb. A total of 219,239 SSR markers with primer pairs were generated *in silico*, from which 160,025 unique markers were identified. These comprised 95% of designed primers from the available SSR loci detected. *In silico* polymorphism scoring generated a range of 1 to 469 alleles. We filtered the markers having less than 10 alleles and ranked based on the variation of the allele sizes. From these, the top 160 markers were selected and synthesized. Subsequently, a subset of 70 markers was used in validating representative strawberry cultivars from the core collection.

Validation and characterization of SSR markers

All the markers synthesized showed amplification for the entire DNA tested. Of the 70 SSRs, 67 primer pairs showed polymorphism (96%) while three marker pairs generated monomorphic bands. Figure 3 shows the representative profile of 10 strawberry cultivars using three polymorphic markers. An average of six alleles per locus was detected across test varieties, and ranges from 2 alleles up to 10 alleles per locus were identified. More than 90% of the synthesized markers generated more than two alleles. The highest number of alleles generated is six with a total of 19 markers (27%), followed by eight alleles (22%). Only three markers generated two

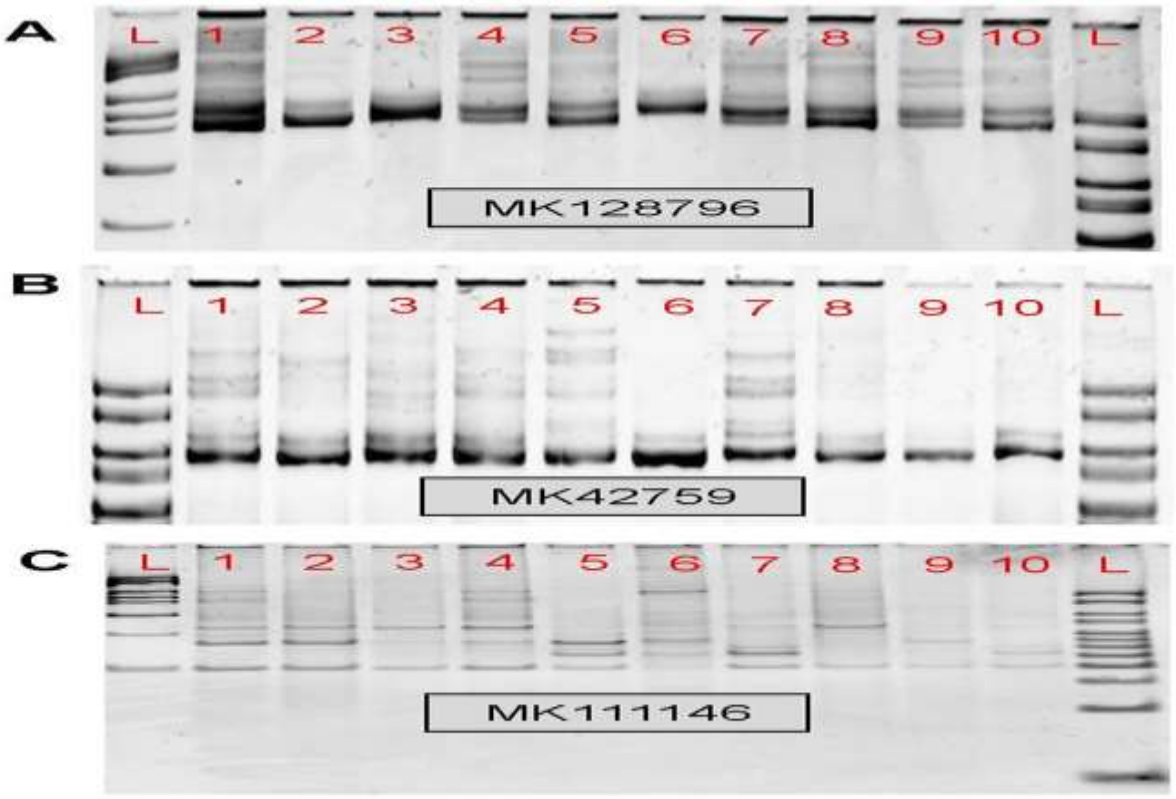

Figure 3. Representative polyacrylamide gel electrophoresis (PAGE) profile. PAGE image of SSR marker 128796 (A), SSR marker 42759 (B), and SSR marker 111146 (C). Lane 1: Sweet Charlie (Argentina); Lane 2: Sweet Charlie (Washington); Lane 3: Benihoppe; Lane 4: Honeoye; Lane 5: Festival; Lane 6: *Fragaria vesca*; Lane 7: Whitney; Lane 8: Sweet Angel Bohol; Lane 9: Earlibrite; Lane 10: Snow White; and L: Ladder.

alleles per locus from only 4% of the available markers validated. In total, there were 440 alleles generated for all the test entries using the synthesized SSR markers. The marker-generated bands confer the expected size of the amplicons based on ePCR. Table 3 summarizes the characteristics of selected SSR markers which includes marker name, repeat motifs, primer sequence, pseudomolecule ID, annealing temperature, product size range, number of alleles and heterozygosity index and polymorphic information content. Furthermore, effective multiplex ratio (E), average heterozygosity (H. ave), Marker Index, Discriminating power (D) and Resolving power (R) were also determined (Table 4). The average heterozygosity index computed for all the markers tested is 0.469, Marker MK26759 have the highest heterozygosity index of 0.637 with 4 alleles generated while markers MK13305 and MK21825 have the least amount of heterozygosity index with 0.0905 with (5 alleles). Polymorphic information content (PIC) computed for all markers has an average value of 0.38. Marker with highest PIC value is MK26759 with a value of 0.564, while markers MK13305 and MIK21825 have the lowest PIC score of 0.090.

