Full Paper

Identification of genome-wide single-nucleotide polymorphisms among geographically diverse radish accessions

Hiroto Kobayashi[1], Kenta Shirasawa [2], Nobuko Fukino[3†], Hideki Hirakawa[2], Takashi Akanuma[1], and Hiroyasu Kitashiba[1]*

[1]Graduate School of Agricultural Science, Tohoku University, Aoba-ku, Sendai 980-8572, Japan, [2]Kazusa DNA Research Institute, Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan, and [3]Institute of Vegetable and Floriculture Science, NARO, Ano, Tsu 514-2392, Japan

*To whom correspondence should be addressed. Tel. +81 22 757 4268. Fax. +81 22 757 4270. Email: hiroyasu.kitashiba.c7@tohoku.ac.jp

†Present address: Strategic Planning Headquarters, NARO, 3-1-1 Kannondai, Tsukuba, Ibaraki 305-8517, Japan.

Abstract

Radish (*Raphanus sativus* L.) is cultivated around the world as a vegetable crop and exhibits diverse morphological and physiological features. DNA polymorphisms are responsible for differences in traits among cultivars. In this study, we determined genome-wide single-nucleotide polymorphisms (SNPs) among geographically diverse radish accessions using the double-digest restriction site-associated DNA sequencing (ddRAD-Seq) method. A total of 52,559 SNPs was identified in a collection of over 500 radish accessions (cultivated and wild) from East Asia, South and Southeast Asia, and the Occident and Near East. In addition, 2,624 SNP sites without missing data (referred to as common SNP sites) were identified among 510 accessions. Genetic diversity analyses, based on the common SNP sites, divided the cultivated radish accessions into four main groups, each derived from four geographical areas (Japan, East Asia, South and Southeast Asia, and the Occident and Near East). Furthermore, we discuss the origin of cultivated radish and its migration from the West to East Asia. SNP data generated in this work will facilitate further genetic studies on the radish breeding and production of DNA markers.

Key words: radish accessions, genome-wide SNPs, ddRAD-Seq

1. Introduction

Radish (*Raphanus sativus* L.; $2n = 18$) is a member of the Brassicaceae family and is closely related to *Brassica rapa* and *Brassica oleracea*. Radish is grown as a vegetable crop around the world, and different plant organs in different shapes and sizes are consumed in various regions. For example, radish is consumed as a big long root in East Asia, small root in Europe and the USA, and immature edible seed pod in South and Southeast Asia.

In recent years, many studies have been conducted on the radish genome. The whole genome sequence of radish has been determined and released by four research groups since 2014.[1–4] In addition, several linkage maps have been constructed using various DNA markers such as restriction fragment length polymorphisms, amplified fragment length polymorphisms, random amplified polymorphic DNAs, simple sequence repeats (SSRs), cleaved amplified polymorphic sequences (CAPS), and single-nucleotide polymorphisms

(SNPs).[5] A genetic map with 7,480 SNPs[5] has been reported, in addition to a high-density linkage map with a total of 2,553 DNA markers, including SNPs, CAPS, and SSRs.[2] These DNA polymorphisms and genome sequences serve as effective tools for the identification of loci and genes responsible for agronomic traits. For example, quantitative trait loci (QTLs) affecting clubroot resistance,[6] *Fusarium* wilt resistance,[7] early bolting and flowering,[8–10] and root glucosinolate content[11,12] have been identified. Kakizaki et al.[13] conducted map-based cloning using a mutant radish accession lacking glucoraphasatin, resulting in the discovery of a novel gene responsible for the synthesis of glucoraphasatin, which is not found in other *Brassica* species.

Studies have also revealed genome-wide DNA polymorphisms among radish accessions. SNPs at 2,880 loci among eight accessions were identified using an amplicon sequencing method.[2,12] Although the exact number of loci was unknown, a large number of genome-wide SNPs and insertions/deletions (Indels) were identified among 17 accessions via genome re-sequencing using next-generation sequencing (NGS) technologies.[14] Furthermore, SSR polymorphisms have been determined at 52 loci among 93 radish accessions,[15] although the number of DNA markers reported in this study was remarkably lower than those reported in the two aforementioned reports. More recently, advanced NGS methods have been developed for the identification of genome-wide SNPs such as restriction site-associated DNA sequencing (RAD-seq),[16] double-digest RAD-seq (ddRAD-Seq),[17,18] multiplexed inter-simple sequence repeat sequencing (MIG-seq),[19] and genotyping by random amplicon sequencing-direct (http://newsroom.toyota.co.jp/en/detail/13715116, accessed 22 February 2020). These methods enable the analysis of several samples at once and therefore are expected to be a cost-effective and time-saving approach alternative to expensive NGS approaches.

Cultivated radish accessions exhibit diverse morphological and physiological characteristics, and are taxonomically categorized as distinct varieties. For example, *R. sativus* var. *sativus* L. is a small radish cultivated mainly in the Occident (Europe, USA, and Mediterranean region) and Near East; *R. sativus* var. *hortensis* Becker is distributed in East Asia including Japan, China, and Korea, and is characterized by big and long taproots; *R. sativus* var. *caudatus* Hooker & Anderson produces an edible seed pod and is commonly called rat-tailed radish; *R. sativus* var. *oleifer* Metzg or var. *chinensis* Gallizioli is mainly used as an oilseed radish. Wild radish (*R. sativus* var. *raphanistroides* Makino) is native to the coast of Japan and Korea. Three wild species of the genus *Raphanus* are known, including *Raphanus raphanistrum* L., *Raphanus maritimus* Smith, and *Raphanus landra* Moretti ex DC.[20,21] Knowledge of genome-wide DNA polymorphisms among these varieties and species could serve as a very valuable resource for the genetics and breeding of radish and for understanding the evolution of the genus *Raphanus*. Information on genome-wide SNPs among cultivars belonging to these varieties is expected to significantly advance the genetics and breeding of radish.

