Developing single nucleotide polymorphism markers for the identification of pineapple (Ananas comosus) germplasm

Lin Zhou, Tracie Matsumoto, Hua-Wei Tan, Lyndel W Meinhardt, Sue Mischke, Boyi Wang and Dapeng Zhang

Pineapple (Ananas comosus [L.] Merr.) is the third most important tropical fruit in the world after banana and mango. As a crop with vegetative propagation, genetic redundancy is a major challenge for efficient genebank management and in breeding. Using expressed sequence tag and nucleotide sequences from public databases, we developed 213 single nucleotide polymorphism (SNP) markers and validated 96 SNPs by genotyping the United States Department of Agriculture - Agricultural Research Service pineapple germplasm collection, maintained in Hilo, Hawaii. The validation resulted in designation of a set of 57 polymorphic SNP markers that revealed a high rate of duplicates in this pineapple collection. Twenty-four groups of duplicates were detected, encompassing 130 of the total 170 A comosus accessions. The results show that somatic mutation has been the main source of intra-cultivar variations in pineapple. Multivariate clustering and a model-based population stratification suggest that the modern pineapple cultivars are comprised of progenies that are derived from different wild Ananas botanical varieties. Parentage analysis further revealed that both A. comosus var. bracteatus and A. comosus var. ananassoides are likely progenitors of pineapple cultivars. However, the traditional classification of cultivated pineapple into horticultural groups (e.g. ‘Cayenne’, ‘Spanish’, ‘Queen’) was not well supported by the present study. These SNP markers provide robust and universally comparable DNA fingerprints; thus, they can serve as an efficient genotyping tool to assist pineapple germplasm management, propagation of planting material, and pineapple cultivar protection. The high rate of genetic redundancy detected in this pineapple collection suggests the potential impact of applying this technology on other clonally propagated perennial crops.

Horticulture Research (2015) 2, 15056; doi:10.1038/hortres.2015.56; published online 25 November 2015

INTRODUCTION

Pineapple, Ananas comosus (L.) Merr., is a perennial herbaceous fruit crop belonging to the family Bromeliaceae. The crop is cultivated in all tropical and subtropical regions and ranks third in production among noncitrus tropical fruits, following banana (including plantain) and mango. The annual worldwide production reached 21.9 million metric tons in 2012 and the top seven producers (Brazil, Philippines, Thailand, Costa Rica, Indonesia, India, and China) jointly accounted for 90% of the global production (FAO, 2014). The pineapple plant is indigenous to South America.4,5 The putative center of origin is located in the Paraná–Paraguay River drainages between southern Brazil and Paraguay, based on the diversity distribution of related species and botanical varieties of pineapple in this region.4,6 However, the eastern part of the Guiana shield has also been hypothesized as the center of domestication for pineapple, based on the variation of chloroplast and nuclear DNA markers, the high level of phenotypic diversity, and the large number of primitive cultigens in this area.7,8 Pineapple was widely cultivated in tropical Americas before the arrival of Christopher Columbus, the first European to see this fruit, in 1493.9

The introduction of pineapple into Asia and the Pacific began with the Spanish in the early sixteenth century and pineapple reached Africa in mid-sixteenth century.5 Since then, there have been multiple introductions and exchanges of germplasm among the pineapple producing countries.5 Although many landraces and traditional cultivars exist in the Americas, only a few cultivars have been displaced to Asia and Africa for use in commercial production.4,10 About 70% of the world’s production comes from a single cultivar, Smooth Cayenne,10 which is a highly productive pineapple excellent for canning.11 The current fresh fruit pineapple market is largely comprised of two cultivars bred by the Pineapple Research Institute, CO-2 and MD-2.11 Developing new cultivars with desirable resistance and postharvest traits will depend on the available germplasm of this species. The United States Department of Agriculture (USDA) - Agricultural Research Service pineapple germplasm collection in Hilo, Hawaii, is one of the major collections in the world, along with the collections maintained by EMBRAPA/CNPMF in Cruz das Almas, Brazil, and by CIRAD-FLHOR in Martinique. As part of theARS National Clonal Germplasm Repository for Tropical and Subtropical Fruit and Nut Crops, the collection at Hilo currently maintains over 180 accessions of pineapple cultivars and their wild relatives.

As with many other tropical perennial crops, pineapple germplasm is almost exclusively maintained by vegetative propagation, by crowns, slips, suckers, or in vitro culture. Vegetative propagation has allowed the exchange of germplasm as clones among regions, countries, and continents. However, the exchange of vegetative planting materials has also resulted in problems for conservators of pineapple germplasm because records and labels of the cultivars have not always followed the same naming conventions, and accessions have limited information about their correct identity. Therefore, homonyms and synonyms are common among the names of pineapple cultivars and that restricts the sharing of information and materials among pineapple researchers and

1Sustainable Perennial Crops Laboratory, USDA-ARS, Beltsville Agricultural Research Center, Beltsville, MD 20705, USA; 2Daniel K. Inouye Pacific Basin Agricultural Research Center, USDA-ARS, Hilo, HI 96720, USA; 3College of Horticulture, Nanjing Agricultural University, Nanjing 210095, Jiangsu, China and Yunnan Forestry Technological College, Kunming 650224, Yunnan, China
Correspondence: DP Zhang (Dapeng.Zhang@ars.usda.gov)
Received 30 September 2015; revised 24 October 2015; accepted 25 October 2015
hampers the use of pineapple germplasm in breeding. Another major challenge for pineapple cultivar identification is that the protracted vegetative propagation has led to the accumulation of somatic mutations. Some mutations caused noticeable phenotypic effects and created intra-cultivar variation, which became the target of clonal selection. While these selected mutants are important in horticultural production, it is necessary to identify them so that breeders and genebank curators can efficiently conserve and use these genetic materials.

The utilization of biochemical and DNA molecular markers for pineapple germplasm management has been recently reviewed. Using isozyme markers, Aradhya et al. studied pineapple germplasm and found considerable variation within and between species of Ananas. In the Hawaii pineapple collection, they identified 66 distinctzymotypes that were able to differentiate all species and botanical varieties. Their results also suggested that, rather than genetic divergence due to reproductive isolating barriers, differentiation among the species of Ananas may be due to ecological isolation, and therefore may represent a species complex.

Both dominant DNA markers (amplified fragment length polymorphism, AFLP) and co-dominant markers (restriction fragment length Polymorphism simple sequence repeat, SSR) have been used to assist pineapple cultivar identification and germplasm management. In spite of the significant progress in marker-assisted germplasm management over the last 20 years, cultivar identification in pineapple remains a challenging task. Using AFLP markers, Kato et al. characterized 148 A. comosus accessions maintained in the USDA pineapple collection in Hilo, Hawaii. They showed that a unique profile for major groups that had been classified by morphological traits, such as 'Cayenne', 'Spanish', and 'Queen', could not be established using AFLP-based DNA fingerprints. SSR markers likewise lacked congruence between phenotype and molecular marker-based classification in pineapple. Moreover, neither AFLP nor SSR are the most suitable marker tool for detection of duplicates in the pineapple germplasm.

Single nucleotide polymorphisms (SNPs) are the most abundant class of polymorphisms in plant genomes. Compared to SSR markers, SNP analysis can be done without requiring DNA separation by size and can, therefore, be automated in high-throughput assay formats. The diallelic nature of SNPs facilitates a much lower error rate in allele calling and promotes compatibility between laboratories. These advantages have resulted in the increasing use of SNPs as the markers of choice for accurate genotype identification and diversity analysis in perennial crops, as recently demonstrated in Citrus maxima, grapevine (Vitis vinifera), pummelo (Citrus maxima), strawberry (Fragaria spp), tea (Camellia sinensis), and longan (Dimocarpus longan). Like other perennial horticulture crops, DNA fingerprinting that uses a small set of SNP markers is in great demand by the pineapple community for a broad range of research and field applications. These applications include, but are not limited to, identification of mislabeled accessions, parentage, and sibship analysis for quality control in breeding and seeds programs, and authentication and traceability to support the production of high-value clones for premium markets. Nonetheless, this most powerful tool for germplasm management has not been applied to pineapple germplasm management.