Genetic relationship among strawberry cultivars

Genetic similarities among the 10 cultivars were calculated based on UPGMA and a dendrogram was generated which categorized them into three groups. The grouping clearly showed the separation of *F. vesca (Group 1)* which is a diploid cultivar from the *F. x*

Table 3a. Characteristics of selected SSR primers designed using GMATA.

Seq No	Marker Name	Motifs	Primer Sequence	Pseudo molecule ID	Calculated T_M	Product Size	No. of alleles	Polymorphism	H_0[a]	PIC_0[b]
1	MK111405	(TA)35	F:CCATTTGATCCCAAGCAACT R:AGGCCGATTGTGTTTAGCAG	Fvb4-3	59.933	215	6	P	0.4994	0.3747
2	MK128796	(AT)7	F:CCCAGCTGATGGATGTTGTA R:TGGAATCAAAGGGAACTGATG	Fvb5-1	59.522	397	3	P	0.4444	0.3457
3	MK93317	(GA)6	F:CCATCCATGGCTTCTCTGTT R:GGTGCTTCAGTTTTGGGATG	Fvb3-4	60.073	270	6	P	0.5711	0.4862
4	MK149611	(GA)21	F:CATGGTGAATGCAACTGGAA R:AAGAGCGAGGCAAAATCAAA	Fvb5-4	60.517	238	8	P	0.5650	0.4824
5	MK42759	(ACA)6	F:AGGTGCTCCTTGAACGTAGG R:CGACATGAAGAACGATCGAG	Fvb2-2	59.353	299	8	P	0.4988	0.3744
6	MK115479	(GA)8	F:GCATAGTACCACGCCCTGTC R:TCATTATTCGGCTCCTCCTC	Fvb4-3	60.547	246	8	P	0.6150	0.5339
7	MK84308	(TG)5	F:TTCGCCAATATCATCCTTCC R:CAATCACAACCCTGTTGCAC	Fvb3-3	59.862	314	4	P	0.3988	0.3192
8	MK130011	(CGA)5	F:TTGGATTCCCATCTTCATCC R:GAAGCCTAAGGAAGCGAGGT	Fvb5-1	59.685	311	5	P	0.4968	0.3734
9	MK158614	(AGC)6	F:GGTAATTATGGCGGAAACCA R:CACCTGCGATTGTAGAGCAA	Fvb6-1	59.66	376	6	P	0.4994	0.3747
10	MK60218	(GGA)6	F:ATCTCTTTTGACGCGCACTT R:AGCAGAGTTCATGGGGTTTG	Fvb3-1	60.022	278	2	P	0.4850	0.4064
11	MK46112	(TA)6	F:GCAAACTTGCAAACAATGGA R:ATTGAGGCCCAAATCAAGG	Fvb2-3	59.712	400	6	P	0.5244	0.4403
12	MK124848	(TA)5	F:TCTGCTTGGGTGCTTCCTAT R:GGGAGATCATAGCCTTGAAAGA	Fvb5-1	59.836	387	6	P	0.3394	0.2818
13	MK117593	(AG)6	F:TCAGAATCCGATGCAAAACA R:ACCAAACCCAACCAAAATGA	Fvb4-4	60.197	312	8	P	0.1997	0.1797
14	MK58078	(TC)6	F:CATGGCAAGGAAGGAAGAAA R:GCCTCGCTCTGATCTTCAAC	Fvb2-4	60.184	387	5	P	0.5688	0.4848
15	MK110487	(TC)5	F:GTGGAAGTGATCGGAGGTGT R:ACCTCATCCTCCCCAATTCT	Fvb4-3	59.969	276	7	M	monomorphic	
