Breeding Science 73: 269–277 (2023)
doi: 10.1270/jsbbs.22071

Research Paper

Elucidation of genetic variation and population structure of melon genetic resources in the NARO Genebank, and construction of the World Melon Core Collection

Gentaro Shigita[1,2], Tran Phuong Dung[1], Mst. Naznin Pervin[1], Thanh-Thuy Duong[1,3], Odirich Nnennaya Imoh[1], Yuki Monden[1], Hidetaka Nishida[1], Katsunori Tanaka[4], Mitsuhiro Sugiyama[5], Yoichi Kawazu[5], Norihiko Tomooka[6] and Kenji Kato*[1]

[1] *Graduate School of Environmental and Life Science, Okayama University*, 1-1-1 Tsushima-Naka, Kita-ku, Okayama 700-8530, Japan
[2] *Department of Life Science Systems, Technical University of Munich*, Emil-Ramann Strasse 2, Freising 85354, Germany
[3] *Faculty of Agronomy, University of Agriculture and Forestry, Hue University*, 102 Phung Hung Street, Hue City, Vietnam
[4] *Faculty of Agriculture and Life Science, Hirosaki University*, 3 Bunkyo, Hirosaki, Aomori 036-8561, Japan
[5] *Institute of Vegetable and Floriculture Science, National Agriculture and Food Research Organization (NARO)*, 360 Kusawa, Ano, Tsu, Mie 514-2392, Japan
[6] *Research Center of Genetic Resources, National Agriculture and Food Research Organization (NARO)*, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan

Numerous genetic resources of major crops have been introduced from around the world and deposited in Japanese National Agriculture and Food Research Organization (NARO) Genebank. Understanding their genetic variation and selecting a representative subset (“core collection”) are essential for optimal management and efficient use of genetic resources. In this study, we conducted genotyping-by-sequencing (GBS) to characterize the genetic relationships and population structure in 755 accessions of melon genetic resources. The GBS identified 39,324 single-nucleotide polymorphisms (SNPs) that are distributed throughout the melon genome with high density (one SNP/10.6 kb). The phylogenetic relationships and population structure inferred using this SNP dataset are highly associated with the cytoplasm type and geographical origin. Our results strongly support the recent hypothesis that cultivated melon was established in Africa and India through multiple independent domestication events. Finally, we constructed a World Melon Core Collection that covers at least 82% of the genetic diversity and has a wide range of geographical origins and fruit morphology. The genome-wide SNP dataset, phylogenetic relationships, population structure, and the core collection provided in this study should largely contribute to genetic research, breeding, and genetic resource preservation in melon.

Key Words: *Cucumis melo*, Cucurbitaceae, genotyping-by-sequencing, genetic resource, genetic diversity, crop origin, core collection.

Introduction

Melon (*Cucumis melo* L.), one of the most important crops in the Cucurbitaceae family, is cultivated in over one million hectares to produce over 28 million tons worldwide (FAO). Melon has been classified into two subspecies (subsp. *melo* and subsp. *agrestis*) based on ovary pubescence, and into 19 horticultural varieties based on botanical features (Pitrat 2016). As most wild *Cucumis* species are native in Africa and their chromosome number is the same as melon (2n = 24), it had been thought that melon was domesticated in Africa. However, several studies with controversial results have proposed that the closest wild relatives of melon are found in Australia (Sebastian *et al.* 2010) and India (Endl *et al.* 2018). Therefore, the origin of melon and its domestication history have been a longstanding question. To clarify these complicated intraspecific relationships and to uncover the origin of melon, several molecular biological studies have been conducted. Based on the sequence polymorphisms in the melon chloroplast genome, it was revealed that cultivated melon landraces in the world consist of three distinct maternal lineages (cytoplasm types Ia, Ib, and Ic), indicating their

Communicated by Ryuji Ishikawa
Received August 24, 2022. Accepted February 25, 2023.
First Published Online in J-STAGE on June 15, 2023.
*Corresponding author (e-mail: kenkato@okayama-u.ac.jp)

polyphylogenetic origin(s) (Tanaka *et al.* 2013). A clear and distinct geographical distribution of the three cytoplasm types was identified as follows: Ia type occurs mainly in South Asia, Southeast Asia, East Asia, and a part of Southern Africa; Ib type occurs in South Asia, Central/West Asia, Europe, America, and the northern part of Africa; and Ic type occurs exclusively in sub-Saharan Africa. A relationship between cytoplasm types and seed length has also been observed. The Ia and Ic types tend to have small seeds (<9 mm), whereas the Ib type tends to have relatively large seeds (≥9 mm). Recently, it was suggested that three independent domestication events (two in India and one in Africa) occurred in melon (Zhao *et al.* 2019). Moreover, this hypothesis is supported by subsequent studies (Liu *et al.* 2020, Wang *et al.* 2021).

A core collection is a subset of genetic resources that represents genetic diversity in an entire collection (Frankel and Brown 1984). Construction of core collection enables the efficient use and preservation of crop genetic resources. In addition, by sharing the core collection for breeding and research works, multidimensional information will be accumulated using common materials. National Agriculture and Food Research Organization (NARO) Genebank has developed core collections of several crops, including rice (*Oryza sativa*), maize (*Zea mays*), azuki bean (*Vigna angularis*), wheat (*Triticum aestivum*), soybean (*Glycine max*), sorghum (*Sorghum bicolor*), mungbean (*Vigna radiata*), eggplant (*Solanum melongena*), and foxtail millet (*Setaria italica*); they are publicly distributed for researchers and breeders upon request (Ebana *et al.* 2008, Hirano *et al.* 2011, Kaga *et al.* 2012, Kojima *et al.* 2005, Miyatake *et al.* 2019, Sangiri *et al.* 2007, Shehzad *et al.* 2009, Xu *et al.* 2008). In particular, the World Rice Core Collection has been used in more than 300 research projects since its establishment in 2005, demonstrating the demand for core collection construction (Tanaka *et al.* 2020).

