Breeding Science 70: 415–421 (2020)
doi: 10.1270/jsbbs.19158

Note

SSR markers developed using next-generation sequencing technology in pineapple, *Ananas comosus* (L.) Merr.

Kenji Nashima*[1)], Fumiko Hosaka[2)], Shingo Terakami[2)], Miyuki Kunihisa[2)], Chikako Nishitani[2)], Chie Moromizato[3)], Makoto Takeuchi[3)], Moriyuki Shoda[4)], Kazuhiko Tarora[4)], Naoya Urasaki[4)] and Toshiya Yamamoto[2)]

[1)] *College of Bioresource Sciences, Nihon University*, 1866 Kameino, Fujisawa, Kanagawa 252-0880, Japan
[2)] *Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization*, 2-1 Fujimoto, Tsukuba, Ibaraki 305-0852, Japan
[3)] *Okinawa Prefectural Agricultural Research Center Nago Branch*, 4605-3 Nago, Nago, Okinawa 905-0012, Japan
[4)] *Okinawa Prefectural Agricultural Research Center*, 820 Makabe, Itoman, Okinawa 901-0336, Japan

Simple sequence repeat (SSR) markers provide a reliable tool for the identification of accessions and the construction of genetic linkage maps because of their co-dominant inheritance. In the present study, we developed new SSR markers with next-generation sequencing using the Roche 454 GS FLX+ platform. Five hundred SSR primer sets were tested for the genetic identification of pineapple, including 100 each for the di-, tri-, tetra-, penta-, and hexa-nucleotide motif SSRs. In total, 160 SSR markers successfully amplified fragments and exhibited polymorphism among accessions. The SSR markers revealed the number of alleles per locus (ranging from 2 to 13), the expected heterozygosity (ranging from 0.041 to 0.823), and the observed heterozygosity (ranging from 0 to 0.875). A total of 117 SSR markers with tri- or greater nucleotide motifs were shown to be effective at facilitating accurate genotyping. Using the SSR markers, 25 accessions were distinguished genetically, with the exception of accessions 'MD-2' and 'Yonekura'. The developed SSR markers could facilitate the establishment of efficient and accurate genetic identification systems and the construction of genetic linkage maps in the future.

Key Words: *Ananas comosus*, NGS, SSR marker, identification of accessions.

Introduction

The pineapple, *Ananas comosus* (L.) Merr., is one of the most economically important fruit species globally. Pineapples are cultivated primarily in tropical and subtropical regions, and global production was approximately 27.4 million tonnes (Mt) in 2017 (http://fao.org/faostat/). Countries with the highest pineapple production included Costa Rica (3.05 Mt), the Philippines (2.67 Mt), and Brazil (2.25 Mt), followed by other countries in tropical or subtropical regions. The pineapple belongs to the family Bromeliaceae, which includes approximately 2,000 species, most of which are epiphytic or ornamental plants (Morton 1987). Prior to European involvement, genus *Ananas* was distributed around northern South America. In the 19th century, the 'Smooth Cayenne' cultivar was introduced to Europe from French Guiana and subsequently distributed to tropical and subtropical regions around the globe. Currently, 'Smooth Cayenne' has been displaced in the international fresh produce market by the variety 'MD-2' (Coppens d' Eeckenbrugge *et al.* 2018). In Japan, breeding program has sought to produce a pineapple variety for fresh consumption. To date, eight cultivars have been bred at Okinawa Prefectural Agricultural Research Center (OPARC) (Ogata *et al.* 2016). Breeding of elite cultivars in the future will require the improvement of agronomic traits and characteristics such as fruit quality. The establishment of advanced breeding systems will be critical to the efficient development of new elite cultivars (Ogata *et al.* 2016). DNA profiling is expected to be useful in the development of efficient breeding systems in pineapples, particularly to protect the rights of plant breeders and to accelerate breeding via DNA marker-assisted selection (MAS).

To date, several types of DNA markers have been developed, including restriction fragment length polymorphism (RFLP) marker (Duval *et al.* 2001, 2003), random-amplified polymorphic DNA (RAPD) marker (Ruas *et al.* 1995, 2001), amplified fragment length polymorphism (AFLP) marker (Kato *et al.* 2004, Paz *et al.* 2005), and

Communicated by Masafumi Yagi
Received November 25, 2019. Accepted March 12, 2020.
First Published Online in J-STAGE on May 30, 2020.
*Corresponding author (e-mail: nashima.kenji@nihon-u.ac.jp)

simple sequence repeat, microsatellite (SSR) marker (Carlier *et al.* 2012, Feng *et al.* 2013, Shoda *et al.* 2012). Among the DNA marker types listed above, SSR markers have several key advantages due to their high levels of polymorphism, multiple alleles, and co-dominance (Ellegren 2004). SSR markers have been applied extensively in population genetics, molecular breeding, and paternity test studies (Ellegren 2004). Numerous markers have been developed, including 18 EST-SSR markers developed by Wöhrmann and Weising (2011), 18 SSR markers developed by Shoda *et al.* (2012), 32 SSR markers developed by Carlier *et al.* (2012), 18 SSR markers developed by Feng *et al.* (2013), and 2 SSR markers developed by Urasaki *et al.* (2015).

