Introduction

Since its establishment in 1985 as a department of the National Institute of Agrobiological Resources (NIAR) (currently, the National Agriculture and Food Research Organization [NARO]), the Genebank Project has allowed the strategic performance of exploration, collection and introduction of genetic resources, followed by their characterization, reproduction and distribution. At present, the Institute of Vegetable and Floriculture Science, NARO, in collaboration with the Genetic Resources Center, NARO, is the main player of the project regarding genetic resources of vegetable crops. Nearly 1,000 accessions of eggplant (Solanum melongena L.) and its wild relative species are respectively registered and deposited at present. Among them, some accessions have been utilized in breeding practice as valuable sources of unique and useful traits. By using the accessions derived from the Genebank, a number of leading varieties has been developed, including the rootstock varieties ‘Torvum Vigor’ (Yamakawa 1981) that shows composite resistance to bacterial wilt, Verticillium wilt, Fusarium wilt and nematodes; ‘Daitaro’ (Monma et al. 1997) and ‘Daizaburo’ (Yoshida et al. 2004) with high resistance to bacterial wilt and Fusarium wilt; and the unique parthenocarpic varieties ‘Anominori’ (Saito et al. 2009) and ‘Anominori 2 go’ (Saito et al. 2015). Genetic resources containing wide genetic variation are widely considered to be indispensable materials with latent potential for future eggplant breeding. However, the status of genetic resources in eggplant (and probably in other species as well) is not completely organized. There are a significant number of confusing cases, for instance, the same variety and/or germplasm may be registered more than once under the same name, the same germplasm can be registered under different names, and the same name can be used for the registration of different germplasms. These situations must be appropriately reorganized. Additionally, for fruit vegetables such as eggplant, the physical effort and field area required to grow and characterize a plant are definitely larger than those for field crops such as wheat or rice. Therefore, the construction of a relatively small subset that consists of a limited number of accessions but still retains...
Construction of an eggplant core collection

the range of genetic variation of all the genetic resources as much as possible is useful for their efficient and pragmatic use. For this purpose, the construction of an eggplant core collection based on molecular genetic information is urgently required. Although the construction of a core collection using molecular genetic information, i.e., DNA marker genotypes, has been reported in several field crop species in Japan, including rice (Ebara et al. 2008), soybean (Kaga et al. 2012) and rapeseeds (Chen et al. 2017), there are only a few reports on vegetable crops. Outside Japan, Gangopadhyay et al. (2010), Kumar et al. (2008), and Mao et al. (2008) have conducted studies to develop an eggplant core collection, but their studies are mainly based on regional sources and phenotype data. As one of the few examples, Cericola et al. (2013) reported that, using the genotype data obtained from 24 SSR markers, a core set consisting of 48 lines could be built to cover all 140 SSR alleles found in the 191 germplasms examined. This study provides some novel information. However, the genetic resources in each country would differ from each other in content and size and therefore it is important to develop individualistic core collections from domestic genetic resources.

Here we report the construction of a unique eggplant core collection that was accomplished using the genotype data of 831 SNPs and 50 SSRs obtained from 893 eggplant accessions collected from across the world, mainly Asia. Unlike the previous examples, this collection was prepared for distribution, and it can be procured from the NARO Genebank with a simple procedure, in addition, the genotype and phenotype data are freely available from the VegMarks database.

Materials and Methods

When the experiment was started, the number of eggplant accessions registered in the NARO Genebank was around 1,000. The seeds of all accessions were sown in a greenhouse and young leaves were sampled from each plant of the 938 accessions that showed good germination and initial growth. The total DNA was extracted using the DNeasy Plant DNA Extraction Kit (Qiagen, Valencia, CA, USA) and 1,536 SNP genotypes were collected for each accession using the GoldenGate Assay Kit (Illumina, San Diego, CA, USA) constructed by Hirakawa et al. (2014). Initially, we removed unreliable markers among the 1,536 SNPs based on the stability of the genotype and the rate of missing data (>10%). Then, two hundred accessions were selected as members of the tentative core collection using the Core Hunter II (De Beukelaer et al. 2012) software adopting the Mixed Replica algorithm method with the default parameter settings suggested in the instruction manual. Further, 111 microsatellite markers assigned to each of the 12 chromosomes (Fukuoka et al. 2012) were genotyped against the tentative core collection, and removed unreliable genotype calls based on the minor allele frequencies (<0.05) and a rate of missing data (>15%). The core collection was then reconstructed using Core Hunter II with a combined dataset of biallelic SNP genotypes and multiallelic microsatellite genotypes. Model-based Bayesian clustering analysis of whole eggplant materials and the core collection was performed using STRUCTURE 2.3.4 (Pritchard et al. 2000) software with the admixture-non F model. The GGT 2.0 (van Berloo 2008) program was used to formulate Jaccard similarity coefficient and an unrooted Unweighted Pair Group Method with Arithmetic mean (UPGMA) tree was constructed using MEGA program (version 6) based on the distance matrix, with 1,000 bootstrap replicates (Tamura et al. 2013). Genetic diversity indices were defined using GenAlex 6.5 (Peakall and Smouse 2012) and PowerMarker 3.25 (Liu and Muse 2005).

Results and Discussion

SNP genotyping of whole accessions and initial construction of core collection

By performing the GoldenGate genotyping assay with the DNA samples of the 938 accessions, the data of SNPs were obtained. Firstly, based on polymorphism and stability of the multiple genotype data of four standard eggplant lines (‘AE-P03’, ‘LS1934’, ‘Nakate-Shinkiro’ and ‘WCGR112-8’) included in the above 938 accessions, 987 SNPs were selected as ‘reliable markers’. Secondly, 893 accessions that exhibited a rate of less than 10% missing data for the 987 ‘reliable’ SNP markers were selected as candidate accessions for constructing of the core collection. Two hundred accessions were selected based on SNP genotypes using Core Hunter II program. Subsequently, possible duplications, judging from the SNP genotype, were removed and 176 accessions were selected as independent members of a tentative core collection.