In this study, we surveyed genome-wide SNPs in a larger collection of radish accessions (more than 500 cultivated and wild accessions) collected from East Asia, South and Southeast Asia, and the Occident and Near East using the ddRAD-Seq method. In addition, genetic diversity analyses were conducted using ~2,600 SNPs without missing data (hereafter referred to as common SNPs). Based on our results, we also discuss the origin of cultivated radish and its migration from the West to East Asia.

2. Materials and methods

2.1. Materials

A total of 520 accessions were surveyed in this study (Supplementary Table S1). These accessions were obtained from the Tohoku University Brassica Seed Bank (http://www.agri.tohoku.ac.jp/pbreed/ Seed_Stock_DB/SeedStock-top.html) and Genebank Project of NARO, or purchased from the market. According to the taxonomic classification (species and variety) and region of cultivation, these accessions were categorized into nine groups (Table 1): (i) Japanese landraces of *R. sativus*; (ii) *R. sativus* var. *raphanistroides* accessions collected from the Hokuriku area of Japan; (iii) *R. sativus* F_1 cultivars bred in Japan; (iv) *R. sativus* accessions collected from East Asia (China, Korea, and Taiwan); (v) *R. sativus* accessions collected from South and Southeast Asia; (vi) *R. sativus* accessions collected from the Occident and Near East; (vii) *R. raphanistrum* accessions; (viii) *R. maritimus* accessions; and (ix) accessions from an unknown area.

2.2. Genome sequencing and SNP detection

To evaluate SNP heterogeneity in a single accession, leaves were sampled evenly from five to eight plants of each radish accession and ground to a fine powder. Genomic DNA was extracted using the DNeasy Plant Mini Kit (QIAGEN, the Netherlands) and subjected to ddRAD-Seq analysis, as described previously.[5,18] To perform ddRAD-Seq, DNA libraries of each accession were constructed using two restriction enzymes, *Pst*I and *Msp*I. The digested DNA fragments were ligated to adapters, and the ligated DNA samples were purified to eliminate short DNA fragments (<300 bp). The purified DNA samples were amplified by PCR using primers including index sequences.[18] The resulting amplicons were pooled, and 300–900 bp DNA fragments were purified. The purified fragments were sequenced on the

Table 1. Summary of radish (*Raphanus* spp.) accessions used in this study

Species	Cultivated area	Type	Number of accessions	Population symbol
R. sativus	Japan	Landraces	258	A1
		Wild radish (var. *raphanistroides*)	10	A2
		F_1 cultivars	64	A3
	East Asia	Landraces	90	B
	South and Southeast Asia	Landraces	24	C
	Occident, Near East	Landraces, wild radish	56	D
R. raphanistrum			13	RR
R. maritimus			3	RM
Unknown			2	
Total			520	

HiSeq2000 or HiSeq4000 (Illumina) platform to generate 93 or 101 bp paired-end reads. The data generated were submitted to the DDBJ Sequence Read Archive (DRA) database under the accession numbers DRA008624, DRA008625, DRA008636, and DRA008637. Low-quality sequences were removed, and adapters were trimmed using PRINSEQ (–trim_right 1 –trim_qual_right 10 –min_len 100 – derep) and fastx_clipper (–a AGATCGGAAGAGC –l 100 –M 10 –n) in the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit; version 0.10.1). The filtered reads were mapped onto radish scaffold sequences (RSA_r1.0)[2] using Bowtie 2 (version 2.1.0)[22] using the following parameters: —minins100 —no-mixed. The resulting sequence alignment/map (SAM) format files were converted to binary sequence alignment/map format files and subjected to SNP calling using the mpileup option of SAMtools (version 0.1.19; parameter: –Duf)[23] and the view option of BCFtools (parameter: –vcg). Heterozygous genotype calls might include the SNP heterogeneity in a single accession as well as heterozygous SNP genotypes. In both cases, it was meant that two different alleles existed in the accession. Variant call format (VCF) files were filtered using VCFtools (version 0.1.11; parameters: –max-alleles 2 –min-alleles 2 –minDP 5–minQ 999– max-missing 0.5 –remove-indels)[24] and then SNPs with a minor allele frequency (MAF) $\geq$0.01 were extracted using TASSEL ver. 5.2.50.[25] After excluding accessions with a low proportion of SNP sites (<40%), common SNPs were identified among the remaining 510 accessions (Supplementary Table S1) and counted.

2.3. Genetic diversity analysis using common SNP sites

Common SNPs (i.e. SNPs without missing data in all accessions including F$_1$ cultivars) were converted to a biallelic representation, e.g. mix base code 'M' to single base 'A' and 'C' and code 'A' to 'A' and 'A', at all sites. The converted SNPs were subjected to principal coordinate analysis (PCoA) and molecular phylogenetic analysis.