Ample genomic resources have been developed for pineapple. The premier online database, 'PineappleDB' (http://genet.imb.uq.edu.au/Pineapple/index.html), includes more than 5,600 expressed sequence tags (ESTs) with 3,383 consensus sequences. The comprehensive sequence, bioinformatics, and functional classification of EST resources are available for text or sequence-based searches. A draft genome of A. comosus has been developed, which covers about 375 Mb (62%) of the estimated 526 Mb genome of this species. These readily available genomic resources provide opportunities for mining new markers to use for pineapple germplasm management and breeding. The objectives of the present study were to develop SNP markers through the data mining of ESTs and transcriptome data and to assess their potential application for pineapple cultivar identification. The results reported herein represent the first validation study of SNPs in pineapple, demonstrating the utility of a transcriptome as an approach for rapid development of a high-quality genotyping tool. These SNP markers, as well as the genotyping method, will be particularly useful for intellectual property rights in varietal protection, germplasm management, and pineapple breeding programs.

MATERIALS AND METHODS

Mining of putative SNPs from EST and nucleotide sequences

All available nucleotide sequences of Ananas spp. were downloaded from NCBI GenBank (http://www.ncbi.nlm.nih.gov, 4 October 2014). Redundant entries were examined and excluded using the CD-HIT program with a 95% sequence similarity threshold. The FASTA-formatted files of pineapple were manipulated into a single data set for further data mining. A total of 26 EST-SNPs were detected using the QualitySNP program. All of these selected clusters included a minimum of six EST sequences, whereas both the minimum redundancy threshold and minimal confidence score required by QualitySNP was set at three. In order to meet the requirements and constraints for primer design, all candidates for SNP markers with less than 50 nucleotides between two neighboring SNPs were removed. A subset of 96 identified SNP sequences was then chosen for design and manufacture of SNP assay.

Validation of putative SNPs

To evaluate the putative SNP markers for suitability of cultivar identification, we used a nanofluidic genotyping system and validated the SNPs for 170 pineapple accessions (Table 1; Supplementary Table S1). The pineapple germplasm samples were from the pineapple collection maintained by the USDA-ARS Tropical Plant Genetic Resources and Disease Unit, at the National Plant Germplasm Repository in Hilo, Hawaii (http://www.ars.usda.gov/main/site_main.htm?modecode=20-40-05-10) and dried in silica gel. DNA was extracted from dried pineapple leaves with the DNeasy Plant Mini kit (Qiagen Inc., Valencia, CA, USA), which is based on the use of silica as an affinity matrix. The dry leaf tissue was placed in a 2-mL microcentrifuge tube with one quarter-inch ceramic sphere and 0.15 g garnet matrix (Lysing Matrix A; MP Biomedicals. Solon, OH, USA). The leaf samples were disrupted by high-speed shaking in a TissueLyser II (Qiagen Inc.) at 30 Hz for 1 min. Lysis solution (DNeasy kit buffer AP1 containing 25 mg/mL polyvinylpyrrolidone), along with RNase A, was added to the powdered leaf samples and the mixture was incubated at 65°C, as specified in the kit instructions. The remainder of the extraction method followed the manufacturer's suggestions. DNA was eluted from the silica column with two washes of 50 μL buffer AE, which were pooled, resulting in 100 μL DNA solution. Using a NanoDrop spectrophotometer (Thermo Scientific, Waltham, MA, USA), DNA concentration was determined by absorbance at 260 nm. DNA purity was estimated by the 260:280 ratio and the 260:230 ratio.

Ninety-six putative SNP sequences were submitted to the Assay Design Group at Fluidigm Corp. (South San Francisco, CA, USA) for design and manufacture of primers for a SNPtype genotyping panel. The assays were based on competitive allele-specific PCR, and they enable bi-allelic scoring of SNPs at specific loci (KBioscience Ltd, Huddesdon, UK).

Table 1. Species categories of 170 A. comosus accessions in USDA-ARS pineapple collection at Hilo, Hawaii.

| Species | No. of accessions |
|---------|-------------------|
| A. comosus var. comosus | 145 |
| A. comosus var. ananasoides | 6 |
| A. comosus var. bracteatus | 8 |
| A. comosus var. erectofolius | 2 |
| A. comosus hybrid | 5 |
| Ananas species | 4 |
| Total | 170 |
The Fluidigm SNPtype Genotyping Reagent Kit was used according to the manufacturer’s instructions.24,25 Using these primers, the isolated DNAs were subjected to Specific Target Amplification in order to enrich the SNP sequences of interest.24 Genotyping was performed on a nanofluidic 96.96 Dynamic Array IFC (Integrated Fluidic Circuit; Fluidigm Corp.). This chip automatically assembles PCRs, enabling simultaneous testing of up to 96 samples with 96 SNP markers. The use of a 96.96 Dynamic Array IFC for SNP genotyping of human samples has been described by Wang et al.13 End-point fluorescent images of the 96.96 IFC were acquired on an EP1 imager (Fluidigm Corp.). The data were analyzed with Fluidigm Genotyping Analysis Software (Fluidigm Corp.).26

Data analysis
Duplicate accessions were identified using pairwise multilocus matching among all individual samples. DNA samples that were fully matched at the genotyped SNP loci were declared the same cultivar or clone. The program GenAlEx 6.5 (2006, 2012) was used for computation.27,28 After duplicate identification, the redundant samples were removed and descriptive statistics for measuring the informativeness of the SNP markers were calculated based on the remaining distinctive cultivars. The key descriptive statistics included minor allele frequency, observed heterozygosity, expected heterozygosity, Shannon’s information index, and inbreeding coefficient. Computations were carried out using the same program. A cluster analysis using the neighbor-joining (NJ) method was used to further examine the genetic relationship among accessions. Kinship coefficient was chosen as genetic distance measurement of shared ancestry among the individual accessions. The computation was executed using MICROSATellite ANALYZER (MSA, 2003).29 A dendrogram was generated from the resulting distance matrix using the NJ algorithm available in PHYLIP.30,31 The unrooted tree was visualized using the web-based tool Interactive Tree of Life v2 (http://itol.embl.de/).32

A model-based clustering algorithm implemented in the STRUCTURE software program was applied to the SNP data.33 This algorithm attempted to identify genetically distinct subpopulations based on allele frequencies. The admixture model was applied and the number of clusters (K-value) indicating the number of subpopulations the program attempted to find, was set from 1 to 10. The analyses were carried out without assuming any prior information about the genetic group or geographic origin of the samples. Ten independent runs were assessed for each fixed number of clusters (K), each consisting of 1 × 10⁶ iterations after a burn-in of 2 × 10⁵ iterations. The Ak value was used to detect the most probable number of clusters and the computation was performed using the online program STRUCTURE HARVESTER.34 Of the 10 independent runs, the one with the highest Ln Pr (X|K) value (log probability or log likelihood) was chosen and represented as bar plots.

