

Research Paper

RAD-Seq analysis of typical and minor *Citrus* accessions, including Bhutanese varieties

Tshering Penjor^{1,2}, Takashi Mimura¹, Nobuhiro Kotoda¹, Ryoji Matsumoto¹, Atsushi J. Nagano^{3,4,5}, Mie N. Honjo³, Hiroshi Kudoh³, Masashi Yamamoto^{†6} and Yukio Nagano^{*†7}

¹ Faculty of Agriculture, Saga University, 1 Honjo-machi, Saga 840-8502, Japan

² Renewable Natural Resources Research Centre Wengkharr, Mongar, Bhutan

³ Center for Ecological Research, Kyoto University, 509-3 2-chome, Hirano, Otsu, Shiga 520-2113, Japan

⁴ JST PRESTO, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012, Japan

⁵ Faculty of Agriculture, Ryukoku University, 1-5 Yokotani, Seta Oe-cho, Otsu, Shiga 520-2194, Japan

⁶ Faculty of Agriculture, Kagoshima University, 1-21-35 Korimoto, Kagoshima 890-0065, Japan

⁷ Analytical Research Center for Experimental Sciences, Saga University, 1 Honjo-machi, Saga 840-8502, Japan

We analyzed the reduced-representation genome sequences of *Citrus* species by double-digest restriction site-associated DNA sequencing (ddRAD-Seq) using 44 accessions, including typical and minor accessions, such as Bhutanese varieties. The results of this analysis using typical accessions were consistent with previous reports that citron, papada, pummelo, and mandarin are ancestral species, and that most *Citrus* species are derivatives or hybrids of these four species. *Citrus* varieties often reproduce asexually and heterozygosity is highly conserved within each variety. Because this approach could readily detect conservation of heterozygosity, it was able to discriminate citrus varieties such as satsuma mandarin from closely related species. Thus, this method provides an inexpensive way to protect citrus varieties from unintended introduction and to prevent the provision of incorrect nursery stocks to customers. One *Citrus* variety in Bhutan was morphologically similar to Mexican lime and was designated as Himalayan lime. The current analysis confirmed the previous proposition that Mexican lime is a hybrid between papada and citron, and also suggested that Himalayan lime is a probable hybrid between mandarin and citron. In addition to Himalayan lime, current analysis suggested that several accessions were formed by previously undescribed combinations.

Key Words: Bhutan, *Citrus*, genetic relationship, high-throughput sequencing, lime, RAD-Seq.

Introduction

Citrus species are economically important fruit trees and are typical diploid ($2n = 2x = 18$) plants. Therefore, it is worthwhile to elucidate their genetic relationships and determine the parentage of cultivated varieties. The definition of the “species” of *Citrus* and its relatives is unique. Species belonging to five genera (*Citrus*, *Fortunella*, *Poncirus*, *Eremocitrus*, and *Microcitrus*) can often cross with each other (Iwamasa *et al.* 1988), i.e., the genus *Citrus* is only one of several cross-compatible genera. Based on morphological studies, Swingle and Reece (1967) classified *Citrus* into 16 species, and Tanaka classified them into 162 species

(Tanaka 1977). However, based on DNA sequence analysis using DNA markers, Sanger sequencing or high-throughput sequencing, recent studies proposed that *C. medica* L. ‘citron’, *C. micrantha* Wester ‘papada’, *C. maxima* (Burm.) Merr. ‘pummelo’, and *C. reticulata* Blanco ‘mandarin’ are ancestral species, and that most *Citrus* species, especially commercially cultivated varieties, are derivatives or hybrids of these four species (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000). Thus, although *Citrus* species may have extensive morphological diversity, their genetic relationships seem to be simple. However, analyzing previously uncharacterized accessions of *Citrus* may enable us to find more novel derivatives and hybrids of these four ancestral species than were previously thought to exist.

Restriction site-associated DNA sequencing (RAD-Seq) (Baird *et al.* 2008) is a method to analyze the reduced-representation genome using high-throughput sequencing and is used to identify and genotype DNA sequence

Communicated by Norihiko Tomooka

Received March 29, 2016. Accepted October 11, 2016.

First Published Online in J-STAGE on November 29, 2016.

*Corresponding author (e-mail: nagano@cc.saga-u.ac.jp)

† These authors contributed equally to this work

polymorphisms simultaneously. Among several varieties of RAD-Seq, double-digest RAD-Seq (ddRAD-Seq) is one of the most inexpensive methods, and is suitable for large numbers of individuals (Peterson *et al.* 2012), although the proportions of analyzable genome regions are smaller than those of traditional RAD-Seq methods (e.g., single-end RAD-Seq using the 6-base cutter *EcoRI*). The cost per sample of ddRAD-Seq is under \$10 (Peterson *et al.* 2012), which is often less expensive than methods based on PCR or Sanger sequencing. The ddRAD-Seq can potentially detect the genetic relationships of certain *Citrus* species. However, there is the possibility that ddRAD-Seq is not applicable for studies into the genetic relationships of other *Citrus* species, namely those that possess more genetic variation than the method can cope with. Therefore, by using ddRAD-Seq, it is worthwhile to confirm previous reports that citron, papeda, pummelo, and mandarin are ancestral species, and that most *Citrus* species are derivatives or hybrids of these four species (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000).

New varieties of *Citrus* species often have been developed through crossing, and these trees were propagated asexually through polyembryony or grafting. In the citrus industry, it is important to develop a simple method to identify each variety. This is because the introduction or exportation of citrus varieties without following local industry protocols can lead to serious intellectual property issues. Additionally, for the producers and distributors of citrus nursery stocks, the sale of inaccurately identified cultivars to customers is an economically important problem. In asexually reproduced plants derived from a single tree, the heterozygous sites are well conserved because a single cross event should result in a drastic change in heterozygosity. Therefore, examining conservation of heterozygosity among individuals is a simple way to identify each variety. In our previous study using traditional RAD-Seq (Tshering Penjor *et al.* 2014a), we examined the conservation of heterozygosity of limes to determine a single tree origin (i.e., asexual reproduction), and showed that they could be separated into two types, each of which should be derived from a single tree. It is important to validate whether inexpensive ddRAD-Seq can be applied to testing single tree origins. This validation will lead to the solution of the above two problems: accidental or potentially illegal introduction of commercial varieties to unintended areas and the sale of incorrectly identified nursery stock to commercial customers. In our previous study (Tshering Penjor *et al.* 2014a), we identified genetic differences within asexually reproduced trees, i.e., one of two types of lime was subdivided into two subtypes. It is important to reproduce this result in order to demonstrate the reliability of the ddRAD-Seq method.

If the ddRAD-Seq method confirms other previous results (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000), it will provide an inexpensive method to reveal the genetic relationships of several *Citrus* species. Some fruits of *Citrus* species, such

as sweet oranges (*C. sinensis* Osbeck), mandarins, lemons (*C. limon* (L.) Burm. f.), limes, and grapefruits (*C. paradisi* Macfad.), are commercially available in many countries. However, various accessions of *Citrus* species are only used in a limited area or are not used by humans. The characterization of these minor accessions may provide novel genetic resources useful as new varieties or for future breeding programs. Previously, we examined the phylogenetic relationships of *Citrus* based on the plastidic *rbcL* (Tshering Penjor *et al.* 2010) and *matK* (Tshering Penjor *et al.* 2013) gene sequences. These previous studies included citrus accessions that are not cultivated worldwide but are preserved in Japan, and it is therefore necessary to genetically characterize additional minor accessions of *Citrus* species.

The region from northeastern India (e.g., Assam) to southwestern China (e.g., Yunnan) is likely where some *Citrus* species originated from (Gmitter and Hu 1990, Tanaka 1959). As Bhutan borders northeastern India, it may also belong to this region of origin, and possibly harbor unique species of *Citrus* available nowhere else in the world. We therefore explored Bhutan in 2007 and 2009 to characterize new accessions of *Citrus* species (Tshering Penjor *et al.* 2014a, 2014b). From our exploration in 2007, we reported on morphological and genetic characteristics of *Citrus* species native to Bhutan in Tshering Penjor *et al.* (2014b). However, because we only sequenced the plastidic *matK* gene, the available genetic information was limited. In our following 2009 expedition (Tshering Penjor *et al.* 2014a), we then found genetically interesting accessions in native Bhutanese lime trees using traditional RAD-Seq (Baird *et al.* 2008). We found the limes in Bhutan were clearly separated into two types based on DNA sequence information. Based on morphological features, we concluded that both species were Mexican limes (*C. aurantifolia* (Christm.) Swingle), and that genetic variance contributed to only slight morphological differences between the two types. However, there remains the possibility that the parents of each type were different, as we could not resolve the genetic position of each type of lime among other *Citrus* species. Therefore, it is worthwhile to genetically reanalyze the trees collected in 2007 and 2009 and compare them with typical species of *Citrus* using high-throughput sequencing.