Principal coordinate analysis

Pairwise distance matrix was calculated by counting the total number of alleles identical between any two accessions using MEGA ver. 7.0.21.[26] Using the obtained distance matrix, PCoA was conducted with the cmdscale function in the stats package of R ver. 3.5.0 (https://www.r-project.org). The results were presented in a chart using Microsoft Office Excel 2016 (Microsoft Corporation, Redmond, USA).

Phylogenetic analysis

Phylogenetic analysis for all accessions was conducted using the concatenated alignments of common SNP sites. Phylogenetic trees were constructed using the maximum likelihood method and Find Best DNA/Protein Model in MEGA version 7.0.21,[26] with 1,000 bootstrap pseudo replicates.

For construction of an outline tree, accessions were selected as evenly as possible in the original tree. In particular, accessions from 'Kameido', 'Ninengo', 'Shigatsu', 'Shijunichi', 'Houryo', 'Miyashige', 'Shogoin', 'Awa Bansei', 'Nerima', 'Minowase', and 'Shiroagari', which are Japanese traditional varietal groups, were included together with several major regional cultivars and several accessions from Nansei Islands (Rs2017_016, Rs2017_018, and Rs2017_201) and southwestern area of Japan, i.e. Kyushu area (Rs2017_142, Rs2017_148, and Rs2017_158).

3. Results

3.1. SNPs identified among 520 radish accessions

Genome sequencing of 520 radish accessions using the ddRAD-Seq method generated 56.9 Gb of data. A total of 76,598 SNPs was identified among these accessions. Of these 520 accessions, 510 accessions excluding 10 that included 8 accessions with a remarkably low proportion of SNP sites (<40%) [Rs2017_065 (Takakura), Rs2017_073 (Bansei Miura), Rs2017_081 (Mera Daikon), Rs2017_089 (Kuroba Soubutori Minowase), Rs2017_204 (Kijikashira), Rs2017_317 (Tibet Kougen Daikon), Rs2017_323 (Everest), and Rs2017_445 (PAK-10445)] (Supplementary Table S1) and two accessions of unknown origin (Rs2017_458 and Rs2017_459) were used for further analyses. Among the remaining 510 accessions, 52,559 SNPs with MAF $\geq$0.01 were identified (Supplementary Table S2). We calculated the ratio of obtained SNP sites in each accession relative to the total number of SNP sites (52,559) (Supplementary Table S3). The average value of this ratio was 80.2%, although it ranged from 44.4% to 95.4%.

The 52,559 SNPs were located on 5,138 scaffolds of RSA_r1.0.[2] Of these 5,138 scaffolds, 903 scaffolds containing 19,114 SNP sites (highlighted in green in Supplementary Table S2) were mapped onto the high-density *R. sativus* restriction site-associated DNA sequencing (Rs-RAD) genetic map constructed by Shirasawa and Kitashiba[5] (Supplementary Fig. S1). The mapped SNPs were distributed on nine linkage groups (Rs1–9), with 1,285–3,054 SNPs per linkage group (Table 2). In addition, 2,624 common SNP sites were identified among 510 accessions (highlighted in yellow in Supplementary Table S4). The common SNP sites were also distributed across each linkage group, with 76–244 SNPs per linkage group (Table 2).

The ratio of a single homozygous allele to the total number of common SNP sites was calculated for each accession. This ratio was 76.0% on average and varied from a maximum of 97.0% in Rs2017_266 (Ankiku 1) to a minimum of 58.1% in Rs2017_413 (Shiromaru) (Supplementary Table S3). Among the radish populations of Japan without F$_1$ cultivars ($n = 263$), East Asia ($n = 88$), South and Southeast Asia ($n = 23$), and the Occident and Near East ($n = 56$), the ratios of homozygous allele to the total number of common SNP sites were 75.9%, 75.3%, 68.8%, and 72.6%, respectively (Supplementary Table S5).

Table 2. Number of SNPs mapped onto the Rs-RAD-based genetic map[5] in each of the nine linkage groups (Rs1–9)

Linkage group	Total mapped SNPs[a]	Common mapped SNPs[b]
Rs1	1,453	139
Rs2	2,425	201
Rs3	1,285	76
Rs4	2,201	152
Rs5	2,664	180
Rs6	3,054	249
Rs7	1,870	84
Rs8	2,259	95
Rs9	1,903	140
Total	19,114	1,316
Average	2,124	146

[a]SNPs mapped on to the Rs-RAD genetic map out of the total 52,559 SNPs identified in this study.

[b]SNPs mapped on to the Rs-RAD genetic map out of 2,624 common SNPs identified in this study.

Of the 2,624 common SNP sites, 2,262 (86.2%) were detected among Japanese *R. sativus* accessions. SNPs of 2,034 (77.5%), 2,122 (80.9%), 2,498 (95.2%), 2,283 (87.0%), and 1,770 (67.5%) were detected among *R. sativus* accessions from East Asia, South and Southeast Asia, and the Occident and Near East and among *R. raphanistrum* and *R. maritimus* accessions, respectively. In addition, the Japanese population shared 5,850 sites of the total 52,559 SNP sites, without missing sites, and SNPs were identified at 5,046 sites.