16	MK92914	(CA)6	F:TTGCCCCAATTAACAAAAGG R:AGAAGATCCTCAAGCCCACA	Fvb3-4	59.805	367	2	P	0.3750	0.3047
17	MK37301	(TA)7	F:CATACCGACTCCGAAACCAT R:TCAAAATCCAAACGCCTCTT	Fvb2-1	59.813	388	8	P	0.4622	0.3554
18	MK21825	(CCT)5	F:CAGCATGCAAAACTCCTTCA R:TCTTGGGTCCCTTCTTGTTG	Fvb1-3	59.988	371	4	P	0.0950	0.0905
19	MK26759	(CT)9	F:TGGCCTTTAAATTGGTCGAT R:CGTGAATGGCCTATTTGGTT	Fvb1-4	59.411	329	5	P	0.6376	0.5639
20	MK12513	(GAA)11	F:TGAATGACCTTTCGCAACAA R:CCCTGCAGTCATGTTGGTTA	Fvb1-2	60.234	353	4	P	0.5588	0.4656
21	MK115000	(AG)5	F:TACGGGCAAACAAATCACAA R:CGGCATTGGTCTGTGTCTTA	Fvb4-3	59.969	352	6	P	0.6161	0.5391
22	MK103594	(TTG)6	F:TGGAAAACCTGGACCATAGAG R:GGAAATGGGAAACCCCTAGA	Fvb4-2	59.038	210	4	P	0.6163	0.5347
23	MK89689	(GT)6	F:GGAAATGGGAAACCCCTAGA R:TCCGCTAGGCATTACAAGGT	Fvb3-4	60.124	384	8	P	0.5738	0.4879
24	MK88433	(TA)5	F:TGCAGCTCAACTCGTCTTTC R:TGTGGTAGTTCGACGCTGTG	Fvb3-4	59.318	319	8	P	0.4988	0.3744
25	MK138601	(AT)6	F:TCTTCCTCCTCATCCGAGAA R:TGACCCCAAAATGAAGAAGC	Fvb5-2	59.879	180	5	P	0.2408	0.2118
26	MK92984	(TC)15	F:GGTCCAAATAAAAGAGCAACCA R:TGGGCAAGAAATAAGGGACA	Fvb3-4	60.344	396	5	P	0.3848	0.3108
27	MK53549	(AT)8	F:TCACTTACCATGAGCCGTTG R:TGGGCAAATCTTGGGTTAAG	Fvb2-4	59.716	388	7	P	0.4935	0.3717
28	MK10802	(AG)13	F:GAGATTTTGAAGGCCCTGTG R:TTCTTTGGCTTTTGCGATCT	Fvb1-2	59.67	303	9	P	0.3805	0.3081
29	MK30802	(GGGGTC)6	F:TGTTGCTGGGGAATACAAGA R:CTGACCAACATCACCCTCCT	Fvb2-1	59.123	275	10	P	0.6104	0.5276
30	MK22054	(AG)5	F:CTGACCAACATCACCCTCCT R:TGTTGCTGGGGAATACAAGA	Fvb1-3	59.962	289	10	P	0.4968	0.3734
31	MK27298	(AG)10	F:TTTGCTATGATGCGGAGTTG R:TTTCTCCGGTGGTGATTAGG	Fvb1-4	59.833	389	3	P	0.4800	0.3648
32	MK152026	(AT)9	F:GAGAGTGGATCGCCTTCTTG R:GTTGACTGTTGGAGGGGTGT	Fvb5-4	59.95	273	10	P	0.4950	0.3725
33	MK149849	(AT)9	F:GTTGACTGTTGGAGGGGTGT R:GAGAGTGGATCGCCTTCTTG	Fvb5-4	59.859	279	8	P	0.5150	0.4345
34	MK5668	(AT)11	F:GGCGCTAGACAAGTCCTTTG R:GCGGTACTTGGAGCATGAAT	Fvb1-1	60.015	392	8	P	0.5697	0.4792
35	MK134130	(AG)6	F:GAAGGGTGGAGTGAATTGGA R:TTGGAAGGAAGCTGAATTGC	Fvb5-2	59.903	258	6	P	0.3200	0.2688