In this study, 44 accessions, including typical and minor varieties, were genetically characterized using ddRAD-Seq. We determined whether our classification of *Citrus* species was consistent with the findings of other similar studies (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000), and also whether ddRAD-Seq was able to test for a single tree origin, thus providing a simple method to identify each variety. Finally, new knowledge about previously undescribed hybrid accessions, such as Himalayan lime, is discussed using new data derived from ddRAD-seq.

Materials and Methods

RAD-Seq analysis

Bhutanese accessions used in this study have been described previously in Tshering Penjor *et al.* (2014a) and Tshering Penjor *et al.* (2014b). The DNA purification procedure used in this study is described in Tshering Penjor *et al.* (2014a). The library for ddRAD-Seq (Peterson *et al.* 2012) was created with slight modifications (Sakaguchi *et al.* 2015), in which *Bg*III was used as the first restriction site adjacent to the binding site of the primer to read a single-end sequence, and *Eco*RI was used as the second restriction site adjacent to the binding site to read an index sequence. The library was sequenced with 49 bp single-end reads in one lane of an Illumina HiSeq2000 (Illumina, San Diego, CA, USA) by BGI Hongkong. Sequences are available at the DDBJ Sequence Read Archive (http://trace.ddbj.nig.ac.jp/dra/index_e.shtml; Accession no. DRA004200).

The data were quality-filtered using the `process_shortreads` program (with the `-c -q` options) within the software Stacks (Catchen *et al.* 2011, 2013). The numbers of quality-filtered reads are shown in **Supplemental Table 1**. The data were aligned with the reference genome (Xu *et al.* 2012) using `bowtie` (Langmead *et al.* 2009) with the `-n 3 -k 10 --best --chunkmbs 1024` options. The data of the reference genome contain the information of the nuclear sequences, but not chloroplast and mitochondrial sequences. The `pstacks` program of Stacks was used to analyze the aligned data with the `-m 10` option (minimum depth of coverage required to report a stack is 10). The `pstacks` program can extract putative loci of a single origin. Using the `cstacks` program of Stacks, the data obtained from `pstacks` were analyzed with the `-n 1` option (the number of mismatches allowed between sample tags when generating the catalog is 1). Using the `sstacks` program of Stacks, the data from `pstacks` and `cstacks` were analyzed with no options selected. The `populations` program of the Stacks package was used to create a variant call format (VCF) file (Danecek *et al.* 2011) with the `-p 23` option (that is, the minimum number of populations where a locus must be present to process that locus is 23). In this step, to reduce the effect of accidental reads and increase reproducibility, the locus common to more than half of the samples was extracted. Based on the VCF file, the mean sequencing depth for each sample was calculated using `vcftools` (Danecek *et al.* 2011) (**Supplemental Table 2**). Principal component analysis (PCA) and multi-dimensional scaling (MDS) analyses were conducted based on this VCF file using the `SNPrelate` program (Zheng *et al.* 2012). The `populations` program of the Stacks package was also used to create multiple alignments within the cluster. In this case, the `-p` option was set to the total sample number within the cluster. For the admixture analysis, the `populations` program of Stacks was used to create a PLINK file (Purcell *et al.* 2007) with the `-p 23` option. In this case, each asexually reproducing accession was treated as a single population. The admixture analysis was conducted based on

this PLINK file using ADMIXTURE (Alexander *et al.* 2009).

Phylogenetic analysis based on matK gene sequences

PCR primers and sequencing primers are described in Tshering Penjor *et al.* (2013). The purified DNA fragments were sequenced in both directions in an Applied Biosystems 3130 Genetic Analyzer (Life Technologies, Carlsbad, CA, USA) with a BigDye Terminator Cycle Sequencing Ready Reaction Kit v. 3.1 (Life Technologies), as described in Tshering Penjor *et al.* (2013). Sequence data were submitted to DDBJ/GenBank/EBI and were assigned accession numbers LC102227 to LC102230. The maximum likelihood methods from the MEGA program (Tamura *et al.* 2011) were used to create phylogenetic trees. The reliability of each branch was tested by bootstrap analysis with 1,000 replicates.

Results

Overview of current analysis

Forty-four accessions of *Citrus* species were used in this study (**Table 1**). By ddRAD-Seq analysis, a VCF file containing information for 5434 sites was created (**Supplemental Table 3**). Based on this VCF file, we conducted a PCA (**Fig. 1**, **Supplemental Table 4**), and also obtained 3-dimensional data from an MDS analysis (**Fig. 2**, **Supplemental Table 5**). These results matched well with those published previously (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000) as discussed later.

In this analysis, three technical replicates (No. 44–46, *Citrus unshiu* (Swingle) Marcow. (Aoshima Unshu) ‘satsuma mandarin’) were included to confirm reproducibility, and we found that there was little difference in the genetic positions among these samples (**Figs. 1, 2**). The multiple alignments (**Supplemental Fig. 1**) showed that 562 of 597 sites were conserved heterozygous sites within technical replicates (94.1%) thus reflecting the reproducibility of this method.

To compare the current ddRAD-Seq analysis with the previous RAD-Seq analysis (Tshering Penjor *et al.* 2014a), multiple alignments within 8 accessions (No. 5–8 and 19–22), using the current study’s and previous study’s data, were created by the method used in the current study. The analysis using the current data extracted 1600 useful sites for genotyping (**Supplemental Fig. 2**), and 61,699 sites were extracted using the previous data (**Supplemental Fig. 3**). As shown below, this ddRAD-Seq analysis again confirmed that Mexican limes can be subdivided into two subclusters (Tshering Penjor *et al.* 2014a). Thus, although the proportions of the analyzable genome regions are smaller, this ddRAD-Seq analysis provided useful information.

For supporting the PCA and MDS analyses, the admixture analysis (**Fig. 3** (K = 4) and **Supplemental Fig. 4** (K = 5–9)) was conducted. Among the possible K values,