To test the hypothesis that vars. *ananassoides*, *bracteatus*, and *erectifolius* are the putative progenitors for cultivated pineapples, we applied parentage analyses for which parentage analyses were carried out. A. *comosus* vars. *comosus* (as labeled in Table 1). These cultivars or breeding lines were considered as ‘offspring’ for which parentage analyses were carried out. A. *comosus* vars. *ananassoides*, *bracteatus*, and *erectifolius* were used as candidate parents. A likelihood-based method implemented in the program CERVUS 3.0 was used for computation.45,46 For each parent–offspring pair, the natural log-arithm of the likelihood ratio (LOD score) was calculated. Critical LOD scores were determined for the assignment of parentage to a group of individuals without knowing the maternity or paternity. Simulations were run for 10 000 cycles, assuming that 10% of candidate parents were sampled, a total of 90% of loci was typed with a 1% typing error rate. The most probable single mother (or father) for each offspring was identified on the basis of the critical difference in LOD scores (D) between the most likely and the next most likely candidate parent at greater than 95 or 80% confidence.45,46

RESULTS

SNP discovery
A total of 13 203 mRNA nucleotide and 5941 EST sequences from pineapple were gathered using methods previously described. After adapter removal, trimming, and quality control, 18 241 higher quality sequences were selected. The program CAP3,47 using default parameters, was used to assemble sequences into 1793 contigs and 11 809 singlets with an average size of 3.59 sequences per contig, among which putative SNPs were detected in only 48 contigs using the QualitySNP program. Each of these selected clusters included a minimum of six EST sequences. In total, we obtained 213 putative SNPs, including 75C/T, 59A/G, 10A/T, 12A/C, 47G, 11C/G, 41 indel, and 1 high tri-allelic polymorphism. To select high-quality SNPs for validation, candidate SNP sites with at least 50 bp before and after the site were filtered. We calculated the number of all sequences in a cluster and the number containing the SNP type in this cluster. We then selected 96 SNPs for validation by genotyping the 170 pineapple accessions in the USDA-ARS pineapple collection.

Screening for polymorphic SNP markers
Out of the chosen 96 SNP markers, 80 were successful for genotyping. The failure of the remaining 16 SNPs was likely due to the sequence complexity or the presence of polymorphisms within the flanking sequences. However, among the 80 successful SNPs, 23 were monomorphic across the 170 pineapple samples (i.e. only one SNP variant was identified in all individuals). These monomorphic markers may have resulted from errors in transcriptome sequencing, which then led to the incorrect identification of SNP. It is also possible that some of these SNPs may correspond to rare alleles that were not present in the set of pineapple accessions we analyzed. A total of 57 polymorphic SNPs were retained for further analysis of this sample set. These 57 SNPs were reliably scored across the validation panel and thus were considered true SNPs. The flanking sequences of these 57 SNPs are listed in Table 2.

Cultivar identification
SNP profiles of the multiple accessions from the same pineapple cultivar showed that genotyping results were highly consistent (Table 3). Multilocus matching of SNP fingerprints revealed a high rate of duplicates in this pineapple collection. A total of 130 accessions could be classified into 24 synonymous groups (Table 4). The largest synonymous group, which includes 36 accessions, was found in cultivar Cayenne. It is also noticeable that some accessions within the same synonymous group have apparent morphological differences, despite matching SNP profiles, indicating somaclonal mutation within the synonymous group. For example, Cayenne 7898 QC has atypical yellow flesh color, whereas Cayenne 7898 4N has a white color, but their SNP profiles are the same (Figure 1).

Descriptive statistics and clustering analysis of 64 distinctive pineapple accessions
From each of the synonymous groups, only one accession was retained and used for subsequent diversity analysis. Among the 170 genotyped accessions, there were 64 accessions with a unique SNP profile. Descriptive statistics were then computed for the 57 polymorphic SNPs across the 64 pineapple accessions with a unique SNP profile and the result is presented in Table 5. The minor allele frequencies of these 57 SNPs ranged from 0.090 to 0.495 with an average of 0.324. The mean information index was 0.601, ranging from 0.304 to 0.693. The observed heterozygosity ranged from 0.110 to 0.935 with an average of 0.520, whereas the mean expected heterozygosity was 0.414 ranging from 0.164 to 0.500 (Table 5).

The unrooted NJ tree grouped the 64 accessions into three main clusters (Figure 2). The clustering patterns presented relationships among accessions based on the different botanical varieties or origins from different geographical regions. The first cluster includes all the accessions of *A. comosus* vars. *ananassoides*, *bracteatus*, and *erectifolius*, as well as the hybrids derived from these related botanical varieties. Within this cluster, vars. *ananassoides* *bracteatus* and *erectifolius* are clearly separated. This cluster also included several cultivated pineapple clones, such as Bogota, Pina Lisa, and Criolla from Colombia, Bermuda from Barbados, Cayenne Lot 520 from Hawaii, Cabezon from Puerto Rico, and Trinidad from Trinidad. The proximity between these cultivars and the two related botanical varieties indicates that these cultivars are either selected...
Molecular characterization of pineapple germplasm using SNP

L Zhou et al

Table 2. Flanking sequences and SNPs of the 57 polymorphic markers.

| SNP ID | Flanking sequences and SNPs of the 57 polymorphic markers. |
|--------|----------------------------------------------------------|
| Ac4    | TAAGCCGAGCGCGGTTAATGGACCCCTGTAGGATGCTGCGGCAACTATAACCTGGAATGTGTTTGGGCCTCT |
| Ac5    | AAGGAGCAAGGACCTCTGCGATAGCTGGGTACTGGTTCGAGAAGATCGACTAATCGGCTCTATAGGGAAAGGAGGACTTTGAGG |
| Ac10   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac13   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac15   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac18   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac21   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac23   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac25   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac27   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac30   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac33   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac35   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac37   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac39   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac41   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |
| Ac43   | CAAAATTTGGGAAAGAATTGAAAGAATGTAATGATCTTTTTCATGAGGTGACTATGAGGTTTTCACTGATGAGGTTTTGGATGG |

or derived from var. ananassoides and bracteatus. The two Bolivian accessions (N94-92 Short Fruit1 and N94-92 Long Fruit2) were labeled as Ananas species in their passport record data. The cluster result showed that they should be A. comosus var. ananassoides or hybrids derived from A. comosus var. ananassoides.

The second cluster comprised of exclusively A. comosus var. comosus, including several well-known cultivars such as Cayenne Hilo, Mauritius, and Antigua. Since these three cultivars represent the reference horticultural groups of 'Cayenne' (Cayenne Hilo) and 'Queen' (Mauritius and Antigua), respectively, their grouping here...
Table 3. Examples of DNA fingerprints based on the array of 57 SNPs for pineapple genotype identification (showing truncated profiles).

| Ac20 | Ac30 | Ac41 | Ac59 | Ac84 | Ac89 | Ac15 | Ac48 | Ac50 | Ac53 | Ac56 | Ac64 | Ac65 | Ac78 | Ac92 | Ac20 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
|      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |      |

PERNAMBUCO Singapore
ABACAXI Brazil
SUGAR LOAF Zaire
BERMUDA Barbados
MONTE LIRIO Guatemala
CAMBRAY Philippines
MO REZENDE Brazil

Table 4. List of 24 synonymous groups, including 130 accessions, identified by SNP markers in USDA pineapple collection, Hilo, Hawaii.