A survey of nucleotide substitution patterns in 510 radish accessions showed that the number of transitions (1,554) was higher than that of transversions (1,070), and the ratio of transitions to transversions was 1.45. This ratio was almost the same as that reported previously among 17 *Raphanus* accessions[14] and similar to that reported in *Brassica napus* (1.39), but quite different from that reported in *B. rapa* (1.0).[27]

3.2. Genetic diversity analysis
Principal coordinate analysis

PCoA was conducted using the common SNP sites identified among 510 accessions. Scores indicated by the first and second components (24.3% of the cumulative variance) for each sample were plotted on a chart (Fig. 1a). Although the 510 accessions were widely scattered on the chart, they could be grouped into four populations (clades): population A (*R. sativus* landraces and F$_1$ cultivars, and *R. sativus*
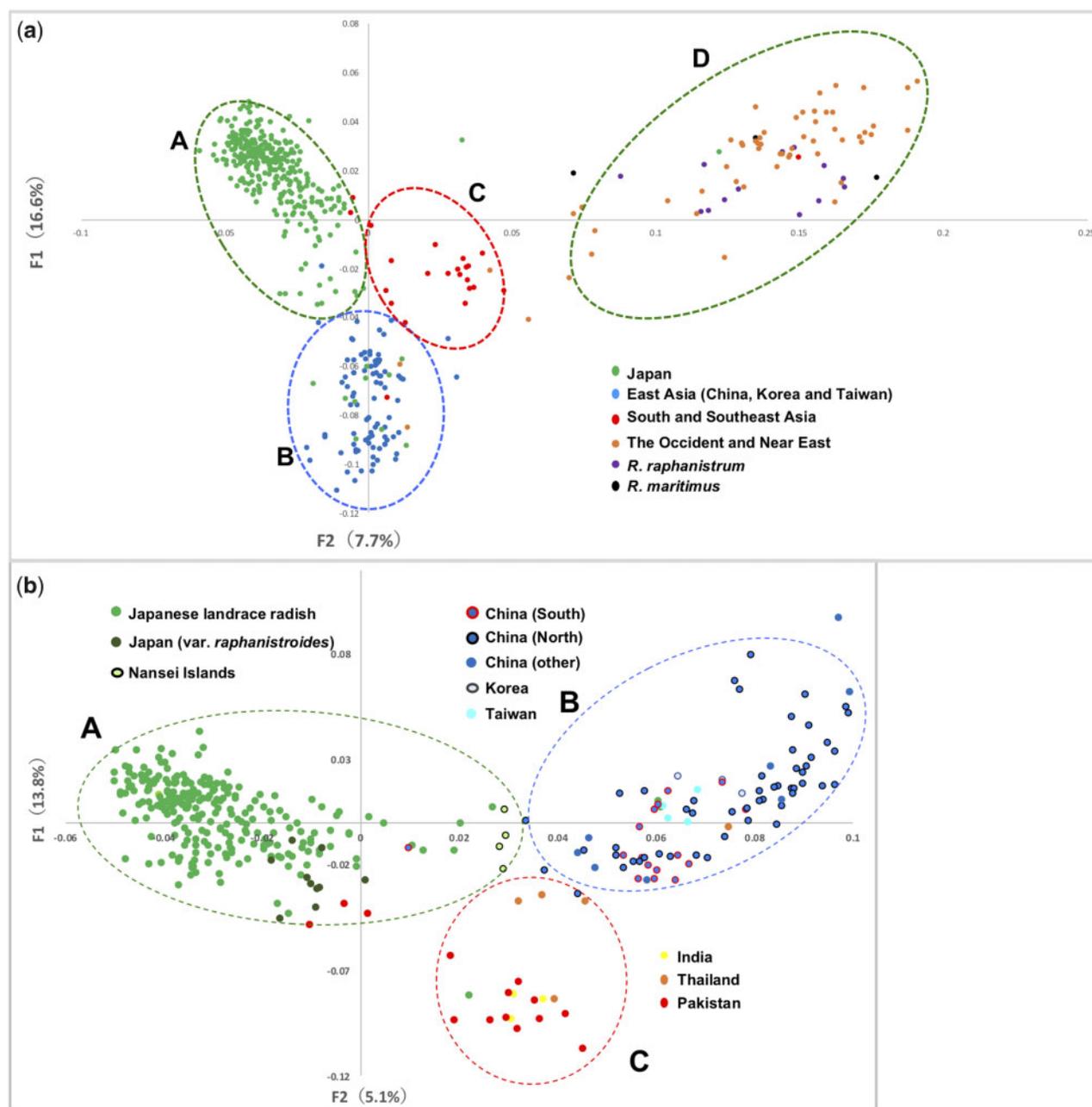