DNA markers have previously been used to construct genetic linkage maps in pineapples (Carlier *et al.* 2004, 2006, 2012, Sousa *et al.* 2013). These genetic linkage maps are composed of 157 to 741 DNA markers, with 32% to 86% map coverage (Leitao 2018). Although the number of markers used and the coverage of such genetic linkage maps are extensive, co-dominant type DNA markers such as SSR and cleaved amplified polymorphic sequence (CAPS) remain scarce. For example, 37 SSR markers and 8 CAPS markers were reported in 741 DNA markers by Sousa *et al.* (2013) based on a genetic linkage map. The DNA markers were primarily dominant types, such as RAPD, sequence-characterized amplified region (SCAR), AFLP, and inter-simple sequence repeat (ISSR) types, which all exhibit relatively low reproducibility and transferability. The development of additional SSR markers would facilitate the construction of more accurate and widely transferable genetic linkage maps. Such genetic linkage maps are also expected to facilitate quantitative trait loci (QTL) analyses and the development of selection markers for DNA MAS. To date, piping (*P*) and spiny-tip (*S*) loci have been revealed as phenotype-determining loci in pineapples, and their selective markers (PLST1_SSR and STLST1-CAP) are the only DNA markers applicable in MAS (Urasaki *et al.* 2015). To facilitate further development of DNA markers linked to other useful MAS traits, novel QTLs or responsible gene loci of qualitative traits must be identified. Consequently, additional DNA markers necessary for the construction of dense linkage maps for QTL analysis will need to be developed.

It is critical to develop SSR markers with motifs of tri- or greater nucleotide repeats with high polymorphism among current pineapple accessions for the identification of future accessions. SSR markers with di-nucleotide motifs often exhibit stutter fragments (which are generated by the slippage of Taq DNA polymerase during PCR) and present challenges to the scoring of alleles (Diwan and Cregan 1997, Harker 2001, Litt *et al.* 1993). The use of tri- or more nucleotide motif SSR markers with greater polymorphism is therefore likely to provide more accurate and efficient identification of accessions. Although SSR markers have previously been used to assess genetic diversity or identify pineapple accessions (Lin *et al.* 2015, Shoda *et al.* 2012, Wang *et al.* 2017, Wöhrmann and Weising 2011), most of the SSR markers used were of di-nucleotides.

In recent years, Next Generation Sequencing (NGS) analyses have provided large-scale sequence information, and NGS has facilitated efficient SSR identification and DNA marker development (Zalapa *et al.* 2012). In the present study, we performed large-scale development of SSR markers for genetic identification using NGS data. The characteristics of the SSR markers were developed using di- to hexa-nucleotide motifs and were examined using 25 different pineapple accessions.

Materials and Methods

Plant material and DNA extraction

The 25 accessions used in this study included seven cultivars from OPARC, Nago Branch, Okinawa, Japan, and 18 foreign accessions introduced from the USA, Brazil, Taiwan, and Australia (**Table 1**). Among the accessions, 22 were *A. comosus* var. *comosus* accessions, 2 were wild varieties (*A. comosus* var. *bracteatus* and *A. comosus* var. *ananassoides* introduced from Brazil), and one was a hybrid variety between *A. comosus* var. *ananassoides* and *A. comosus* var. *comosus*. DNA extraction was performed using DNeasy plant mini kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

NGS analysis

The genomic DNA of 'N67-10' was sheared through nebulization (600 to 900-bp length) and amplified by emulsion PCR. Two single-read pyrosequencing runs were performed using 454 GS FLX+ genome sequencer (Roche Diagnostics, Basel, Switzerland). The sequencing data obtained through this process were registered as DRA accession no. DRR174973 and used for SSR marker development.

SSR detection and SSR marker development

SSR detection was performed using a Microsatellite Identification Tool (Thiel *et al.* 2003). SSRs were defined as more than 10 to 20 repeats for the di-nucleotide motif, 5 to 15 repeats for the tri-nucleotide motif, and 5 to 10 repeats for tetra-, hexa-, and penta-nucleotide motifs. For the di-nucleotide motif, 19 SSRs of AC/GT motif, 24 of AT/TA motif, and 57 of AG/CT motif were randomly selected and applied for PCR primer designing. For greater than tri-nucleotide motifs, 100 PCR primers were designed based on randomly selected SSRs. PCR primers were designed using BatchPrimer3 (http://batchprimer3.bioinformatics.ucdavis.edu) using the default settings for picking a primer (You *et al.* 2008). PCR primers were designed on the sides of SSRs with 10 to 20 repeats for di-nucleotide motifs, 5 to 15 repeats for tri-nucleotide motifs, and 5 to 10 repeats for tetra-, penta- and hexa-nucleotide motifs. Subsequently, amplified regions for the designed SSR markers were searched against each other using BLAST to identify and