| Grp | Code | Name |
|-----|------|------|
| 1   | HANA 10 | Cayenne Hilo Hawaii, USA |
| 1   | HANA 100 | Cayenne M95 4N Hawaii, USA |
| 1   | HANA 111 | Cayenne M111 Seedy Fruit Hawaii, USA |
| 1   | HANA 112 | Cayenne Paper Leaf Hawaii, USA |
| 1   | HANA 114 | Cayenne Bottleneck Hawaii, USA |
| 1   | HANA 115 | Cayenne M236 Nubby Hawaii, USA |
| 1   | HANA 119 | Cayenne M267 Dry Sweet Hawaii, USA |
| 1   | HANA 133 | Kew Philippines |
| 1   | HANA 92 | Cayenne M73 Hawaii, USA |
| 1   | HANA 94 | Cayenne Clone 9 Hawaii, USA |
| 1   | HANA 95 | Cayenne 1069 Hawaii, USA |
| 1   | HANA 99 | Cayenne M34 4N Hawaii, USA |
| 1   | N90-92 | N90-92 Hawaii, USA |
| 1   | HANA 179 | 3621 M03-23 Hawaii, USA |
| 1   | HANA 162 | Cayenne John Teves Hawaii, USA |
| 1   | HANA 101 | Cayenne M4W Hawaii, USA |
| 1   | HANA 167 | 32419 N91-15 Florida, USA |
| 1   | HANA 103 | Cayenne M61 Low Bloom Hawaii, USA |
| 1   | HANA 104 | Cayenne M63 Plus Bloom Hawaii, USA |
| 1   | HANA 110 | Cayenne M109-5 Hawaii, USA |
| 1   | HANA 117 | Cayenne M35 Hawaii, USA |
| 1   | HANA 123 | Red Spanish Balboa, Panama |
| 1   | HANA 170 | N91-34 Hawaii, USA |
| 1   | HANA 178 | Chimpanze M03-22 Hawaii, USA |
| 1   | HANA 93 | Cayenne M66 Hawaii, USA |
| 1   | HANA 102 | Cayenne M24 Hawaii, USA |
| 1   | HANA 105 | Cayenne M91 Big Eye Type Hawaii, USA |
|     |       | Big Eye #3 Hawaii, USA |
| 1   | HANA 106 | Cayenne M92 Big Eye Johnson Hawaii, USA |
| 1   | HANA 107 | Cayenne M106 Big Eye Excess Hawaii, USA |
| 1   | HANA 108 | Cayenne Seedy #24 Hawaii, USA |
| 1   | HANA 116 | Cayenne CPC Big Eye Hawaii, USA |
| 1   | HANA 139 | Cayenne Azores Azores, Portugal |
| 1   | HANA 163 | N91-05 Chiang Rai, Thailand |
| 1   | HANA 180 | 153 N03-24 Hawaii, USA |
| 1   | HANA 98 | Cayenne 45 #4N Hawaii, USA |
| 1   | HANA 176 | N00-10 Dole Cayenne Hawaii, USA |
| 2   | HANA 53 | British Samoa P1 British Samoa |
| 2   | HANA 11 | Columbia Variety No. 1 Colombia |
| 2   | HANA 54 | British Samoa P5 British Samoa |
| 2   | HANA 55 | Apaporis Apaporis, Colombia |
| 2   | HANA 56 | Apaporis P1 Apaporis, Colombia |
| 2   | HANA 62 | Rio Kananari Río Kananari, East Colombia |
| 3   | HANA 13 | Spanish Samoa American Samoa |
| 3   | HANA 17 | Natal Natal, South Africa |
| 3   | HANA 24 | Montserrat Philippines |
| 3   | HANA 25 | Macgregor Philippines |
| 3   | HANA 28 | Dacca Philippines |
| 3   | HANA 164 | N91-06 Chumphon, Thailand |
| 4   | HANA 14 | Pernambuco Singapore |
| 4   | HANA 46 | Sugar Loaf Zaire |
| 4   | HANA 21 | Abacaxi Brazil |
| 5   | HANA 16 | Bermuda Barbados, BWI |
| 5   | HANA 137 | Redonda Pantanillo, Venezuela |
| 5   | HANA 19 | Sarawak Malaysia |
| 6   | HANA 18 | Mauritius Taiwan |
| 6   | HANA 29 | Sugar Loaf Philippines |
| 6   | HANA 125 | Jamaica Sugar Jamaica, WI |
| 6   | HANA 23 | Sam Clarke Plot West Indies, Jamaica |
| 6   | HANA 43 | Mo Vietnam |
in one main cluster, demonstrated that the differences used to designate membership to these two horticultural groups are relatively small, in comparison with the other botanical varieties. The third cluster includes 26 cultivated pineapples that formed a large and diverse group. Within this large cluster there are several important pineapple cultivars such as Criolla from Mexico, Montelirio from Guatemala, and Pernambuco and Manzana from Brazil. The majority of the accessions in this cluster seemed mainly cultivated in South America.

Assignment test by STRUCTURE
Population stratification of the 64 accessions, based on $D_K$ value computed by STRUCTURE HARVESTER, revealed two clusters as the most probable number of $K$ (Figures 3 and 4) and this partitioning was largely compatible with the cluster analysis (Figure 2). All the

| Grp | Code | Name | Country/Region |
|-----|------|------|----------------|
| 6   | HANA 26 | Philippine Red | Philippines |
| 6   | HANA 30 | Sylhet Jaldubi | Philippines |
| 6   | HANA 33 | Ananas Kendal | Java, Indonesia |
| 6   | HANA 41 | Pho Lang Tsang | Vietnam |
| 6   | HANA 42 | Sg Baton Red | Vietnam |
| 6   | HANA 44 | Moe | Vietnam |
| 6   | HANA 122 | Uh | Taiwan |
| 6   | HANA 127 | Kahi | Taiwan |
| 6   | HANA 134 | Kumta | Karnataka, India |
| 6   | HANA 154 | CB 67 | Brazil |
| 6   | HANA 128 | Spiny Anpi | Taiwan |
| 6   | HANA 129 | Philippine Green | Philippines |
| 7   | HANA 37 | Criolla | Guadalajara Jalisco, Mexico |
| 7   | HANA 60 | Spanish Guatemala | Guatemala |
| 7   | HANA 48 | Mexican Criolla | Tamazunchale, Mexico |
| 8   | HANA 64 | CB 5 | Tatuhy, Brazil |
| 8   | HANA 68 | CB 11 | Mogi-Mirim, Brazil |
| 9   | HANA 12 | Congo | Africa |
| 9   | HANA 15 | Ruby | Singapore |
| 9   | HANA 27 | Wild Kailua | Manoa, Hawaii, USA |
| 9   | HANA 31 | Black Antigua | Philippines |
| 9   | HANA 35 | Amalsad | Gujarat, India |
| 9   | HANA 45 | Nep | Vietnam |
| 9   | HANA 71 | CB 18 | Puerto Bertoni, Paraguay |
| 9   | HANA 78 | CB 36 | Dourados, Brazil |
| 9   | HANA 85 | Fazenda Moura | Rio de Janeiro, Brazil |
| 9   | HANA 90 | Prazeres | Prazeres, Brazil |
| 9   | HANA 118 | ACC. 253 | Hawaii, USA |
| 9   | HANA 120 | Los Banos | Hawaii, USA |
| 9   | HANA 130 | Kajatan | Buitenzorg, Java, Indonesia |
| 9   | HANA 131 | Ananas Merah | Java, Indonesia |
| 9   | HANA 140 | Pakse | Vietnam |
| 9   | HANA 142 | Den | Vietnam |
| 9   | HANA 148 | CB 24 | San Lorenzo, Paraguay |
| 9   | HANA 149 | CB 33 | Bella Vista, Brazil |
| 9   | HANA 153 | CB 65 | Rio de Janeiro, Brazil |
| 10  | HANA 145 | Cabezona | Puerto Rico |
| 10  | HANA 165 | N91-13 31358 | Florida, USA |
| 11  | HANA 146 | Antigua | Antigua, Guatemala |
| 11  | HANA 121 | Amarillo | Brazil |
| 11  | HANA 150 | CB 38 | Ponta Pora, Brazil |
| 12  | HANA 20 | P 347 (PI 56907) | Brazil |
| 12  | HANA 73 | CB 20 | Puerto Bertoni, Paraguay |
| 12  | HANA 74 | CB 21 | Puerto Bertoni, Paraguay |
| 12  | HANA 185 | N04-8 | Brazil |
| 12  | HANA 75 | CB 23 | San Ignacio, Argentina |
| 13  | HANA 3 | 61-2223 (PI 536881) | Hawaii, USA |
| 13  | HANA 61 | Unknown Foreign Variety | N/A |
| 14  | HANA 86 | Jandaira | Doura, Brazil |
| 14  | HANA 87 | Rezende | Rezende, Brazil |
| 15  | HANA 96 | Cayenne 7898 Qc | Hawaii, USA |
| 15  | HANA 97 | Cayenne 7898 4n | Hawaii, USA |
| 16  | HANA 172 | N94-92 Short Fruit #1 NGRL 33 | Santa Cruz, Bolivia |
| 16  | HANA 174 | N94-92 Long Fruit #1 NGRL 33 | Santa Cruz, Bolivia |
| 16  | HANA 173 | N94-92 Short Fruit #2 NGRL 33 | Santa Cruz, Bolivia |
| 17  | HANA 136 | Spanish Criolla Red | El Hatillo, Venezuela |
| 17  | HANA 138 | Red Spanish Pina Lisa | Pantanalito, Venezuela |
| 18  | HANA 161 | S8-474 (PI 536979) | Hawaii, USA |
| 18  | HANA 158 | 57-503 (PI 536977) | Hawaii, USA |
| 19  | HANA 51 | Ananas De Vaupes | Bogota, Colombia |
| 19  | HANA 169 | 32424 N01-17 | Florida, USA |
| 20  | HANA 63 | CB 2 | Rio de Janeiro, Brazil |
| 20  | HANA 66 | CB 9 | Amazon, Brazil |
| 21  | HANA 81 | Rondon | Campinas, Brazil |

