Evaluation of Japanese potatoes using single nucleotide polymorphisms (SNPs)

Toshiya Igarashi[1] ・ **Mutsuo Tsuyama**[1] ・ **Keiichi Ogawa**[1] ・ **Emiko Koizumi**[1] ・ **Rena Sanetomo**[2] ・ **Kazuyoshi Hosaka**[2]

T. Igarashi・M. Tsuyama・K. Ogawa・E Koizumi

Potato Research Institute, Calbee Potato Inc., Memuro, Hokkaido 082-0006, Japan

R. Sanetomo・K. Hosaka

Potato Germplasm Enhancement Laboratory, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

Running head: Evaluation of Japanese potatoes using SNPs

K. Hosaka (Corresponding author)

Potato Germplasm Enhancement Laboratory, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

E-mail: spudman@obihiro.ac.jp

Tel: +81-155-49-5494

Fax: +81-155-49-5577

Abstract

An understanding of the genetic diversity and structure of a breeding population is fundamental information for breeders to develop strategies for variety improvement. The genetic diversity of 250 diverse North American potatoes was previously characterized using an 8K single nucleotide polymorphism (SNP) array. In this study, using a 12K SNP array, 164 Japanese potatoes including 70 breeding clones for chip processing were characterized and compared with North American and European potatoes. A cluster analysis using 5972 polymorphic SNP loci grouped recently released Japanese chip processing varieties and breeding clones with North American and European chip processing varieties. The other Japanese potatoes, including those for starch processing, were not clearly differentiated to specific market types. A unique group of double-cropping varieties was identified, suggesting that strong selection pressure was applied to the cultivation system (spring-season product is used for fall-season cropping). The inter-populational distance between Japanese and foreign genotypes (mean Euclidean distance of 48.4) was significantly lower than the within-populational distance of foreign genotypes (49.3), indicating that the Japanese potatoes are not uniquely differentiated but are included within the genetic diversity of foreign genotypes. This study demonstrates the usefulness of SNP array technology for performing a cost-effective, robust and direct comparison of genetic diversity among different gene pools.

Keywords Potato, Diversity, Breeding population, Single nucleotide polymorphism (SNP)

Introduction

Genetic diversity and variability are fundamental for breeding because the larger the genetic variation in a base population, the greater the potential breeding achievements can be. Thus, it is important for breeders to understand the extent of genetic diversity and the genetic structure of populations of interest. Since the advent of molecular markers such as restriction fragment polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), and simple sequence repeat (SSR, or microsatellite) markers, various crop species have been evaluated for genetic variation. Currently, with the advent of next-generation-sequencing and single nucleotide polymorphism (SNP) array technologies, genotype data from thousands of SNP loci can be obtained with

54 a relatively low cost and low labor input (Myles et al. 2010; Hamilton et al. 2011; Song et al. 2013;
55 Uitdewilligen et al. 2013; Torkamaneh et al. 2018).

56 Potato has undergone extensive changes since its first introduction from South America to Europe
57 in the early 16th century. Due to devastating epidemics caused by late blight (*Phytophthora infestans*)
58 in the 1840s, almost all varieties of European potatoes grown at that time disappeared. As a result, a
59 Chilean potato, Rough Purple Chili, was introduced by C. Goodrich in 1851, and Early Rose, a
60 grandchild of Rough Purple Chili, contributed greatly to later potato breeding because of its earliness,
61 which significantly reduced the severity of late blight (Plaisted and Hoopes 1989). Consequently, Early
62 Rose is the basis of the genetic variability of potato worldwide, resulting in a narrow genetic base of
63 modern potato varieties (Mendoza and Haynes 1974). Since R. N. Salaman first found late blight
64 resistance in *Solanum edinense* Berth. (a natural hybrid of *S. demissum* Lindl. × *S. tuberosum* L.) in 1906,
65 11 *R* genes (*R1–R11*) have been identified from the progenies of *S. demissum* (Ross 1986). In addition,
66 exotic germplasm, such as that of Andean primitive varieties and their closely related wild species, has
67 been frequently utilized in potato breeding since the beginning of the 20th century (Plaisted and Hoopes
68 1989; Ross 1986; Vos et al. 2015).