Figure 1. PCoA of radish accessions using common SNPs. PCoA of 510 accessions (a) and 372 Asian accessions (b).
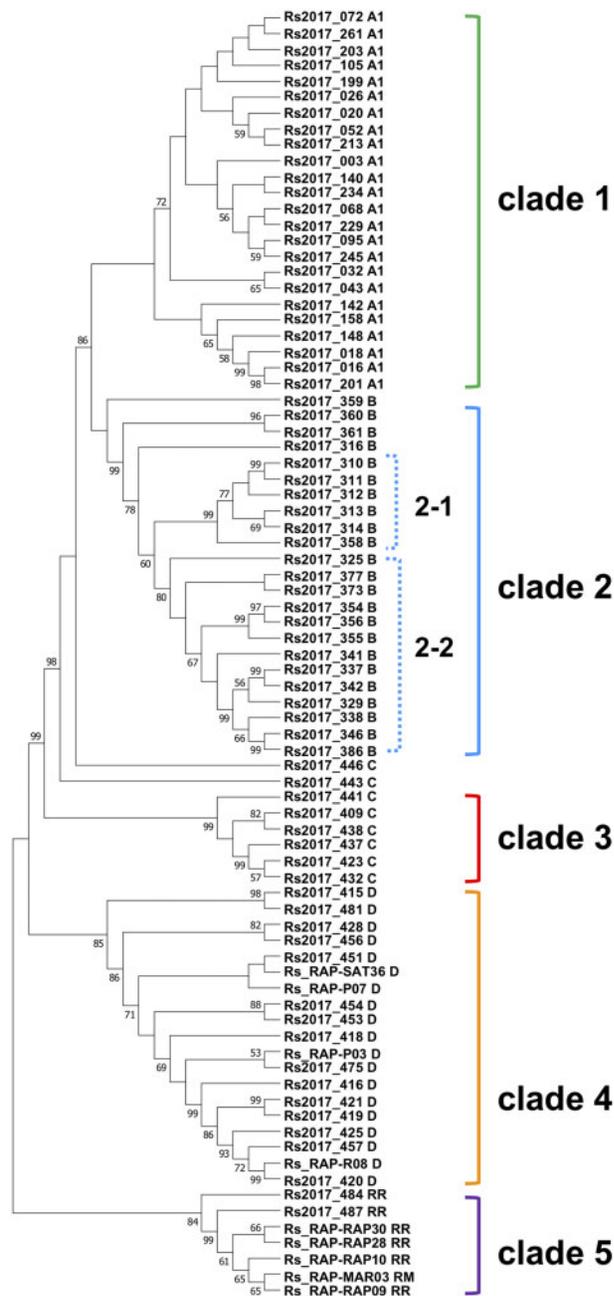
Figure 2. Phylogenetic tree of 81 representative radish accessions constructed using the maximum likelihood method.

var. *raphanistroides* accessions mainly from Japan), population B (*R. sativus* landraces mainly from East Asia), population C (*R. sativus* landraces mainly from South and Southeast Asia), and population D (*R. sativus* landraces mainly from the Occident and Near East as well as *R. raphanistrum* and *R. maritimus* accessions). Another PCoA of 372 Asian accessions, excluding F$_1$ cultivars from Japan, revealed three populations (A, B, and C) (Fig. 1b) similar to those shown in Fig. 1a.

Phylogenetic analysis and SNP ratio

Phylogenetic analyses of 510 accessions were conducted using the common SNP sites with the maximum likelihood method. The GTR+G model was determined as the best nucleotide substitution model in MEGA7. Phylogenetic tree constructed using common SNPs among all 510 accessions is shown in Supplementary Fig. S2. An outline tree was reconstructed using the SNPs of 81 accessions enclosed by rectangles as shown in Supplementary Fig. S2 (Fig. 2).The examined *Raphanus* accessions were divided into five clades (Supplementary Fig. S2 and Fig. 2). Clade 1 included accessions mainly categorized as population A in Fig. 1a. Clades 2 and 3 included accessions of populations B and C, respectively, shown in Fig. 1a. Clade 4 consisted of *R. sativus* accessions mainly from the Occident and Near East. Clade 5 included accessions of the wild species, *R. raphanistrum* and *R. maritimus*. Interestingly, clades 1 and 2 were separated from clade 3 with a high bootstrap value. Clade 2 was subdivided into two sub-clades (2-1 and 2-2). The clade 2-1 consisted of accessions mainly from south area of East Asia including Taiwanese accessions, Rs2017_312, Rs2017_313, and Rs2017_314. On the other hand, accessions of the north area including northern part of China and Korea were included in clade 2-2.

Next, we examined the coverage of 2,624 common SNP sites among the 81 accessions. Of the 2,624 common SNPs, 1,790 (68.2%) were detected in the Japanese clade. When landraces from other clades were added, the coverage of SNPs increased to 75.9% (clade 1 + clade 2) and 80.2% (clade 1 + clade 2 + clade 3). Drastic increases in SNP coverage were observed in groups comprising clades 1–4 (98.1%) and clades 1–5 (99.0%). In addition, these 81 accessions shared 5,211 sites of the total SNP sites (52,559), without missing data, and SNPs at most of the sites (98.8%) were detected.

4. Discussion

4.1. Large number of SNPs from geographically diverse radish accessions

To date, nucleotide polymorphisms among radish accessions have been analysed in several studies. Shen et al.[28] identified 28,758 SNPs among 18 radish accessions, based on the analysis of expressed sequence tags (ESTs) available in the NCBI database of ESTs. Kim et al.[14] revealed a total of 4,033,588 SNPs and Indels among 17 radish accessions using their own re-sequencing data and publicly available genome sequence information, although the number of the polymorphic sites was not reported. From the viewpoint of the number of radish accessions, Wang et al.[15] reported DNA polymorphisms among 93 radish accessions using 559 EST–SSR markers. Previously, we detected 17,341 SNPs among seven accessions from Japan and one accession from Southeast Asia.[2,12] In the present study, we identified SNPs in a much larger collection of radish accessions with a wide geographical distribution and diverse morphological features and end uses (phenotypic data are available in Genebank project, NARO). While 52,559 SNP sites were identified among 510 radish accessions, the total number of SNPs among 510 accessions was estimated as 21,444,072 SNPs (52,559 sites × 0.80 [average ratio of the obtained SNP sites] × 510 accessions). Thus, our study adds valuable SNP information to the databases described above, in terms of both the number of SNPs and number of accessions. We have also constructed a new URL (http://radish.kazusa.or.jp), where comprehensive SNP data will be publicly available.