SSR genotyping and second-round construction of the core collection—World Eggplant Core collection

With a view to practically utilize the collection, we set the final size of the core collection to 100 and made further refined the collection. To investigate the genetic background of the 176 accessions in depth, the genotype data were collected from 111 microsatellite markers that were polymorphic among the four standard eggplant lines and analyzed (Nunome et al. 2009). To remove poorly reliable genotype calls, the data obtained from alleles of which the frequency was lower than 0.05 and/or the data called as heterozygous were envisaged as missing data. Fifty microsatellites were found to produce genotype data with less than 15% missing data, and therefore the data obtained from these 50 microsatellites were determined to be reliable enough and used for further analyses (Supplemental Table 1). Regarding the 50 microsatellites investigated, among the 176 accessions, the number of alleles of each microsatellite seemed reasonable, ranging from 2 to 8 (4.5 on average, 227 in total). Similarly, 831 SNP markers that showed minor allele
A cluster analysis based on the genetic-distance matrix obtained from the SNP genotype data of the WEC collection lines suggested a cluster structure reflecting their geographic origins (Supplemental Fig. 1). With the STRUCTURE analysis, the understanding of the cluster structure could become more clear and accurate. The optimum cluster number \((K)\) was suggested to be \(K=2\) according to Evanno’s method (Evanno et al. 2005); however, judging from the transition of the cluster structure with an increase in the value of \(K\) and taking the origin of the lines and dendrogram structure into account, \(K=4\) might be more appropriate (Supplemental Fig. 1). Each cluster (S1–S4) corresponds to geographical group. The S1 cluster originated from European, American and African countries, the S2 cluster originated from East Asian countries (mainly Japan); the S3 cluster originated from the Southeast Asian countries frequencies of 0.05 or more were selected from the 987 markers used for the selection of the 176 accessions. The genotype data from the carefully selected 881 markers (831 SNPs and 50 microsatellites) were used to select 100 accessions for the final core collection, World Eggplant Core (WEC) collection, using the Core Hunter II program (Fig. 1, Supplemental Table 2). The genetic diversity indices for each country are summarized in Table 1. Generally, the values of expected heterozygosity (He), Shannon’s information index (I) and polymorphism information content (PIC) were higher in Asian countries (especially India and Malaysia) than in other countries. The WEC collection consisted of 3 Africa (among 22), 4 American (among 27), 80 Asian (among 695), 8 Europe (among 87) and 5 unknown accessions. Based on the examined accessions and diversity indices, a large number of accessions from Malaysia (21), Lao PDR (11) and Japan (10) were collected (Table 1). The WEC collection constructed using the Core Hunter II program had a high retention ratio (97.5%; Supplemental Table 3), indicating that it has almost all alleles observed in the whole collection. Additionally, the diversity indices, except observed heterozygosity (Ho), were no significantly different among the whole collection and WEC collection (Supplemental Table 3). These results suggest that the WEC collection maintains most of the genetic diversity in whole collection.

**Basic characterization of the WEC collection**

A cluster analysis based on the genetic-distance matrix obtained from the SNP genotype data of the WEC collection lines suggested a cluster structure reflecting their geographic origins (Supplemental Fig. 1). With the STRUCTURE analysis, the understanding of the cluster structure could become more clear and accurate. The optimum cluster number \((K)\) was suggested to be \(K=2\) according to Evanno’s method (Evanno et al. 2005); however, judging from the transition of the cluster structure with an increase in the value of \(K\) and taking the origin of the lines and dendrogram structure into account, \(K=4\) might be more appropriate (Supplemental Fig. 1). Each cluster (S1–S4) corresponds to geographical group. The S1 cluster originated from European, American and African countries, the S2 cluster originated from East Asian countries (mainly Japan); the S3 cluster originated from the Southeast Asian countries.

---

**Fig. 1.** Pictures of mature (Left) and immature (Right) fruit of 100 accessions constituting the World Eggplant Core collection (WEC).
In conclusion, the eggplant core collection, WEC collection, represents the genetic diversity of a large collection with a pragmatic size of 100 accessions. Hence, it will enable easy access to eggplant genetic resources and accelerate its utilization. In addition, molecular genetic information of the WEC collection will help strategic planning of research in the future. Basic trait-based characterization of the core collection is underway, which will contribute to the validation and utilization of the collection. The WEC collection will be a valuable source for developing new breeding materials to improve important and complicated traits such as biotic and abiotic stress tolerance, plant architecture, and stable productivity against today’s worsening environmental conditions.

Data and material availability

Basic data of the WEC collection, such as provenance
and phenotype data listed in Supplemental Table 4, are published through the ‘Genetic resources’ menu of the VegMarks database (https://vegmarks.nivot.affrc.go.jp/resource/) and the DNA marker genotypes (831 SNP data and 50 SSR fragment length) are published on the same site. Seeds of the WEC collection will be distributed by the NARO Genebank (https://www.gene.affrc.go.jp/index_en.php).

Acknowledgments

This study was supported by the NARO Genebank Project, JAPAN. We express our thanks to Dr. Daisuke Sekine, Mr. Toshihiko Uemura, Mr. Hisayoshi Maruyama, Mr. Katsuyuki Yamauchi, Mr. Hiroshi Saito, Ms. Satomi Kizaki for their insightful suggestions and skillful technical assistance. Furthermore, we would also like to thank Editage (www.editage.jp) for English language editing.

Literature Cited