High-throughput sequencing technology has enabled quick and labor-saving acquisition of genome-wide polymorphism data from a large number of samples. Genotyping-by-sequencing (GBS), also known as restriction site-associated DNA sequencing (RAD-seq), is a promising approach for high-throughput SNP discovery and genotyping by using a high-throughput sequencer (Elshire *et al.* 2011). This method enables highly multiplexing of many samples by confining the sequencing region to a few percent of a whole genome. Since its effectiveness was first demonstrated in maize and barley in 2011, GBS has been used in various plant species for various purposes, including phylogenetic study, development of a high-density genetic linkage map, and genome-wide association study (GWAS). GBS has also been applied in cucurbit crops, and its effectiveness for constructing core collections has been demonstrated in cucumber (*Cucumis sativus*) (Wang *et al.* 2018) and melon (Wang *et al.* 2021).

The melon core collection developed by Wang *et al.* (2021) consists of 383 accessions representing more than 98% of genetic variation in 2,083 accessions deposited in the U.S. Department of Agriculture (USDA) National Plant Germplasm System. Their sampling considerably focused on South Asia (708 accessions), Central/West Asia (478 accessions), Europe (339 accessions), and Turkey (204 accessions), which together account for more than 80% of the accessions used. In contrast, the melon genetic resources of NARO Genebank are largely from East and Southeast Asian countries and unique in their area coverage. Therefore, melon core collection of the NARO Genebank would complement the genetic variation that potentially has been underrepresented in the core collection of other genebanks.

In this study, we investigated the genetic variation and population structure of 755 accessions of worldwide melon genetic resources deposited in the NARO Genebank, Japan. Our analyses, based on genome-wide SNP data identified by GBS, revealed that their genetic relationships and population structure were highly associated with the cytoplasm type and geographical origin and supported the hypothesis on the origin and domestication history of melon proposed in previous studies. Finally, we constructed the World Melon Core Collection, which consists of 100 accessions representing at least 82% of the genetic diversity of the entire NARO collection.

Materials and Methods

Plant materials and DNA extraction

A total of 755 melon accessions of *C. melo* were used in this study, comprising 667 accessions deposited in the NARO Genebank, 76 accessions maintained at Okayama University originally introduced from the USDA National Plant Germplasm System, and 12 Japanese commercial cultivars. These accessions were selected to represent the broad geographical distribution of the species: 288 from East Asia, 136 from Europe, 114 from South Asia, 67 from the Americas, 56 from Central/West Asia, 47 from Southeast Asia, 43 from Africa and 4 of unknown origin (**Fig. 1**). In addition, two accessions of *C. sagittatus* (PI 374208 and PI 374209) and one accession of *C. ficifolius* (PI 273648), introduced from the USDA, were used as the outgroup. Detailed information on all accessions used in this study is available in **Supplemental Table 1**.

Young leaves were collected from 1–5 seedlings of each accession and ground in liquid nitrogen for DNA extraction. Genomic DNA (gDNA) was extracted using the cetyltrimethylammonium bromide method (Murray and Thompson 1980) with minor modifications. DNA concentration and purity were measured using NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA).

Cytoplasm type determination

The cytoplasm type of each accession was determined based on genotypes of three SNPs in the melon chloroplast genome. Each SNP was genotyped by a PCR-based method developed in a previous study (Tanaka *et al.* 2016a).

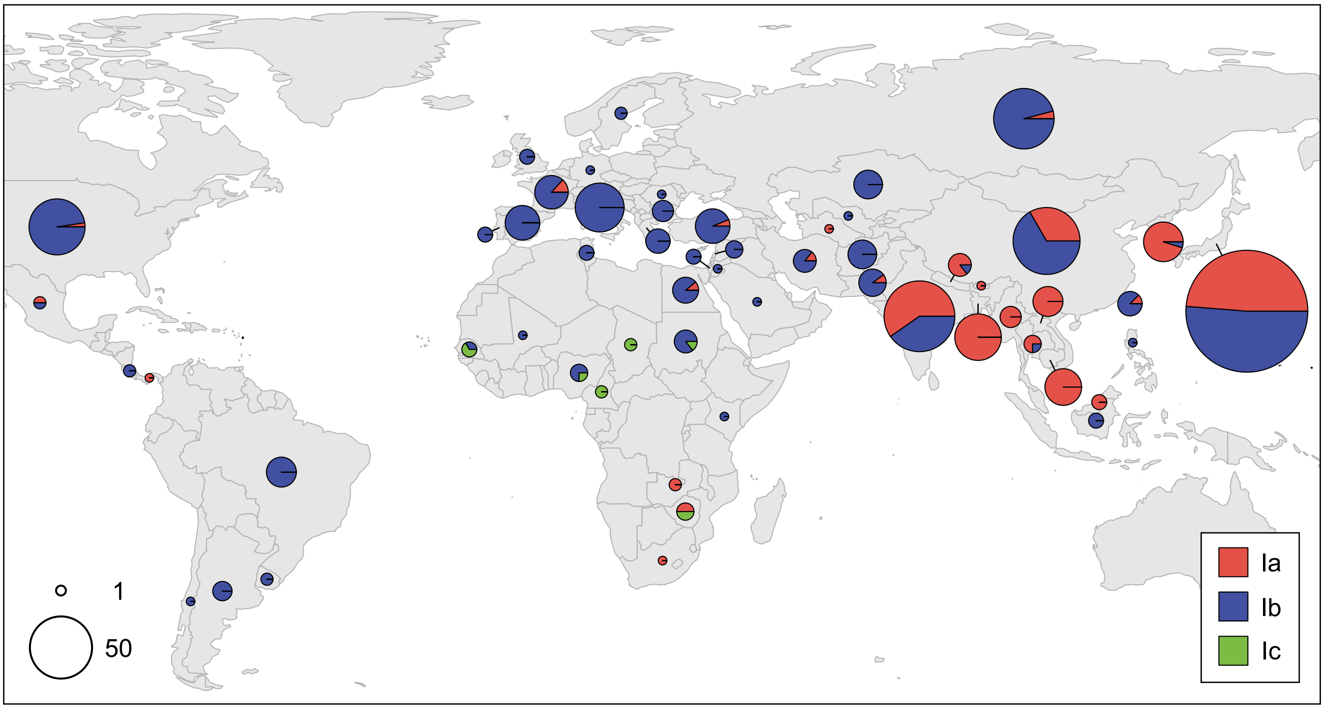

Fig. 1. Geographical origin and cytoplasm types of the 755 melon accessions. The pie charts and their sizes represent the compositions of the three cytoplasm types and the number of used accessions in each country, respectively. Eight accessions with unknown geographical origin are not shown.