Previously, amplicon sequencing[2] and genome re-sequencing[14] approaches were used to detect genome-wide SNPs in radish. Recently, ddRAD-Seq has proven to be highly effective for the detection of genome-wide SNPs. In tomato, a large number of SNPs (19,969) were identified between two inbred lines[18] and used for the

construction of a genetic map. The ddRAD-Seq method has also been used to detect SNPs in other crops, such as sweet potato, sweet cherry, and ornamental cherry[29–31] and to produce a high-density genetic map in radish.[5] Moreover, this technique successfully detected a large number of genome-wide SNPs in several radish accessions in the present study. These data suggest that the ddRAD-Seq is one of the most effective tools for genome analysis.

4.2. Genetic diversity in the genus *Raphanus*

PCoA identified four populations (A–D; Fig. 1a) among the 510 radish accessions examined in this study. Clades 1–3 detected by phylogenetic analysis almost exactly corresponded to populations A, B, and C detected by PCoA, respectively. Accessions in clade 4 were present in population D. Therefore, according to the results of genetic analyses, *R. sativus* accessions were grouped into four populations. These four populations showed a strong association with their geographic distribution; (i) Japan, (ii) China and Korea, (iii) South and Southeast Asia, and (iv) the Occident and Near East. According to plant morphology and end uses, *R. sativus* is further categorized into several varieties, such as var. *sativus*, var. *niger*, var. *caudatus*, and var. *hortensis*.[20,21] Most of the European small radishes and black radishes are categorized as *R. sativus* var. *sativus* and var. *niger*, respectively.[20,21] In the present study, both varieties were included in population D detected by PCoA. Rat-tailed radish cultivated in South and Southeast Asia is categorized as *R. sativus* var. *caudatus*,[20,21] and most of the accessions belonging to var. *caudatus* were included in population C. Most of the big radishes including thick taproot, large taproot, long taproot, and round taproot found in East Asia and Japan were categorized as *R. sativus* var. *hortensis*. Most accessions of *R. sativus* var. *hortensis* used in this study were included in either population A or population B. Thus, genetic diversity in radish accessions showed a strong relationship with the variety and geographical distribution.

4.3. Origin and migration of cultivated radish

Among the five clades detected by phylogenetic analysis, four clades comprised cultivated radish accessions, while one clade comprised wild radish species such as *R. raphanistrum* and *R. maritimus*. In addition, cultivated radish was separated from the wild species (Fig. 2). A similar result was reported by Wang et al.[15] in a study on 93 radish accessions and 50 SSR markers. Although the number of samples examined by Shen et al.[28] and Kim et al.[14] was limited (<22 samples), phylogenetic analysis in these studies also showed similar divergence of cultivated radish from wild species. These results suggest that cultivated radish species were derived either from one of the wild *Raphanus* species (*R. raphanistrum*, *R. maritimus*, or another wild species) or from a common ancestor of the genus *Raphanus*. Furthermore, since wild *Raphanus* species are thought to have originated in the Mediterranean region,[32] it would be reasonable to speculate that cultivated radish (*R. sativus*) also originated around the same area.

It remains unclear how cultivated radish transferred from the West to East Asia. One possibility, as suggested by Kitamura,[32] Huh and Ohnishi,[33] and Wang et al.,[34] is that the cultivated species (*R. sativus*) originated in the Mediterranean region and migrated towards the Black Sea and Caspian Sea, and then further to the East, i.e. East Asia, along a certain road such as the Silk Road. During migration, new varieties such as var. *caudatus* and var. *hortensis* probably originated, and then migrated to and domesticated in South Asia, Southeast Asia, and East Asia (Fig. 3a, hypothesis-a).
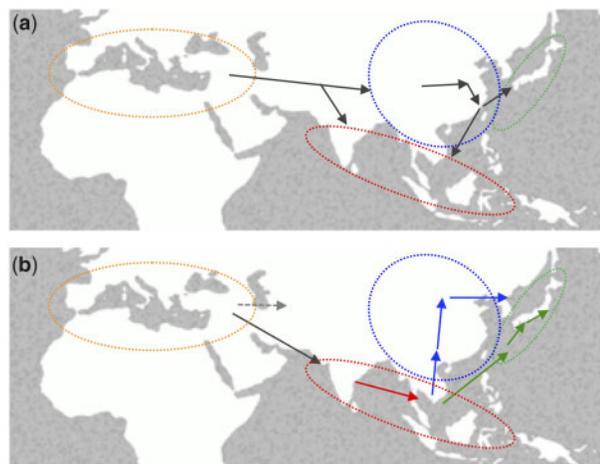

Figure 3. Possible migration route of cultivated *Raphnaus* species from the West to East Asia. (a) Conventional hypothesis, based on previous studies. (b) Proposed hypothesis, based on the results of this study. Arrow means direction of the migration.

Furthermore, *R. sativus* var. *hortensis* is thought to have been transported to Japan via Southern China.