Table 1. Citrus accessions used in this study

| No. | Latin name (Tanaka) | Latin name (Swingle & Reece) | Common name | Accession | Source | Accession number | Embryony | Maternal type based on <i>matK</i> sequences |
|-----------------|--|--|------------------------|-----------------|---|--------------------------|----------|--|
| 1 | <i>C. medica</i> L. | <i>C. medica</i> L. | Citron | Maru busshukan | Saga University | 5001 | Mono | Citron |
| 2 | <i>C. medica</i> var. <i>sarcodactylis</i> Swingle | <i>C. medica</i> var. <i>sarcodactylis</i> Swingle | Citron | Fingered Citron | Saga University | 5008 | Mono | Citron |
| 3 | <i>C. medica</i> L. | <i>C. medica</i> L. | Citron | | Bhutan | Bhutan-07015 (B07015) | Mono | Citron-relative |
| 4 | <i>C. aurantifolia</i> (Cristm.) Swingle | <i>C. aurantifolia</i> (Cristm.) Swingle | Lime | Mexican | Saga University | 5115 | Poly | Papeda |
| 5 | <i>C. aurantifolia</i> (Cristm.) Swingle | <i>C. aurantifolia</i> (Cristm.) Swingle | Lime | Mexican | Bhutan | Bhutan-09005 (B09005) | Poly | Papeda |
| 6 | <i>C. aurantifolia</i> (Cristm.) Swingle | <i>C. aurantifolia</i> (Cristm.) Swingle | Lime | Mexican | Saga University | Indonesia-88035 (I88035) | Poly | Papeda |
| 7 | <i>C. aurantifolia</i> (Cristm.) Swingle | <i>C. aurantifolia</i> (Cristm.) Swingle | Lime | Mexican | Saga University | Indonesia-88045 (I88045) | Poly | Papeda |
| 8 | <i>C. aurantifolia</i> (Cristm.) Swingle | <i>C. aurantifolia</i> (Cristm.) Swingle | Lime | Mexican | Saga University | Indonesia-88065 (I88065) | Poly | Papeda |
| 9 | <i>C. excelsa</i> Webster | <i>Citrus</i> sp. | Limon real | | Kagoshima University | - | No info. | Papeda ^b |
| 10 | <i>C. macroptera</i> Montrouz. | <i>C. macroptera</i> Montrouz. | Melanesian papeda | | Saga University | 7006 | Mono | Papeda |
| 11 | <i>C. micrantha</i> Wester | <i>C. micrantha</i> Wester | Biasong | | Saga University | 7004 | Mono | Papeda |
| 12 | <i>C. hystrix</i> DC. | <i>C. hystrix</i> DC. | Mauritius papeda | | Saga University | 7003 | Mono | Papeda |
| 13 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lemon | Villafranca | Bhutan | Bhutan-07004 (B07004) | Poly | Lemon/Myrtle-leaf orange |
| 14 | <i>C. limon</i> (L.) Burm. f. | <i>C. limon</i> (L.) Burm. f. | Bergamot | | Saga University | 5207 | Poly | Lemon/Myrtle-leaf orange |
| 15 | <i>C. bergamia</i> Risso & Poir. | <i>Citrus</i> sp. | Sweet lemon | | Kagoshima University | - | Poly | Lemon/Myrtle-leaf orange ^b |
| 16 | <i>C. limetta</i> Risso | <i>Citrus</i> sp. | Sweet lime | | Saga University | 5218 | Poly | Lemon/Myrtle-leaf orange ^b |
| 17 | <i>C. limetoides</i> Tanaka | <i>C. aurantifolia</i> (Cristm.) Swingle | Sweet lime | | Saga Prefectural Fruit Tree Research Station | 5638 | Poly | Sweet orange/Kunenbo |
| 18 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-07008 (B07008) | Poly | Sweet orange/Kunenbo |
| 19 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-09015 (B09015) | Poly | Ponkam/Satsuma mandarin |
| 20 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-09024 (B09024) | Poly | Ponkam/Satsuma mandarin |
| 21 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-09027 (B09027) | Poly | Ponkam/Satsuma mandarin |
| 22 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-09030 (B09030) | Poly | Ponkam/Satsuma mandarin |
| 23 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-07007 (B07007) | Poly | Ponkam/Satsuma mandarin |
| 24 | <i>Citrus</i> sp. | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-07009 (B07009) | Poly | Ponkam/Satsuma mandarin |
| 25 | <i>C. montana</i> Tanaka | <i>Citrus</i> sp. | Lime | | Bhutan | Bhutan-07006 (B07006) | Poly | Ponkam/Satsuma mandarin |
| 26 | <i>C. montana</i> Tanaka | <i>Citrus</i> sp. | Bilolo | | Kagoshima University | - | Poly | Papeda ^b |
| 27 | <i>C. maxima</i> (Burm.) Merr. | <i>C. maxima</i> (Burm.) Merr. | Pummelo | Mato Buntan | Saga University | 3202 | Mono | Pummelo |
| 28 | <i>C. maxima</i> (Burm.) Merr. | <i>C. maxima</i> (Burm.) Merr. | Pummelo | Suisho Buntan | Saga University | 3301 | Mono | Pummelo |
| 29 | <i>C. maxima</i> (Burm.) Merr. | <i>C. maxima</i> (Burm.) Merr. | Pummelo | Bampeiyu | Saga University | 3206 | Mono | Pummelo |
| 30 | <i>C. paradisi</i> Macfad. | <i>C. paradisi</i> Macfad. | Grapefruit | Star Ruby | Saga University | 3103 | Poly | Not analyzed |
| 31 | <i>C. paradisi</i> Macfad. | <i>C. paradisi</i> Macfad. | Grapefruit | Sagan Ruby | Saga University | 3107 | Poly | Not analyzed |
| 32 | <i>C. sinensis</i> Osbeck | <i>C. sinensis</i> Osbeck | Sweet orange | Valencia | Saga University | 2200 | Poly | Sweet orange/Kunenbo |
| 33 | <i>C. sinensis</i> Osbeck | <i>C. sinensis</i> Osbeck | Sweet orange | Ohtsu Valencia | Saga University | 2203 | Poly | Sweet orange/Kunenbo |
| 34 | <i>C. sinensis</i> Osbeck | <i>C. sinensis</i> Osbeck | Sweet (Blood) orange | Tarocco | Saga University | 2402 | Poly | Sweet orange/Kunenbo |
| 35 | <i>C. sinensis</i> Osbeck | <i>C. sinensis</i> Osbeck | Sweet orange | Fukuhara | Saga University | 2100 | Poly | Sweet orange/Kunenbo |
| 36 | <i>C. reticulata</i> Blanco | <i>C. reticulata</i> Blanco | Ponkan | Yoshida Ponkan | Kagoshima University | - | Poly | Ponkam/Satsuma mandarin |
| 37 | <i>C. genshokan</i> hort. ex Tanaka | <i>C. reticulata</i> Blanco | Genshokan | | Kagoshima University | - | Poly | Ponkam/Satsuma mandarin |
| 38 | <i>C. deliciosa</i> Ten. | <i>C. reticulata</i> Blanco | Mediterranean mandarin | | National Institute of Fruit Tree Science, Japan | JP117393 | Poly | Ponkam/Satsuma mandarin |
| 39 | <i>C. tangerina</i> Tanaka | <i>C. reticulata</i> Blanco | Dancy | | National Institute of Fruit Tree Science, Japan | JP117396 | Poly | Ponkam/Satsuma mandarin |
| 40 | <i>C. sunki</i> hort. ex Tanaka | <i>C. reticulata</i> Blanco | Sunki | | Kagoshima University | - | Poly | Ponkam/Satsuma mandarin |
| 41 | <i>C. retini</i> hort. ex Tanaka | <i>C. reticulata</i> Blanco | Cleopatra | | Kagoshima University | - | Poly | Ponkam/Satsuma mandarin |
| 42 | <i>C. aurantium</i> L. | <i>C. aurantium</i> L. | Sour orange | Kabusu | Kagoshima University | - | Poly | Sour orange |
| 43 | <i>Citrus unshiu</i> (Swingle) Marcow. | <i>C. reticulata</i> Blanco | Satsuma mandarin | Original strain | Saga University | 1300 | Poly | Ponkam/Satsuma mandarin |
| 44 ^a | <i>Citrus unshiu</i> (Swingle) Marcow. | <i>C. reticulata</i> Blanco | Satsuma mandarin | Aoshima Unshu | Saga Prefectural Fruit Tree Research Station | 1401 | Poly | Ponkam/Satsuma mandarin |
| 45 ^a | <i>Citrus unshiu</i> (Swingle) Marcow. | <i>C. reticulata</i> Blanco | Satsuma mandarin | Aoshima Unshu | Saga Prefectural Fruit Tree Research Station | 1401 | Poly | Ponkam/Satsuma mandarin |
| 46 ^a | <i>Citrus unshiu</i> (Swingle) Marcow. | <i>C. reticulata</i> Blanco | Satsuma mandarin | Aoshima Unshu | Saga Prefectural Fruit Tree Research Station | 1401 | Poly | Ponkam/Satsuma mandarin |

^a Technical replicates.

^b The accessions whose *matK* sequences were determined in this study.