Briefly, PCR amplifications were performed in 15 µl reaction volumes comprising 75 ng of gDNA, 1 × PCR buffer (Sigma-Aldrich, Burlington, MA, USA), 1.5 mM $MgCl_2$, 0.8 mM dNTP, 0.5 µM of each primer, and 0.25 U *Taq* DNA polymerase (Sigma-Aldrich). PCRs were performed using iCycler (Bio-Rad, Hercules, CA, USA) with the following program: initial denaturing step at 95°C for 3 min, followed by 35 cycles of 95°C for 1 min, annealing for 2 min, and 72°C for 2 min, and final extension at 72°C for 5 min. The annealing temperature was optimized for each primer pair. The amplicons were digested by a restriction enzyme at 37°C for 3 hours in a 11.3 µL mixture consisted of 10 µL of amplicon, 1 µL of 10 × CutSmart buffer (New England Biolabs [NEB], Ipswich, MA, USA), and 0.3 µL (3 or 6 U) of restriction enzyme appropriate for each primer pair. The primer sequences, annealing temperatures, and restriction enzymes are shown in **Supplemental Table 2**. The digested DNA fragments were electrophoresed on 3% agarose gel at a constant voltage of 100 V, stained with ethidium bromide, and visualized under UV exposure. The genotype of each SNP was determined based on the length of the digested DNA fragment, and the cytoplasm type was determined as shown in **Supplemental Table 3**.

GBS library preparation

To eliminate RNA in the extracted gDNA solutions, 7 µg of gDNA of each accession were digested with RNase I_f (NEB) at 37°C for 1 hour in 35 µL volumes containing 1 × NEBuffer 3 and 50 U of RNase I_f. Then, gDNA was fragmented with a restriction enzyme *Hsp* 92II (Promega, Madison, WI, USA) or *Nla* III (NEB). Both restriction enzymes are isoschizomers that cleave the same "CATG" sequence in the same way. The fragmentation was carried out in 25 µL volumes containing 140 ng of gDNA and 10 U of restriction enzyme, for overnight (ca. 16 hours) at 37°C. Next, adapters were ligated to both ends of the restriction fragments. Each adapter consists of an annealing site for sequencing primer, an 8 bp-barcode sequence for demultiplexing, and a P5 or P7 sequence for binding to Illumina flowcell. Adapter ligation was performed in a total volume of 30 µL containing 25 µL of digested gDNA, 1 mM ATP, 0.5 µL of each adapter, and 400 U of T4 DNA ligase at 20°C for 30 min, followed by 65°C for 10 min. Purification and fragment size selection were conducted between each enzymatic reaction using AMPure XP magnetic beads (Beckman Coulter, Carlsbad, CA, USA). The library was quantified by Qubit dsDNA HS assay (Thermo Fisher Scientific). For library concentration enrichment, libraries of 16 accessions were equally pooled (3 ng each) and PCR amplified in 50 µL volumes containing 1 × Q5 Reaction Buffer, 1 mM dNTP, 1.2 µM each adapter primer, 48 ng of library pool, and 1 U of Q5 High-Fidelity DNA Polymerase (NEB), with the following program: initial denaturing step at 98°C for 30 sec, followed by 35 cycles of 98°C for 10 sec, 65°C for 30 sec and 72°C for 1 min, and final extension at 72°C for 15 min. From the amplified libraries, 350–700 bp fragments were selected by 1.5% agarose gel electrophoresis and purified using MonoFas DNA I kit (GL

Sciences, Tokyo, Japan) according to the manufacturer's protocol. The 12 enriched and size-selected libraries were further pooled in equal amounts of DNA. The resulting 192-plex (16 × 12) libraries were sequenced on the HiSeq 4000 system (Illumina, San Diego, CA, USA) with the paired-end mode and read length of 101 bp each. All adapter oligos and PCR primers used for the preparation and sequencing of the GBS libraries are given in **Supplemental Table 4**.

SNP calling

The raw reads were demultiplexed into each accession according to the combination of barcode sequences in both adapters. Adapter sequences and low-quality bases were removed from the demultiplexed reads using fastp (Chen *et al.* 2018) with the parameters "-3 -l 20 -a = AGATCG GAAGAGC". The preprocessed reads of each accession are deposited in the DDBJ Sequence Read Archive (DRA) under accession numbers DRR404032–DRR404789. Then, the preprocessed reads of each accession were mapped to the DHL92 melon reference genome (v.3.6.1) (Ruggieri *et al.* 2018) using BWA-MEM (Li and Durbin 2009) with default parameters. Among several reference genomes of different melon varieties published to date, DHL92 was chosen because it is a doubled-haploid line derived from an inter-subspecific cross, and thus was expected to allow unbiased mapping of the reads from diverse accessions.

From the pileups generated by the mapping, raw SNPs were called using SAMtools and BCFtools (Danecek *et al.* 2021). This analysis workflow was executed using a modified GB-eaSy pipeline (Wickland *et al.* 2017). The raw SNPs were filtered based on read depth and missing rate using VCFtools (Danecek *et al.* 2011), and the filtered SNPs were used for subsequent analyses.

Phylogenetic and population structure analyses

To infer the genetic relationships and population structure in the melon accessions, phylogenetic tree reconstruction and model-based clustering were performed using the filtered SNPs with read depth ≥3 and missing rate <80%. A phylogenetic tree was reconstructed using the maximum likelihood method implemented in IQ-TREE (Minh *et al.* 2020) using the GTR + ASC model. The phylogenetic tree was rooted using the three wild *Cucumis* accessions. Model-based clustering was performed using ADMIXTURE (Alexander *et al.* 2009). The major mode of 10 independent runs for each K ($K = 2$–4) was visualized using Pong (Behr *et al.* 2016).

Selection and evaluation of core collection

GenoCore (Jeong *et al.* 2017) was used to select a subset of accessions that represent the allelic diversity of melon accessions genotyped by GBS, with the parameter "-cv 100". Seventy-three F_1 cultivars were excluded for GenoCore analysis. We constructed the final core collection containing 100 accessions, of which 84 were selected by GenoCore. The remaining 16 accessions were selected because of their historical importance and disease resistance.