69 As part of the USDA SolCAP project, an Infinium 8303 SNP array was developed based on SNPs
70 in six potato varieties: Atlantic, Bintje, Kennebec, Premier Russet, Shepody, and Snowden (Hamilton et
71 al. 2011; Felcher et al. 2012). The 8K SNP array revealed that within diverse North American varieties,
72 the chip processing market class was the most distinct, being clearly separated from all other market
73 classes. Further clustering identified the French fry processing and table russet group, the yellow and
74 round white table group, and the pigmented group (Hamilton et al. 2011; Hirsch et al. 2013). The same
75 8K SNP array was used by Kolech et al. (2016) to evaluate the genetic diversity of Ethiopian potato
76 cultivars, and to assess their relationships with germplasms from North America, Europe and the
77 International Potato Center (CIP). They found that most of the Ethiopian local cultivars were duplicates
78 and that most local cultivars in the northwest originated in Europe, whereas the predominant southern
79 cultivars were most closely related to germplasm developed by CIP. Vos et al. (2015) developed a 20K
80 SNP array and genotyped a total of 569 potato genotypes. They found that 96% of the genetic variants
81 present in cultivars released before 1945 remained polymorphic in modern cultivars and that new genetic
82 variation introduced in the last decades caused an increase of genetic variation in the potato gene pool.

Potato was first introduced into Japan during the Keicho period (1596−1614) (Kawakami 1948; Asama 1978) or in between 1609 and 1615 (Laufer 1938). It is believed that Dutch traders brought potatoes to Nagasaki. Through repeated famines, potatoes spread over the country as a hardy crop, particularly into cooler mountain regions in central and northern Japan. For approximately 200 years until 1854, Japan had been closed to the outer world except to Chinese and Dutch traders. Since 1873, new materials were successively introduced from European countries and the United States of America (Tsukikawa 1990). Among these new materials, Irish Cobbler and May Queen were selected and continue to be leading varieties in present day. Inter-varietal cross-breeding started in 1916, and in 1939, the first interspecific crosses were made (Takase 1977). In 1938, three decades after the first use of *S. demissum* for late blight resistance breeding in Germany, *S. demissum* was introduced to Japan. In 1940, selfed seeds labeled "*S. demissum* forma atrocyaneum" were obtained from S. M. Bukasov. This accession was used to derive the first late blight resistant varieties, Yoraku and Rishiri (released in 1958 and 1960, respectively).

Potato chips have been industrialized since 1970s. The chip processing varieties, Waseshiro and Toyoshiro (a great grandchild and a second great grandchild of Irish Cobbler, respectively), were released in 1974 and 1976, respectively. The high yield and good quality of Toyoshiro promoted the rapid growth of the potato chipping industry. Approximately 35% of the potatoes in Japan are used for starch production (Ministry of Agriculture, Forestry and Fisheries 2018). The current major variety for starch processing is Konafubuki, released in 1981. The unfortunate arrival of the golden cyst nematode (*Globodera rostochiensis*) occurred in 1972. Rapid spread of the cyst nematodes has threatened potato cultivation and forced the breeding of cyst-nematode-resistant varieties. None of the top four varieties, Irish Cobbler, May Queen, Toyoshiro, and Konafubuki, are resistant to golden cyst nematodes. Thus, the last three to four decades have been a period of struggle against cyst nematodes (Mori et al. 2007). Genetic variation of these Japanese potato varieties was previously analyzed by RAPD marker analysis, which revealed that Japanese potatoes are an interesting collection of genetic diversity including North American, European, and Andean germplasms and relic potatoes (Hosaka et al. 1994).

Toyoshiro is the most important chip processing variety in Japan. However, it cannot be stored for long periods because reducing sugar contents increase after cold storage (cold sweetening), resulting in an undesirable dark chip color. In addition, Toyoshiro does not have cyst nematode resistance.

112 Aiming to achieve a year-round supply of high-quality potatoes, Calbee Potato Inc. (hereafter,
113 abbreviated 'CP'), a subsidiary company of Calbee Inc., the largest chipping company in Japan, started
114 breeding for chip processing varieties in 1984. First, they introduced superior chip processing varieties
115 from the United States of America (US). Among many varieties such as Atlantic, Yankee Chipper,
116 Kanona, and others, Snowden and Norking Russet were successfully adopted for their longer storability
117 and commercialized. Inter-varietal cross-breeding at CP was started in 2003. Poroshiri was the first
118 variety released from the CP potato breeding program in 2017. It is a chip processing variety with
119 extreme resistance to cyst nematodes and fairly good resistance to common scab. Recently, European
120 varieties have been incorporated into the CP gene pool.
121 As mentioned above, Japanese potatoes have a long history of cultivation and breeding. In
122 contrast, the CP potato breeding program emerged recently as a unique program focused on the breeding
123 of chip processing varieties. Although we introduced many foreign varieties in the past and used them
124 as parents, it is important to know the extent of genetic diversity and the genetic relationships among
125 Japanese varieties and between Japanese and foreign varieties, which can facilitate the development of
126 strategies for future breeding. In this study, we obtained 12K potato V2 SNP array data for Japanese
127 potatoes, including parental clones of the CP potato breeding program, and compared them with 8K
128 potato SNP array data on foreign varieties, mostly US varieties, investigated by Hirsch et al. (2013). In
129 addition, cytoplasmic genome types were determined for the CP clones because they are often associated
130 with different types of male sterility (Hosaka and Sanetomo 2012).
131
132
133 **Materials and Methods**
134
135 Plant materials
136 A total of 432 genotypes, designated the population 'All', was used, which included Japanese potatoes
137 and those studied by Hirsch et al. (2013) (Table 1). One hundred sixty-four were Japanese-bred
138 genotypes, designated the population 'Japan', which included 92 named varieties (60 table potatoes, 19
139 for starch processing, 9 for chip processing, and 4 for French fry processing), two advanced breeding
140 clones (for table potatoes), and 70 advanced breeding clones bred by Calbee Potato Inc. (referred to as CP