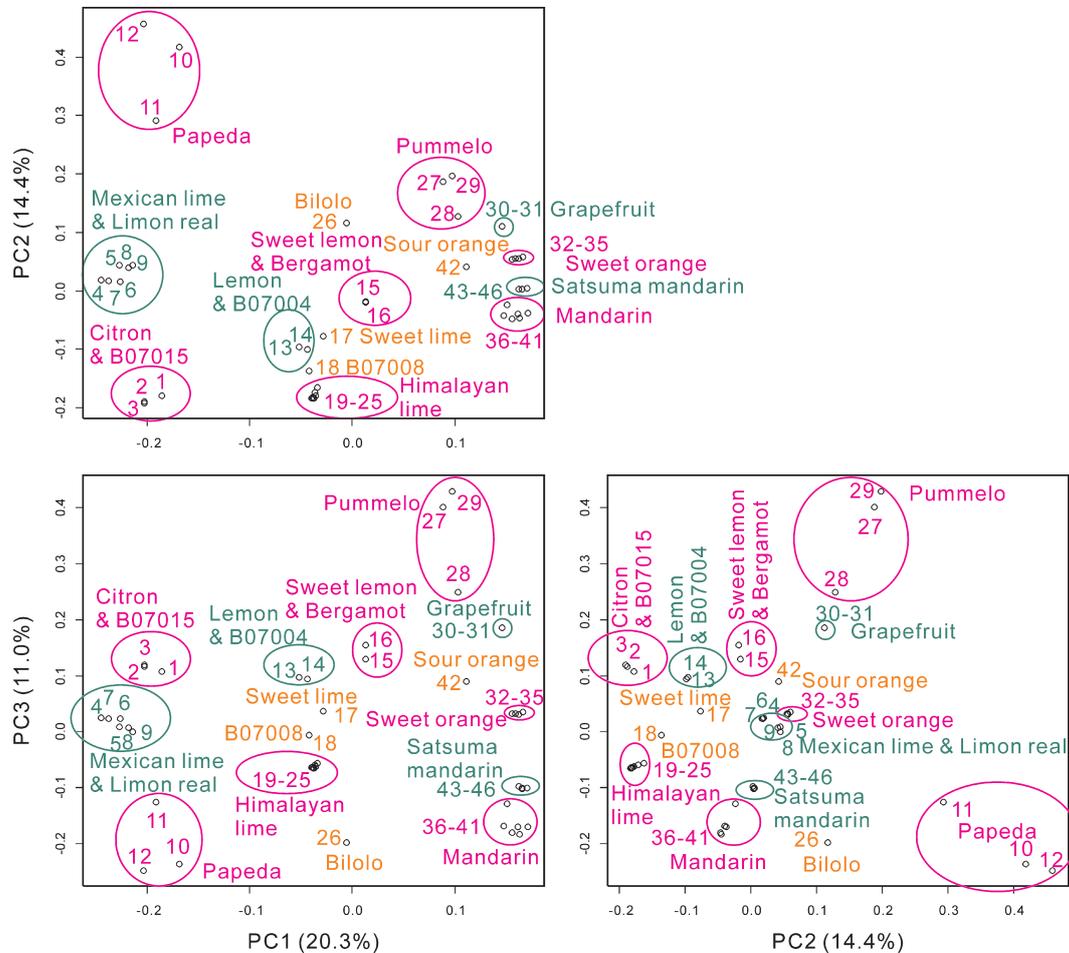


Fig. 1. Principal component analysis (PCA) representation of the accessions used in this study. The results of the first three components are shown. Three-dimensional data was shown by three two-dimensional data. The contribution rate of each component is shown in parentheses. Colors were used to distinguish between clusters.

the value 4 is important because of the following reasons: 1) previous reports proposed that citron, papeda, pummelo, and mandarin are ancestral species, and that most *Citrus* species are derivatives or hybrids of these four species, which is consistent with the current analysis. 2) Admixture history was not predicted in the case of $K = 5-9$. For example, in the case of $K = 5-8$, the admixture history of Mexican lime was not predicted, and in the case of $K = 9$, the admixture history of sweet orange was not predicted. 3) The cross-validation (CV) errors were calculated to estimate the possible K values according to the manual of ADMIXTURE software (Alexander *et al.* 2009), which describes that an acceptable value of K will exhibit a low CV error compared to other K values. The CV error of $K = 4$ was not the lowest value, but was at the bottom of the graph (Supplemental Fig. 4).

ddRAD-Seq readily detects conservation of heterozygosity

To validate whether inexpensive ddRAD-Seq can be applied to determine a single tree origin, conservation of heterozygosity was analyzed. As shown below, we describe the

cases of grapefruit, sweet orange, satsuma mandarin, Mexican lime, and Himalayan lime, each of which is considered to be derived from a single tree by asexual reproduction.

The heterozygous sites of two grapefruit accessions were found to be conserved (Supplemental Fig. 5; 1037 of 1089 sites are conserved heterozygous sites (95.2%)). The heterozygous sites of four sweet orange accessions were conserved (Supplemental Fig. 6; 894 of 949 sites were conserved heterozygous sites (94.2%)). In the analysis of satsuma mandarin, two accessions (No. 44–46 which were technical replicates of the same cultivar, and No. 43 which was an original tree) were used. The heterozygous sites within satsuma mandarin were conserved (Supplemental Fig. 7; 562 sites among 597 sites were conserved heterozygous sites (94.1%)), reflecting historical records (Tanaka 1932).

In our previous analyses (Tshering Penjor *et al.* 2014a) using traditional RAD-Seq (single-digest RAD-Seq using the *EcoRI*) (Baird *et al.* 2008), the limes were classified into two distinct genetic clusters, Indonesian and Bhutanese clusters. Three accessions from Indonesia (No. 6–8) and

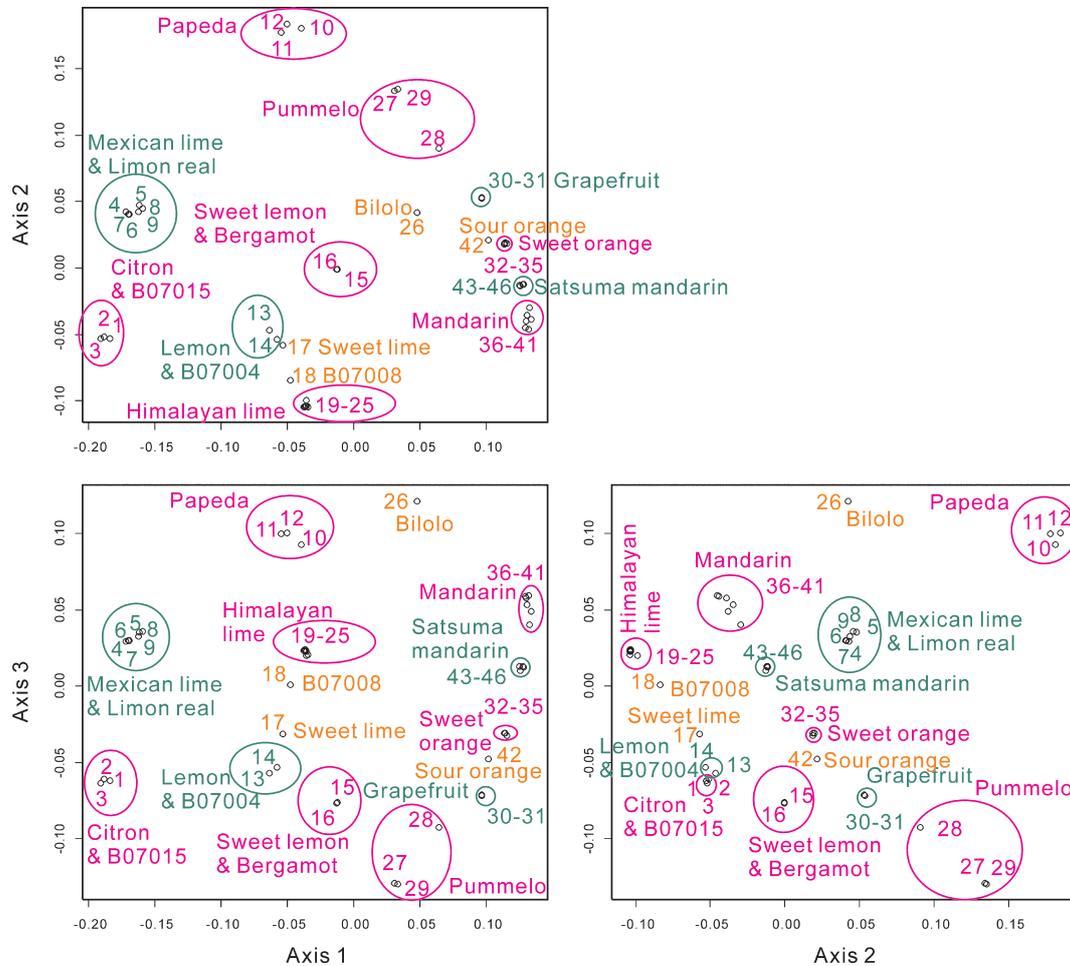


Fig. 2. Multidimensional scaling (MDS) representation of the accessions used in this study. Three-dimensional data were obtained in this analysis. Three-dimensional data was shown by three two-dimensional data. Colors were used to distinguish between the clusters.