Results

Geographical distribution of the three cytoplasm types

The cytoplasm type determined based on three SNPs in the chloroplast genome showed distinct geographical distribution (**Fig. 1**). Ia type and Ic type are mostly distributed within their native cultivation areas: the Ia type is generally distributed in South Asia, Southeast Asia, East Asia, and South Africa, whereas the Ic type is exclusively distributed in sub-Saharan Africa. In contrast, the Ib type is globally distributed not only in its native cultivation areas of Europe, Americas, Central/West Asia, and Northern Africa, but also in East Asian countries such as China, Taiwan, and Japan.

Genotyping-by-sequencing

GBS generated a total of approximately 1.5 billion reads, which contained 147.1 Gb of data. The number of reads for each accession ranged from 106,366 to 19,283,624, with an average of 1,921,782 reads. The proportion of preprocessed reads mapped to the DHL92 genome was 93.2%. On average, approximately 2.4% (9.0 Mb out of 375.4 Mb) of the DHL92 genome was covered by the aligned reads, with an average depth of 13.4× in each accession. Based on the pileups, we identified a total of 3,182,277 raw SNPs. After retaining genotypes with read depth ≥3 and SNPs with <80% missing rate, 39,324 SNPs were obtained as a final data set. These SNPs were distributed throughout the 12 melon chromosomes, with an average of 1 SNP per 10.6 kb (**Table 1**).

Phylogenetic relationships and population structure of the melon genetic resources

Based on the 39,324 SNPs dataset, we reconstructed a rooted maximum likelihood tree to infer phylogenetic relationships among the 755 melon accessions (**Fig. 2A**). The topology of the phylogenetic tree allowed us to divide them into three major groups (groups I, II, and III). They mostly corresponded to the classification by cytoplasm type. Group I consisted of 14 accessions. Among them, 9 (64%) accessions were Ic type, and 10 (71%) and 3 (21%) accessions were of African and South Asian origin, respectively. Groups II and III mostly consisted of Ia type and Ib type, respectively; 99% (199/201) of group II was Ia type, while 86% (464/540) of group III was Ib type. Exceptionally, the basal part of group III was mainly composed of Ia type. South Asian, Southeast Asian, and East Asian accessions accounted for 95% (190/201) of group II. On the other hand, group III was composed of accessions from wider geographical regions.

To investigate the population structure in the 755 melon accessions, model-based clustering implemented in the ADMIXTURE program was performed (**Fig. 2B**). Combining the results from the phylogenetic tree and population

Table 1. Summary statistics of the SNPs identified by GBS across each melon chromosome

Chromosome	Size (bp)	No. of raw SNPs	Average raw SNP distance (bp)	No. of filtered SNPs (depth ≥3× & missing rate <80%)	Average filtered SNP distance (bp)
chr00[a]	41,641,883	140,075	297	15,573	2,674
chr01	37,037,532	311,544	119	1,541	24,035
chr02	27,064,691	237,677	114	2,541	10,651
chr03	31,666,927	260,031	122	1,587	19,954
chr04	34,318,044	270,230	127	2,293	14,966
chr05	29,324,171	232,125	126	1,747	16,785
chr06	38,297,372	304,612	126	2,182	17,551
chr07	28,958,359	224,949	129	2,215	13,074
chr08	34,765,488	276,110	126	1,790	19,422
chr09	25,243,276	190,330	133	2,195	11,500
chr10	26,663,822	227,276	117	1,881	14,175
chr11	34,457,057	274,527	126	1,640	21,010
chr12	27,563,660	232,791	118	2,139	12,886
Total	417,002,282	3,182,277	131	39,324	10,604

[a] Contigs and scaffolds unanchored to chromosomes.

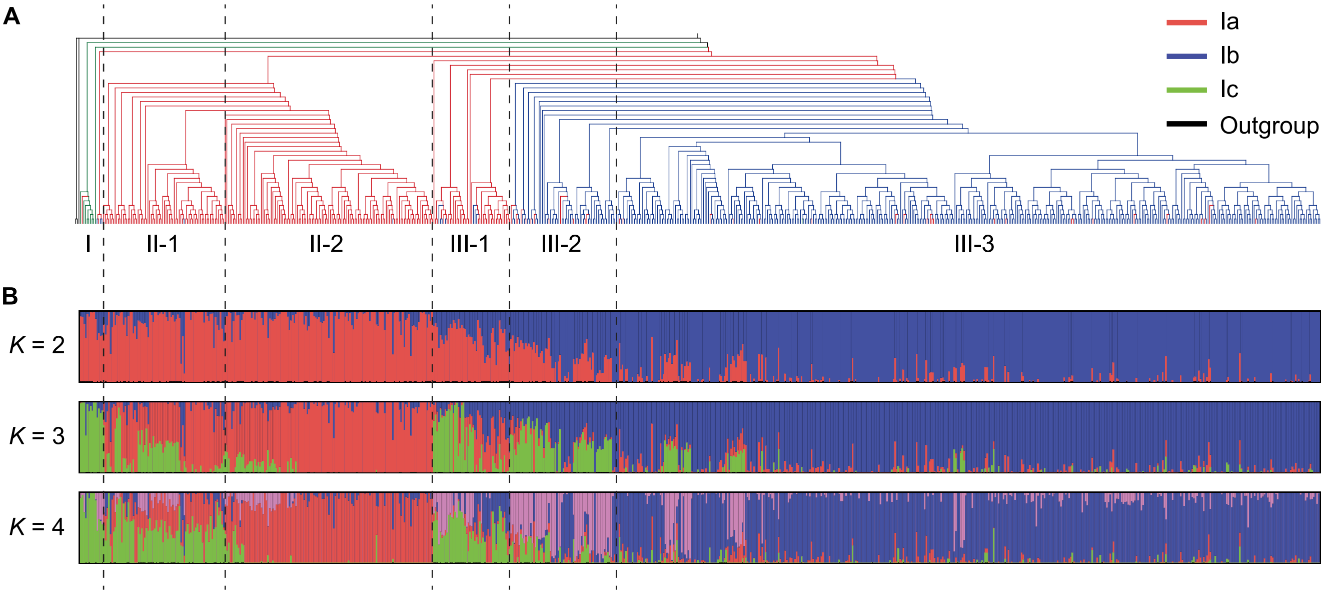

Fig. 2. Genetic relationships and population structure of the 755 melon accessions based on 39,324 SNPs. (A) Maximum-likelihood phylogenetic tree rooted using the three wild *Cucumis* accessions. Branch colors represent different cytoplasm types. A tree with scaled branch lengths is available in **Supplemental Fig. 1**. (B) Model-based clustering with different numbers of putative populations (*K* = 2–4). The memberships of each accession to each putative population are shown as stacked-bar plots, with different colors indicating different putative populations. The *y*-axes quantify the relative membership to each population, and the *x*-axes list the different accessions. The orders and positions of accessions on the *x*-axes are consistent with those in the phylogenetic tree.

