The cover page:

Corresponding author

Naoko ISHIKAWA

e-mail: naishikawa@gmail.com

Number of tables, black-and-white figures, and color figures.

tables 8

black-and-white figures 3

color figures 4

Electronic Supplementary Material

figures 2

Title page:

Title: Allotetraploid origin and putative ancient introgression in *Plantago hakusanensis* (Plantaginaceae)

Naoko Ishikawa1\*, Shota Sakaguchi2, Chikako Hasekura3, Alexey Shipunov4, Ayumi Matsuo1, Yoshihisa Suyama1, Hirokazu Tsukaya5, Hiroshi Ikeda6, Motomi Ito7,

1. Kawatabi Field Science Center, Graduate School of Agricultural Science, Tohoku University, Yomogida 232-3, Osaki, Miyagi 989-6711, Japan.

2. Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan

3. Faculty of Agriculture, Tokyo University of Agriculture，Funako 1737， Atsugi， Kanagawa 243-0034, Japan

4. Department of Biology, Minot State University, Minot, North Dakota, USA 58707

5. Graduate School of Science, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

6. The University Museum, The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-0033, Japan

7. Graduate School of Arts and Sciences, The University of Tokyo, Komaba 3-8-1, Meguro-ku, Tokyo, 153-8902, Japan

Corresponding author

\*Naoko ISHIKAWA

e-mail: naishikawa@gmail.com

**Abstract**

*Plantago hakusanensis* (2n = 4x = 24) is an endangered endemic species that occurs in subalpine zones in Japan. To clarify the unresolved taxonomic status of *P*. *hakusanensis* within the subgenus *Plantago*, we conducted a phylogenetic analysis based on the nuclear-encoded single-copy gene sucrose–proton symporter 1(*SUC1*) using 60 previously reported alleles from 24 taxa in the subgenus *Plantago*. We found that *P*. *hakusanensis* was closely related to *P*. *asiatica* var. *densiuscula*. The phylogenetic relationships between *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula* were examined by analyses of the *SUC1* nuclear regions and the internal transcribed spacer (ITS) of rDNA, genome-wide single-nucleotide polymorphism genotyping (via multiplexed inter-simple sequence repeat genotyping by sequencing), as well as additional analyses of three chloroplast (cp) regions (*trnL-F*, *ndhF-rpl32*, and *rpl32-trnL*) in 25 individuals of *P*. *hakusanensis* and 53 individuals of *P*. *asiatica* var. *densiuscula*. Monophyly of *P*. *hakusanensis* was suggested by the nuclear marker analyses, whereas the cp haplotypes of *P*. *hakusanensis* were shared with *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* in China. The disparity between the nuclear and cp data may be explained by introgression of the cp genome (cp capture) during Quaternary climate changes. Our results provide (i) a molecular phylogenetic basis for the taxonomy and (ii) insight into the intraspecific diversification history of *P*. *hakusanensis*.

**Keywords**

Allopolyploid, chloroplast capture, endangered species, introgression, *Plantago*, phylogeographical inference

**Introduction**

*Plantago hakusanensis* Koidz. (Plantaginaceae) (2n = 4x = 24) is a perennial herb endemic to Japan (Yamazaki 1993). *P. hakusanensis* is distributed around snow patches and wet fields in subalpine zones in Honshu (*ca*. 1500–2300 m above sea level; Koidzumi 1930; Yamazaki 1992, 1993). The species has been found on 12 mountains located between Mt. Moriyoshi (Akita Prefecture) to the north and Mt. Hakusan (Ishikawa Prefecture) to the south (Yamada and Satomi 1975). Due to their rarity, *P*. *hakusanensis* and its hairless form, *P*. *hakusanensis* f. *glabra* T.Yamaz.,have been recognized as category I endangered taxa in Akita and Nagano Prefectures, and as category II vulnerable taxa in Fukushima, Gunma, Ishikawa, and Gifu Prefectures (Search System of Japanese Red Data 2020). Based on morphological features, *P*. *hakusanensis* has been assigned to the subgenus *Plantago*, which comprises *ca*. 131 species in five sections (Rahn 1996). Rahn (1996) suggested that *P*. *hakusanensis* belongs to sect. *Plantago* andis related to *P*. *asiatica* L. But the section assignment of *P*. *hakusanensis* has been debated by some authors. For example, Yamazaki (1992) included *P*. *hakusanensis* in sect. *Gentianoides* Pilg. and considered it as the central Asian species closest to *P*. *gentianoides* Sm. However, no adequate morphological or molecular phylogenetic analyses to judge which is correct have been reported.

*P*. *hakusanensis* has been identified as a conservation target on Mt. Hakusan because of concerns that it may be genetically polluted via hybridization with *P*. *asiatica* L.var. *densiuscula* Pilg. (2n = 4x = 24) (Nogami 2001; Nakayama et al. 2006, 2008; Sano et al.2016, 2019). *P*. *asiatica* var. *densiuscula* is a perennial herb distributed from Central China (Kiangsu and Hunan Provinces) to the Japanese archipelago. It commonly grows in sunny locations, such as roadsides and unpaved parking lots, at low elevations (Pilger 1922; Yamazaki et al. 1993; Ishikawa et al. 2006; Ishikawa et al. 2009). *P*. *asiatica* var. *densiuscula* is morphologically similar to *P*. *hakusanensis*, but the two taxa differ in several aspects, including seed number per fruit, seed morphology, and leaf shape. *P*. *hakusanensis* has 1–2 seeds per fruit, whereas *P*. *asiatica* var. *densiuscula* has 4–7 seeds per fruit (Pilger 1922; Yamazaki et al. 1993; Nakayama et al. 2008; Ohashi et al. 2017). *P*. *asiatica* var. *densiuscula* has invaded native habitats of *P*. *hakusanensis* on Mt. Hakusan. The sticky wet seeds of *P*. *asiatica* var. *densiuscula* may have been transferred to the habitats of *P*. *hakusanensis* via the shoes of hikers (Nogami 2001). The two taxa are genetically compatible, and fertile F1 hybrids have been obtained by artificial pollination (Sano et al.2016). The occurrence of protogynous and wind-pollinated flowers in the two taxa increases their out-crossing rate, and moreover, they had overlapping flowering periods on Mt. Hakusan during 3 out of the 4 years between 2011 and 2014 (Sano et al.2019). Putative hybrids with an intermediate leaf shape have been found on Mt. Hakusan at locations where the taxa are sympatric (Nakayama et al. 2008). The government ministries of Japan (Ministry of Agriculture, Forestry and Fisheries of Japan et al. 2015) led efforts to remove *P*. *asiatica* var. *densiuscula* from Mt. Hakusan, where it is regarded as an exotic taxon. Genetic pollution is also a concern in other locations. Invasions of *P*. *asiatica* var. *densiuscula* have been reported on Mt. Gassan (Yokoyama 2015). Although *P*. *hakusanensis* is (i) an endemic species with limited distribution and (ii) threatened by hybridization with *P*. *asiatica* var. *densiuscula*, its phylogenetic relationships and evolutionary origins have not been investigated adequately.