Table 1. Pineapple accessions used in this study

Accession name	Parentage	Variant	Origin	Type
N67-10	selection from Hawaiian Smooth Cayenne	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Soft Touch	Hawaiian Smooth Cayenne × I-43-880	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Haney Bright	Mitsubishi Smooth Cayenne × I-43-908	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Summer Gold	Cream Pineapple × McGregor ST-1	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Yugafu	Cream Pineapple × HI101	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Gold Barrel	Cream Pineapple × McGregor ST-1	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Julio Star	N67-10 × Cream Pineapple	*A. comocsus* var. *comosus*	bred by OPARC-Nago[a]	cultivar
Tainung No. 11	(Smooth Cayenne × Mouritius) × Smooth Cayenne	*A. comocsus* var. *comosus*	introduced from Taiwan	cultivar
Tainung No. 17	Smooth Cayenne × Rough	*A. comocsus* var. *comosus*	introduced from Taiwan	cultivar
MD-2	58-1184 × 59-443	*A. comocsus* var. *comosus*	introduced from USA	cultivar
A882	Ripely Queen × Puerto Rico	*A. comocsus* var. *comosus*	unknown	breeding line
Bogor	Smooth Cayenne × Singapore Spanish	*A. comocsus* var. *comosus*	unknown	breeding line
Yonekura	unknown	*A. comocsus* var. *comosus*	unknown	breeding line
HI101	unknown	*A. comocsus* var. *comosus*	introduced from USA	breeding line
Red Spanish	unknown	*A. comocsus* var. *comosus*	introduced from Brazil	indigenous
McGregorST-1	unknown	*A. comocsus* var. *comosus*	introduced from Australia	indigenous
Seijyo Cayenne	unknown	*A. comocsus* var. *comosus*	intoduced from Taiwan	indigenous
Cream Pineapple	unknown	*A. comocsus* var. *comosus*	introduced from USA	indigenous
Cheese Pine	unknown	*A. comocsus* var. *comosus*	introduced from USA	indigenous
Papuri Vaupes Colombia	unknown	*A. comocsus* var. *comosus*	introduced from USA	indigenous
I-43-880	unknown	*A. comocsus* var. *comosus*	introduced from Brazil	indigenous
Santa Marta No. 1	unknown	*A. comocsus* var. *comosus*	introduced from USA	indigenous
A.comosus var. *ananassoides*	unknown	*A. comocsus* var. *ananasoides*	introduced from Brazil	indigenous
A.comosus var. *bracteatus*	unknown	*A. comocsus* var. *bracteatus*	introduced from Brazil	indigenous
A. comosus var. *ananassoides* × Rondon	*A. comosus* var. *ananassoides* × Rondon	*A. comocsus* var. *ananasoides* × *A. comocsus* var. *comosus*	introduced from USA	indigenous

[a] OPARC-Nago: Okinawa Prefectural Agricultural Research Center Nago.

eliminate redundant SSR markers (https://www.blast.ncbi.nlm.nih.gov/Blast.cgi).

Genotyping on the 25 accessions in **Table 1** was performed for 500 SSR markers, which consisted of 100 SSR markers for each di- to hexa-nucleotide motif. SSR-PCR amplification was performed in a 10-μL reaction mixture containing 5 μL of GoTaq Master Mix, including GoTaq DNA Polymerase (Promega, USA), 5 pmol FAM-labeled universal primer (5′-FAM-GCTACGGACTGACCTCGGAC-3′), 2.5 pmol forward and reverse primer (unlabeled), and 5 ng of template DNA. The "GCTACGGACTGACCTCGGAC" nucleotide sequence was added to the 5′ ends of the forward primers as a universal label in order to obtain FAM-labeled PCR products. For reverse primers, the "GTTTCTT" nucleotide sequence was added to the 5′ end of reverse primers as pig-tailing (Brownstein *et al.* 1996) to enhance adenylation and to facilitate accurate genotyping. The DNA was amplified in 35 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 2 min, and a final extension of 5 min at 72°C. The amplified PCR products were separated and detected in a PRISM 3130xl DNA sequencer (Applied Biosystems, USA). The sizes of the amplified bands were scored against internal-standard DNA (400HD-ROX, Applied Biosystems, USA) using GeneMapper (Applied Biosystems, USA). When fragment patterns exhibited two or more peaks detected among 25 accessions, no amplification pattern in less than 6 accessions, and polymorphism detected among 25 accessions, they were considered to be successful SSR markers.