structure analysis, we divided groups II and III into two and three subgroups, respectively. In both groups, basal subgroups (II-1, III-1, and III-2) showed a high degree of genetic admixtures compared with the other subgroups (II-2 and III-3). More than 85% of subgroup II-1 was composed of South Asian and Southeast Asian accessions, accounting for 51% (38/74) and 34% (25/74), respectively (**Table 2**). On the other hand, 89% (113/127) of subgroup II-2 consisted of East Asian accessions, with only 2 South Asian and 8 Southeast Asian accessions were included in this subgroup. Of the 9 Ia-type accessions from South Africa, 5 accessions were placed in subgroup II-1, while the rest were scattered among the other groups (1 accession each in group I and subgroup III-1, and 2 accessions in subgroup III-3). In subgroup III-1, the Ia type accounted for 85% (40/47), and most of them were native to South Asia (72%; 34/47). In contrast, the Ib type accounted for 81% (52/64) and 94% (405/429) in subgroups III-2 and III-3, respectively.

Table 2. Cytoplasm type and geographical origin of the accessions classified into six groups by the phylogenetic tree and population structure analysis

Group	No. of accessions	Cytoplasm type			Geographical origin							
		Ia	Ib	Ic	Africa	South Asia	Southeast Asia	East Asia	Central/ West Asia	Europe	Americas	Unknown
I	14	2	3	9	10	3	–	–	1	–	–	–
II-1	74	73	1	–	5	38	25	4	1	1	–	–
II-2	127	126	1	–	–	2	8	113	–	2	2	–
III-1	47	40	7	–	1	34	4	6	1	–	1	–
III-2	64	12	52	–	8	17	1	17	10	9	2	–
III-3	429	23	405	1	19	20	9	148	43	124	62	4
Total	755	276	469	10	43	114	47	288	56	136	67	4

Construction of the World Melon Core Collection

A total of 682 accessions, excluding F_1 cultivars, were analyzed with GenoCore, and representative accessions were selected in order of allelic coverage of the genome-wide SNPs. The results showed that 53, 93, 160, 281 and 674 accessions represented 80%, 85%, 90%, 95% and 100% of the allelic diversity of 39,324 SNPs within the 682 accessions, respectively (**Fig. 3A**). Considering the balance between allelic coverage and collection size, we selected 67 accessions as the primary core collection, which covered 82% of the allelic diversity in the entire collection. In addition, we selected an additional 33 accessions based on geographical origin, disease resistance, and historical importance, for a total of 100 accessions to construct the World Melon Core Collection, of which 84 accessions have GBS data. The 84 accessions were widely distributed throughout the phylogenetic tree, with a modest concentration in group I and subgroup III-1 (**Fig. 3B**). Their geographical origins were diverse: 26 from South Asia; 21 from Africa; 21 from East Asia; 12 from Southeast Asia, including Cambodia, Laos and Vietnam; 11 from Central/West Asia including the Kyrgyz Republic; 5 from the Americas; and 4 from Europe (**Supplemental Tables 5**, **6**). In addition, the fruit morphology of the World Melon Core Collection was highly diverse in various traits, such as shape, size, skin color, and flesh color (**Fig. 4**). These results demonstrated that the World Melon Core Collection established in this study covers most of the genetic diversity in the genetic resource collection of the NARO Genebank. A full list of the World Melon Core Collection is available in **Supplemental Table 6**.

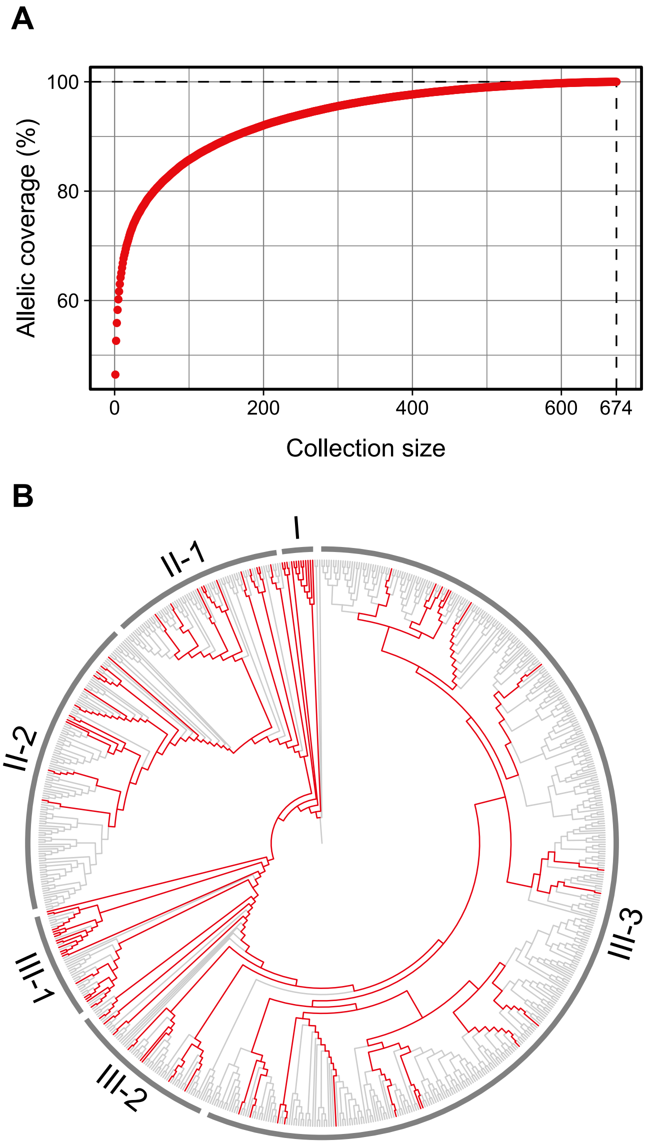

Fig. 3. Selection and evaluation of the World Melon Core Collection. (A) Coverage of allelic diversity of 39,324 SNPs (y-axis) versus the number of selected accessions (x-axis). (B) Same maximum-likelihood phylogenetic tree as in **Fig. 2A**, with the 84 accessions in the core collection highlighted in red.