Polyploidy is frequent in the subgenus *Plantago* (67% of species, as shown by chromosome counts, Rahn 1996), and allopolyploidy has been identified by molecular phylogenetic evidence (Ishikawa et al. 2009). Although next-generation sequencing allows processing of large numbers of DNA sequences, it is still not easy to obtain molecular data that can resolve the relationships among polyploids. Hence, we aimed to determine the phylogenetic position of tetraploid *P*. *hakusanensis* within the subgenus *Plantago* via phylogenetic analyses based on the nuclear-encoded single-copy gene sucrose–proton symporter 1(*SUC1*) using 60 previously reported alleles from 24 representative taxa in the subgenus *Plantago*. We obtained the DNA sequence of *SUC1* in *P*. *hakusanensis* using either cloning or allele (homoeolog)-specific PCR amplification. To identify possible hybridizations between *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula*, we conducted phylogenetic analyses of the nuclear-encoded rDNA internal transcribed spacer (ITS) regions and genome-wide single-nucleotide polymorphism (SNP) genotyping (via multiplexed inter-simple sequence repeat genotyping by sequencing [MIG-seq]) (Suyama and Matsuki 2015); three chloroplast (cp) regions (*trnL-F*, *ndhF-rpl32*, and *rpl32-trnL*). We found that (i) *P*. *hakusanensis* was phylogenetically closest among examined species to *P*. *asiatica* var. *densiuscula*, and (ii) the phylogenies of nuclear and cp DNA were incongruent among populations of the two taxa. The disparity between nuclear and cp DNA phylogenies may be explained by cp genome introgressions (cp capture) during Quaternary climate changes, although the possibility of incomplete lineage sorting of ancestral polymorphisms cannot be ruled out.

**Materials and methods**

**Taxon sampling and DNA isolation**

To determine the taxonomic position of *P*. *hakusanensis* in the subgenus *Plantago*, we performed a phylogenetic analysis based on (i) the region extending from exon 1 to exon 2 of the nuclear-encoded single-copy *SUC1* gene collected from two individuals of *P*. *hakusanensis* growing on Mt. Hakusan, and (ii) 60 previously reported alleles in 24 taxa representing all five sections of the subgenus *Plantago* (Ishikawa et al. 2009, Table 2). Alleles were isolated from one individual of each of these 24 taxa, except in the case of diploid *P*. *major* L., from which we collected two individuals. The numbers of alleles obtained from each individual varied from one to seven, depending on the levels of ploidy and/or heterozygosity (Ishikawa et al. 2009). We selected *P*. *tenuiflola* Waldst. et Kit, one of the 24 taxa evaluated by Ishikawa et al. (2009), as an outgroup based on previously determined phylogenetic relationships within subg. *Plantago* (Rønsted et al. 2002; Ishikawa et al. 2009; Iwanycki Ahlstrand et al. 2019). Note that *P*. *formosana* Tateishi et Masam. is sometimes considered to be conspecific with *P*. *asiatica* or *P*. *major* (Hatusima 1971; Shimabuku 1997).

The phylogenetic analysis showed that *P*. *hakusanensis* is closely related to *P*. *asiatica* var. *densiuscula*. Therefore, to further investigate differences between *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula*, we conducted additional analyses using(i)genetic polymorphisms obtained fromthe same *SUC1* region (extending from exon 1 to exon 2), (ii) the nuclear-encoded rDNA ITS regions, and (iii) three cp regions (*trnL-F*, *ndhF-rpl32*, and *rpl32-trnL*). We also used MIG-seq to identify SNPs (Suyama and Matsuki 2015). We used 25 individuals of *P*. *hakusanensis*, and 53 individualsof *P. asiatica* var. *densiuscula*, including *P*. *asiatica* var. *densiuscula* f. *yakusimensis* (Masam) N.Ishikawa et al. The samples of *P*. *hakusanensis* were collected from Mt. Hakusan (4 individuals), Mt. Chokai (3 populations, 17 individuals), Mt. Gassan (1 individual), and Mt. Asahi (3 individuals), thereby covering the main distribution range of the taxon. Mt. Hakusan is the lectotype locality, and both Mt. Chokai and Mt. Gassan were listed as known localities of *P*. *hakusanensis* specimens in the original description (Koidzumi 1930). The samples of *P*. *asiatica* var. *densiuscula* comprised 46 individuals collected from Taiwan, Cheju Island in Korea, and a broad geographical range across the Japanese archipelago. We also included seven individuals from the subalpine zone on Mt. Chokai to investigate hybridization between *P*. *hakusanensis* and *P*. *asiatica var*. *densiuscula* invaders in the subalpine zone. Both taxa were found on Mt. Chokai along a mountain trail at elevations above *ca*. 1200 m. *P*. *hakusanensis* plants were distributed around snow patches, whereas *P*. *asiatica* var. *densiuscula* plants were found around formerly and currently used mountain huts (Table 1).

Data for the phylogenetic analysis based on the *SUC1* region were obtained from 11 individuals of *P*. *hakusanensis*, 1 individual of *P*. *asiatica* var. *densiuscula*, and1 individual of *P*. *asiatica* var. *densiuscula* f. *yakusimensis*. The North American putative tetraploid *P*. *rugelii* Decne. was added as an outgroup. Data for the phylogenetic analysis based on ITS sequences were obtained from 25 individuals of *P*. *hakusanensis*, 48 individuals of *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* var. *densiuscula* f. *yakusimensis* (including 38 previously reported individuals; Ishikawa et al.2006, Table 1), and 1 individual of *P*. *asiatica* from China (Rønsted et al. 2002). We also included *P*. *camtschatica* Link., *P*. *major*, and *Plantago major* var. *japonica* (Franch. et Sav.) Miyabe as related taxa. Phylogenetic analysis of the MIG-seq data was performed using five representative individuals of *P*. *hakusanensis*, five of *P*. *asiatica* var. *densiuscula*, and one of *P*. *asiatica* var. *densiuscula* f. *yakusimensis*. The selected samples spanned the geographical range of each taxon. MIG-seq is a PCR-based method that concentrates and isolates inter-simple sequence repeat regions located mainly in the nuclear genome (Suyama and Matsuki 2015). The cp phylogenetic analysis included 24 individuals of *P*. *hakusanensis* from the four mountains and 37 specimens of *P*. *asiatica* var. *densiuscula* that included plants collected from the subalpine zone on Mt. Chokai (Table 1). We also included the cp sequences of one individual of *P*. *asiatica* from China (Rønsted et al. 2002; Iwanycki Ahlstrand et al. 2019; Kew DNA bank ID=9585, K) and two individuals of *P*. *camtschatica* as related taxa (Iwanycki Ahlstrand 2019). Although the *ITS* analysis included *P*. *major*, the species was excluded from the cpDNA analysis because excessive numbers of polymorphisms were found in *P*. *major* relative to *P*. *hakusanensis* (compared with the other taxa included in this analysis).

Total genomic DNA was extracted from fresh or dried leaves using a slightly modified cetyltrimethylammonium bromide method (Murray and Thompson 1980).

**Chromosome observations**

We checked the chromosome number of *P*. *hakusanensis* to confirm the previous reports (Yamazaki 1983, Ohashi 2017) which could not be traced the original data. Three individuals of *P*. *hakusanensis* were transplanted from Mt. Chokai to a nursery in the University of Tokyo, then used for cytological observation. Fresh root tips were pretreated in 2 mM 8-hydroxyquinoline solution for 1 h at 20°C and then stored at 4°C for 15 h. We subsequently fixed them in Newcomer’s fluid (6:3:1:1:1 isopropanol, propionic acid, petroleum ether, acetone, 1,4-dioxane). The root tips were macerated in 1 N HCl at 60°C for 10 min, then stained with 2% lacto-propionic orcein, and squashed for cytological observation.