However, our results of phylogenetic analysis and PCoA led us to another hypothesis (Fig. 3b, hypothesis-b). Populations A–D detected by PCoA corresponded almost exactly to the phylogenetic clades 1–4. Based on the results of phylogenetic analysis, cultivated *Raphanus* species in the Asian group are inferred to have diverged from those in the Western group. Within the Asian group, the ancestors of East Asian and Japanese groups are inferred to have diverged from the South and Southeast Asian group. In the East Asian group, ancestral cultivars of some present accessions such as Rs2017_359, _360 and _361 in the southern area diverged from those in the South and Southeast Asian group. Furthermore, it is deduced that a part of the ancestral cultivars spread in southern area, while another part of the ancestral ones, which might have acquired cold tolerance possibly due to increase in genetic variation, migrated to the northern area. In addition, within the Japanese group, radish accessions in southwestern Japan (Nansei Islands), e.g. Rs2017_016, and in Kyushu area, e.g. Rs2017_158, diverge from those in the South and Southeast Asian group and then the others of mainland diverge. This implies that a part of ancestral cultivated varieties of the South and Southeast Asian group migrated towards the main island of Japan through southwestern area of Japan. Meanwhile, genetic variation might increase due to nucleotide substitutions and/or crosses between different cultivars. In summary, it is suggested that the cultivated species, *R. sativus*, originated in the West and migrated to the South and Southeast Asia, although a part of the cultivated species may have migrated further east from the West as suggested by Kitamura[32] (a broken grey arrow, in Fig. 3b). Furthermore, those probably spread independently to area in China and to area in Japan from South and Southeast Asia (Fig. 3b; hypothesis-b).

In the analyses of mitochondrial DNA, five different haplotypes of the *cox1–orfB* region detected among wild radish species have also been observed among populations from three different areas: Europe, Asia including China, Korea, and countries of the Southeast Asia, and Japan.[35] However, the observed haplotypes showed a strong association with the geographical areas. For example, the European population contained one haplotype, whereas Asian

populations contained four to five haplotypes. In addition, haplotypes within the Asian populations showed different frequencies,[35] indicating that radishes were domesticated in the respective areas; this is supported by the results of a similar analysis of chloroplast genes.[36] In addition, phylogenetic analysis of the chloroplast *trnK/matK* region conducted by Lü et al.[21] indicated that three kinds of cultivated radishes, including European small radish, Asian big radish, and black radish, belong to different clusters, suggesting that the domestication of cultivated radish occurred independently in the respective areas. To further understand the route of migration of radishes in more detail, haplotype analyses of mitochondrial or chloroplast genes are needed in a large number of geographically diverse accessions.

In genetic analyses such as QTL analysis and genome-wide association study, genetic distance between cultivars greatly affects the efficiency of analysis. A high degree of DNA polymorphisms among cultivars is expected to discover trait-associated loci or genes at a high resolution. SNP data revealed in this study will facilitate the selection of cultivars based on genetic analysis. Furthermore, SNP-related information will enable the production of DNA markers for the accurate identification of cultivars and will greatly contribute to the genetic analyses and breeding of radish cultivars.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at DNARES online.