Discussion

Numerous studies have been conducted to elucidate the origin and domestication history of melon. To provide new insights into this long-standing question, we investigated the genetic variation and population structure among melon genetic resources deposited mainly in the NARO Genebank. The cytoplasm type of each accession was determined based on the three SNPs in the chloroplast genome. As a result of investigating the composition of the cytoplasm types

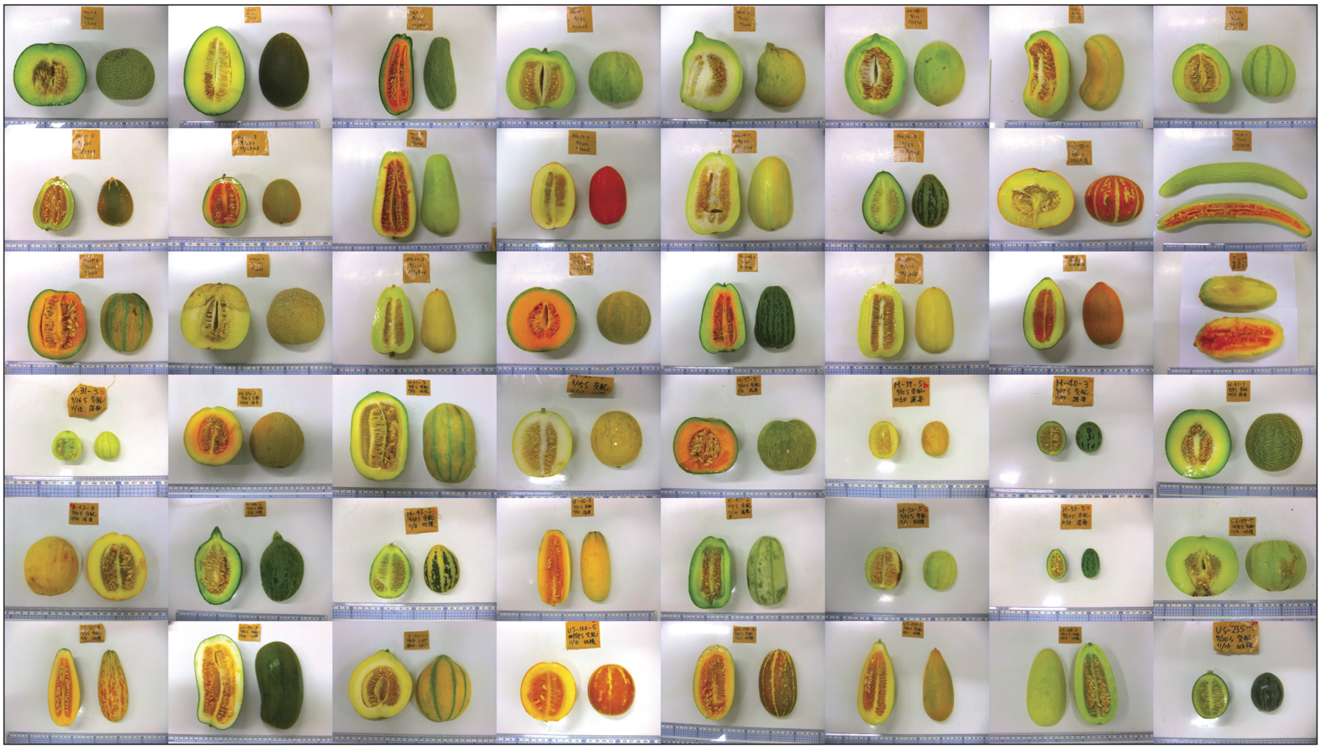

Fig. 4. Diversity in the fruit morphology of the World Melon Core Collection (partial).

in each country, Ia and Ic types were generally restricted within their traditional cultivation areas, while the Ib type was widespread in the world (**Fig. 1**). This result may reflect the introduction and rising demand for melon cultivars of var. *cantalupensis* and var. *inodorus* (both typically Ib type) in East Asian countries such as China, Taiwan, and Japan, except Xinjiang (Xinjiang Uyghur Autonomous Region, China) where Hami melon of Ib type has been traditionally cultivated.

The GBS of the 755 melon genetic resources resulted in the identification of the SNPs densely distributed throughout the melon genome (**Table 1**). The phylogenetic analysis based on this SNP dataset revealed the genetic relationship and the population structure among the accessions that largely corresponded to the cytoplasm types and geographical origins (**Fig. 2**, **Table 2**). Group I was the most ancestral group in the phylogenetic tree and mainly consisted of the accessions from African and South Asian countries, considered to be the origins of cultivated melon. Of the 10 Ic-type accessions used in this study, 9 accessions clustered in this group. Groups II and III were mostly composed of Ia-type and Ib-type accessions, respectively, and appeared to correspond to the classification of two subspecies, subsp. *agrestis* and subsp. *melo*. Taken together with the association between the cytoplasm types and seed length observed in a previous study (Tanaka *et al.* 2013), this result implies that seed length, in addition to ovary pubescence, could be useful as a key trait to distinguish between the two subspecies. In both groups II and III, the proportion of South Asian accessions was notably higher in the basal subgroups (II-1, III-1, and III-2) than in the other subgroups (II-2 and III-3). In Group III, subgroup III-1 was distinct from the other two subgroups in that most of the accessions are Ia type. On the other hand, subgroup III-2 was mostly Ib type, but was distinguished from subgroup III-3 by its higher genetic admixture. These two subgroups were considered to be either (1) ancestral populations from which subgroup III-3 was selected, (2) inter-subspecific hybrids between Group II and subgroup III-3, or (3) mixtures of both. In any case, the above results support the recent hypothesis that cultivated melon was domesticated in Africa and two independent domestication events for subsp. *melo* and subsp. *agrestis* occurred in India (Zhao *et al.* 2019). All three cytoplasm types are distributed on the African continent, and Ia-type and Ib-type accessions in Africa were grouped together with accessions from other regions with their respective cytoplasm types. This result suggests that those accessions were reverse-introduced to the African continent after domestication events occurred in India. We speculate on possible routes and timings of reverse-introduction as follows. As for Ia-type accessions in Southern Africa, they may have been introduced from South Asia through the maritime route connecting the Cape of Good Hope and India pioneered during the Age of Discovery. On the other hand, Ib-type accessions in Northern Africa may have been introduced during dispersal of cultivated melon from India to Western countries and/or through the European exploration of Africa started in the 15th century. To verify this hypothesis, further studies with a larger diversity panel of African landraces, as well as with archaeological seed remains from each region, would be needed. In that case, special attention should be paid to the world's oldest melon seed remains from Lower Egypt (the fertile Nile Delta area) dating from 3700–3500 BC (van Zeist and de Roller 1993).