**PCR amplification and DNA sequencing of the *SUC1* region**

Preliminary analysis indicated that determination of the nucleotide sequence of the *SUC1* region by direct sequencing of the PCR product would be difficult, presumably because of the allotetraploid origin of *P*. *hakusanensis*. Thus, we applied two methods to efficiently determine two homoeologs of *SUC1* (i.e., Homoeolog L and Homoeolog I) of *P*. *hakusanensis*. First, the nucleotide sequences of two individuals collected from Mt. Hakusan were determined by the cloning and sequencing method described by Ishikawa et al. (2009). In this procedure, the PCR conditions were optimized to avoid over-amplification, which potentially produces artificial recombinants among multiple alleles (homoeologs) of the polyploid (Bradley and Hillis 1997; Judo et al. 1998; Kanagawa 2003). Second, the *SUC1* sequences of nine *P*. *hakusanensis* individuals were obtained by homoeolog-specific PCR amplification and direct sequencing because the cloning and sequencing method is excessively laborious. Primers for the specific PCR were designed using polymorphic sites between the homoeologs obtained by the cloning and sequencing method. Each of two *SUC1* homoeologs(Homoeolog L and Homoeolog I) was amplified as two overlapping regions and assembled into a continuous sequence (Fig. 1). For example, Homoeolog L was amplified using two primer sets: (i) SUC1-F11 (5′-ATGGGTGAATTGTCAGGAATTGAA-3′) and SUC1-hJ-R1 (5′-TCAAACAAATTCTGAAGTC-3′) and (ii) SUC1-hJ-F1 (5′-GATCCGTTCAATACTGATAGATCCA-3′) and SUC1-R4 (5′-GAGCCACCATGTCTTAG-3′). Homoeolog I was amplified using the following primer sets: (iii) SUC1-F11 and SUC1-hM-R2 (5′-CGATGTATACCTCTTCTATG-3′) and (iv) SUC1-hM-F2: CATGGTACGGACATGGAAATGG and SUC1-R4. The homoeolog-specific PCR parameters were as follows: incubation at 94°C for 1 min, 20 cycles of touchdown PCR (denaturation at 94°C for 30 s, annealing at 60°C for 30 s with a temperature reduction of 0.5°C per cycle, and extension at 72°C for 1 min), 20 cycles of non-touchdown PCR (denaturation at 94°C for 30 s, annealing at 55°C for 30 s, and extension at 72°C for 1 min), and final extension at 72°C for 7 min. PCR was performed in a 20-μl volume using TaKaRa Ex Taq polymerase (Takara Bio Inc., Shiga, Japan). The PCR products were purified with ExoSAP-IT™ reagent (Thermo Fisher Scientific K. K., Tokyo, Japan) following the manufacturer’s instructions. The nucleotide sequences of the PCR products were sequenced by Fasmac Co., Ltd. (Kanagawa, Japan) using the Sanger method. The GenBank accession numbers of the *SUC1* alleles are listed in Table 3. All sequencing chromatograms obtained by the Sanger method were visually checked for quality and heterozygous sites using CLC Genomics Workbench v10.0.1 software (Filgen, Nagoya, Japan).

**PCR amplification and DNA sequencing of the ITS and cp regions**

The ITS region was amplified using AB101 and AB102 primers (Douzery et al. 1999). The PCR conditions were as follows: incubation at 94°C for 1 min, 25 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, extension at 72°C for 1 min, followed by final extension for 7 min. Three cp regions (*trnL-trnF*, *ndhF-rpl32*, and *rpl32-trnL*) were chosen based on a previous study showing that they are useful for phylogenetic studies of the subgenus *Plantago* (Dunbar-Co et al. 2008; Iwanycki Ahlstrandet al. 2019). The cp regions were amplified using universal primers (Taberlet et al. 1991; Shaw et al. 2007). PCR and sequencing procedures were identical to those used for the *SUC1* homoeolog-specific PCR.

**Phylogenetic analysis of the *SUC1* region**

We assembled forward and reverse reads using CLC Genomics Workbench v10.0.1 software (Filgen, Nagoya, Japan). All sequence alignments were performed using MAFFT v7 software (Katoh et al. 2019) and manually edited using CLC Genomics Workbench v10.0.1 software. The phylogenetic relationships of the *SUC1* region were inferred using the maximum parsimony (MP), neighbor-joining (NJ; Saitou and Nei 1987), and maximum likelihood (ML; Felsenstein 1981) procedures using PAUP\* 4.0a software (Swofford 2003). In the MP analysis, parsimony informative indels were coded as binary (present or absent) characters, “gapmode” was set as missing, and all characters were weighted equally. The analysis was performed via a heuristic search using the tree bisection–reconnection branch-swapping option. One hundred rounds of random additions were performed to identify multiple islands of equally most parsimonious trees (Maddison 1991). The search setting used to find MP trees was applied to 1000 bootstrap replications (Felsenstein 1985). The distance option in our NJ analysis was set to ML; the substitution rate classes and gamma shape parameter were estimated following the PAUP\* procedures manual (model correspondence = GTR+G) (Swofford and Sullivan 2009, Swofford and Bell 2017). This search setting was applied to 1000 bootstrap replications. In our ML tree search, the nucleotide evolution model was selected using PAUP\* 4.0a software (model correspondence = GTR+I+G) (Swofford and Sullivan 2009), and the analysis was performed via a heuristic search using the tree bisection–reconnection branch-swapping option (addseq = random, nreps = 10). This search setting was applied to 100 bootstrap replications.

**Median-joining (MJ) network analysis of the ITS and cp regions**

The MJ network was constructed using PopART v1.7 software (Bandelt et al. 1999; Leigh and Bryant 2015) to determine the relationships of ITS genotypes and cp haplotypes with ε = 0. All sites with only ambiguous base states (e.g., Y, R, and K) were excluded, and indels were coded as binary (present or absent) characters. A mononucleotide repeat region detected in *rpl32-trnL* (from 1844 bp to 1865 bp in a concatenated alignment) was removed before data analysis because of its significant homoplasy. There was a sequencing gap between 1947 bp and 1948 bp in the *rpl32-trnL* region of the concatenated alignment because sequencing of the middle sector of the *rpl32-trnL* region was not feasible for either *P*. *hakusanensis* or *P*. *asiatica* var. *densiuscula* due to the poor quality of the raw data. The GenBank accession numbers of the ITS genotypes and cp haplotypes are listed in Table 4 and 5, respectively.

**Preparation of the MIG-seq library, high-throughput sequencing, and phylogenetic inference**

In brief, we used a two-step amplification procedure based on the protocol of Suyama and Matsuki (2015), but we changed the annealing temperature from 48°C to 38°C in the first PCR of this protocol. The amplicons were purified and sequenced on the Illumina MiSeq sequencer (Illumina, San Diego, CA, USA). Primer regions, anchors, and low-quality reads were removed using the FASTX Toolkit package (http://hannonlab.cshl.edu/fastx\_toolkit/). To remove reads derived from extremely short library entries, we searched for the sequences of the primer-targeted regions within the sequences of reads 1 and 2, and reads containing the searched sequences were removed using TagDust software (Lassmann et al. 2009).