clones). Four varieties (Kintoki-imo, Nemuromurasaki, Seinaiji-ki-imo, and Kobo-imo) were old varieties of unknown origin and were treated as Japanese landraces and table potatoes. The table potatoes also included five diploid (Inca Gold, Inca-no-hitomi, Inca-no-mezame, Inca Rouge, and Benihisashi) varieties and one triploid (Neodelicious) variety. Note that Inca Gold is different from a US-bred tetraploid variety Inca Gold. Irish Cobbler and some other foreign varieties and breeding clones have significantly contributed to the Japanese potatoes. Thus, to evaluate the Japanese potatoes, 38 important foreign varieties and breeding clones and four species, *S. acaule*, *S. tuberosum* Andigenum Group (referred to as *S. tuberosum* ssp. *andigena* by Hawkes 1990, hereafter Andigena), *S. pinnatisectum*, and *S. phureja* DM, as an outgroup (using the SNP data from Hirsch et al. 2013) were added to the population 'Japan' to form the population 'Japan Plus' (Supplementary Table 1). Some European varieties and breeding clones that Calbee Potato Inc. introduced were not available for the original identities and were therefore described by the CP identity numbers C44–C69. Hirsch et al. (2013) used 250 genotypes. However, a survey of their SNP data revealed that Russet Burbank and Burbank were similar at all available SNPs, as described by Bethke et al. (2014), and three Russet Nokotah clones with different sources were all similar. Consequently, 247 genotypes formed the population 'US'.

Determination of cytoplasm types

Hosaka and Sanetomo (2012) classified potato cytoplasm into six distinct types (M, P, A, W, T, and D types). The A-type cytoplasm is the most prevalent type in Andigena, whereas the T-type cytoplasm is the most prevalent type in the common potato (referred to as *S. tuberosum* ssp. *tuberosum* by Hawkes 1990, hereinafter Tuberosum). The D-type and W/γ-type cytoplasms were introduced from *S. demissum* and *S. stoloniferum*, respectively, into the common potato gene pool and cause functionally male sterility (Dionne 1961; Lössl et al. 2000; Sanetomo and Hosaka 2011). The procedures used to determine the cytoplasm types are described in Hosaka and Sanetomo (2012).

SNP data

Total DNA was extracted from fresh leaves by either the Doyle and Doyle (1987) or Hosaka and Hanneman (1998) methods. Five micrograms of dried DNA from each of 164 Japanese and 26 foreign

varieties were sent to GeneSeek (Neogen Corporation, NE, US) to obtain 12K potato V2 SNP array data. For the population 'US', all of the 8K SNP array data were available from the literature (Hirsch et al. 2013). All informative loci of the 8K SNP array are included among the 12K SNP loci. Using a diploid genotype model, AA, AB, and BB were converted to the ordinal variables 0, 1, and 2, respectively. Hierarchical cluster analysis was performed by the Ward method (Ward 1963) using the software JMP Pro 14.0.0 (SAS Institute Inc.). To evaluate the extent of within and between populational differences, the genotype data 0, 1, and 2 were treated as categorical variables, and the dissimilarities between genotypes were calculated as Euclidean distances using the software JMP Pro. An overall mean of pairwise Euclidean distances was obtained for both within and between populations.

Results

We obtained 12K potato V2 SNP array data for 190 genotypes, of which SNP genotypes for 94 Japanese varieties and breeding clones are provided in Supplementary Table 2. Four genotypes (Andover, Pike, NY115, and NY121) were duplicated between our 12K SNP array data and the 8K SNP array data of Hirsch et al. (2013). For all comparable SNP loci, genotypes were perfectly identical between the duplicates. Our data contained fewer missing values, so they were used for subsequent analysis. Inca-no-mezame and Inca Rouge (a sport of Inca-no-mezame) had perfectly identical genotypes at all SNP loci. Thus, Inca Rouge was discarded from further analyses, and the size of the population 'Japan' was reduced to 163 (Table 1). In each population, monomorphic SNP loci and those with more than 10% missing values were discarded. Consequently, 10810 of 12720 SNPs (85.0%) were used to calculate the percentage of heterozygous SNP loci of the population 'Japan' (Table 2). The 6373 SNP loci reported in the literature (Hirsch et al. 2013) were identified in the 12K SNP array data, of which 83.3%, 96.9%, and 93.7% were polymorphic with less than 10% missing values in the populations 'US', 'Japan Plus' and 'All', respectively (Table 2).