Using a comprehensive approach that accounts for genetic diversity, historical importance, and disease resistance, we selected the 100 accessions as the World Melon Core Collection, which represent at least 82% of the allelic variation of the genome-wide SNPs identified by GBS. The representativeness of the World Melon Core Collection was confirmed by the following three results: (1) the accessions were selected from all cytoplasm types and from wide geographical origins without major deviations (**Supplemental Table 5**); (2) they were widely distributed throughout the phylogenetic tree (**Fig. 3B**); and (3) the fruit morphology of the core collection was highly diverse (**Fig. 4**). Note that 21 African and 26 South Asian accessions were selected and together accounted for nearly half of the core collection, despite relatively small number of accessions from those regions (43 from Africa and 114 from South Asia) were used in this study (**Table 2**, **Supplemental Table 5**). This result may reflect the fact that accessions from African and South Asian countries retain higher genetic diversity than accessions from other regions (Wang *et al.* 2021).

Recently, another core collection of melon was developed from the USDA National Plant Germplasm System using a similar GBS-based approach (Wang *et al.* 2021). Their core collection consists of 383 accessions representing more than 98% of allelic variation in 27,471 SNPs identified from 2,083 accessions. The allelic coverage of our core collection (82%) is not as high as that of their core collection, probably because we selected fewer accessions while accounting for more SNPs than they did. However, it still represents a great majority of genetic variation within the species, and the compact collection size will fit a wide range of research scales, from simple screening for a trait to intensive pan-genomic studies. In addition, our core collection is characterized by the inclusion of accessions from Southeast and Central Asian countries, including Vietnam, Laos, Cambodia, and the Kyrgyz Republic. The accessions from those countries were introduced to the NARO Genebank through recent explorations conducted under the PGRAsia project (Matsunaga *et al.* 2015, Tanaka *et al.* 2016b, 2017, 2019, Yoshioka *et al.* 2020), and are rarely found in major genebanks in other countries, such as USDA, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), and the National Institute for Agricultural Research (INRA). Recent work by Imoh *et al.* (in preparation) showed the richness of melon landraces with virus disease resistance in Southeast Asian countries, and thus the core collection constructed in this study is considered a unique and novel set of melon genetic resources.

The World Melon Core Collection is currently undergoing seed multiplication, agronomic trait evaluation, and whole genome sequencing using a long-read sequencer. This will provide a more comprehensive picture of genomic variation, and that will enable us to perform GWAS and pan-genomic studies. The World Melon Core Collection will be publicly available from the NARO Genebank (https://www.gene.affrc.go.jp/databases-core_collections_en.php) in 2024, along with the agronomic trait data and genomic information. The genome-wide SNP dataset, phylogenetic relationships, population structure, and the World Melon Core Collection provided in this study will be valuable resources for genetic research, breeding, and genetic resource preservation of melon.

Author Contribution Statement

KK, NT, and YK conceived the project; MS and KK provided materials; GS, TPD, MP, TTD, and OI performed the experiments; GS analyzed the data, prepared figures, and wrote a draft of the manuscript; and YM, HN, and KT provided advice on the experimental implementation and helped draft the manuscript.

Acknowledgments

This study was supported by the PGRAsia project (https://sumire.gene.affrc.go.jp/pgrasia/index_en.php) from the Ministry of Agriculture, Forestry and Fisheries of Japan. We also thank Dr. Koichiro Ushijima (Okayama University, Japan) for his helpful advice on establishing a protocol for GBS library preparation.

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.

Behr, A.A., K.Z. Liu, G. Liu-Fang, P. Nakka and S. Ramachandran (2016) pong: Fast analysis and visualization of latent clusters in population genetic data. Bioinformatics 32: 2817–2823.

Chen, S., Y. Zhou, Y. Chen and J. Gu (2018) fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34: i884–i890.

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.

Danecek, P., J.K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M.O. Pollard, A. Whitwham, T. Keane, S.A. McCarthy, R.M. Davies *et al.* (2021) Twelve years of SAMtools and BCFtools. Gigascience 10: giab008.

Ebana, K., Y. Kojima, S. Fukuoka, T. Nagamine and M. Kawase (2008) Development of mini core collection of Japanese rice landrace. Breed Sci 58: 281–291.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler and S.E. Mitchell (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6: e19379.

Endl, J., E.G. Achigan-Dako, A.K. Pandey, A.J. Monforte, B. Pico and H. Schaefer (2018) Repeated domestication of melon (*Cucumis melo*) in Africa and Asia and a new close relative from India. Am J Bot 105: 1662–1671.

FAO. FAOSTAT. License: CC BY-NC-SA 3.0 IGO. Extracted from: https://www.fao.org/faostat/en/#data/QCL. Date of access: 17-08-2022.

Frankel, O.H. and A.H.D. Brown (1984) Plant genetic resources today: A critical appraisal. *In:* Holden, J.H.W. and J.T. Williams (eds.) Crop genetic resources: Conservation and evaluation, George Allan and Unwin, London, pp. 249–257.

Hirano, R., K. Naito, K. Fukunaga, K.N. Watanabe, R. Ohsawa and M. Kawase (2011) Genetic structure of landraces in foxtail millet (*Setaria italica* (L.) P. Beauv.) revealed with transposon display and interpretation to crop evolution of foxtail millet. Genome 54: 498–506.

Jeong, S., J.Y. Kim, S.C. Jeong, S.T. Kang, J.K. Moon and N. Kim (2017) GenoCore: A simple and fast algorithm for core subset selection from large genotype datasets. PLoS One 12: e0181420.

Kaga, A., T. Shimizu, S. Watanabe, Y. Tsubokura, Y. Katayose, K. Harada, D.A. Vaughan and N. Tomooka (2012) Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. Breed Sci 61: 566–592.