one accession from Bhutan (No. 5) belonged to the former cluster. In this study, we reanalyzed these four accessions using ddRAD-Seq, and found that these four accessions (No. 5–8 in **Figs. 1, 2**) grouped with the Saga University standard strain of Mexican lime (No. 4). The multiple sequence alignments created for five accessions (**Supplemental Fig. 8**) of Mexican lime showed that 827 of 1044 sites were conserved heterozygous sites (79.2%). In the previous analysis (Tshering Penjor *et al.* 2014a), we classified the Indonesian cluster into two subclusters, Indonesian subcluster 1, containing accessions B09005 (No. 5) and I88065 (No. 8), and subcluster 2, containing I88035 (No. 6) and I88045 (No. 7). The standard strain of Mexican lime (No. 4) was found to belong to the latter subcluster (No. 4, 6, and 7 in **Figs. 1, 2**). The multiple alignments (**Supplemental Figs. 9, 10**) showed that 1086 of 1144 sites were conserved heterozygous sites (94.9%) within subcluster 1, and 1033 of 1104 sites were conserved heterozygous sites (93.6%) within subcluster 2. Therefore, heterozygous sites were well conserved within a cluster or within a subcluster. Each cluster/subcluster was probably derived from a single tree by asexual reproduction, as proposed previously (Tshering

Penjor *et al.* 2014a). Current ddRAD-Seq analysis again confirmed that Mexican limes can be subdivided into two subclusters.

The previous study assigned four accessions from Bhutan (B09015, B09024, B09027, and B09030) (No. 19–22) to the Bhutanese lime cluster (Tshering Penjor *et al.* 2014a). The additional three accessions (B07006, B07007, and B07009) (No. 23–25) (Tshering Penjor *et al.* 2014b) that were newly analyzed by ddRAD-Seq in this study also belonged to this group. We designated these trees as Himalayan lime, and the probable genetic positions of Himalayan lime are described in the next section. As shown in **Supplemental Fig. 11**, 1034 of 1093 sites were conserved heterozygous sites (94.6%).

Probable genetic positions of Bhutanese and several minor accessions

Because the results of the analysis using typical accessions supported previous reports (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000), the genetic positions of Bhutanese and several minor accessions were investigated.

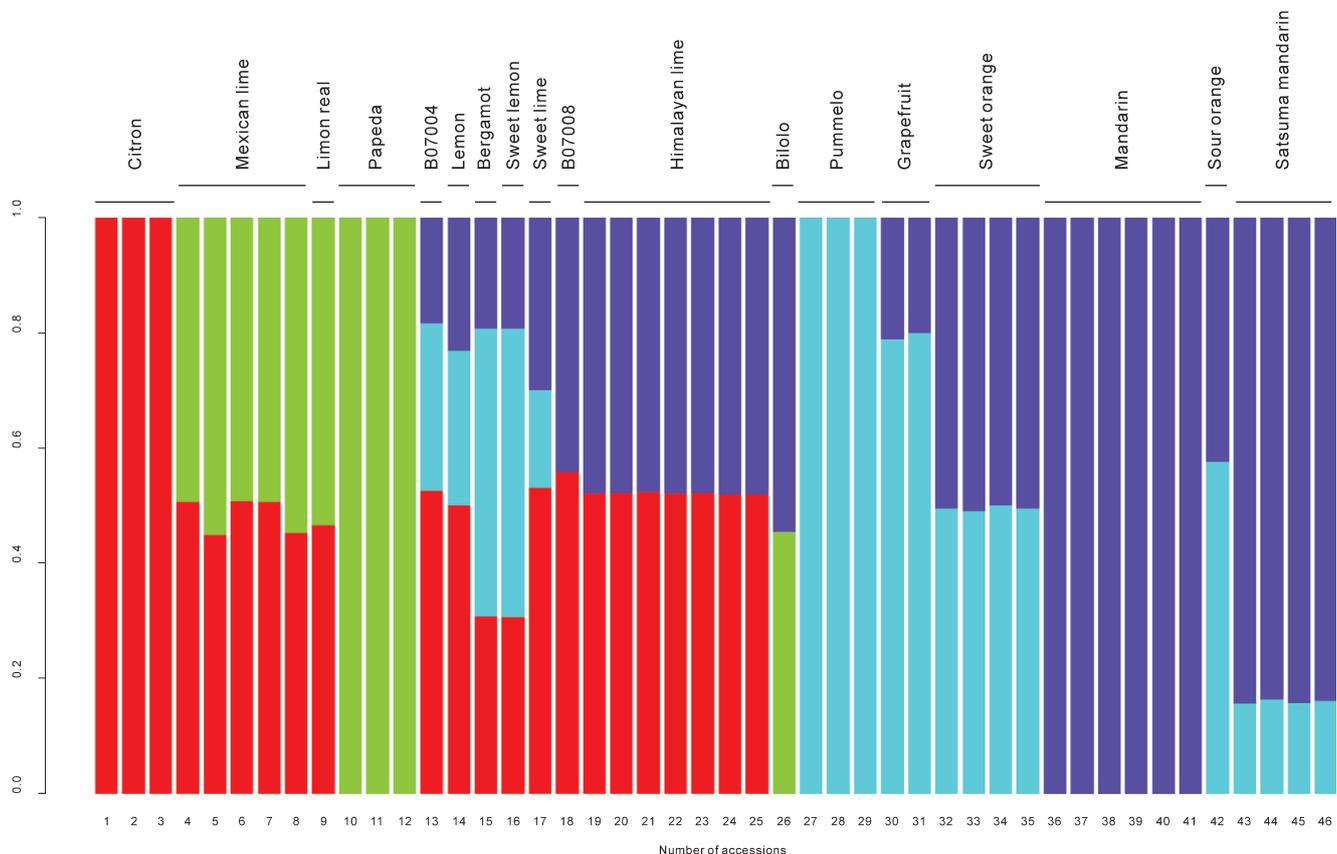


Fig. 3. Admixture analysis of the accessions used in this study. The number of populations (K) was set to 4.

The genetic positions of the seven accessions (No. 19–25 in **Figs. 1, 2**) of Himalayan lime were clearly different from those of the Mexican lime, despite these seven plants possessing the usual morphological features of limes (Tshering Penjor *et al.* 2014a). This type of lime was found to be genetically intermediate between mandarin and citron (**Figs. 1, 2**), suggesting that it is a hybrid. Admixture analysis ($K = 4$) also supported this possibility (**Fig. 3**). Therefore, Himalayan lime and Mexican lime are morphologically similar but genetically dissimilar accessions, and the genetic origin of Himalayan lime is different from that of Mexican lime. As shown previously in Tshering Penjor *et al.* (2014b), plastidic *matK* sequences of Himalayan limes were identical to those of typical mandarins, including ponkan (*C. reticulata* Blanco) and satsuma mandarin (**Table 1, Supplemental Fig. 12**), suggesting a mandarin maternal origin.

Bhutanese accession B07008 is morphologically similar to Himalayan lime, although the fruit size is slightly larger than that of Himalayan lime (Tshering Penjor *et al.* 2014b). However, its genetic position was slightly different from the Himalayan lime cluster (No. 18 in **Figs. 1, 2**), although admixture analysis ($K = 4$) suggested that B07008 was a hybrid between mandarin and citron (**Fig. 3**). The multiple sequence alignments (**Supplemental Fig. 13**) showed that 773 of 1264 sites were conserved heterozygous sites (61.2%) within the cluster containing B07008 and Himala-

yan limes. This level of conservation does not definitively indicate that B07008 and Himalayan limes were derived from a single tree by asexual reproduction. The plastidic *matK* sequence of B07008 (**Table 1, Supplemental Fig. 12**) was not identical to that of mandarin, but identical to those of sweet orange and some types of mandarin (e.g., *C. nobilis* Lour. (Kunenbo)), as shown previously in Tshering Penjor *et al.* (2014b). Therefore, the maternal origin of B07008 is different from that of Himalayan lime.