SSR marker genotyping data analysis

The number of alleles (*Na*), expected heterozygosity (*He*), and observed heterozygosity (*Ho*) at single-locus SSR markers in the tested pineapple cultivars were calculated using Marker Toolkit (Fujii *et al.* 2008). *He* was calculated from the allele frequencies using the unbiased formula $1 - \Sigma^{pi2}$ $(1 \leq i \leq m)$, where m is the number of alleles at the target locus and pi is the allele frequency of the i^{th} allele at the target locus. *Ho* was calculated as the ratio of the heterozygous genotypes scored at each locus. MinimalMarker (Fujii *et al.* 2013) was used to identify minimal marker subsets for distinguishing the 25 accessions.

Correspondence with pineapple genome SSR marker sequences

Each SSR marker was matched to a corresponding pineapple genome sequence (Ming *et al.* 2015) using a BLASTn search for SSR marker amplified sequences against pineapple genome sequences (Ming *et al.* 2015) with an E-value < 1.0E-4.

Results

Identification of SSRs from Roche 454 GS FLX+ data

We obtained sequence data for identifying SSRs and developing SSR markers. The obtained sequence data included 1,340,605 reads with an average length of 1,023 bp and a total base of 1.37 Gb, approximately 2.6-fold greater than the estimated size of the pineapple genome, 520 Mb. For the development of SSR markers, 10 to 20 repeats for di-nucleotide motifs, 5 to 15 for tri-nucleotide motifs, and 5 to 10 for tetra- to hexa-nucleotide motifs were the target SSRs for SSR marker development. We identified 111,671 target SSRs from a total of 454 GS FLX+ sequences (**Table 2**, **Supplemental Tables 1–5**). The AG/CT motif was dominant among the detected SSRs, and the GC/CG motif was scarce in di-nucleotide SSRs. As the number of repeats decreased, the number of SSRs detected increased in di- to hexa-nucleotide motifs. This tendency was predominant in the greater nucleotide motifs. Notably, the majority of the SSRs detected in the hexa-nucleotide motif were of 5 to 6 repeats, accounting for 93% of all hexa-nucleotide motif SSRs.

Table 2. Identified target SSR numbers in FLX+ sequence in pineapple

Repeat motif	Number of target repeats	No. of target SSRs in FLX+ sequence
Di-nucleotide	10–20	44,564
Tri-nucleotide	5–15	56,313
Tetra-nucleotide	5–10	7,578
Penta-nucleotide	5–10	1,540
Hexa-nucleotide	5–10	1,676
Total		111,671

SSR marker development and calculation of the number of alleles, expected heterozygosity, and observed heterozygosity

We designed 500 SSR markers from the 454 GS FLX+ sequence data, including 100 SSR markers each for the di- to the hexa-nucleotide motifs. The amplification stability and the polymorphisms of the SSR markers were examined against 25 accessions, revealing 160 relevant SSR markers. These showed stable amplification with polymorphisms across 25 accessions. Conversely, 301 markers exhibited unstable or no amplification, and 39 markers did not show polymorphism across 25 accessions. The 160 SSR marker characteristics and the established genotypes are listed in **Supplemental Tables 6** and **7**. The average repeat numbers in established SSR markers were 13.4 in di-nucleotide motif, 6.5 in tri-nucleotide motif, 5.4 in tetra-nucleotide motif, 5.4 in penta-nucleotide motif, and 5.3 in hexa-nucleotide motif. Established SSR markers were distributed throughout the pineapple genome (**Supplemental Fig. 1**). SSR markers with di-nucleotide motifs showed the highest success rate among the established SSR markers, as well as the greatest mean *Na*, *He*, and mean fragment length difference among alleles (**Table 3**). *Na* ranged from 2 at 13 of the loci (TsuAc238, TsuAc240, TsuAc253, TsuAc255, TsuAc257, TsuAc260, TsuAc266, TsuAc274, TsuAc276, TsuAc283, TsuAc287, TsuAc296, and TsuAc338) to 12 at one of the loci (TsuAc264), with an average value of 4.63 (**Supplemental Table 6**). The value of *He* ranged from 0.041 at TsuAc257 and TsuAc274, to 0.823 at TsuAc264, with an average value of 0.513 (**Supplemental Table 6**). The value of *Ho* ranged from 0 at 3 loci (TsuAC216, TsuAC253, and TsuAC276) to 0.875 at 2 loci (TsuAC201 and TsuAC290), with an average value of 0.476 (**Supplemental Table 6**). Variance in mean fragment length among alleles was relatively low in di- and tri-nucleotide motifs, and high in tetra- to hexa-nucleotide motifs (**Table 3**). Null alleles were found in 19 SSR markers (**Supplemental Table 7**). Since null alleles could only be detected when homozygous, it is possible that null alleles might have been included in homozygous of some alleles. Among the tested accessions, 'Yonekura' and 'MD-2' could not be distinguished from each other because all of the SSR marker genotypes between 'Yonekura' and 'MD-2' were identical (**Supplemental Table 7**). All of the 25 tested accessions

Table 3. Characteristics of newly developed SSR markers for each repeat motif. *Na*: number of alleles, *He*: expected heterozygosity, *Ho*: observed heterozygosity