Kojima, Y., K. Ebana, S. Fukuoka, T. Nagamine and M. Kawase (2005) Development of an RFLP-based rice diversity research set of germplasm. Breed Sci 55: 431–440.

Li, H. and R. Durbin (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754–1760.

Liu, S., P. Gao, Q. Zhu, Z. Zhu, H. Liu, X. Wang, Y. Weng, M. Gao and F. Luan (2020) Resequencing of 297 melon accessions reveals the genomic history of improvement and loci related to fruit traits in melon. Plant Biotechnol J 18: 2545–2558.

Matsunaga, H., K. Matsushima, K. Tanaka, S. Theavy, S. Layheng, T. Channa, Y. Takahashi and N. Tomooka (2015) Collaborative exploration of the Solanaceae and Cucurbitaceae vegetable genetic resources in Cambodia, 2014. AREIPGR 31: 169–187.

Minh, B.Q., H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler and R. Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37: 1530–1534.

Miyatake, K., Y. Shinmura, H. Matsunaga, H. Fukuoka and T. Saito (2019) Construction of a core collection of eggplant (*Solanum melongena* L.) based on genome-wide SNP and SSR genotypes. Breed Sci 69: 498–502.

Murray, M.G. and W.F. Thompson (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8: 4321–4325.

Pitrat, M. (2016) Melon genetic resources: Phenotypic diversity and horticultural taxonomy. *In:* Grumet, R., N. Katzir and J. Garcia-Mas (eds.) Genetics and genomics of Cucurbitaceae, Springer, Cham, pp. 25–60.

Ruggieri, V., K.G. Alexiou, J. Morata, J. Argyris, M. Pujol, R. Yano, S. Nonaka, H. Ezura, D. Latrasse, A. Boualem *et al.* (2018) An improved assembly and annotation of the melon (*Cucumis melo* L.) reference genome. Sci Rep 8: 8088.

Sangiri, C., A. Kaga, N. Tomooka, D. Vaughan and P. Srinives (2007) Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) genepool on the basis of microsatellite analysis. Aust J Bot 55: 837–847.

Sebastian, P., H. Schaefer, I.R.H. Telford and S.S. Renner (2010) Cucumber (*Cucumis sativus*) and melon (*C. melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc Natl Acad Sci USA 107: 14269–14273.

Shehzad, T., H. Okuizumi, M. Kawase and K. Okuno (2009) Development of SSR-based sorghum (*Sorghum bicolor* (L.) Moench) diversity research set of germplasm and its evaluation by morphological traits. Genet Resour Crop Evol 56: 809–827.

Tanaka, K., Y. Akashi, K. Fukunaga, T. Yamamoto, Y. Aierken, H. Nishida, C.L. Long, H. Yoshino, Y. Sato and K. Kato (2013) Diversification and genetic differentiation of cultivated melon inferred from sequence polymorphism in the chloroplast genome. Breed Sci 63: 183–196.

Tanaka, K., C.J. Stevens, S. Iwasaki, Y. Akashi, E. Yamamoto, T.P. Dung, H. Nishida, D.Q. Fuller and K. Kato (2016a) Seed size and chloroplast DNA of modern and ancient seeds explain the establishment of Japanese cultivated melon (*Cucumis melo* L.) by introduction and selection. Genet Resour Crop Evol 63: 1237–1254.

Tanaka, K., T.T. Duong, H. Yamashita, S. Lay Heng, S. Sophany and K. Kato (2016b) Collection of cucurbit crops (Cucurbitaceae) from eastern Cambodia, 2015. AREIPGR 32: 109–137.

Tanaka, K., G. Shigita, Y. Sophea, V. Thun, S. Sophany and K. Kato (2017) Collection of melon and other Cucurbitaceous crops in Cambodia in 2016. AREIPGR 33: 175–205.

Tanaka, K., G. Shigita, T.P. Dung, Y. Sophea, V. Thun, S. Sophany and K. Kato (2019) Collection of melon and other Cucurbitaceous crops in Cambodia in 2017. AREIPGR 35: 121–146.

Tanaka, N., M. Shenton, Y. Kawahara, M. Kumagai, H. Sakai, H. Kanamori, J. Yonemaru, S. Fukuoka, K. Sugimoto, M. Ishimoto *et al.* (2020) Whole-genome sequencing of the NARO World Rice Core Collection (WRC) as the basis for diversity and association studies. Plant Cell Physiol 61: 922–932.

van Zeist, W. and G.J. de Roller (1993) Plant remains from Maadi, a predynastic site in lower Egypt. Veg Hist Archaeobot 2: 1–14.

Wang, X., K. Bao, U.K. Reddy, Y. Bai, S.A. Hammar, C. Jiao, T.C. Wehner, A.O. Ramírez-Madera, Y. Weng, R. Grumet *et al.* (2018) The USDA cucumber (*Cucumis sativus* L.) collection: Genetic diversity, population structure, genome-wide association studies, and core collection development. Hortic Res 5: 64.

Wang, X., K. Ando, S. Wu, U.K. Reddy, P. Tamang, K. Bao, S.A. Hammar, R. Grumet, J.D. McCreight and Z. Fei (2021) Genetic characterization of melon accessions in the U.S. National Plant Germplasm System and construction of a melon core collection. Molecular Horticulture 1: 11.

Wickland, D.P., G. Battu, K.A. Hudson, B.W. Diers and M.E. Hudson (2017) A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. BMC Bioinformatics 18: 586.

Xu, H.X., T. Jing, N. Tomooka, A. Kaga, T. Isemura and D.A. Vaughan (2008) Genetic diversity of the azuki bean (*Vigna angularis* (Willd.) Ohwi & Ohashi) gene pool as assessed by SSR markers. Genome 51: 728–738.

Yoshioka, Y., D. Kami, T. Kakizaki, K. Tanaka, N. Zhumakadyrova, B. Imanbaeva and A. Usupbaev (2020) Collaborative exploration of vegetable genetic resources in Kyrgyz in 2019. AREIPGR 36: 203–225.

Zhao, G.W., Q. Lian, Z.H. Zhang, Q.S. Fu, Y.H. He, S. Ma, V. Ruggieri, A.J. Monforte, P.Y. Wang, I. Julca *et al.* (2019) A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat Genet 51: 1607–1615.