Table 3b. Characteristics of selected SSR primers designed using GMATA.

Seq No	Marker Name	Motifs	Primer Sequence	Pseudo molecule ID	Calculated T_M	Product Size	No. of alleles	Polymorphism	H_0[a]	PIC_0[b]
36	MK155633	(AG)14	F:TTGAATGCCTTAACGATCACC R:AACGCAGTTTTGAGTGCCTTA	Fvb5-4	59.952	371	6	P	0.4444	0.3457
37	MK134343	(TA)19	F:AGCAAGGGTTAACATTCCAGT R:TAGCTTTCCAGCAAACATGC	Fvb5-2	57.737	335	6	P	0.5150	0.4345
38	MK134350	(AG)8	F:TAGCTTTCCAGCAAACATGC R:AGCAAGGGTTAACATTCCAGT	Fvb5-2	59.066	340	6	M	monomorphic	
39	MK126936	(AG)10	F:ACCCACTTTATGCACCGTTT R:CAAGAGGTGTTGGTGGCTCT	Fvb5-1	59.359	376	8	P	0.3047	0.2583
40	MK136424	(GTG)5	F:AACATTCCTGGCCAATTCAC R:TGCTTATCATTGCTGCATCC	Fvb5-2	59.797	392	9	P	0.4583	0.3533
41	MK111069	(CT)5	F:GACTCCGCCTATGTTTTGGA R:GCCGGCAGCCATATATAGAA	Fvb4-3	60.074	135	8	P	0.4997	0.3748
42	MK35652	(TA)8	F:TGACCAGACCCACACAAGAA R:TTGGATTTGCAGTCCATCAA	Fvb2-1	60.129	384	8	P	0.5722	0.4807
43	MK114975	(AT)22	F:ATGAAAATCCGCACCAAAAC R:GAAATTCGATCGGGAAGTGA	Fvb4-3	59.807	214	6	P	0.5178	0.4513
44	MK23605	(AT)6	F:AGGAAAGGGTCAAACCTCGT R:AAAGTGAGGGACACGGACAC	Fvb1-4	59.972	213	6	P	0.5178	0.4513
45	MK114211	(GA)8	F:CCTACAATGGCTCGAGGAAA R:TATTTCTCCAAACCCCACCA	Fvb4-3	60.206	395	6	P	0.4644	0.3566
46	MK114249	(TA)11	F:TATTTCTCCAAACCCCACCA R:CCTACAATGGCTCGAGGAAA	Fvb4-3	60.162	395	8	P	0.4297	0.3374
47	MK66995	(TC)5	F:AATCATTCTCCCCTCCCTTTT R:CAAGCAGCATCACAATCACC	Fvb3-1	60.136	380	7	P	0.4800	0.3648
48	MK24581	(AG)12	F:GGCGCTTAGAAAACGTGAAG R:TCTCAGGAGAAGCCATGGAG	Fvb1-4	60.018	326	3	P	0.5867	0.5094
49	MK28186	(AT)13	F:TCTCAGGAGAAGCCATGGAG R:GGCGCTTAGAAAACGTGAAG	Fvb1-4	60.488	350	5	P	0.4968	0.3734
50	MK111146	(AC)6	F:AATTCGATCGGAAAGTGGTG R:TCGGCTAAAGTGCTTGTCCT	Fvb4-3	59.933	161	9	P	0.4990	0.3745
51	MK55518	(AT)20	F:AGGATGACGGCGAGATATTG R:TGTTGTAGCAGTTGCCCTTG	Fvb2-4	60.059	244	8	P	0.4286	0.3368
52	MK132227	(AT)8	F:GGGTGGAAATAGCCTCATCC R:AAATGGGTAAAGGAGGCAAGA	Fvb5-1	60.662	376	6	P	0.4911	0.3705
53	MK131750	(AC)5	F:AAATGGGTAAAGGAGGCAAGA R:GGGTGGAAATAGCCTCATCC	Fvb5-1	59.952	377	3	P	0.4511	0.3929
54	MK79369	(AT)7	F:TCGTTTGGGGTTGTTTCTTT R:CCTGCATGAGAAGGTTAGGG	Fvb3-3	59.448	398	9	P	0.4978	0.3739
55	MK18475	(TC)6	F:AACGCAGTTTTGAGTGCCTTA R:AACGCAGTTTTGAGTGCCTTA	Fvb1-3	59.934	394	10	P	0.6166	0.5378
56	MK43115	(AT)16	F:CCTGCTCACTGTCTCCATCA R:TGCGATACTTGATCCCACAA	Fvb2-2	59.98	335	10	P	0.4950	0.3725
57	MK116596	(AT)5	F:TAGATCGCAACGAGTGCAAC R:ACGGAAGCAGAGGAAGCATA	Fvb4-3	60.019	326	6	P	0.4550	0.3515
58	MK89686	(CCT)5	F:CGTCAAAATGCTTTGGAAGA R:GAGCGGGTGATTTAGGACAA	Fvb3-4	58.892	362	6	P	0.4911	0.3705
59	MK78964	(GT)7	F:ACCGCTTCTGGGTTTCTTTT R:TACGTGAGCTTCGAACATGG	Fvb3-3	60.11	296	5	P	0.5312	0.4516
60	MK118282	(GT)6	F:TTTTGATGAATGTGGCTTGG R:CACTGGCACACAAAGTAGCAA	Fvb4-4	59.518	380	6	P	0.6200	0.5419
61	MK43841	(CA)5	F:TCCTTCAACCATTTGGCTTC R:GGTGGGACATTTCTGCTGTT	Fvb2-2	60.051	323	6	M	monomorphic	
62	MK21052	(AG)8	F:TTTGCACGCTGATTGTGACT R:TGCATCTTGACTCCATTCCA	Fvb1-3	60.459	320	4	P	0.3488	0.2879
63	MK65432	(AT)10	F:TGAAACATAGGCTCTCATTTGC R:CGTCTCTCTTGGGGCACTTA	Fvb3-1	59.365	341	6	P	0.4728	0.3610
64	MK21538	(TC)7	F:AAACGCCCTAACAACAAACG R:GACGCACGAACTCAAGAAAA	Fvb1-3	60.032	300	4	P	0.4388	0.3425
65	MK59001	(TA)15	F:GTCGGCAGAGGAATTTACGA R:GGTGCACATTCGACCAATTA	Fvb2-4	60.214	194	5	P	0.4352	0.3405
66	MK101641	(AG)15	F:TCGATCGTGCATTTCTTGAG R:GGATTTTCGCAGTCGTCTTT	Fvb4-2	59.948	340	8	P	0.4747	0.3620
67	MK91772	(AG)14	F:GGATTTTCGCAGTCGTCTTT R:TCGATCGTGCATTTCTTGAG	Fvb3-4	59.316	355	8	P	0.2888	0.2471
68	MK124032	(AT)5	F:CCACACAACACAGAACCAGAG R:CCCACTTGCTGTAATGGTTG	Fvb4-4	59.212	371	4	P	0.5000	0.3750
69	MK13305	(GGGTCG)6	F:TATCGCAATCCTTCACCACA R:CTCCTCTTCCATCCAGGACA	Fvb1-2	60.073	226	2	P	0.0950	0.0905
70	MK125700	(CT)5	F:TGAGGGTGCACTGAGATTGT R:CCTTATTATGGTGGTGGGAAT	Fvb5-1	59.261	388	5	P	0.3848	0.3108