Repeat motif	No. of examined markers	No. of established markers	*Na* (mean)	*He* (mean)	*Ho* (mean)	Mean fragment length difference among allele (bp)
Di-nucleotide	100	43	5.72	0.603	0.526	3.32
Tri-nucleotide	100	35	4.17	0.445	0.393	3.16
Tetra-nucleotide	100	31	4.03	0.465	0.448	3.91
Penta-nucleotide	100	27	4.48	0.513	0.474	3.97
Hexa-nucleotide	100	24	4.75	0.537	0.537	4.97
Total	500	160	4.63	0.513	0.476	3.76

could be differentiated from each other with the using of five combinations of two markers based on at least one difference in SSR genotype (TsuAc191 and TsuAc282, TsuAc205 and TsuAc264, TsuAc229 and TsuAc292, TsuAc230 and TsuAc290, and TsuAc282 and TsuAc343), with the exception of 'Yonekura' and 'MD-2'.

Correspondence of SSR markers for pineapple genome

Each SSR marker was aligned to the pineapple genome sequence; 147 markers were aligned between LG01 and LG25 (pseudo-molecules of the pineapple genome), 11 markers were aligned to scaffolds, and 2 markers, TsuAC303 and TsuAC321, were not aligned to the pineapple genome (**Supplemental Table 6**).

Discussion

In the present study, we developed SSR markers using the Roche 454 GS FLX+ platform and used this data to identify accessions. In total, we developed 160 SSR markers with di- to hexa-nucleotide motifs, revealing their *Na*, *He*, *Ho*, and we developed minimal subsets for distinguishing 25 accessions with no distinction between the 'Yonekura' and 'MD-2' accession.

Detected SSRs from 454 GS FLX+ sequences showed lower frequencies of AT/TA motifs relative to the pineapple genome sequence (Ming *et al.* 2015). The genome sequence of pineapple included 25,367 AT/TA motif SSRs which occupied about 51% of di-nucleotide motifs, while AT/TA motifs occupied about 34% of di-nucleotide motifs in this study. This difference could have been induced by sequencing bias of 454 GS FLX+. In this study, the AT/TA motif displayed a low success rate (6 out of 24 markers, 25%) of SSR marker establishment while the AG/CT motif displayed a high success rate (33 out of 57 markers, 59%). This result might be due to PCR amplification stability, because AT rich regions tend to be difficult to amplify by PCR. Similarly, because emulsion PCR was performed during the construction of the sequencing library, amplification bias of di-nucleotide microsatellite regions might be occurred and biased the relative frequencies of the motifs.

Di-nucleotide and tri-nucleotide motif frequencies in monocotyledon species (Qin *et al.* 2015) were similar to those of pineapple. Among the di-nucleotide motifs, the AT/TA motif showed a rising trend with the increase of repeat number in plants. Frequencies of repeat motifs of monocotyledon species with 10 repeats were 30 to 55% of AT/TA (40% in pineapple), 30 to 55% of AG/GC (53% in pineapple), 10 to 20% of AC/GT (6% in pineapple), and 0 to 1% of GC/CG (0.4% in pineapple) (Qin *et al.* 2015). In dicotyledon species, 60 to 70% of AT/TA and 20 to 30% of AG/CT motif frequencies were observed. In pineapple, although AC/GT repeats had a relatively low frequency when compared to other monocotyledon species, the SSR motif frequency trend was similar to that of other monocotyledons.

Previously, SSR markers were primarily developed using di-nucleotide motifs (Carlier *et al.* 2012, Feng *et al.* 2013, Shoda *et al.* 2012, Urasaki *et al.* 2015, Wöhrmann and Weising 2011). Tri- or more nucleotide motif SSR markers often yield fewer stutter fragments, and their neighbor alleles are more easily separated from each other as compared to di-nucleotide motif SSR markers (Cipriani *et al.* 2008, Diwan and Cregan 1997). We observed a similar tendency in developed SSR markers. While di-nucleotide motifs amplified unexpected stutter fragments, tri- to hexa-nucleotide SSR markers clearly amplified the target fragments (**Supplemental Fig. 2**). In addition, SSR markers with motifs of tetra- to hexa-nucleotide repeats showed larger mean fragment length differences among alleles. These characteristics have been noted to be preferable for accurate genotyping. In the interest of establishing accurate and efficient systems for genetic identification, several key factors were considered. First, null alleles were not observed in accessions, because none of the amplification results were indistinguishable from PCR failure. We did, however, find homozygous null alleles when no amplification was observed. Specific homozygous alleles could not be distinguished from a heterozygous expression of that allele and a null allele. Cross-breeding progeny of accessions would likely occur homozygous for the null allele, and although breeding was not used in this study, we indicate nonetheless that SSR markers including null alleles should be avoided. Secondly, large differences in fragment lengths among alleles were observed due to the ease of distinguishing alleles. Third, the *Na* and *He* values were high because these markers facilitate efficient genetic identification. Therefore, we hypothesized that SSR markers with the following characteristics were suitable candidates for establishing an efficient genetic identification system: 1. Tri- or more nucleotide motifs; 2. Null alleles not observed in accessions; 3. High *Na* values (more than 4); 4. High *Ho* values (more than 0.5); and 5. Large differences in fragment lengths among alleles (mean amplified fragment distance greater than 4.0). Based on the criteria above, 16 of the SSR markers were selected as optimal SSR markers, including two tri-nucleotide motifs (TsuAC235 and TsuAc244), 3 tetra-nucleotide motifs (TsuAc269, TsuAc278, and TsuAc284), 5 penta-nucleotide motifs (TsuAc299, TsuAc300, TsuAC313, TsuAc317, and TsuAc319) and 6 hexa-nucleotide motifs (TsuAc334, TsuAc335, TsuAc336, TsuAc341, TsuAc342, and TsuAc346). We confirmed that the 25 examined accessions were successfully differentiated based on 17 combinations of 3 optimal SSR markers based on at least one difference in SSR genotype (e.g., TsuAC244, TsuAC269, and TsuAC319, **Supplemental Table 8**), with the exception of 'Yonekura' and 'MD-2'. Because the three SSR markers amplified different ranges of PCR fragments, multiplexed PCR analyses of the markers could efficiently identify the accessions simultaneously.