Accession B07004 from Bhutan possesses the morphological features of citron, although it has polyembryonic seeds (Tshering Penjor *et al.* 2014b). However, this accession was genetically similar to lemon (No. 13 in **Figs. 1, 2**). As shown previously (Tshering Penjor *et al.* 2014b), the plastidic *matK* sequence of B07004 was also identical to that of lemon (**Table 1, Supplemental Fig. 12**), suggesting that the maternal origin of B07004 was identical or similar to that of lemon. The maternal origin of lemon may be from a sour orange relative (**Table 1, Supplemental Fig. 12**). Thus, B07004 and lemon are genetically similar but morphologically dissimilar accessions. In addition, B07004 and citron are morphologically similar but genetically dissimilar accessions. The conservation of the heterozygous sites was absent between B07004 and lemon (**Supplemental Fig. 14**).

Accession B07015, also collected in Bhutan, possesses the morphological features of citron, yet has polyembryonic

seeds (Tshering Penjor *et al.* 2014b). This accession belonged to the cluster containing citrons (No. 3 in **Figs. 1, 2**). The plastidic *matK* sequence of B07015 (**Table 1, Supplemental Fig. 12**) was not identical to the two citron accessions tested, but was instead found to belong to the clade containing the two accessions, as shown previously in Tshering Penjor *et al.* (2014b).

Genetic analysis revealed that limon real (*C. excelsa* Webster; No. 9 in **Figs. 1, 2**) belonged to the cluster containing Mexican lime. The leaves of limon real are similar to those of Mexican lime, although their wings are slightly larger than those of Mexican lime (**Supplemental Fig. 15**). However, available literature and web information for the accession for limon real is sparse and the specimen preserved in Kagoshima University has not yet produced fruit. The multiple alignments (**Supplemental Fig. 16**) showed that 617 of 1186 sites were conserved heterozygous sites (52.0%) within the cluster containing both limon real and Mexican limes. Therefore, the positions of the heterozygous sites of limon real are similar to those of Mexican lime, but this level of conservation does not definitively indicate that both were derived from a single tree by asexual reproduction. The *matK* sequence was identical to those of papeda and Mexican lime (**Table 1, Supplemental Fig. 12**). Therefore, the maternal origin of limon real is similar to that of Mexican lime.

Genetic analysis revealed sweet lime (*C. limettioides* Tanaka; No. 17 in **Figs. 1, 2**) was somewhat similar to lemon and B07004. The heterozygous sites among the three plants were not conserved (**Supplemental Fig. 14**). Admixture analysis ($K = 4$) (**Fig. 3**) suggested that sweet lime is a hybrid between citron and a plant containing the genomes of both mandarin and pummelo (e.g., sour orange (*C. aurantium* L.) and sweet orange). Indeed, in our PCA and MDS results sweet lime was located intermediate between citron and sour orange, or between citron and sweet oranges (**Figs. 1, 2**). As shown previously (Tshering Penjor *et al.* 2013), the plastidic *matK* sequence of sweet lime is identical to that of sweet orange and some types of mandarin (e.g., *C. nobilis* Lour. (Kunenbo)) (**Table 1, Supplemental Fig. 12**), showing that the maternal origin of sweet lime is different from that of lemon and B07004, and is more similar to plants containing the genomes of both mandarin and pummelo (e.g., sweet orange).

Sweet lemon (*C. limetta* Risso; No. 16 in **Figs. 1, 2**) is served as juice in South Asian countries, and bergamot (*C. bergamia* Risso & Poit.; No. 15 in **Figs. 1, 2**) is used for cosmetics and Earl Grey tea. These two accessions were genetically similar (**Figs. 1, 2**), and therefore are classed as genetically similar but morphologically dissimilar accessions. The heterozygous sites were highly conserved between the two plants (**Supplemental Fig. 17**) with 955 of 1002 sites being conserved (95.3%). This conservation suggests that both were derived from a single tree by asexual reproduction. Admixture analysis ($K = 4$) (**Fig. 3**) suggested that the ratio of the contribution of pummelo, citron, and

mandarin to their genome is 2:1:1, respectively. Both the PCA and MDS analyses matched the result of the admixture analysis (**Figs. 1, 2**). The plastidic *matK* sequences of sweet lemon and bergamot were identical to each other, and were identical to those of lemon, Bearss lime (*Citrus latifolia* (Yu.Tanaka) Tanaka), myrtle-leaf orange (*C. myrtifolia* Raf.), and B07004 (**Table 1, Supplemental Fig. 12**). Although pummelo and mandarin do not typically carry this type of *matK* sequence, some of them may carry this type of maternal sequence.

Bilolo (*C. montana* Tanaka; No. 26 in **Figs. 1, 2**) (Tanaka 1948) was located between papeda and mandarin and may be a hybrid between papeda and mandarin. Admixture analysis ($K = 4$) supported this hypothesis (**Fig. 3**). The *matK* sequence was identical to those of papedas (**Table 1, Supplemental Fig. 12**), suggesting a papeda maternal origin. The fruit of this plant was round and had a slightly rough exterior, although the degree of roughness was less than that of papeda (**Supplemental Fig. 15**). The color of fruit pulp was light yellow, and the size of the leaf wing was small, which are morphological features similar to those of mandarin rather than those of papeda.

Discussion

Both PCA and MDS analyses showed that each of four ancestral species are clustered together (**Figs. 1, 2**) as follows; two accessions (No. 1 and 2) of citrons, three accessions of papedas (No. 10–12), three accessions (No. 27–29) of pummelo, and six accessions of mandarin (No. 36–41). Heterozygous sites are not conserved within the members of each cluster (**Supplemental Figs. 18–21**), contrasting with the other species, such as grapefruit, sweet orange, satsuma mandarin, Mexican lime, and Himalayan lime. Each cluster is located at one of four corners, which supports the notion that these four are ancestral species.

Some *Citrus* species are considered hybrids of these four ancestral species (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000), and the current analysis supports this thought. As shown below, we describe the genetic origins of Mexican lime, grapefruit, sweet orange, sour orange, and lemon.

Mexican lime has been proposed as a probable hybrid between papedas and citron (Barkley *et al.* 2006, Bayer *et al.* 2009, Nicolosi *et al.* 2000), and the current analysis supports this understanding as the five accessions are found to be intermediates between citron and papeda (No. 4–8 in **Figs. 1, 2**). Admixture analysis ($K = 4$) (**Fig. 3**) also supports the hybrid status of Mexican lime. As shown previously (Tshering Penjor *et al.* 2013), the plastid *matK* sequence of Mexican lime is identical to those of papedas (**Table 1, Supplemental Fig. 12**), suggestive of a papeda maternal origin.

Two accessions of grapefruit (No. 30 and 31 in **Figs. 1, 2**) cluster together. Grapefruits have been proposed to be a hybrid between pummelo and sweet orange (Barrett and

Rhodes 1976, Scora *et al.* 1982), and the current data (Figs. 1, 2) show that the two accessions are located intermediate between pummelo and sweet orange. Admixture analysis ($K = 4$) (Fig. 3) also supports this interpretation.

Four accessions of sweet orange (No. 32–35 in Figs. 1, 2) cluster together. Sweet orange is a probable hybrid between pummelo and mandarin (Barkley *et al.* 2006, Barrett and Rhodes 1976, Berhow *et al.* 2000, Fang and Roose 1997, Nicolosi *et al.* 2000, Torres *et al.* 1978). The current analyses (Figs. 1–3) suggest that the genomes of pummelo and mandarin contribute to the sweet orange lineage, supporting the previous proposition. However, it is difficult to analyze the details of their contributions as we only analyzed the reduced genome. Several hypotheses are discussed by other studies in which whole genome sequencing was conducted (Wu *et al.* 2014, Xu *et al.* 2012).

Sour orange is also considered a probable hybrid between pummelo and mandarin (Barkley *et al.* 2006, Berhow *et al.* 2000, Gulsen and Roose 2001, Moore 2001, Nicolosi *et al.* 2000, Scora 1975). The current analysis shows that the genomes of pummelo and mandarin may have contributed to that of the sour orange (No. 42 in Figs. 1–3), although it is difficult to determine the details of their contributions due to the restrictions of analyzing the reduced genome, as was the case with sweet orange.