Table 4. Average, minimum, and maximum values of polymorphism indices generated using iMEC for the polymorphic markers genotyped using the 10 strawberry cultivars.

Polymorphism Indices	Average	Min	Max
Heterozygosity Index (H)	0.469	0.095	0.638
Polymorphic Information Content (PIC)	0.380	0.090	0.564
Effective Multiplex Ratio (EMR)	2.997	1.000	7.100
Average Heterozygosity (H. ave)	0.198	0.002	0.638
Marker Index (MI)	0.210	0.009	0.638
Discriminating power (D)	0.546	0.099	0.857
Resolving power (R)	2.918	0.200	6.400

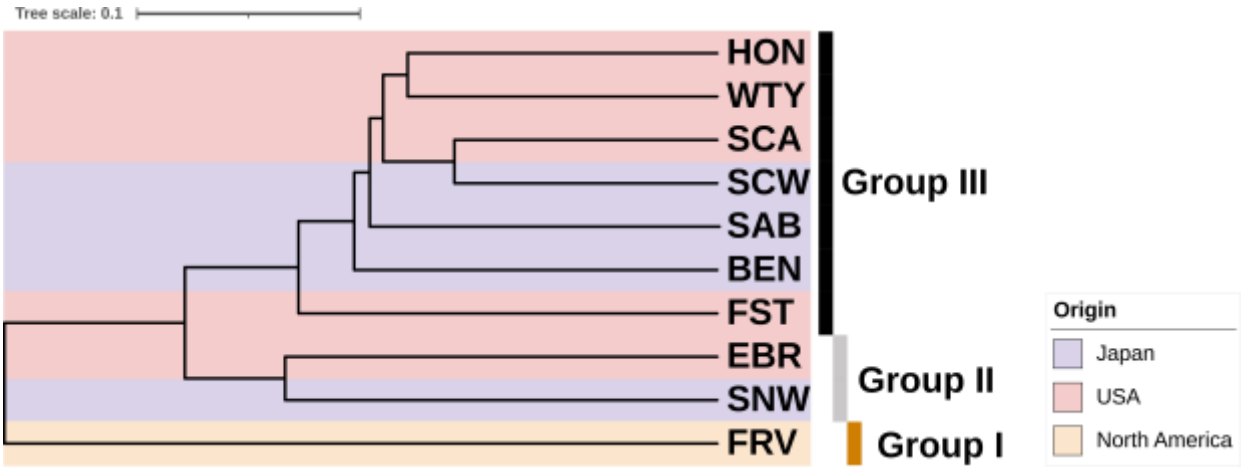

Figure 4. Classification of 10 strawberry cultivars based on their molecular diversity. Genetic similarities were calculated by UPGMA with the dendrogram constructed using FigTree. SCA: 'Sweet Charlie Argentina', SCW: 'Sweet Charlie Washington', HON: 'Honeoye', WTY: 'Whitney', SAB: 'Sweet Angel Bohol', BEN: 'Benihoppe', FST: 'Festival', EBR: 'Earlibrite', SNW: 'Snow White', FRV: '*Fragaria vesca.'* Color ranges denote known origin of each accession.

ananassa cultivars which is an octoploid. The *F. vesca* was least similar to Snow white with a similarity coefficient of 0.313, and most similar to Sweet Charlie-Washington with the similarity coefficient of 0.392. Generally, the *F. vesca* was 30% similar to other *F. x ananassa* cultivars. Group 2 consists of 2 cultivars: Earlibrite, which originated from the United States of America (USA) and Snow white which originated from Japan, with a similarity index of 60%. Earlibrite was observed to have an average of 55% similarity with other *F. x ananassa* cultivars, having Sweet Angel Bohol as the closest relationship with 68%. Snow white has an average similarity of 40% with other strawberry cultivars. The third group comprises 7 cultivars: Sweet Charlie Argentina, Sweet Charlie Washington, Honeoye, Whitney and Festival (all originating from different parts of the USA, while, Benihoppe and Sweet Angel Bohol originated from Japan (Edger, *et al*., 2019). The most similar cultivars from this group is the two sweet Charlie coming from Washington and Argentina with genetic similarity of 78%, followed by Sweet Charlie Washington and Honeoye with a similarity of 72%. The least similar in this group is Festival and Benihoppe with a genetic similarity of 60%. The average similarity within this group is 67% (Figure 4).