References

1. Jeong, Y.M., Kim, N., Ahn, B.O., et al. 2016, Elucidating the triplicated ancestral genome structure of radish based on chromosome-level comparison with the *Brassica* genomes, *Theor. Appl. Genet.*, **129**, 1357–72.
2. Kitashiba, H., Li, F., Hirakawa, H., et al. 2014, Draft sequences of the radish (*Raphanus sativus* L.) genome, *DNA Res.*, **21**, 481–90.
3. Mitsui, Y., Shimomura, M., Komatsu, K., et al. 2015, The radish genome and comprehensive gene expression profile of tuberous root formation and development, *Sci. Rep.*, **5**, 10835.
4. Zhang, X., Yue, Z., Mei, S., et al. 2015, A *de novo* genome of a Chinese Radish cultivar, *Hortic. Plant J.*, **1**, 155–64.
5. Shirasawa, K. and Kitashiba, H. 2017, Genetic maps and whole genome sequences of Radish, In: Nishio, T. and Kitashiba, H. (eds) *The Radish Genome*, Springer International Publishing, Switzerland, pp. 31–42.
6. Kamei, A., Tsuro, M., Kubo, N., et al. 2010, QTL mapping of clubroot resistance in radish (*Raphanus sativus* L.), *Theor. Appl. Genet.*, **120**, 1021–7.
7. Yu, X., Choi, S.R., Ramchiary, N., et al. 2013, Comparative mapping of *Raphanus sativus* genome using *Brassica* markers and quantitative trait loci analysis for the Fusarium wilt resistance trait, *Theor. Appl. Genet.*, **126**, 2553–62.
8. Kitashiba, H. and Yokoi, S. 2017, Genes for bolting and flowering, In: Nishio, T. and Kitashiba, H. (eds) *The Radish Genome*, Springer International Publishing, Switzerland, pp. 151–163.
9. Wang, Q., Zhang, Y. and Zhang, L. 2018, A naturally occurring insertion in the *RsFLC2* gene associated with late-bolting trait in radish (*Raphanus sativus* L.), *Mol. Breed.*, **38**, 137.
10. Yi, G., Park, H., Kim, J.S., Chae, W.B., Park, S. and Huh, J.H. 2014, Identification of three *FLOWERING LOCUS C* genes responsible for vernalization response in radish (*Raphanus sativus* L.), *Hortic. Environ. Biotechnol.*, **55**, 548–56.
11. Ishida, M., Kakizaki, T., Morimitsu, Y., et al. 2015, Novel glucosinolate composition lacking 4-methylthio-3-butenyl glucosinolate in Japanese white radish (*Raphanus sativus* L.), *Theor. Appl. Genet.*, **128**, 2037–46.
12. Zou, Z., Ishida, M., Li, F., et al. 2013, QTL analysis using SNP markers developed by next-generation sequencing for identification of candidate genes controlling 4-methylthio-3-butenyl glucosinolate contents in roots of radish, *Raphanus sativus* L, *PLoS One*, **8**, e53541.
13. Kakizaki, T., Kitashiba, H., Zou, Z., et al. 2017, A 2-oxoglutarate-dependent dioxygenase mediates the biosynthesis of glucoraphasatin in radish, *Plant Physiol.*, **173**, 1583–93.
14. Kim, N., Jeong, Y.M., Jeong, S., et al. 2016, Identification of candidate domestication regions in the radish genome based on high-depth resequencing analysis of 17 genotypes, *Theor. Appl. Genet.*, **129**, 1797–814.
15. Wang, Q., Zhang, L. and Zheng, P. 2015, Genetic diversity and evolutionary relationship analyses within and among *Raphanus* species using EST-SSR markers, *Mol. Breed.*, **35**, 62.
16. Baird, N.A., Etter, P.D., Atwood, T.S., et al. 2008, Rapid SNP discovery and genetic mapping using sequenced RAD markers, *PLoS One*, **3**, e3376.
17. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. and Hoekstra, H.E. 2012, Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species, *PLoS One*, **7**, e37135.
18. Shirasawa, K., Hirakawa, H. and Isobe, S. 2016, Analytical workflow of double-digest restriction site-associated DNA sequencing based on empirical and *in silico* optimization in tomato, *DNA Res.*, **23**, 145–53.
19. Suyama, Y. and Matsuki, Y. 2015, MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform, *Sci. Rep.*, **5**, 16963.
20. Yamagishi, H. 2017, Speciation and diversification of radish, In: Nishio, T. and Kitashiba, H. (eds) *The Radish Genome*, Springer International Publishing, Switzerland, pp. 11–30.
21. Lü, N., Yamane, K. and Ohnishi, O. 2008, Genetic diversity of cultivated and wild radish and phylogenetic relationships among *Raphanus* and *Brassica* species revealed by the analysis of *trnK/matK* sequence, *Breed. Sci.*, **58**, 15–22.
22. Langmead, B. and Salzberg, S.L. 2012, Fast gapped-read alignment with Bowtie 2, *Nat. Methods*, **9**, 357–9.
23. Li, H., Handsaker, B., Wysoker, A., et al. 2009, The sequence alignment/map format and SAMtools, *Bioinformatics*, **25**, 2078–9.
24. Danecek, P., Auton, A., Abecasis, G., et al. 2011, The variant call format and VCFtools, *Bioinformatics*, **27**, 2156–8.
25. Bradbury, P.J., Zhang, Z., Kroon, D.E., et al. 2007, TASSEL: software for association mapping of complex traits in diverse samples, *Bioinformatics*, **23**, 2633–5.
26. Kumar, S., Stecher, G. and Tamura, K. 2016, MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets, *Mol. Biol. Evol.*, **33**, 1870–4.
27. Park, S., Yu, H.J., Mun, J.H. and Lee, S.C. 2010, Genome-wide discovery of DNA polymorphism in *Brassica rapa*, *Mol. Genet. Genomics*, **283**, 135–45.
28. Shen, D., Sun, H., Huang, M., et al. 2013, Comprehensive analysis of expressed sequence tags from cultivated and wild radish (*Raphanus* spp.), *BMC Genomics*, **14**, 721.

29. Shirasaswa, K., Isuzugawa, K., Ikenaga, M., et al. 2017, The genome sequence of sweet cherry (*Prunus avium*) for use in genomics-assisted breeding, *DNA Res.*, **24**, 499–508.

30. Shirasawa, K., Tanaka, M., Takahata, Y., et al. 2017, A high-density SNP genetic map consisting of a complete set of homologous groups in autohexaploid sweet potato (*Ipomoea batatas*), *Sci. Rep.*, **7**, 44207.

31. Shiraswa, K., Esumi, T., Hirakawa, H., et al. 2019, Phased genome sequence of an interspecific hybrid flowering cherry, 'Somei-Yoshino' (*Cerasus × yedoensis*), DNA Res., **26**, 379–89.

32. Kitamura, S. 1958, Varieties of radish and their transition, In: Nishiyama, I. (ed.) *Japanese Radish*, Jap. Sci. Soc. Press, Tokyo, pp. 1–19.

33. Huh, M.K. and Ohnishi, O. 2002, Genetic diversity and genetic relationships of east Asia natural population of wild radish revealed by AFLP, *Breed. Sci.*, **52**, 79–88.

34. Wang, N., Kitamoto, N., Ohsawa, R. and Fujimura, T. 2008, Genetic diversity of radish (*Raphanus sativus*) germplasms and relationships among worldwide accessions analyzed with AFLP markers, *Breed. Sci.*, **58**, 107–12.

35. Yamagishi, H. and Terachi, T. 2003, Multiple origins of cultivated radishes as evidenced by a comparison of the structural variations in mitochondrial DNA of *Raphanus*, *Genome*, **46**, 89–94.

36. Yamagishi, H., Terachi, T., Ozaki, A. and Ishibashi, A. 2009, Inter- and intraspecific sequence variations of the chloroplast genome in wild and cultivated *Raphanus*, *Plant Breed.*, **128**, 172–7.