SSR marker generally exhibits the polymorphism due to the difference of the number of repeat motifs, and their

mean fragment length difference among alleles is thus expected to be the same or larger than that of the repeat motif length. However, SSR markers with tetra- to hexa-nucleotide repeats displayed smaller fragment length differences among alleles than their repeat motif length. This suggests that both repeat number and small indels around SSRs should be involved in fragment length polymorphisms of SSR markers. In addition, in this study established SSR markers tended to have low repeat numbers. High repeat numbers of SSR markers tended to show high mean fragment length differences among alleles in di- and tri-nucleotide motifs. This is likely because larger repeat number SSRs frequently contain repeat number mutations, and the possibility of including alleles with two or more differences in repeat numbers would be increased if this were the case. Since SSRs in 454 GS FLX+ sequences were more abundant in smaller numbers of repeats, randomly selected SSRs for PCR primer design tended to be low in repeat number. Low mean fragment length differences among alleles in tetra- or greater nucleotide motifs could be attributed to the low repeat numbers of SSRs. To develop optimal SSR markers for genetic identification, higher repeat numbers would be preferable.

The 'Yonekura' and 'MD-2' pineapple cultivars could not be distinguished in the present study due to the identical genotypes among all of the SSR markers. The derivation of 'Yonekura' has not yet been clearly defined, but 'Yonekura' has nonetheless been used in cross-breeding trials at OPARC in Japan, and the breeding population derived from 'Yonekura' has been used to develop DNA markers for leaf margin phenotype (Urasaki *et al.* 2015). The cultivar 'MD-2' was developed by Del Monte Fresh Produce International and has gained a major market share, particularly on the international flesh fruit market (Chan *et al.* 2003). Breeding selection in pineapples is largely conducted through clonal selection, in which superior individual plants are selected from the field and eventually multiplied as new clones (Chan *et al.* 2003). Considering the genetic similarity between 'MD-2' and 'Yonekura', it is probable that a commercial grower (or growers) selected superior individuals in an 'MD-2' field and then named the resulting clone 'Yonekura'. Additionally, qualitative traits (leaf margin phenotype and fruit color) were the same and quantitative traits (fruit weight, harvest day, soluble solid content and acidity) were very similar between 'MD-2' and 'Yonekura' (data not shown). We therefore indicate that 'Yonekura' is a clonal selection of 'MD-2'.

Developed SSR markers are also useful for the construction of genetic linkage maps and QTL analyses. Previous genetic linkage maps were primarily constructed using dominant markers derived from DNA fingerprinting techniques such as AFLP, RAPD, and ISSR (Carlier *et al.* 2004, 2006, 2012, Sousa *et al.* 2013). These markers are associated with limitations including limited reproducibility and difficulty in application to other mapping populations (Edwards and McCouch 2007). Because SSR markers provide reliable and informative genotyping data due to their co-dominance, our novel SSR markers will aid in the construction of informative and more widely applicable genetic linkage maps. Map coverage and marker density can be improved through genotyping-by-sequencing (GBS) analyses, including restriction site-associated DNA sequencing (RAD-Seq) (Davey *et al.* 2011). High-density genetic linkage maps have been constructed for several crops using the RAD-Seq technique with SSR markers (Shirasawa *et al.* 2017, Yagi *et al.* 2017). Similar genetic linkage maps for pineapple would be applicable in QTL analyses, and effective QTLs could be detected for MAS in breeding programs. Although QTL analysis could be performed using non-dense-genetic linkage maps, Stange *et al.* (2013) suggested that high-density genetic linkage maps can improve the precision of QTL localization and the resolution of linked QTL, which enables the localization of two linked QTL separately.