DISCUSSION

The advent of next-generation sequencing (NGS) technologies makes genome sequencing of commercial and specialty crops more affordable and accessible for breeders. Reference genomes generated using NGS technologies have been used to mine and develop robust SSR markers in wild tomato, carrot, spinach, and coconut (Mangat *et al*., 2020; Uncu *et al*., 2020; Bhattarai *et al*., 2021; Caro *et al.*, 2022). Here, a near-complete chromosome scale assembly for cultivated octoploid strawberry (*F. x ananassa* cultivar 'Camarosa') retrieval from GDR and was utilized for mining and development of robust SSR markers for molecular marker-assisted breeding. This high-quality reference genome is 805,488,706 bp in length, distributed across 28 chromosome-level pseudomolecules (Edger *et al*., 2019). The availability of a strawberry reference genome has enabled us to discover a sufficient number of polymorphic loci. Furthermore, this study uncovered many genome-wide SSR markers with known physical location in the chromosome-level pseudomolecules of cultivated strawberry.

The density of SSRs in plants largely differs in a wide variety of plant genomes, which points out to the genome's adaptable tolerance to SSRs. As genome size increases, the density of SSRs decreases, which suggests a negative correlation between genome size and SSR density (Morgante *et al*., 2002). This trend is also evident in our data because we found that the SSR density in cultivated strawberry genome is about 286.47 SSRs/Mb (Table 2) compared to the SSR density in wild strawberry which is 405.58 SSRs/Mb (Guan*, et al*., 2013). Generally, plants with large genomes like corn and wheat have a lower SSR density compared to those with smaller genomes such as cucumber and dates (Cavagnaro *et al*., 2010; Zhao *et al.*, 2017).

The newly designed and developed markers had the PCR amplification verification using selected primers on 10 cultivars as a representative of the strawberry core collection of the breeding program. The SSR primers were characterized based on their ability to amplify all the samples and based on the size of the amplicons. The observation of an average of 6 alleles per locus across test varieties, with a range from 2 to 10 alleles per locus, suggests a high degree of genetic diversity within the tested strawberry cultivars. The finding that over 90% of the markers synthesized generated more than two alleles further supported this diversity. The high frequency of markers generating six and eight alleles per locus suggests the presence of common alleles within the tested cultivars similar to those generated by the study of Cipriani *et al*. (2018). The overall success of the synthesized SSR markers in generating bands of the expected size further supports their utility for genetic studies in strawberry cultivars (Diaz-Martin *et al.*, 2012, *Yoon et al*., 2012).

The generation of several SSR markers would provide a pool for selection but may not be completely utilized in marker synthesis because of laboratory limitations. Hence, selection and filtering of SSRs is necessary to optimize characterization of available germplasm with the smallest number of markers for practical purposes (Caro *et al.*, 2022). The criteria used in selecting markers for wet lab validation are their base pair size, number of alleles and the variation within the expected allele sizes for each of the markers. These components would translate to the visibility of bands, polymorphism, and ease of scoring for the test cultivars (Li *et al.*, 2019; Song *et al.*, 2021).

Allelic variation may be correlated with the number of repeats within a particular microsatellite locus. In other words, the repeat length may be associated with the polymorphism information content. PIC (polymorphic information content) value is an important parameter that is used to measure the level of polymorphism of a molecular marker (Amiryousefi *et al*., 2018). It represents the probability that two randomly selected alleles from a population will be different. A higher PIC value indicates higher genetic diversity within the population and makes the marker more informative for genetic analysis. Using the criteria developed by Botstein in the early 80s, markers with PIC

values greater than 0.5 are very informative, values between 0.25 and 0.5 are somewhat informative, while values lower than 0.25 are not very informative. Based on these, there are 8 markers generated in this study that are considered to be highly polymorphic with PIC value ranging from 0.509 to 0.563. These are MK26759, MK118282, MK115000, MK18475, MK103594, MK115479, MK30802 and MK24581. Additionally, 56 out of the 70 markers (80%) showed a moderate polymorphism ranging from 0.258 to 0.487.The PIC values generated in this experiment are relatively higher than those reported in SSR generated in other studies which generated PIC values ranging from 0.285 to 0.384 (Lim *et al.,* 2017; Zhang, 2021).

This study evaluated the variation and genetic relationship among cultivars to further investigate the effectiveness of the SSR developed. It can be observed that the cultivars were divided into 3 groups based on UPGMA. It can be seen that the *F. vesca* was completely separated from the other *F. x ananassa* cultivars. The dendrogram created from this grouping showed that there was a clear separation between diploid and octoploid cultivars, with *F. vesca* being the only diploid cultivar in the study. These results are congruent with the study of Lim *et al.* (2017) and Aristya *et al.* (2019), which showed the highly discriminatory nature of the SSR markers. Similarly, the data presented by Lebedev *et al.* (2020) are also congruent, showing the efficiency of SSR markers in correctly identifying and discriminating among strawberry cultivars.