To obtain genotypes at SNPs, we used the Universal Network Enabled Analysis Kit (UNEAK) pipeline (Lu et al. 2013) to assemble 80-bp clean reads. UNEAK is a non-reference, network-based pipeline that has been successfully used for genotype calling in polypoid species (e.g. Clark et al. 2015; Li et al. 2014). A default setting was used in the read assembly by UNEAK. SNPs were exported in HapMap format and then filtered using TASSEL 5.0 software (Bradbury et al. 2007) with the following parameters: siteMinCount 10, MinAlleleFreq 0.05, and maxHetero 0.5. The loci genotyped in all samples were retained in the final dataset.

We used RAxML v.8.2.10 software (Stamatakis 2014) to infer a ML phylogenetic tree. In this analysis, we specified the GTRGAMMA model as the nucleotide evolution model and performed 1000 bootstrap iterations to assess the node support values. The GenBank accession numbers of the MIG-seq raw data are listed in Table 6.

**Results**

**Chromosome numbers**

We counted chromosome numbers of 2n = 24 in three individualsof *P*. *hakusanensis* collected on Mt. Chokai (Fig. 2). This number was reported previously by Yamazaki (1993) and Ohashi(2017), but we were unable to locate original data from either study showing the collection site or photographs of chromosomes. We confirmed that the chromosome number of *P*. *hakusanensis* is 2n = 24, and that the species is a tetraploid with a basic chromosome number x = 6 (Rahn 1996).

**Phylogenetic analysis based on *SUC1* showed that *P. hakusanensis* is a close relative of *P***. ***asiatica* var. *densiuscula* within sect. *Plantago***

The PCR-amplified *SUC1* region of *P*. *hakusanensis* was *ca*. 1.25 kb long. Two distinct *SUC1* alleles were obtained from each of the two individuals from Mt. Hakusan. After removing the entire intron 1 from the ambiguous alignment, the aligned matrix of all 63 unique alleles from 25 taxa from subg. *Plantago* (17 polyploids, including *P*. *tenuiflora* used as an outgroup) was 835 bp long, with 180 variable and 111 parsimony-informative sites. In the MP analysis, we obtained 59 most parsimonious trees with 311 steps. The overall consistency index (Kluge and Farris 1969) was 0.723, the overall retention index (Farris 1989) was 0.855, and the overall rescaled index (Farris 1989) was 0.618. The topologies produced from the MP, NJ, and ML procedures were mostly similar. The MP tree is shown in Fig. 3. The NJ and ML trees are shown in Online Resources 1 and 2, respectively.

We compared our results with those of Ishikawa et al. (2009) and found that the addition of three *P*. *hakusanensis* alleles and a change in outgroup identity had little effect on the major topologies of the trees. Thus, we present below a brief summary of our results with a particular focus on *P*. *hakusanensis*. The alleles of the ingroups from four sections of the subg. *Plantago* fell into two sister clades: clade 1 (MP/NJ/ML support values = 63/65/60) and clade 2 (support values = 93/94/87). Three subclades (B–D) were recognized in clade 1. Three alleles (E, G, and M) and 11 subclades (F, H–L, and N–R) were found in clade 2 (Fig. 3). Two alleles obtained from a *P*. *hakusanensis* individual were represented in the two distantly separated subclades I (support values = 93/95/93) and L (support values = 82/82/87) (Fig. 3). The alleles found in the distinct subclades were considered homoeologs originating from allotetraploidization that occurred through interspecific hybridization between two diploids followed by full duplication of the hybrid genome (Ramsey and Schemske 1998; Glover et al. 2016). A set of two homoeologs in subclades I and L was also found in *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* var. *densiuscula* f. *yakusimensis*; hence, *P*. *hakusanensis* is an allotetraploid originating from ancestral lineages shared with *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* var. *densiuscula* f. *yakusimensis*.

**Monophyly of *P. hakusanensis* inferred by analyses based on *SUC1*, the ITS region, and MIG-seq**

Two homoeologs of the *SUC1* were successfully amplified by homoeolog-specific PCR amplification in all 10 individuals of *P*. *hakusanensis* that we investigated. The homoelogs were successfully sequenced with one exception: one homoeolog in subclade L in an individual collected from Mt. Hakusan could not be sequenced, likely due to the presence of indels between two biparentally inherited alleles. The topologies of the trees inferred from each homoeolog of the I and L lineages were consistent (data not shown). We therefore concatenated the two homoeologs in the phylogenetic analysis. The aligned matrix of all 14 sequences comprised 2589 characters with 40 variable and 20 parsimony-informative sites. In our MP analysis, we obtained nine most parsimonious trees after 63 steps. The overall consistency index was 0.952, the overall retention index was 0.935, and the overall rescaled index was 0.89. The topologies produced by MP, NJ, and ML were nearly identical, although there were a few slight differences in resolution at the branch tips. Thus we present only the MP tree in Fig. 4. *P*. *hakusanensis* formed a monophyletic clade with high to moderate bootstrap support values (MP/NJ/BI support values = 93/56/78). This clade was sister to a clade comprising *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* f. *yakusimensis* (clade A). The *P*. *hakusanensis* clade was divided into two sister clades: sub-clade H-1 (Mt. Chokai, Mt. Gassan, and Mt. Asahi) and sub-clade H-2 (Mt. Hakusan).

Among 78 individuals examined, substitutions and indels in the ITS regions were recognized at 44 and 4 sites in 742 characters, respectively. There were 25 parsimony-informative sites. The variable positions indicated a total of 12 genotypes (Table 7). The MJ network showed that those genotypes were divided into three taxon-specific groups (I, II, and *P*. *camtschatica*); the remaining members of the group belonged to *P*. *asiatica* from China, *P*. *major*, and *P*. *major* var. *japonica* (Fig. 5, Table 1). Group I contained six genotypes of *P*. *asiatica* var. *densiuscula*, and group II contained two genotypes of *P*. *hakusanensis*; and these two groups were distinctly separated by six mutation steps. Within the MJ network, group I had a star-like configuration in which one major genotype was surrounded by several low-frequency genotypes distinguished by one or two mutation steps. This structure is indicative of a rapid population expansion.

We postulated that the recent hybridization between *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula* might be detectable in mixed ITS sequences of the two taxa via biparental inheritance of nuclear-encoded alleles, but sequencing chromatograms indicated no sign of hybridization in the 17 individuals of *P*. *hakusanensis* and 7 individuals of *P*. *asiatica* var. *densiuscula* from Mt. Chokai.

In MIG-seq analysis, the average number of reads obtained from 10 individuals were 2,561,487. The genotype matrix consisted of 420 SNPs and the genotyping rate at these markers was 1.0. Phylogenetic analyses revealed monophyly of *P*. *hakusanensis* (Fig. 6, bootstrap value = 100). Within the *P*. *hakusanensis* clade was divided into two sister clades (H1 and H2, bootstrap value = 100). Clade H1 comprised individuals from Mt. Chokai and Mt. Gassan. Clade H2 comprised individuals from Mt. Hakusan, similar to the *SUC1* tree (Figs. 4 and 6). Branch lengths from the tips to the nodes at the last common ancestor for each taxon were shorter in *P*. *asiatica* var. *densiuscula*, even though it has a wide distribution across the Japanese archipelago. This suggests that *P*. *asiatica* var. *densiuscula* expanded more rapidly than *P*. *hakusanensis*.