In the present study, at least one SSR marker was aligned to each pseudo molecule of the pineapple genome from LG01 to LG25 (**Supplemental Fig. 1**, **Supplemental Table 6**). Such SSR markers would aid in the construction of dense genetic linkage maps by accurately identifying genotypes.

Author Contribution Statement

KN, FH, ST, MK, CN and TY performed NGS data processing and SSR marker analysis. CM, MT and MS Maintained pineapple accessions and provided information for each accession. KT and NU performed NGS analysis.

Acknowledgments

This study was supported by an Okinawa special promotion grant.

Literature Cited

Brownstein, M.J., J.D. Carpten and J.R. Smith (1996) Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004–1010.

Carlier, J.D., A. Reis, M.F. Duval, G. Coppens d' Eeckenbrugge and J.M. Leitão (2004) Genetic maps of RAPD, AFLP and ISSR markers in *Ananas bracteatus* and *A. comosus* using the pseudo-testcross strategy. Plant Breed. 123: 186–192.

Carlier, J.D., D. Nacheva, G. Coppens d' Eeckenbrugge and J.M. Leitao (2006) Genetic mapping of DNA markers in pineapple. Acta Hortic. 702: 79–86.

Carlier, J.D., N.H. Sousa, T.E. Santo, G. Coppens d' Eeckenbrugge and J.M. Leitão (2012) A genetic map of pineapple (*Ananas comosus* (L.) Merr.) including SCAR, CAPS, SSR and EST-SSR markers. Mol. Breed. 29: 245–260.

Chan, Y.K., G. Coppens d' Eeckenbrugge and G.M. Sanewski (2003) Breeding and variety improvement. *In*: Baryholomew, D.P., R.E. Paull and K.G. Rohrbach (eds.) The pineapple. Botany, production and uses, CAB International, Wallingford, pp. 33–55.

Cipriani, G., M.T. Marrazzo, G. Di Gaspero, A. Pfeiffer, M. Morgante

and R. Testolin (2008) A set of microsatellite markers with long core repeat optimized for grape (*Vitis* spp.) genotyping. BMC Plant Biol. 8: 127.

Coppens d' Eeckenbrugge, G., M.F. Duval and F. Leal (2018) The pineapple success story: from domestication to pantropical diffusion. *In*: Ming, R. (ed.) Genetics and Genomics of pineapple, Springer International Publishing, Cham, pp. 1–25.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen and M.L. Blaxter (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12: 499–510.

Diwan, N. and P.B. Cregan (1997) Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. 95: 723–733.

Duval, M.F., J.L. Noyer, X. Perrier, G. Coppens d' Eeckenbrugge and P. Hamon (2001) Molecular diversity in pineapple assessed by RFLP markers. Theor. Appl. Genet. 102: 83–90.

Duval, M.F., G.S. Buso, F.R. Ferreira, J.L. Noyer, G. Coppens d' Eeckenbrugge, P. Hamon and M.E. Ferreira (2003) Relationships in *Ananas* and other related genera using chloroplast DNA restriction site variation. Genome 46: 990–1004.

Edwards, J.D. and S.R. McCouch (2007) Molecular markers for use in plant molecular breeding and germplasm evaluation. *In*: Guimaraes, E.P., J. Ruane, B.D. Scherf, A. Sonnino and J.D. Dargie (eds.) Marker-assisted selection: Current status and future perspectives in crops, livestock, forestry and fish, Food and agriculture organization of the united nations, Rome, pp. 29–49.

Ellegren, H. (2004) Microsatellites: simple sequences with complex evolution. Nat. Rev. Genet. 5: 435–445.

Feng, S., H. Tong, Y. Chen, J. Wang, Y. Chen, G. Sun, J. He and Y. Wu (2013) Development of pineapple microsatellite markers and germplasm genetic diversity analysis. Biomed Res. Int. 2013: 317912.

Fujii, H., H. Yamashita, T. Shimada, T. Endo, T. Shimizu and T. Yamamoto (2008) MarkerToolKit: an analysis program for data sets consist of DNA marker types obtained from various varieties. DNA Polymorph. 16: 103–107 (In Japanese).

Fujii, H., T. Ogata, T. Shimada, T. Endo, H. Iketani, T. Shimizu, T. Yamamoto and M. Omura (2013) Minimal marker: an algorithm and computer program for the identification of minimal sets of discriminating DNA markers for efficient variety identification. J. Bioinform. Comput. Biol. 11: 1250022.

Harker, N. (2001) Collection, reporting and storage of microsatellite genotype data. *In*: Henry, R.J. (ed.) Plant genotyping: the DNA fingerprinting of plants, CAB International, Wallingford, pp. 251–264.

Kato, C.Y., C. Nagai, P.H. Moore, F. Zee, M.S. Kim, D.L. Steiger and R. Ming (2004) Intra-specific DNA polymorphism in pineapple (*Ananas comosus* (L.) Merr.) assessed by AFLP markers. Genet. Resour. Crop Evol. 51: 815–825.