Additionally, the study found that cultivars from different parts of the world, including the United States and Japan were present in the second and third groups, indicating that there is a genetic diversity among these strawberries that is not solely based on their geographic location. It contrasts with several studies where clustering of cultivars into a specific population group was strongly influenced by geographical origin as well as adaptive environment rather than their known parentage and pedigree (Lim *et al*., 2017; Wada *et al*., 2017). Furthermore, the variation present outside of its geographical origin could be accounted for continuous hybridization and introgression within cultivars.

Generally, SSR markers have conventionally proven effective in distinguishing between cultivars, playing pivotal role in characterizing and fingerprinting of diverse germplasms, which substantiated the findings that we have generated.

CONCLUSIONS

In this study, we have generated a total of 219, 239 SSR markers, from which 73% are unique, using *in silico* mining derived from the *F. x ananassa* 'Camarosa' genome assembly locally established bioinformatics pipeline. It is in response to the limited number of SSR markers available for characterizing strawberry germplasm for tropical environment. Validating a subset of 70 markers based on the in house established criteria and results revealed that 95.7% are polymorphic and 90% have medium to high PIC. These markers were able to discriminate representative strawberry cultivars. The following markers: MK26759, MK118282, MK115000, MK18475, MK103594, MK115479, MK30802 and MK24581 with high PIC value more than 0.50 are priority markers recommended for use in molecular breeding program. The development of novel SSR markers for strawberry will serve as a valuable resource for characterizing the Philippine strawberry germplasm in BSU breeding program and as baseline information for future marker-assisted breeding applications.

ACKNOWLEDGMENTS

This research was supported by the Office of the Vice President for Academic Affairs- Emerging Inter-Disciplinary Research Program (OVPAA-EIDRP-C09-12). The authors gratefully acknowledge the Benguet State University (BSU) HORTI collaborators Drs. Belinda Tad-awan and Darwin Basquial and Ms. Milagros Dumaslan for providing the leaf samples of the strawberry cultivars for molecular characterization; and Plant Molecular Phylogenetics Laboratory group of the Institute of Biology (IB), College of Science for providing the necessary

laboratory materials. We would also like to extend our gratitude to Mary Grace Burgos, Dexter Pahayo, and Lorelie Ollayan for the technical assistance.

REFERENCES

Amiryousefi A, Hyvönen J, Poczai P (2018). iMEC: Online marker efficiency calculator. *Appl. Plant. Sci.* 6(6): 1-4. [AA2]

Aristya GR, Kasiamdari R, Arif MF, Amalia F, Pawestri RD (2019). Comparison of molecular markers ISSR, SSR and RFLP to study the genetic variation of strawberry cultivars. *Am. J. Agric. Biol. Sci.* 14(1): 61-68.

Awad MF, Hassanin AA (2022). In silico profiling of proline biosynthesis and degradation related genes during fruit development of tomato. SABRAO J. Breed. Genet. 54(3): 549-564. http://doi.org/10.54910/ sabrao2022.54.3.8.

Bhattarai G, Shi A, Kandel DR, Solis-Gracia N, da Silva J, Avila C (2021). Genome-wide simple sequence repeats (SSR) markers discovered from whole-genome sequence comparisons of multiple spinach accessions. *Sci. Rep*. 11(1): 1-16. [AA3]

Caro RES, Cagayan J, Gardoce RR, Manohar ANC, Canama-Salinas AO, Rivera RL, Lantican DV, Galvez HF, Reaño CE (2022). Mining and validation of novel simple sequence repeat (SSR) markers derived from coconut (*Cocos nucifera* L.) genome assembly. *J. Genet. Eng. Biotechnol.* 20 (71): 1-14.[AA4]

Cavagnaro PF, Senalik DA, Yang L (2010). Genome-wide characterization of simple sequence repeats in cucumber (*Cucumis sativus L*.). *BMC. Genom.* 11: 1-18. [AA5]

Chen L, Zhong Y, Peng X, Ding Y, Lu Y (2018). In silico identification and experimental validation of SSR markers in sweet potato [*Ipomoea batatas* (L.) Lam.]. *BMC. Genom.* 19(1): 1-10.

Cheng Y, Zhang Q, Wu J (2017). A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity. *Nat. Genet.* 49: 1515-1523.

Choudhary S, Sharma PC, Kumar S, Agarwal S, Tyagi A (2016). Developing SSR markers from EST database of pigeonpea (*Cajanus cajan L.*). *Plant. Cell. Rep.* 35: 1649-1662.

Cipriani G, Testolin R, Rossi G (2018). Development and characterization of simple sequence repeat markers for Fragaria species. *Mol. Ecol. Resour*. 18(2): 363-373.

Diaz-Martin J, Alvarez A, Espinoza C, Pico B (2012). Development of polymorphic microsatellite markers for genetic analysis of strawberry. *J. Agric. Sci.* 4(9): 176-184.

Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR (2019). Origin and evolution of the octoploid strawberry genome. *Nat. Genet*. 51: 541-547.

FAOSTAT (2021). Statistics division. Food and Agriculture Organization of the UN.