Lemon has been proposed as a hybrid between sour orange and citron (Gulsen and Roose 2001, Nicolosi *et al.* 2000). Lemon is located intermediate between citron and sour orange using PCA and MDS analyses (Figs. 1, 2). Admixture analysis ($K = 4$) (Fig. 3) also supports this interpretation, i.e., lemon is a hybrid between citron and a plant containing the genomes of both mandarin and pummelo, and sour orange has the genomes of both mandarin and pummelo as described above. Despite the *matK* sequence of lemon not being identical to that of sour orange, it still belongs to the clade containing sour orange sequences (Table 1, Supplemental Fig. 12) (Tshering Penjor *et al.* 2013). Therefore, the maternal origin of lemon may be from a sour orange relative, which may not be sour orange itself. We would like to note here that the plastidic *matK* sequences of Bearss lime and myrtle-leaf orange are identical to that of lemon (Table 1, Supplemental Fig. 12) (Tshering Penjor *et al.* 2013). Because myrtle-leaf orange resembles sour orange morphologically, myrtle-leaf orange may contain the genomes of both mandarin and pummelo, and therefore could be identical or similar to the maternal parent of lemon. However, as Bearss lime and myrtle-leaf orange remain to be analyzed by RAD-Seq, further study is required to test this hypothesis.

The method of testing for the conservation of heterozygosity can readily detect single tree origins. We would like to note that, for determining varietal identity, it is important to test not only the genetic positions by PCA or MDS, but also the conservation of heterozygosity, because some accessions are located at similar genetic positions, but heterozygous sites may not be conserved. For example, our

mandarin accessions (No. 36–41) clustered together, but their heterozygous sites are not conserved (Supplemental Fig. 21). Furthermore, this method detected the conservation of heterozygosity within each subtype of Mexican lime. Thus, this inexpensive method may be useful to check the variety of trees originating from a single tree, though this may be limited to a case-by-case basis.

We found morphologically similar but genetically dissimilar accessions of *Citrus* species: 1) Mexican lime (probable hybrid between papedas and citron) and Himalayan lime (probable hybrid between mandarin and citron), and 2) accession B07004 and citron. Likewise, we found genetically similar but morphologically dissimilar accessions of *Citrus* species: 1) accession B07004 and lemon, and 2) sweet lemon and bergamot. Therefore, in *Citrus* species morphological similarities are not always predictive of genetic similarities, and vice versa. Consequently, it is worthwhile to analyze minor accessions, such as the local accessions from Bhutan. Until now, *Citrus* species have been classified primarily based on their morphologies. However, the current analyses based on both morphology and genomics suggest that there are more varieties in *Citrus* species than previously thought. Because classification is essential for breeding programs, it is important to identify *Citrus* species using both methods.

Citron, papeda, pummelo, and mandarin have all been proposed to be ancestral species, and most *Citrus* species may be derivatives or hybrids of these four varieties (Curk *et al.* 2014, 2015, Froelicher *et al.* 2011, Garcia-Lor *et al.* 2013, Nicolosi *et al.* 2000). However, citron \times mandarin and papeda \times mandarin were missing links until now. In this study, we found the possible links—Himalayan lime and bilolo, respectively. In addition, a new type of citron \times papeda, limon real, was found. These findings again emphasize the importance of analyzing the minor accessions.

One of the pummelo accessions (Suisho Buntan; No. 28) might have received gene flow from mandarin, as it is proximal to mandarin in the PCA and MDS analyses (Figs. 1, 2). Suisho Buntan seems to be an artificial hybrid derived from the cross between Banokan and Tosa Buntan (Tanaka 1980). Although both parental cultivars are classified as pummelo based on their morphological traits, Banokan might not be a true pummelo since pummelos have monoembryonic seeds, but Banokan has polyembryonic seeds (Ueno *et al.* 1967). Banokan therefore may have originated from a cross between the monoembryonic pummelo and polyembryonic mandarin.

Two accessions of satsuma mandarin belong to the same group and were separated slightly from the cluster containing the other mandarins (Figs. 1, 2). Our analyses (Figs. 1–3) suggest that satsuma mandarin may have received gene flow from pummelo. Gene flow from pummelo to mandarin varieties such as Ponkan has already been suggested by Wu *et al.* (2014), and the amount of gene flow may be greater in satsuma mandarin than in other mandarins.

The conservation of heterozygous sites suggests that sweet lemon and bergamot were derived from a single tree via asexual reproduction. *De novo* mutations might result in the morphological diversification observed in the current study. We have previously proposed that a phenomenon known as the loss of heterozygosity, wherein heterozygous sites somatically change to homozygous sites, contributed to the genetic diversification of limes (Tshering Penjor *et al.* 2014a), especially in the formation of the two subtypes of Mexican lime. This may also contribute to the emergence of observed morphological differences.

The conservation of heterozygous sites is absent between B07004 and lemon (**Supplemental Fig. 14**), yet the two accessions were genetically similar. The current analysis suggests possible instances of self-reproduction or crossing within the cluster, although a confounding factor is that lemon and accession B07004 carry polyembryonic seeds. This factor however may not be diagnostic as lemons usually have an extremely small embryonic number (Ueno *et al.* 1967), or indeed are occasionally monoembryonic. This seed property may be related to sexual reproduction.

It is difficult to interpret the relationships between Himalayan lime and accession B07008, in which about 60% of heterozygous sites are conserved (**Supplemental Fig. 13**). Here, we provide two potential alternative explanations: 1) a previous cross between citron and mandarin might have resulted in the emergence of Himalayan lime. In a similar fashion, a cross between citron and another type of mandarin, such as *C. nobilis* (Kunenbo), or a cross between citron and a mandarin-pummelo hybrid (e.g., sweet orange) might have resulted in the emergence of the ancestor of accession B07008, after which the trees might have reproduced asexually. 2) Previous crosses between types of citron and types of mandarin, with the possibility of gene flow received from pummelo, might have resulted in the emergence of hybrids between mandarin and citron. Although they generally reproduce asexually, the two types of trees that emerged might have rarely crossed. For example, a cross between sweet lime and Himalayan lime may have produced accession B07008. This does not conflict with the observation that the number of heterozygous sites in Himalayan lime is nearly identical to that of accession B07008, or with the results of the PCA and MDS analyses (**Figs. 1, 2**), although the admixture analysis does not predict this possibility (**Fig. 3**).

An issue similar to the case of Himalayan lime and accession B07008 is also present in the relationship between Mexican lime and limon real. Two explanations for Himalayan lime and accession B07008 are possible in the case where mandarin is replaced by papeda in the above-mentioned two possibilities. In addition, because the maternal origins of Mexican lime and limon real are identical, at least in the *matK* sequences, the other following explanation is possible; both Mexican lime and limon real might be derived from a single tree by asexual propagation, and some mechanism such as loss of heterozygosity might have pro-

duced the large genetic differences over time.

In this study, we genetically analyzed small numbers of accessions native to Bhutan. In Bhutan, many citrus accessions, like citron and Himalayan limes, are grown in the wild. Ichang papeda and rough lemon are also grown in the wild, but we did not analyze them in the current study. Although some accessions of these trees can be found growing in backyard orchards of some Bhutanese farmers, these trees are actually brought from the forest by farmers as either seedlings or seeds and grown in their fields. Most Bhutanese do not use Himalayan limes or citrons for cuisine, with the exception of some Nepalese-speaking Bhutanese people. Therefore, it would be interesting to study more accessions native to Bhutan as well as to other countries.

Acknowledgments

We thank Ms. Akiko Sakai for technical assistance, Mr. Lhap Dorji for providing the name Himalayan lime, Mr. Yuichi Tomiyasu for help with the research performed in Bhutan, and Dr. Makoto Ikenaga for useful suggestions. Part of this work was supported by a Grant-in-Aid for Scientific Research (26450039 to Y.N., 19405019 to M.Y.) and RONPAKU (Dissertation PhD) Program (11401 to T.P.) from the Japan Society for the Promotion of Science. The RAD-Seq analysis was supported by the Joint Usage/Research Program of the Center for Ecological Research, Kyoto University. We would like to thank Editage for English language editing.