Figure 1. Somaclonal mutation of clone ‘Cayenne 7898’ showing the difference in dark yellow (Cayenne 7898 4N, HANA 97) and white (Cayenne 7898, HANA96) flesh colors.
Table 5. Minor allele frequency, information index, heterozygosity, and inbreeding coefficient of the 57 SNP loci scored on 64 pineapple accessions.

| SNP ID | Minor allele frequency | Information index | Observed heterozygosity | Expected heterozygosity | Inbreeding coefficient |
|--------|------------------------|-------------------|-------------------------|-------------------------|------------------------|
| Ac1    | 0.440                  | 0.686             | 0.433                   | 0.493                   | 0.122                  |
| Ac2    | 0.313                  | 0.622             | 0.478                   | 0.430                   | 0.110                  |
| Ac3    | 0.313                  | 0.622             | 0.388                   | 0.430                   | 0.098                  |
| Ac4    | 0.159                  | 0.438             | 0.288                   | 0.268                   | 0.076                  |
| Ac5    | 0.239                  | 0.550             | 0.478                   | 0.364                   | 0.314                  |
| Ac6    | 0.246                  | 0.558             | 0.284                   | 0.371                   | 0.236                  |
| Ac7    | 0.440                  | 0.686             | 0.433                   | 0.493                   | 0.122                  |
| Ac8    | 0.379                  | 0.663             | 0.485                   | 0.471                   | 0.030                  |
| Ac9    | 0.366                  | 0.657             | 0.433                   | 0.464                   | 0.067                  |
| Ac10   | 0.119                  | 0.366             | 0.239                   | 0.210                   | 0.136                  |
| Ac11   | 0.431                  | 0.684             | 0.800                   | 0.490                   | 0.631                  |
| Ac12   | 0.254                  | 0.566             | 0.507                   | 0.379                   | 0.340                  |
| Ac13   | 0.067                  | 0.246             | 0.134                   | 0.125                   | 0.072                  |
| Ac14   | 0.396                  | 0.671             | 0.791                   | 0.478                   | 0.654                  |
| Ac15   | 0.418                  | 0.680             | 0.836                   | 0.487                   | 0.718                  |
| Ac16   | 0.090                  | 0.302             | 0.060                   | 0.163                   | 0.634                  |
| Ac17   | 0.336                  | 0.636             | 0.433                   | 0.446                   | 0.030                  |
| Ac18   | 0.357                  | 0.652             | 0.429                   | 0.459                   | 0.067                  |
| Ac19   | 0.100                  | 0.325             | 0.015                   | 0.180                   | 0.915                  |
| Ac20   | 0.157                  | 0.434             | 0.254                   | 0.264                   | 0.040                  |
| Ac21   | 0.119                  | 0.366             | 0.239                   | 0.210                   | 0.136                  |
| Ac22   | 0.269                  | 0.582             | 0.418                   | 0.393                   | 0.063                  |
| Ac23   | 0.362                  | 0.654             | 0.723                   | 0.462                   | 0.566                  |
| Ac24   | 0.453                  | 0.689             | 0.438                   | 0.496                   | 0.117                  |
| Ac25   | 0.112                  | 0.351             | 0.194                   | 0.199                   | 0.024                  |
| Ac26   | 0.265                  | 0.578             | 0.470                   | 0.390                   | 0.205                  |
| Ac27   | 0.300                  | 0.611             | 0.600                   | 0.420                   | 0.429                  |
| Ac28   | 0.396                  | 0.671             | 0.433                   | 0.478                   | 0.095                  |
| Ac29   | 0.396                  | 0.671             | 0.433                   | 0.478                   | 0.095                  |
| Ac30   | 0.119                  | 0.366             | 0.239                   | 0.210                   | 0.136                  |
| Ac31   | 0.400                  | 0.673             | 0.800                   | 0.480                   | 0.667                  |
| Ac32   | 0.077                  | 0.271             | 0.154                   | 0.142                   | 0.083                  |
| Ac33   | 0.321                  | 0.628             | 0.463                   | 0.436                   | 0.062                  |
| Ac34   | 0.410                  | 0.677             | 0.403                   | 0.484                   | 0.167                  |
| Ac35   | 0.227                  | 0.536             | 0.303                   | 0.351                   | 0.137                  |
| Ac36   | 0.302                  | 0.612             | 0.340                   | 0.421                   | 0.171                  |
| Ac37   | 0.439                  | 0.686             | 0.879                   | 0.493                   | 0.784                  |
| Ac38   | 0.261                  | 0.574             | 0.254                   | 0.386                   | 0.343                  |
| Ac39   | 0.127                  | 0.380             | 0.254                   | 0.222                   | 0.145                  |
| Ac40   | 0.276                  | 0.589             | 0.493                   | 0.400                   | 0.232                  |
| Ac41   | 0.336                  | 0.638             | 0.672                   | 0.446                   | 0.506                  |
| Ac42   | 0.402                  | 0.674             | 0.803                   | 0.481                   | 0.671                  |
| Ac43   | 0.216                  | 0.522             | 0.433                   | 0.339                   | 0.276                  |
| Ac44   | 0.470                  | 0.691             | 0.463                   | 0.498                   | 0.071                  |
| Ac45   | 0.313                  | 0.622             | 0.388                   | 0.430                   | 0.098                  |
| Ac46   | 0.455                  | 0.639             | 0.479                   | 0.496                   | 0.192                  |
| Ac47   | 0.493                  | 0.693             | 0.448                   | 0.500                   | 0.104                  |
| Ac48   | 0.468                  | 0.691             | 0.498                   | 0.689                   |
| Ac49   | 0.119                  | 0.366             | 0.179                   | 0.210                   | 0.148                  |
| Ac50   | 0.333                  | 0.637             | 0.667                   | 0.444                   | 0.500                  |
| Ac51   | 0.052                  | 0.205             | 0.104                   | 0.099                   | 0.055                  |
| Ac52   | 0.354                  | 0.650             | 0.708                   | 0.457                   | 0.548                  |
| Ac53   | 0.452                  | 0.689             | 0.841                   | 0.495                   | 0.698                  |
| Ac54   | 0.306                  | 0.616             | 0.484                   | 0.425                   | 0.138                  |
| Ac55   | 0.423                  | 0.681             | 0.846                   | 0.488                   | 0.733                  |
| Ac56   | 0.067                  | 0.246             | 0.104                   | 0.125                   | 0.166                  |
| Ac57   | 0.462                  | 0.690             | 0.924                   | 0.497                   | 0.859                  |
| Mean   | 0.298                  | 0.565             | 0.465                   | 0.385                   | 0.157                  |