Fu Y, Cheng J, Wei S, Zhang Y, Wang Z, Zhao G, Lu J, Han L, Tang X (2016). Development and characterization of EST-SSR markers for switchgrass (*Panicum virgatum*). *Biochem. Syst. Ecol.* 68: 86-92.

Gonzalez M, Frank D, Fernandez M, Bertoldi D (2020). Strawberry quality and health benefits: A review: *Food. Sci. Technol.* 40(1): 1-16.

Guan JF, Huang GQ, Feng XW, Wang Y, Wan BY, Chen and Qiao YS (2013). Survey of simple sequence repeats in woodland strawberry (*Fragaria vesca*). *Genet. Mol. Res.* 12(3): 2637-2651.

Hancock JF (1990). Ecological genetics of natural strawberry species. *HortSci.* 25(8): 869-871.

Haymes KM, Hokanson SC, Salazar K, Maas J (2000). Development of SCAR markers linked to a *Phytophthora fragariae* resistance gene and their assessment in European and North American strawberry genotypes. *J. Am. Soc. Hortic. Sci.* 125: 330-339.

Honjo M, Nunome T, Kataoka S, Yano T, Yamazaki H, Hamano M, Yui S, Morishita M (2011). Strawberry cultivar identification based on hypervariable SSR markers. *Breed Sci.* [AA6][v7]61(4): 420-425.

Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica.* 177: 330-334.

Khaled KAM, Habiba RMM, Bashasha JA, Azzam CR, M.H. Abdel-Aziz MH (2023). In silico and genetic analysis related to tillering ability in maize. SABRAO J. Breed. Genet. 55(1): 156-162. http://doi.org/10.54910/ sabrao2023.55.1.15

Lebedev VG, Subbotina NM, Maluchenko OP, Lebedeva TN, Krutovsky KV, Shestibratov KA (2020). Transferability and polymorphism of SSR markers located in flavonoid pathway genes in *Fragaria* and *Rubus* species. *Genes.* 11(1): 1-24.

Li C, Wang Y, Chen H, Chen Y, Yu H, Ma R (2019). Development of SSR markers in strawberry based on transcriptome sequencing using GA/CT/TC/CG repeats. *BMC Genom*. 20(1): 1-10.

Lim S, Lee J, Lee HJ, Park KH, Kim DS, Min, SR, Jang WS, Kim T Il, Kim HR (2017). The genetic diversity among strawberry breeding resources based on SSRs. *Sci. Agric.* 74(3): 226-234.

Mangat PK, Gannaban RB, Singleton JJ, Angeles-Shim RB (2020). Development of a PCR-based, genetic marker resource for the tomato-like nightshade relative, *Solanum lycopersicoides* using whole genome sequence analysis*. PLoS One*. 15(11): 1-14.[AA8]

Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat. Genet.* 30(2): 194-200.

Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R, Jüngling R, Singh J, Kavi Kishor PB, Sivaramakrishnan S, Hoisington DA, Kahl G, Winter P, Cook DR, Varshney RK (2017). Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.* 130: 27-40.

Padua DP, Kilakil, CM (2011). Strawberry variety improvement in the Philippines. *Philipp. J. Crop. Sci*. *Suppl* 36: p31.

Shulaev V, Sargent D, Crowhurst R (2011). The genome of woodland strawberry *(Fragaria vesca*). *Nat Genet. 43*: 109-116.

Song X, Yang Q, Bai Y, Gong K, Wu T, Yu T, Pei Q, Duan W, Huang Z, Wang Z, Liu Z, Kang X, Zhao W, Ma X (2021). Comprehensive analysis of SSRs and database construction using all complete gene-coding sequences in major horticultural and representative plants. *Hort. Res.* 8(122): 1-17.

Uncu AO, Uncu, AT (2020). High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota L*.) genome with chromosome-anchored markers. *Biotechnol. Equip.* 34(1): 1-9.

Varshney RK, Graner A, Sorrells ME (2013). Genic microsatellite markers in plants: features and applications. *Trends. Biotechnol*. 31(3): 155-163.

Varshney RK, Nayak SN, May GD, Jackson SA (2016). Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol*. 34: 522-530.

Wada T, Noguchi Y, Isobe S, Kunihisa M, Sueyoshi T, Shimomura K (2017). Development of a core collection of strawberry cultivars based on SSR and CAPS marker polymorphisms. *Hort. Res.* 86(3): 365-378.

Wang X, Wang L (2016). GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Front. Plant. Sci.* 7(1350): 1-11

Yoon MY, Moe KT, Kim DY, Rho IR, Kim S, Kim KT, Won MK, Chung JW, Park YJ (2012). Genetic diversity and population structure analysis of strawberry (*Fragaria x ananassa* Duch.) using SSR markers. *Electron. J. Biotechnol.* 15(2): 71-85.

Zhao Y, Keremane M, Prakash CS, He G (2017). Characterization and amplification of gene-based simple sequence repeat (SSR) markers in date palm. *Methods. Mol. Biol.* 1638: 259-271.