**Cp haplotypes shared among *P. hakusanensis*, *P***. ***asiatica* var. *densiuscula*, and *P***. ***asiatica* from China**

Among 64 individuals examined, substitutions and indels were recognized at 46 and 14 sites in 2175 characters, respectively. There were 36 parsimony-informative sites. The variable positions indicated a total of 13 haplotypes (Table 8). Eleven haplotypes were taxon-specific, but two haplotypes were shared by different taxa. The first shared haplotype, H1, was shared by *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula*. The second haplotype, H11, was shared by *P*. *hakusanensis* and *P*. *asiatica* in China (Table 1). MJ network analysis showed that these haplotypes were divided into four groups (E, W, C, and *P*. *camtschatica*, Fig. 7a). *P*. *asiatica* var. *densiuscula* contained groups E (haplotypes H1, H3–H5) and W (haplotypes H6–H9). The distributions of the two groups in *P*. *asiatica* var. *densiuscula* did not overlap, except in the case of one individual on Mt. Chokai (Fig. 7). Individuals with group E haplotypes were found in Taiwan and in the eastern sector of the Japanese archipelago. Group W was distributed mostly within the western sector (Table 1). Haplotypes of *P*. *hakusanensis* consisted of groups E (H1 and H2), W (H10), and C (H11). Haplotypes H2 and H10 were specific to *P*. *hakusanensis*, although they were separated from haplotypes H1 and H8 of *P*. *asiatica* var. *densiuscula*, respectively, by only one substitution (H1 was shared by both *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula*, as indicated above).

**Discussion**

***P. hakusanensis* is an allotetraploid related to *P***. ***asiatica* var. *densiuscula***

Our phylogenetic analyses based on the nuclear-encoded single-copy *SUC1* region showed that *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula* (and *P*. *asiatica* var. *densiuscula* f. *yakusimensis*) are both allotetraploids with two distinctly related homoeologs. Each homoeolog was found to belong to the same subclade with high support values (subclade I and L in Fig. 3). Thus, *P*. *hakusanensis* and *P*. *asiatica* var. *densiuscula* are closely related to one another, and *P*. *hakusanensis* is presumed to have originated either (i) from the same ancestral allotetraploid as *P*. *asiatica* var. *densiuscula*, with subsequent differentiation into distinct taxa in different biomes, or (ii) from independent allopolyploidization via hybridization between the same or closely related parental species of *P*. *asiatica* var. *densiuscula*.

A diploid *P*. *major* (or *P*. *major* var. *japonica*) in subclade I and a diploid species related to *P*. *depressa* and *P*. *camtschatica* in sect. *Mesembrynia* (subclade K in Fig. 3) have been proposed as the parental species in the allopolyploidization of *P*. *asiatica* var. *densiuscula*, based on their phylogenetic positions and current distributions (Ishikawa et al. 2009). The assumptions made for *P. asiatica* var. *densiuscula* would also be applicable to parental species of *P*. *hakusanensis*. In contrast to the *SUC1*, only one lineage of ITS sequences was obtained from both *P. asiatica* var. *densiuscula* and *P*. *hakusanensis*, and those sequences were closely related to *P. major* and *P. major* var. *japonica* (Fig. 5). The result might be explained by (i) PCR failure to amplify ITS sequences related to *P. camtschatica*, or (ii) biparentally inherited homoeologous ITS regions might have been homogenized by both inter-locus and intra-locus concerted evolution biased toward one of two parental lineage (*P. major* or *P. major* var. *japonica*) as reported in other allopolyploids (Wendel et al., 1995; Kovarik et al., 2005). The former possibility might be less likely, because ITS of *P. camtschatica* were successfully amplified in present study, and the set of primers used in this analysis have been shown to work for a wide taxonomic range (Douzery et al. 1999; Sonboli et al. 2011; Kokubugata et al. 2011; Koecke et al. 2013).

**Disparities between phylogenies based on nuclear- and cp-encoded genes in *P. hakusanensis* and *P***. ***asiatica* var. *densiuscula***

Phylogenetic analyses of the two nuclear-encoded regions (*SUC1* and *ITS*) and MIG-seq data revealed the monophyly of *P*. *hakusanensis*, at least in the samples examined in our study (Figs. 3–6). In contrast, cp haplotypes of *P*. *hakusanensis* (haplotypes H1, H2, H10, and H11) were shared or phylogenetically close to *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* in China. In general, shared genetic diversity between closely related taxa is explained by incomplete lineage sorting of ancestral polymorphisms and/or introgression (Rieseberg and Soltis 1991; Comes and Abbott 2001; Dixon et al. 2007). Although it is difficult to completely rule out the possibility of incomplete lineage sorting, the shared and related haplotypes in (i) *P*. *hakusanensis* and (ii) *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* may be best explained by introgressions of the cp genome (cp capture) for the three reasons outlined below. First, incomplete lineage sorting tends to occur shortly after speciation, but we considered the two taxa to be distinctly differentiated from one another based on the results of nuclear marker analyses (Figs. 4–6). Second, ancestral polymorphisms in incomplete lineage sorting are more readily fixed in the cp genome, which has a smaller effective population size than that of the nuclear genome (Schaal et al. 1998), but we did find shared genetic diversity between taxa in the cp markers examined in the present study. Third, secondary contacts between the two taxa were presumably possible under certain conditions, even though the two taxa are currently distributed across different elevation ranges. For example, the establishment of a sympatric distribution may have occurred via (i) invasion of *P*. *asiatica* var. *densiuscula* into the habitat of *P*. *hakusanensis* in subalpine regions, as found on Mt. Chokai (in the present study), Mt. Hakusan, and Mt. Gassan (Nogami 2001; Nakayama et al. 2006, 2008; Yokoyama 2015), or (ii) a possible range expansion of *P*. *hakusanensis* from the subalpine zone to lower elevations during the Quaternary glacial period. Simulation studies have predicted that genomes experiencing reduced gene flow should have elevated rates of introgression (Currat et al. 2008; Excoffier et al.2009). This prediction explains why organelle genomes have higher levels of introgression than does the nuclear genome.