Acknowledgments

This study was supported by the National key research and development project (2016YFD0101200) and the National Natural Science Foundation of China (31571755). We are very grateful to the pineapple germplasm conservationists who provided the samples for genetic diversity analysis. We are also grateful to the pineapple community for the critical feedback and support. This study was supported by the National Natural Science Foundation of China (31571755).
well documented. These somatic mutations are the major source of variation exploited for the selection of new cultivars. For example, the spiny or smooth leaf margins, caused by a single gene, are the signature character for the cultivar group Smooth Cayenne. Such a mutation is difficult to detect when a small set of molecular markers are applied. Similar problems were found in fingerprinting projects dealing with other vegetative propagated crops such as bananas (Musa spp., 2014), bread fruit (2015), and apple (Malus spp., 2012). More comprehensive genomic approaches, such as next-generation sequencing, would be needed to detect which genes or alleles had been changed, thereby causing the phenotypic variation. For this reason, the reduction of identified duplicates in pineapple germplasm genebank needs to be considered on a case-by-case basis. Characterization of phenotypic traits among the synonymous group members is still essential to complement DNA fingerprinting for genotype identification.

Classification of pineapple germplasm

A. comosus is a mostly self-incompatible diploid with $2n=2x=50$ chromosomes. This species includes five botanical varieties of A. comosus: vars. comosus, ananassoides, parguazensis, erectifolius, and bracteatus, based on the revised classification of Coppens d’Eeckenbrugge and Leal. The present results show that out of the 170 Ananas accessions maintained in the USDA pineapple

Figure 2. NJ unrooted tree depicting the relationship among 64 pineapple accessions from USDA-ARS, Pacific Basin Tropical Plant Germplasm Resource Center in Hilo, Hawaii. Identification of accessions corresponds to samples listed in Table 1 and Supplemental Table 1.

Figure 3. Plot of $\Delta K$ (filled circles, solid line) calculated as the mean of the second-order rate of change in likelihood of $K$ divided by the standard deviation of the likelihood of $K$, $m|L'(K)|/s[L(K)]$. 

Horticulture Research (2015) © 2015 Nanjing Agricultural University
| Variety                  | Species                  |
|-------------------------|--------------------------|
| CB 6                    | *A. comosus* var. *ananassoides* |
| CB 19                   | *A. comosus* var. *ananassoides* |
| WILD BRAZIL             | *A. comosus* var. *ananassoides* |
| BOGOTA                  | *A. comosus* var. *comosus* |
| CB 15                   | *A. comosus* var. *ananassoides* |
| CB 61 Dr. C’S 304       | *A. comosus* var. *ananassoides* |
| CB 71                   | *A. comosus* var. *ananassoides* |
| N94-92 SHORT FRUIT #1   | *A. comosus* var. *ananassoides* |
| PINA LISA               | *A. comosus* var. *comosus* |
| CAYENNE LOT 520         | *A. comosus* var. *comosus* |
| N94-92 LONG FRUIT #2    | *A. comosus* var. *comosus* |
| WILD BRAZIL X LOT 520   | *A. comosus* var. *comosus* |
| RONDON Dr. C’S 387      | *A. comosus* var. *comosus* |
| 58-474                  | *A. comosus* var. *comosus* |
| 58-1184                 | *A. comosus* var. *comosus* |
| MONTE LIRIO             | *A. comosus* var. *comosus* |
| CRIOLLA Plot 379        | *A. comosus* var. *comosus* |
| White Jade              | *A. comosus* var. *comosus* |
| COWBOY                  | *A. comosus* var. *comosus* |
| DR C’S 2009             | *A. comosus* var. *comosus* |
| 31722 N91-16            | *A. comosus* var. *comosus* |
| MANZANA                 | *A. comosus* var. *comosus* |
| TAINUNG 9               | *A. comosus* var. *comosus* |
| CAYENNE LANAI           | *A. comosus* var. *comosus* |
| CAYENNE HILO            | *A. comosus* var. *comosus* |
| N91-14 27280            | *A. comosus* var. *comosus* |
| PHU QUI                 | *A. comosus* var. *comosus* |
| 61-2223                 | *A. comosus* var. *comosus* |
| 75-50 N93-66            | *A. comosus* var. *comosus* |
| TABOGA                  | *A. comosus* var. *comosus* |
| PERNAMBUCO              | *A. comosus* var. *comosus* |
| 41-411                  | *A. comosus* var. *comosus* |
| CONGO                   | *A. comosus* var. *comosus* |
| MANDAIRA                | *A. comosus* var. *comosus* |
| MAURITIUS               | *A. comosus* var. *comosus* |
| ANANAS DE VAUPES        | *A. comosus* var. *comosus* |
| PAPURI VAUPES COLOMBIA  | *A. comosus* var. *comosus* |
| PINA CRIOLLA            | *A. comosus* var. *comosus* |
| ALBUS Dr. C’S 315       | *Ananas comosus* var. *bracteatus* |
| 53-116                  | *A. comosus* var. *comosus* |
| BRITISH SAMOA           | *A. comosus* var. *comosus* |
| F1 CAYENNE X P. SAGENARIUS | *Ananas hybrid* |
| CB 9                    | *A. comosus* var. *comosus* |
| 58-696                  | *Ananas comosus* var. *bracteatus* |
| CB 5                    | *A. comosus* var. *comosus* |
| 63-759                  | *A. comosus* var. *comosus* |
| CANTERRA                | *A. comosus* var. *comosus* |
| TRINIDAD                | *A. comosus* var. *comosus* |
| CB 30                   | *A. comosus* var. *comosus* |
| PLOT 347                | *Ananas comosus* var. *bracteatus* |
| SANTA MARTA             | *A. comosus* var. *comosus* |
| CHEESE PINE             | *A. comosus* var. *comosus* |
| PINA DE CASTILLA        | *A. comosus* var. *comosus* |
| ABACAXI VERMELHO        | *A. comosus* var. *comosus* |
| ANTIGUA                 | *A. comosus* var. *comosus* |
| CABEZONA                | *A. comosus* var. *comosus* |
| CAYENNE 7898 4N         | *A. comosus* var. *comosus* |
| SPANISH SAMOA           | *A. comosus* var. *comosus* |
| SPINELESS ANPI 52       | *A. comosus* var. *comosus* |
| N94-43 SINGAPORE        | *A. comosus* var. *comosus* |
| BERMUDA                 | *Ananas comosus* var. *bracteatus* |
| CB 17                   | *A. comosus* var. *comosus* |
| 59-656                  | *A. comosus* var. *comosus* |
| CB 32                   | *A. comosus* var. *comosus* |

**Figure 4.** Inferred clusters in the pineapple accessions varieties using STRUCTURE in the overall analyzed pineapple accessions. Each vertical line represents one individual multilocus genotype. Individuals with multiple colors have admixed genotypes from multiple clusters. Each color represents the most likely ancestry of the cluster from which the genotype or partial genotype was derived. Clusters of individuals are represented by colors.
The second observation is that several cultivated pineapple accessions (Bogota, Pina Lisa, Bermuda, Cayenne Lot 520, Cabezona, and Trinidad) were grouped together with *A. comosus* var. *ananassoides* or *A. comosus* var. *bracteatus*, instead of with the rest of the *A. comosus* var. *comosus* accessions (Figures 2 and 3). This result indicates that the current system that classifies all cultivated pineapple into a single botanical variety (*A. comosus* var. *comosus*) may be questionable. It would be appropriate to consider cultivated pineapple as a complex of different botanical varieties, with possible significant gene flow among them.