Leitao, J.M. (2018) Genetic mapping in pineapple. *In*: Ming, R. (ed.) Genetics and genomics of pineapple, Springer International Publishing, Cham, pp. 61–68.

Lin, Y.S., C.S. Kuan, I.S. Weng and C.C. Tsai (2015) Cultivar identification and genetic relationship of pineapple (*Ananas comosus*) cultivars using SSR markers. Genet. Mol. Res. 14: 15035–15043.

Litt, M., X. Hauge and V. Sharma (1993) Shadow bands seen when typing polymorphic dinucleotide repeats: some causes and cures. BioTechniques 15: 280–284.

Ming, R., R. VanBuren, C.M. Wai, H. Tang, M.C. Schatz, J.E. Bowers, E. Lyons, M.L. Wang, J. Chen, E. Biggers *et al.* (2015) The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. 47: 1435–1442.

Morton, J. (1987) Pineapple. *In*: Morton, J.F. and F.L. Miami (eds.) Fruit of Warm Climates, Creative Resource Systems, Winterville, pp. 18–28.

Ogata, T., S. Yamanaka, M. Shoda, N. Urasaki and T. Yamamoto (2016) Current status of tropical fruit breeding and genetics for three tropical fruit species cultivated in Japan: pineapple, mango, and papaya. Breed. Sci. 66: 69–81.

Paz, E.Y., K. Gil, L. Rebolledo, A. Rebolledo, D. Uriza, O. Martínez, M. Isidrón and J. Simpson (2005) AFLP characterization of the Mexican pineapple germplasm collection. J. Am. Soc. Hortic. Sci. 130: 575–579.

Qin, Z., Y. Wang, Q. Wang, A. Li, F. Hou and L. Zhang (2015) Evolution analysis of simple sequence repeats in plant genome. PLoS ONE 10: e0144108.

Ruas, C.F., P.M. Ruas and J.R.S. Cabral (2001) Assessment of genetic relatedness of the genera *Ananas* and *Pseudananas* confirmed by RAPD markers. Euphytica 119: 245–252.

Ruas, P.M., C.F. Ruas, D.J. Fairbanks, W.R. Andersen and J.S. Cabral (1995) Genetic relationship among four varieties of pineapple, *Ananas comosus*, revealed by random amplified polymorphic DNA (RAPD) analysis. Braz. J. Genet. 18: 413–416.

Shirasawa, K., K. Isuzugawa, M. Ikenaga, Y. Saito, T. Yamamoto, H. Hirakawa and S. Isobe (2017) The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding. DNA Res. 24: 499–508.

Shoda, M., N. Urasaki, S. Sakiyama, S. Terakami, F. Hosaka, N. Shigeta, C. Nishitani and T. Yamamoto (2012) DNA profiling of pineapple cultivars in Japan discriminated by SSR markers. Breed. Sci. 62: 352–359.

Sousa, N., J. Carlier, T. Santo and J. Leitao (2013) An integrated genetic map of pineapple (*Ananas comosus* (L.) Merr.). Sci. Hortic. 157: 113–118.

Stange, M., H.F. Utz, T.A. Schrag, A.E. Melchinger and T. Würschum (2013) High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. Theor. Appl. Genet. 126: 2563–2574.

Thiel, T., W. Michalek, R.K. Varshney and A. Graner (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106: 411–422.

Urasaki, N., S. Goeku, R. Kaneshima, T. Takamine, K. Tarora, M. Takeuchi, C. Moromizato, K. Yonamine, F. Hosaka, S. Terakami *et al.* (2015) Leaf margin phenotype-specific restriction-site-associated DNA-derived markers for pineapple (*Ananas comosus* L.). Breed. Sci. 65: 276–284.

Wang, J.S., J.H. He, H.R. Chen, Y.Y. Chen and F. Qiao (2017) Genetic diversity in various accessions of pineapple [*Ananas comosus* (L.) Merr.] using ISSR and SSR markers. Biochem. Genet. 55: 347–366.

Wöhrmann, T. and K. Weising (2011) In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. Theor. Appl. Genet. 123: 635–647.

Yagi, M., K. Shirasawa, T. Waki, T. Kume, S. Isobe, K. Tanase and H. Yamaguchi (2017) Construction of an SSR and RAD marker-based genetic linkage map for carnation (*Dianthus caryophyllus* L.). Plant Mol. Biol. Rep. 35: 110–117.

You, F.M., N. Huo, Y.Q. Gu, M.C. Luo, Y. Ma, D. Hane, G.R. Lazo, J. Dvorak and O.D. Anderson (2008) BatchPrimer3: a high throughput web application for PCR and sequencing primer design. BMC Bioinformatics 9: 253.

Zalapa, J.E., H. Cuevas, H. Zhu, S. Steffan, D. Senalik, E. Zeldin, B. McCown, R. Harbut and P. Simon (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am. J. Bot. 99: 193–208.