The simulation studies also predicted neutral gene introgression from locally established species to invading taxa (Currat et al. 2008; Excoffier et al.2009). We found the cp haplotypes of *P*. *asiatica* var. *densiuscula* groups W and E in the western and eastern sectors of the Japanese archipelago, respectively (Fig. 7). A similar geographic structure also occurred in *P*. *hakusanensis*. Although the direction of cp genome introgression between *P*. *asiatica* var. *densiuscula* and *P*. *hakusanensis* is unclear, we consider two possible hypotheses below. First, the geographic structure was originally established in *P*. *asiatica* var. *densiuscula*, and two independent cp genome introgressions from *P*. *asiatica* var. *densiuscula* to *P*. *hakusanensis* occurred on Mt. Hakusan (group W) and in the Tohoku area (group E) during the glacial period, when *P*. *hakusanensis* may have been distributed at lower elevations. Partial support for this postulate may be found in the results of an earlier study showing similar western and eastern groups of a nuclear-encoded marker (1.4-kb region at the 5′ upstream region of *SUC1*) in *P*. *asiatica* var. *densiuscula* populations found in the Japanese archipelago (Ishikawa et al. 2006). Nevertheless, this geographic structuring was not supported by the MIG-seq analysis, but the number of *P*. *asiatica* var. *densiuscula* samples may not have been adequate for elucidating phylogeographic structures. Alternatively, the second hypothesis proposes that haplotypes in groups W and E originated from *P*. *asiatica* var. *densiuscula* and *P*. *hakusanensis*, respectively. This proposal is based on two previous studies in the genera *Veratrum* and *Cercidiphyllum* (Kikuchi et al.2010; Qi et al. 2012). Both studies reported ancient introgressions of the cp genome from species in cool-temperate forest/subalpine (subalpine) habitats to species in warm-temperate/low- to mid-elevation (low elevation) locations in the northern sector of central Honshu. In both *Veratrum* and *Cercidiphyllum*, the species occurring at low elevation was distributed from China (or the Eurasian continent) to Japan, and the subalpine species was endemic to mountains in northern central Honshu. Interestingly, *P*. *asiatica* var. *densiuscula* and *P*. *hakusanensis* have distribution ranges that are similar to those of *Veratrum* and *Cercidiphyllum*, and hence, a common phylogeographic scenario may be shared among these taxa. Under this second hypothesis, *P*. *hakusanensis* may have differentiated in the cool climate of central Honshu. Subsequently, the direction of cp genome (Group E) introgression was from *P*. *hakusanensis* to *P*. *asiatica* var. *densiuscula*, which rapidly expanded its distribution range during the Quaternary climate changes from the southwestern sector of the Japanese archipelago toward the northeastern sector. Under this scenario, another cp genome (group W) introgression from *P*. *asiatica* var. *densiuscula* to *P*. *hakusanensis* on Mt. Hakusan would be required to explain the current haplotype distribution (Fig. 7). Although a rapid population expansion of *P*. *asiatica* var. *densiuscula* in the Japanese archipelago (as suggested by both the ITS and MIG-seq analyses) would be consistent with the second hypothesis, more comprehensive studies, including confirmation of recent hybridization and introgression on Mt. Hakusan, would be required to provide proof in support of either hypothesis.

**Taxonomy and the relationship between *P. hakusanensis* and *P***. ***asiatica* in Asia**

*P*. *asiatica* is widely distributed in southern and eastern Asia, and Rahn (1996) listed 16 species, including *P*. *hakusanensis*, that may be closely related to *P*. *asiatica* based on morphology. However, we did not expect the cp haplotype H of *P*. *hakusanensis* to occur in *P*. *asiatica* on the Chinese mainland (Table 1, Fig. 7). The *P*. *asiatica* sample from China shared the same set of three *SUC1* homoeologs with the hexaploid *P*. *formosana* (Fig. 3 and our unpublished data). *P*. *formosana* was originally described in Taiwan (Masamune 1932; Hatusima 1971) and was considered an allohexaploid that originated from hybridization between the tetraploid *P*. *asiatica* var. *densiuscula* and the diploid *P*. *major* (or *P*. *major* var. *japonica*) (Ishikawa et al. 2009). *P*. *formosana* is often considered to be conspecific with *P*. *asiatica* or *P*. *major* (Hatusima 1971; Shimabuku 1997). Hexaploid *P*. *asiatica* and *P*. *major* var. *japonica* have been reported in Niigata and Toyama Prefectures on the coast of Japan (Iwatsubo et al. 2000; Ogino 2001). Thus, we can postulate that a putative hexaploid with haplotype H may have been distributed at low elevations on the coast, and that cp genome introgressions from this hexaploid to *P*. *hakusanensis* occurred. A phylogenetic analysis to elucidate the relationships between *P*. *hakusanensis* and other Asian species related to *P*. *asiatica* could not be readily performed because of taxonomic confusion, putative hybridizations among taxa, and the occurrence of polyploidy. However, investigations using both organelle and nuclear markers with adequate resolution and a confirmation of the ploidy levels might resolve the complicated evolutionary history of these taxa.

**Acknowledgments**

The authors are grateful to Mr. K. Sawa and Dr. Y. Nakayama for providing the plant materials. This research was approved by the Ministry of Environment, Japan (for investigations in Ishikawa, Yamagata, Toyama, and Kagoshima Prefectures). Funding was provided by the Japan Society for the Promotion of Science Grant-in-Aid for Scientiﬁc Research (JSPS KAKENHI 17K07527, 19K06805, 24K09566) and the Environmental Restoration and Conservation Agency of Japan for the Environment Research and Technology Development Fund (4-2001). The computation was performed using Research Center for Computational Science, Okazaki, Japan (Project: NIBB, 24-IMS-C357).

**References**

Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48. doi: 10.1093/oxfordjournals.molbev.a026036

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi: 10.1093/bioinformatics/btm308

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. Molecular Biology and Evolution 14: 592–593. doi: 10.1093/oxfordjournals.molbev.a025797

Clark LV, Stewart JR, Nishiwaki A, Toma Y, Kjeldsen JB, Jørgensen U, Zhao H, Peng J, Yoo JH, Heo K, Yu CY, Yamada T, Sacks EJ (2015) Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. J Exp Bot 66:4213–4225. doi: 10.1093/jxb/eru511

Comes HP, Abbott RJ (2001) Molecular phylogeography, reticulation, and lineage sorting in Mediterranean *Senecio* sect. *Senecio* (Asteraceae). Evolution 55:1943–1962. doi: 10.1111/j.0014-3820.2001.tb01312.x

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: massive introgression by local genes. Evolution 62:1908–1920. doi: 10.1111/j.1558-5646.2008.00413.x

Dixon CJ, Schönswetter P, Schneeweiss GM (2007) Traces of ancient range shifts in a mountain plant group (*Androsace halleri* complex, Primulaceae). Mol Ecol 16:3890–901. doi: 10.1111/j.1365-294X.2007.03342.x

Douzery EJP, Pridgen AM, Kores P, Linde HP, Kurzwell H, Chase MW (1999) Molecular phylogenetics of *Diseae* (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. Am J Bot 86:887–889. doi: 10.2307/2656709

Dunbar-Co S, Wieczorek AM, Morden CW (2008) Molecular phylogeny and adaptive radiation of the endemic Hawaiian *Plantago* species (Plantaginaceae). Am J Bot 95:1177–88. doi: 10.3732/ajb.0800132

Excoffier L, Foll M, Petit RJ (2009) Genetic consequences of range expansions. Annu Rev Ecol Evol Syst 40:481–501. doi: 10.1146/annurev.ecolsys.39.110707.173414

Farris JS (1989) The retention index and the rescaled consistency index. Cladistics 5:417–419. doi: 10.1111/j.1096-0031.1989.tb00573.x

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–76. doi: 10.1007/BF01734359

Felsenstein J (1985) Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783–791. doi: 10.1111/j.1558-5646.1985.tb00420.x

Glover NM, Redestig H, Dessimoz C (2016) Homoeologs: What Are They and How Do We Infer Them? Trends Plant Sci 21:609–621. doi: 10.1016/j.tplants.2016.02.005.