The third observation is about the validity of the horticultural classification of pineapple germplasm. Pineapple cultivars are classified into several horticultural groups. The commonly known groups include ‘Abacaxi’, ‘Cayenne’, ‘Maipure’, or ‘Perolera’, ‘Queen’, and ‘Spanish’.

Despite these horticultural groups having been adopted by many users, little investigation has been done to show a genetic basis to reinforce this categorization. Coppens d’Eeckenbrugge and Leal hypothesized that *var. ananassoides* is the likely progenitor of cultivated pineapple, and it is likely that domestication happened in the Guiana shield. Strong piece of evidence supporting this hypothesis is that all four chloroplast haplotypes that have been identified in cultivated materials are present in the wild *var. ananassoides*. On the other hand, *var. bracteatus* was not considered as a progenitor in this hypothesis, mainly because *var. bracteatus* appeared to be a homogeneous variety with narrow genetic diversity, which is an unlikely basis for diverse domesticated cultigens of pineapple. The present result, however, shows that 11 pineapple cultivars (Canberra, Papuri Vaupes Colombia, CB 30, Pina de Castilla, Rondon, Congo, Phu Qui, Mauritius, Cheese pine), which are dispersed across different clusters as shown in the NJ tree, could have their parentage (either male or female) traced back to *var. bracteatus* (Table 6). *Ananas comosus var. bracteatus* is native to Brazil, Bolivia, Argentina, Paraguay, and Ecuador but not to the Guiana shield. The present result thus indicates the possibility that pineapple could have been domesticated at multiple sites, involving both var. *ananassoides* and var. *bracteatus*. The Paraná-Paraguay river drainage area could be one of the domestication sites, since both var. *bracteatus* and var. *ananassoides* are indigenously distributed in this area. Geographically disparate origins of crop domestication are not uncommon in the Americas, as in the case of common bean (*Phaseolus vulgaris*), chili pepper (*Capsicum spp.*), potato (*Solanum spp.*), and cacao (*T. cacao*), as reviewed by Clement et al.

In conclusion, we conducted a study to develop a set of SNP markers for pineapple and employed them for fingerprinting the USDA’s pineapple collection, using a nanofluidic array. This approach enabled us to generate high-quality SNP profiles for the purpose of pineapple cultivar identification. This is a highly useful tool for genebank management, which will also lead to more efficient crop improvement and, furthermore, has the potential to protect intellectual property rights of breeders. Our result also generated significant insight regarding the origin and domestication of pineapple. Efforts to sequence multiple cultivars from the same synonymous groups with somaclonal mutations are underway, in order to gain a comprehensive understanding about the genetic basis for mutation-based changes in important agronomic traits. This information will be highly useful for verification of pineapple cultivars and will improve the efficiency of pineapple genebank operation. The high rate of genetic redundancy detected in this collection, there are only 64 distinctive genotypes. Clustering analysis and model-based stratiﬁcation both showed that *A. comosus var. ananassoides* differs from *A. comosus var. bracteatus*, thus supporting the revised taxonomy system that classiﬁed *A. comosus var. ananassoides* and var. *bracteatus* as two different botanical varieties. However, accessions of *var. erectifolius* were found to have high similarity and grouping closely together with the accessions of *var. bracteatus*. This result differs with a previous report based on isozyme variation, which showed that *A. comosus var. erectifolius* did not have a distinctive isozyme proﬁle, in comparison with the rest of the *A. comosus var. comosus* cultivars. Nonetheless, the present study only used two accessions of *A. comosus var. erectifolius*, which may be a bias in terms of the sample representation. Additional samples of *A. comosus var. erectifolius* from other genebanks need to be examined and a larger number of SNP markers need to be analyzed to clarify if the classiﬁcation of *A. comosus var. erectifolius* should be revised.

The likelihood assignment of parentage of pineapple (*A. comosus var. comosus*) accessions based on 57 SNP markers with LOD scores above 80% probability.

| Candidate parent | Botanical variety | Offspring | LOD scorea |
|------------------|------------------|-----------|------------|
| CB 15            | var. *ananassoides* | Wild Brazil x Lot 520 | 2.88 |
| CB 15            | var. *ananassoides* | Pina Lisa | 1.25 |
| CB 17            | var. *bracteatus* | Canberra | 2.84 |
| CB 17            | var. *bracteatus* | Papuri Vaupes Colombia | 2.90 |
| CB 17            | var. *bracteatus* | CB 30 | 3.89 |
| CB 17            | var. *bracteatus* | Pina de Castilla | 2.62 |
| CB 19            | var. *ananassoides* | Cayenne Lot 520 | 3.05 |
| CB 5             | var. *bracteatus* | Rondon | 1.43 |
| CB 6             | var. *ananassoides* | Bogota | 18.66 |
| CB 61            | var. *ananassoides* | CB 32 | 2.11 |
| Plot 37 (PI 56907) | var. *bracteatus* | Congo | 2.65 |
| Plot 37 (PI 56907) | var. *bracteatus* | Phu Qui | 2.84 |
| Plot 37 (PI 56907) | var. *bracteatus* | Mauritius | 2.31 |
| Plot 37 (PI 56907) | var. *bracteatus* | Cheese Pine | 0.61 |

4 Putative parental accessions used in the present analysis. Only those with significant LOD score were listed.
5 Critical LOD (the natural logarithm of the likelihood) ratio for assignment of parentage is 2.60 at 95% confidence and 0.24 at 80% confidence.
6 Parentage analysis showed that both var. *bracteatus* and var. *ananassoides* can be progenitors of pineapple cultivars (Table 6). This result is in agreement with the fact that both var. *bracteatus* and var. *ananassoides* can intercross successfully with var. *comosus* to produce fertile offspring. Coppens d’Eeckenbrugge and Leal hypothesized that var. *ananassoides* is the likely progenitor of cultivated pineapple, and it is likely that domestication happened in the Guiana shield. One strong piece of evidence supporting this hypothesis is that all four chloroplast haplotypes that have been identified in cultivated materials are present in the wild var. *ananassoides*. On the other hand, var. *bracteatus* was not considered as a progenitor in this hypothesis, mainly because var. *bracteatus* appeared to be a homogeneous variety with narrow genetic diversity, which is an unlikely basis for diverse domesticated cultigens of pineapple. The present result, however, shows that 11 pineapple cultivars (Canberra, Papuri Vaupes Colombia, CB 30, Pina de Castilla, Rondon, Congo, Phu Qui, Mauritius, Cheese pine), which are dispersed across different clusters as shown in the NJ tree, could have their parentage (either male or female) traced back to var. *bracteatus* (Table 6). *Ananas comosus var. bracteatus* is native to Brazil, Bolivia, Argentina, Paraguay, and Ecuador but not to the Guiana shield. The present result thus indicates the possibility that pineapple could have been domesticated at multiple sites, involving both var. *ananassoides* and var. *bracteatus*. The Paraná-Paraguay river drainage area could be one of the domestication sites, since both var. *bracteatus* and var. *ananassoides* are indigenously distributed in this area. Geographically disparate origins of crop domestication are not uncommon in the Americas, as in the case of common bean (*Phaseolus vulgaris*), chili pepper (*Capsicum spp.*), potato (*Solanum spp.*), and cacao (*T. cacao*), as reviewed by Clement et al.