Literature Cited

- Alexander, D.H., J. Novembre and K. Lange (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19: 1655–1664.
- Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko and E.A. Johnson (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.
- Barkley, N.A., M.L. Roose, R.R. Krueger and C.T. Federici (2006) Assessing genetic diversity and population structure in a citrus germplasm collection utilizing simple sequence repeat markers (SSRs). *Theor. Appl. Genet.* 112: 1519–1531.
- Barrett, H.C. and A.M. Rhodes (1976) A numerical taxonomic study of affinity relationships in cultivated Citrus and its close relatives. *Syst. Bot.* 1: 105–136.
- Bayer, R.J., D.J. Mabberley, C. Morton, C.H. Miller, I.K. Sharma, B.E. Pfeil, S. Rich, R. Hitchcock and S. Sykes (2009) A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. *Am. J. Bot.* 96: 668–685.
- Berhow, M.A., S. Hasegawa, K. Kwan and R.D. Bennett (2000) Limonoids and the chemotaxonomy of Citrus and the Rutaceae family. In: Berhow, M.A., S. Hasegawa and G.D. Manners (eds.) *Citrus Limonoids: Functional Chemicals in Agriculture and Food*, American Chemical Society, Washington DC, pp. 212–228.
- Catchen, J.M., A. Amores, P. Hohenlohe, W. Cresko and J.H. Postlethwait (2011) *Stacks*: building and genotyping loci *de novo*

- from short-read sequences. *G3* (Bethesda) 1: 171–182.
- Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores and W.A. Cresko (2013) Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22: 3124–3140.
- Curk, F., G. Ancillo, A. Garcia-Lor, F. Luro, X. Perrier, J.P. Jacquemoud-Collet, L. Navarro and P. Ollitrault (2014) Next generation haplotyping to decipher nuclear genomic interspecific admixture in *Citrus* species: analysis of chromosome 2. *BMC Genet.* 15: 152.
- Curk, F., G. Ancillo, F. Ollitrault, X. Perrier, J.P. Jacquemoud-Collet, A. Garcia-Lor, L. Navarro and P. Ollitrault (2015) Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern citrus varieties. *PLoS ONE* 10: e0125628.
- Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry *et al.* (2011) The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.
- Fang, D.Q. and M.L. Roose (1997) Identification of closely related citrus cultivars with inter-simple sequence repeat markers. *Theor. Appl. Genet.* 95: 408–417.
- Froelicher, Y., W. Mouhaya, J.B. Bassene, G. Costantino, M. Kamiri, F. Luro, R. Morillon and P. Ollitrault (2011) New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. *Tree Genet. Genomes* 7: 49–61.
- Garcia-Lor, A., F. Curk, H. Snoussi-Trifa, R. Morillon, G. Ancillo, F. Luro, L. Navarro and P. Ollitrault (2013) A nuclear phylogenetic analysis: SNPs, indels and SSRs deliver new insights into the relationships in the “true citrus fruit trees” group (Citrinae, Rutaceae) and the origin of cultivated species. *Ann. Bot.* 111: 1–19.
- Gmitter, F.G. and X. Hu (1990) The possible role of Yunnan, China, in the origin of contemporary citrus species (Rutaceae). *Econ. Bot.* 44: 267–277.
- Gulsen, O. and M.L. Roose (2001) Lemons: diversity and relationships with selected *Citrus* genotypes as measured with nuclear genome markers. *J. Am. Soc. Hortic. Sci.* 126: 309–317.
- Iwamasa, M., N. Nito and J.T. Ling (1988) Intra- and intergeneric hybridization in the orange subfamily, Aurantioideae. *In: Goren, R. and K. Mendel* (eds.) *Proc. Int. Soc. Citricult.* 1. Balaband, Rehovot, Israel and Margrat Publishers, Weikersheim, Germany, pp. 123–130.
- Langmead, B., C. Trapnell, M. Pop and S.L. Salzberg (2009) Ultrafast and memory efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10: R25.
- Moore, G.A. (2001) Oranges and lemons: clues to the taxonomy of *Citrus* from molecular markers. *Trends Genet.* 17: 536–540.
- Nicolosi, E., Z.N. Deng, A. Gentile, S. La Malfa, G. Continella and E. Tribulato (2000) Citrus phylogeny and genetic origin of important species as investigated by molecular markers. *Theor. Appl. Genet.* 100: 1155–1166.
- Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher and H.E. Hoekstra (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7: e37135.
- Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81: 559–575.
- Sakaguchi, S., T. Sugino, Y. Tsumura, M. Ito, M.D. Crisp, D.M.J.S. Bowman, A.J. Nagano, M.N. Honjo, M. Yasugi, H. Kudoh *et al.* (2015) High-throughput linkage mapping of Australian white cypress pine (*Callitris glaucophylla*) and map transferability to related species. *Tree Genet. Genomes* 11: 121.
- Scora, R.W. (1975) On the history and origin of Citrus. *Bull. Torrey Bot. Club.* 102: 369–375.
- Scora, R.W., J. Kumamoto, R.K. Soost and E.M. Nauer (1982) Contribution to the origin of the grapefruit, *Citrus paradisi* (Rutaceae). *Syst. Bot.* 7: 170–177.
- Swingle, W. and P. Reece (1967) The botany of Citrus and its wild relatives. *In: Reuther, W., H. Webber and L. Batchelor* (eds.) *The citrus industry*, University of California, Berkeley, California, pp. 190–430.
- Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731–2739.
- Tanaka, T. (1932) A monograph of the Satsuma orange. *Mem. Fac. Sci. & Agric., Taihoku Imp. Univ.* 4: 1–626.
- Tanaka, T. (1959) A revision of Assam citrus. *Bull. Univ. Osaka Pref. Ser. B.* 9: 29–39.
- Tanaka, T. (1977) Fundamental discussion of Citrus classification. *Stud. Citrol.* 14: 1–6.
- Tanaka, Y. (1948) An iconograph of Japanese citrus fruits. Vol. 2. Yokendo, Tokyo.
- Tanaka, Y. (1980) An iconograph of Japanese citrus fruits: A monographic study of species and varieties of citrus fruits grown in Japan, Vol. III. Yokendo, Tokyo.
- Torres, A.M., R.K. Soost and U. Diedenhofen (1978) Leaf isozymes as genetic markers in Citrus. *Am. J. Bot.* 65: 869–881.
- Tshering Penjor, T. Anai, Y. Nagano, R. Matsumoto and M. Yamamoto (2010) Phylogenetic relationships of *Citrus* and its relatives based on *rbcl* gene sequences. *Tree Genet. Genomes* 6: 931–939.
- Tshering Penjor, M. Yamamoto, M. Uehara, M. Ide, N. Matsumoto, R. Matsumoto and Y. Nagano (2013) Phylogenetic relationships of *Citrus* and its relatives based on *matK* gene sequences. *PLoS ONE* 8: e62574.
- Tshering Penjor, T. Mimura, R. Matsumoto, M. Yamamoto and Y. Nagano (2014a) Characterization of limes (*Citrus aurantifolia*) grown in Bhutan and Indonesia using high-throughput sequencing. *Sci. Rep.* 4: 4853.
- Tshering Penjor, Y. Nagano, T. Mimura, R. Matsumoto and M. Yamamoto (2014b) Exploration of local citrus genetic resources in Bhutan and their chloroplast DNA analysis. *Hort. Res. (Japan)* 13: 307–314.
- Ueno, I., M. Iwamasa and M. Nishiura (1967) Embryo number of various varieties of citrus and its relatives. *Bull. Hort. Res. Sta. Japan.* B7: 11–21.
- Wu, G.A., S. Prochnik, J. Jenkins, J. Salse, U. Hellsten, F. Murat, X. Perrier, M. Ruiz, S. Scalabrin, J. Terol *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat. Biotechnol.* 32: 656–662.
- Xu, Q., L.L. Chen, X. Ruan, D. Chen, A. Zhu, C. Chen, D. Bertrand, W.B. Jiao, B.H. Hao, M.P. Lyon *et al.* (2012) The draft genome of sweet orange (*Citrus sinensis*). *Nat. Genet.* 45: 59–66.
- Zheng, X., D. Levine, J. Shen, S.M. Gogarten, C. Laurie and B.S. Weir (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28: 3326–3328.