Hatusima S (1971) Flora of the Ryukyus (Including Amami Islands, Okinawa Islands, and Sakishima Archipelago). Okinawa Biological Education and Research Society, Naha, Okinawa, Japan [in Japanese]

Ishikawa N, Yokoyama J, Ikeda H, Takabe E, Tsukaya H (2006) Evaluation of morphological and molecular variation in *Plantago asiatica* var. *densiuscula*, with special reference to the systematic treatment of *Plantago asiatica* var. *yakusimensis*. J Plant Res 119:385–395. doi: 10.1007/s10265-006-0286-y

Ishikawa N, Yokoyama J, Tsukaya H (2009) Molecular evidence of reticulate evolution in the subgenus *Plantago* (Plantaginaceae). Am. J. Bot. 96:1627–1635. doi: 10.3732/ajb.0800400

Iwanycki Ahlstrand N, Verstraete B, Hassemer G, Dunbar‐Co S, Hoggard R, Meudt HM, Rønsted N (2019) Ancestral range reconstruction of remote oceanic island species of *Plantago* (Plantaginaceae) reveals differing scales and modes of dispersal. J Biogeogr 46:706–722. doi: 10.1111/jbi.13525

Iwatsubo Y, Ogino K, Kodate G, Nakamura T (2000) Chromosome numbers of *Plantago asiatica* L. (Plantaginaceae) in Toyama Prefecture, central Japan. The journal of phytogeography and taxonomy 48:67 – 70.

Judo MS, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. Nucleic Acids Research 26: 1819–1825. doi: 10.1093/nar/26.7.1819

Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). Journal of Bioscience and Bioengineering 96: 317–323. doi: 10.1016/S1389-1723(03)90130-7

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Bioinformatics 20:1160–1166. doi: 10.1093/bib/bbx108

Kikuchi R, Jae-Hong P, Takahashi H, Maki M (2010) Disjunct distribution of chloroplast DNA haplotypes in the understory perennial *Veratrum album* ssp. *oxysepalum* (Melanthiaceae) in Japan as a result of ancient introgression. New Phytol 188:879–91. doi: 10.1111/j.1469-8137.2010.03398.x

Kluge AG, Farris JS (1969) Quantitative phyletics and the evolution of anurans. Syst Zool 18: 1–32.

Koecke AV, Muellner-Riehl AN, Pennington TD, Schorr G Schnitzler J (2013) Niche evolution through time and across continents: The story of Neotropical *Cedrela* (Meliaceae). Am. J. Bot. 100: 1800–1810. doi: 10.3732/ajb.1300059.

Koizumi G. 1930. *Plantago hakusanensis* Koidz. Fl Symb Orient-Asiat 19.

Kokubugata G, Hirayama Y, Peng CI, Yokota M, Möller M (2011) Phytogeographic aspects of *Lysionotus pauciflorus* sensu lato (Gesneriaceae) in the China, Japan and Taiwan regions: phylogenetic and morphological relationships and taxonomic consequences, Plant Syst Evol 292: 177-188. doi: 10.1007/s00606-010-0410-2

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Mol Biol Evol 35:1547–1549. doi: 10.1093/molbev/msy096

Lassmann T, Hayashizaki Y, Daub CO (2009) TagDust--a program to eliminate artifacts from next generation sequencing data. Bioinformatics 25:2839–40. doi: 10.1093/bioinformatics/btp527

Leigh JW, Bryant D (2015) PopART: Full-feature software for haplotype network construction. Methods Ecol Evol 6:1110–1116. doi: 10.1111/2041-210X.12410

Li X, Wei Y, Acharya A, Jiang Q, Kang J, Brummer EC (2014) A saturated genetic linkage map of autotetraploid alfalfa (*Medicago sativa* L.) developed using genotyping-by-sequencing is highly syntenous with the *Medicago truncatula* genome. G3 (Bethesda). 4:1971–1979. doi: 10.1534/g3.114.012245

Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE (2013) Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet 9:e1003215 doi: 10.1371/journal.pgen.1003215

Maddison WP (1991) The discovery and importance of multiple islands of most-parsimonious trees. Syst Zool 40: 315–328. doi: 10.1093/sysbio/40.3.315

Masamune G (1932) Symbolae florae australi-japonicae. Journal of the Society of Tropical Agriculture 4: 191–197.

Ministry of Agriculture, Forestry and Fisheries of Japan, Ministry of Land, Infrastructure, Transport and Tourism, and Ministry of the Environment, Government of Japan. 2015. Ecosystem Conservation and Recovery Plan in Hakusan National Park. Ministry of the Environment, Government of Japan. <https://warp.da.ndl.go.jp/info:ndljp/pid/11455340/www.env.go.jp/park/system/files/kanri_10_6.pdf> (in Japanese). Accessed April 2020

Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Research 8:4321–4325. doi: 10.1093/nar/8.19.4321

Nakayama Y, Nogami T, Yagyu A (2006) Distribution of lowland plants in alpine and subalpine zone of Mt. Hakusan (5) Weed invasion into Minamiryugabanba and Murodo. Ishikawa Pref Nat Conserv Center Study Rep. 33:15–23. (in Japanese)

Nakayama Y, Nogami T, Yagyu A (2008) Distribution of lowland plants in alpine and subalpine zone of Mt. Hakusan (6) Distribution of natural hybrids between *Plantago asiatica* and *P. hakusanensis* and alien dandelions. Ishikawa Pref Nat Conserv Center Study Rep. 35:17–22. (in Japanese)

Nogami T (2001) Distribution of lowland plants in alpine and subalpine zone of mt. Hakusan. Ishikawa Pref Nat Conserv Center Study Rep 28:1–6. (in Japanese)

Ogino K (2001) http://raicho.sci.u-toyama.ac.jp/~iwatsubo/home%20page/laboratory/student%20list/student%20abst/ogino-m2.html Accessed April 2020

Ohashi H (2017) Plantaginaceae. *In*: Ohashi H, Kadota Y, Kihara H, Murata J, Yonekura K (eds.), Wild Flowers of Japan 5:72–90. Heibonsha, Tokyo

Pilger R (1922) Die Arten der Plantago major-Gruppe in Ostasien. Notizbl bot Gart Mus Berlin-Dahlem 8:104–116.

Qi XS, Chen C, Comes HP, Sakaguchi S, Liu YH, Tanaka N, Sakio H, Qiu YX (2012) Molecular data and ecological niche modelling reveal a highly dynamic evolutionary history of the East Asian Tertiary relict *Cercidiphyllum* (Cercidiphyllaceae). New Phytol 196:617–630. doi: 10.1111/j.1469-8137.2012.04242.x

Rahn K (1996) A phylogenetic study of the Plantaginaceae. Bot J Linn Soc 120:145–198. doi: 10.1111/j.1095-8339.1996.tb00484.x

Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst 29:467–501. doi: 10.1146/annurev.ecolsys.29.1.467

Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5:65–84.