In conclusion, we conducted a study to develop a set of SNP markers for pineapple and employed them for fingerprinting the USDA’s pineapple collection, using a nanofluidic array. This approach enabled us to generate high-quality SNP profiles for the purpose of pineapple cultivar identification. This is a highly useful tool for genebank management, which will also lead to more efficient crop improvement and, furthermore, has the potential to protect intellectual property rights of breeders. Our result also generated significant insight regarding the origin and domestication of pineapple. Efforts to sequence multiple cultivars from the same synonymous groups with somaclonal mutations are underway, in order to gain a comprehensive understanding about the genetic basis for mutation-based changes in important agronomic traits. This information will be highly useful for verification of pineapple cultivars and will improve the efficiency of pineapple genebank operation. The high rate of genetic redundancy detected in this
collection, also suggests the potential impact of applying this technology on other tropical perennial crops.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We would like to give special thanks to Carol Mayo-Riley for sampling the pineapple germplasm and organizing the passport information; Stephen Pinney for assisting SNP genotyping using nonfluoridic array. This work was partially supported by the National Natural Science Foundation of China (30800884, 31730688). References to a company and/or product by the USDA are only for the purposes of information and do not imply approval or recommendation of the product to the exclusion of others that may also be suitable.

REFERENCES

1 Food and Agriculture Organization of the United Nations. Statistical databases. United Nations: FAO, 2014. Available at http://faostat.fao.org/site/567/DesktopDefault.aspx?PageID=567&ancor (accessed December 31, 2014).
2 Baker KF, Collins JL. Notes on the distribution and ecology of Ananas and Pseudananas in South America. Am J Bot 1939; 26: 697–702.
3 Pearsall DM. The origins of plant cultivation in South America. In: Cowan CW, Watson PJ, editors. The Origins of Agriculture: An International Perspective. Washington, DC: Smithsonian Institution Press; 1992, pp 173–205.
4 Bartholomew DP, Paull RE, Rohrbach KG (eds). The Pineapple: Botany, Production and Uses. Wallingford: CABI Publishing; 2003.
5 Pursseglove JW. Tropical Crops. Monocotyledons. London: Longman; 1972, pp 75–91.
6 Loison-Cabot C. Origin, phylogeny and evolution of pineapple species. Fruits 1992; 47: 25–32.
7 Coppens d’Eekenbrugge G, Leal F. “Chapter 2: Morphology, Anatomy, and Taxonomy.” In: Bartholomew DP, Paull RE, Rohrbach KG, editors. The Pineapple: Botany, Production, and Uses. Wallingford: CABI Publishing; 2003, pp 13-32.
8 Coppens d’Eekenbrugge G, Sanewski GM, Smith MK et al. Ananas: In: Kole C editor. Wild Crop Relatives: Genomic and Breeding Resources, Tropical and Subtropical Fruits. Berlin and Heidelberg: Springer-Verlag; 2011, pp 21-41.
9 Morrison SE. Journals and Other Documents of the Life and Voyages of Christopher Columbus. New York: New York Heritage Press; 1973.
10 Collins JL. The Pineapple, Botany, Utilisation. Cultivation. London: Leonard Hill Ltd; 1960. p294.
11 Bartholomew DP, Hawkins RA, Lopez JA. Hawaii pineapple: the rise and fall of an industry. HortScience 2012; 47: 1390–1398.
12 Hidayat T, Abdullah FJ, Kuppusamy C, Samad AA, Wagiran A. Molecular identification of Malaysian pineapple cultivar based on internal transcribed spacer region. APCBEE Procedia 2012; 4: 146–151.
13 Sripanayaya S, Marchant R, Power JB, Davey MR. Herbicide-tolerant transgenic pineapple (Ananas comosus L.) root development. Plant Sci 2002; 160: 1011–1035.
14 Ong WD, Yoo L-YC, Kumar VS. De Novo assembly, characterization and functional annotation of pineapple fruit transcriptome through massively parallel sequencing. PLoS One 2012; 7: e66937.
15 Tang J, Vosman B, Voorrips RE, van der Linden CG, Leunissen JA. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid species. BMC Bioinformatics 2006; 7: 438.
16 Zhang HN, Wei YZ, Shen JY et al. Transcriptomic analysis of floral initiation in litchi (Litchi chinensis Sonn.) based on de novo RNA sequencing. Plant Cell Report 2013; 33: 1723–1735.
17 Platé RK, Jain M, NGS QC Toolkit: a quality control tool for next generation sequencing data. PLoS One 2012; 7: e30619.
18 Wang J, Lim M, Crenshaw A et al. High-throughput single nucleotide polymorphism genotyping using nonfluoridic dynamic arrays. BMC Genomics 2009; 10: 561.
19 Fluidigm. Fluidigm SNP Genotyping User Guide Rev H1, PN 68000998. San Francisco, CA: Fluidigm Corporation; 2011.
20 Peñalva R, Smouse PE. Genomics: Theor Appl Genet 2006; 113: 288–295.
21 Peñalva R, Smouse PE. GenAx: 6.5; Genetic analysis in excel. Population genetic software for teaching and research. Mol Ecol Notes 2006; 6: 288–295.
22 Earl DA, vonHoldt BM. Structure Harvester: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 2012; 4: 359–361.
23 Marshall TC, Slave J, Kruuk LEB, Pemberton JM. Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 1998; 7: 639–655.
24 Kalembowski SJ, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 2011; 16: 1099–1106.
25 Huanga X, Madan A, CAP3: A DNA sequence assembly program. Genome Res 1999; 9: 868–877.
26 Noyer JL. Preliminary study of genetic diversity of the genus Ananas by RFLP. Fruits (France) 1991; 46: 372–375.
27 Conyers JA, Nesi M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987; 4: 406–425.
28 Felsenstein J. PHYLIP: phylogeny inference package (version 3.2). Cladistics 1989; 5: 164–166.
29 Letunic I, Bork P. Interactive tree of life v2: Online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 2011; 39: W478–W478.
30 Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000; 155: 945–959.
31 Earl DA, vonHoldt BM. Structure Harvester: website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 2012; 4: 359–361.
Molecular characterization of pineapple germplasm using SNP+
L Zhou et al

52 Marchant CJ. Chromosome evolution in the Bromeliaceae. Kew Bulletin 1967; 21: 161–168.
53 Brown GK, Palaci CA, Luther HE. Chromosome numbers in Bromeliaceae. Am J Bot 1997; 76: 85–88.
54 Leal F, Soule J. Maipure, a new spineless group of pineapple cultivars. HortScience 1977; 12: 393–403.
55 Wee YC, Thongtham MLC. Ananas comosus (L.) Merr. In: Verheij EWM, Coronel RE, editors. Plant Resources of South-East Asia No. 2 Edible Fruits and Nuts. Wageningen: The Netherlands, Pudoc; 1991. pp66–71.
56 Coppens d'Eeckenbrugge G, Leal F, Duval MF. Germplasm resources of pineapple. Hort Rev 1997; 21: 133–175.
57 Clement CR, Cristo-Araújo MD, Coppens D'Eeckenbrugge G, Pereira AA, Picanço-Rodrigues D. Origin and domestication of native amazonian crops. Diversity 2010; 2: 72–106.

Supplemental Information for this article can be found on the Horticulture Research website (http://www.nature.com/hortres).