Rønsted N, Chase MW, Albach DC, Bello MA (2002) Phylogenetic relationships within *Plantago* (Plantaginaceae): evidence from nuclear ribosomal *ITS* and plastid *trnL-F* sequence data. Bot J Linn Soc 139:323–338. doi: 10.1046/j.1095-8339.2002.00070.x

Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. doi: 10.1093/oxfordjournals.molbev.a040454

Sano S, Nakayama Y, Ohigashi K, Nogami T, Yagyu A (2016) Flowering behaviors of the inflorescences of an alien plant (*Plantago asiatica*), an alpine plant (*Plantago hakusanensis*), and their hybrids on Mt. Hakusan, Japan, Weed Biol Manag 16:108–118. doi: 10.1111/wbm.12098

Sano S, Nakayama Y, Nogami T, Yagyu A (2019) Flowering phenology and seed production of an alpine plant (*Plantago hakusanensis*), a domestic alien plant (*Plantago asiatica*), and their hybrids on Mt. Hakusan, Japan. J Weed Sci Tech 64:73–84. doi: 10.3719/weed.64.73

Schaal, BA, Hayworth DA, Olsen KM, Rauscher JT, Smith WA (1998) Phylogenetic studies in plants: Problems and prospects. Mol Ecol 7:465–474.

Search System of Japanese Red Data <http://jpnrdb.com/>. accessed on April 2020

Shaw J, Lickey EB, Schilling EE, Small RL (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: The tortoise and the hare III. Am J Bot 94:275–288. doi: 10.3732/ajb.94.3.275

Shimabuku K (1997) Check list of the vascular flora of the Ryukyu Islands. Kyushu University Press, Fukuoka, Fukuoka, Japan [in Japanese]

Sonboli A, Kazempour Osaloo S, Vallès J, Oberprieler C (2011). Systematic status and phylogenetic relationships of the enigmatic *Tanacetum paradoxum* Bornm. (Asteraceae, Anthemideae): evidences from nr DNA ITS, micromorphological, and cytological data. Plant Syst Evol 292: 85–93. doi: 10.1007/s00606-010-0415-x

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. doi: 10.1093/bioinformatics/btu033

Suyama Y, 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. doi: 10.1038/srep16963

Swofford DL (2003) PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford DL, Bell CD (2017) http://phylosolutions.com/paup-documentation/paupmanual.pdf. Accessed 30 April 2020

Swofford DL, Sullivan J (2009) Phylogenetic inference based on parsimony and other methods using PAUP\*. In: Lemey P, Salemi M, Vandamme AM (eds) The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge University Press, Cambridge, pp 267-288

Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol Biol 17:1105–1109. doi: 10.1007/BF00037152

Yamada K, Satomi N (1975) The plant community at the distributional boundaries in Japan: The *Plantago hakusanensis* community at Mt. Hakusan and Mt. Tateyama. Ishikawa Pref Nat Conserv Center Study Rep 2:47–53. (in Japanese)

Yamazaki T (1992) On *Plantago hakusanensis* Koidz. Journal of Japanese Botany 67:116–117. (in Japanese with English summary)

Yamazaki T (1993) Plantaginaceae. *In*: Iwatsuki K, Yamazaki T, Boufford DE, Ohba H (eds.), Flora of Japan, Vol. IIIa: Angiospermae, Dicotyledoneae, Sympet al.ae (a), pp. 384–386. Kodansha, Tokyo

Yokoyama J. (2015) Heisei26 Gairaiseibustu-chosa-hokokusho [A report of research on alien species in 2014]. Yamagata prefecture (in Japanese). [https://www.pref.yamagata.jp/ou/kankyoenergy/050011/sizenkankyo/gairaiseibutu/H26gairaityousahoukokusyo.pdf　Accessed April 2020](https://www.pref.yamagata.jp/ou/kankyoenergy/050011/sizenkankyo/gairaiseibutu/H26gairaityousahoukokusyo.pdf　Accessed%20April%202020)

**Figure legends**

Fig. 1.

Schematic of thesucrose–proton symporter 1 (*SUC1*) gene structure and positions of the primers. Open rectangles represent exons.

Fig. 2.

Photograph of the somatic chromosomes in a *Plantago hakusanensis* collected on Mt. Chokai.

Fig. 3.

Phylogenetic position of *Plantago hakusanensis* among representative species of the subgenus *Plantago*. One of 59 most parsimonious trees based on the sucrose–proton symporter 1 (*SUC1*) sequences is presented. Bootstrap values for maximum parsimony, neighbor-joining, and maximum likelihood analyses are located adjacent to the branches (values <50 are shown as hyphens). The alleles of *P*. *hakusanensis* are enclosed in gray boxes. The alleles of *P*. *asiatica* var. *densiuscula* and *P*. *asiatica* var. *densiuscula* f. *yakusimensis* are indicated by black arrow heads. Each allele is associated with a taxon name, which is followed by an underscore, the putative ploidy level, an underscore, and the GenBank accession number. Diploid species are identified by red asterisks. Color codes for Rahn’s (1996) classification of *Plantago* sections are provided in the bottom left corner.

Fig. 4. Phylogenetic tree of *Plantago hakusanensis* and *P*. *asiatica* var. *densiuscula* based on concatenated homoeologs of *sucrose–proton symporter 1* (*SUC1*) sequences. Bootstrap values for the maximum parsimony, neighbor-joining, and maximum likelihood analyses are presented adjacent to the branches (values <50 are shown as hyphens). Numbers in parentheses after each sample name refer to the “Population no.” in Table 1.

Fig. 5.

**a** Median-joining (MJ) network of the rDNA internal transcribed spacer (ITS) genotypes of *Plantago hakusanensis* and related taxa. Black bars indicate substitutions or indels. The sizes of the circles within the MJ network are proportional to the sample size (n): the smallest circle represents a sample size of 1, and the largest circle for genotype A represents a sample size of 41. Unlabeled small black dots at the nodes represent inferred genotypes.

**b** Geographic distributions of ITS genotypes in *P*. *asiatica* var. *densiuscula* and *P*. *hakusanensis*. The sample sizes at the representative localities are indicated in parentheses. The smallest circle represents a sample size of 1, and the largest circle for Mt. Chokai represents a sample size of 17. The sizes of the circles within the maps are proportional to the sample sizes (n). The colors of the circles conform to the colors in panel **a**.

Fig. 6.

A maximum likelihood tree of *Plantago hakusanensi*s and *P*. *asiatica* var. *densiuscula* based on the single-nucleotide polymorphisms obtained by MIG-seq. Bootstrap values are presented adjacent to the branches. The scale bar represents the mean number of nucleotide substitutions per site. Boldface numbers in parentheses adjacent to the sample names refer to the “Population no.” in Table 1.

Fig. 7.

**a** Median-joining (MJ) network of chloroplast (cp) haplotypes in *Plantago hakusanensis*, *P*. *asiatica* var. *densiuscula*, and *P*. *camtschatica*. Each black bar indicates a substitution or an indel, depending on the site. The sizes of the circles within the network are proportional to the sample size (n): the smallest circle represents a sample size of n = 1, and the largest circle for haplotype H1 represents a sample size of n = 17. Unlabeled small black dots at the nodes represent inferred genotypes.

**b** Distribution of cp haplotypes in *P*. *asiatica* var. *densiuscula* and *P*. *hakusanensis*. Sample sizes at the representative localities are shown in parentheses. The smallest circle represents a sample size of 1; the largest circle (Mt. Chokai) represents a sample size of 16. The sizes of the circles within the maps are proportional to the sample size (n). The colors of the circles conform to those in panel **a**.

**Data Accessibility**

The DNA sequences generated in the present study have been deposited in NCBI, and the GenBank accession numbers are listed in Tables 3, 4, 5, and 6.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

The English in this document has been checked by at least two professional editors, both native speakers of English. For a certificate, please see:

http://www.textcheck.com/certificate/Tec0Gv