Evaluation of the population 'Japan Plus'

Of the 94 Japanese varieties, landraces, and advanced breeding clones, 63 (67.0%) had T-type, 21 (22.3%) had D-type, seven (7.4%) had P-type, one had A-type (1.1%), and two (2.1%) had W/γ-type cytoplasm, as mostly published previously in Hosaka and Sanetomo (2012). In contrast, D-type

199 cytoplasm was the major cytoplasm among CP clones (54.3%; Supplementary Table 1). T- and W/γ-
200 type cytoplasm was found in 37.1% and 7.1%, respectively, of the CP clones. One P-type cytoplasm
201 was found among the CP clones, whereas the M- and A-type cytoplasms were not found among the CP
202 clones.

203 The population 'Japan Plus' was analyzed using 6172 polymorphic SNP loci with a diploid
204 genotype model (AA, AB and BB). A total of 205 genotypes were separated into four distinct clusters
205 (Fig. 1). Cluster 4 was composed of the outgroup of wild species and all of the diploid varieties.
206 Cluster 1 contained the oldest variety of Early Rose and the other named varieties, whereas Clusters 2 and
207 3 were formed mostly by chip processing varieties and CP clones.

208 Within Cluster 1, the European chip processing varieties and their derived CP clones formed a
209 closely related subcluster with the old variety group. Nagasaki-bred varieties formed four distinct
210 clusters, each containing a parent and their progeny. Most subclusters within the Cluster 1 did not
211 reflect market classes (table stock, chip, French fry or starch processing varieties) but indicated parentage
212 relationships: parents and their immediate progenies were grouped together. For example, Toyoshiro
213 (chip) was first united with its parent Eniwa (starch); then, it was clustered with the child Hokkaikogane
214 (French fry) and then with the grandchild North Chip (chip). Musamaru was bred from a maternal
215 parent Tunika, which is a German variety frequently used in Japan as a source of a resistance gene (*H1*)
216 against golden cyst nematode (*Globodera rostochiensis*) and has D-type cytoplasm. Musamaru (French
217 fry) was then used as the maternal parent for Pearl Starch (starch), Koganemaru (French fry), and
218 Natsufubuki (starch) and as the grandmother for Konayutaka (starch). These five varieties formed one
219 subcluster and had the D-type cytoplasm in common. C55 has the W/γ-type cytoplasm, which causes
220 complete male sterility, known as tetrad sterility (Hosaka and Sanetomo 2012). Four of five CP clones
221 derived maternally from C55 clustered with C55 and had the W/γ-type cytoplasm in common. The
222 remaining one CP clone derived from C55 clustered with the paternal parent Kitahime in Cluster 2.

223 Clusters 2 and 3 consisted of 89 genotypes, of which 84 (94.4%) were chip processing clones.
224 Except for Lilas Chip, the recently released Japanese chipping varieties were present either in Cluster 2 or
225 Cluster 3 together with their immediate parents: ND860-2, Norking Russet, Pike, or Atlantic. For
226 example, Poroshiri was bred from Norking Russet × Pike.

227 Of the 70 CP clones, 30 (42.9%) were included in each of Cluster 2 and Cluster 3. For the 70
228 CP clones, the most frequently used parent was NY115 (for 13 clones), followed by Kitahime, a
229 grandchild of ND860-2 (for 10 clones), C44 and Monticello (for 9 clones each), Pike (for 8 clones), and
230 Andover (for 6 clones). These parents and their progeny clones were united together and formed
231 distinct subclusters. D-type cytoplasm was incorporated into the CP clones most frequently from
232 Monticello (in 9 clones as the mother and in 4 as the grandmother), C44 (in 5 clones as the mother and in
233 3 as the grandmother), and MN12828 (in 4 clones as the mother and in 2 as the grandmother).
234
235 Re-evaluation of the population 'US'
236 The population 'US' is similar to the population analyzed by Hirsch et al. (2013). They used 3763 SNP
237 markers with a tetraploid genotype model (AAAA, AAAB, AABB, ABBB and BBBB) and an
238 unweighted pair group method with arithmetic mean (UPGMA) for clustering. They defined market
239 classes within cultivated potatoes (chip processing, French fry processing, pigmented, table russet, round
240 white table, and yellow). The French fry processing and table russet group, the yellow and round white
241 table group, the pigmented group, and the three chip processing groups, derived from Lenape, ND860-2,
242 and S440, were classified in distinct clusters.
243 We used a diploid genotype model (AA, AB and BB) for 5306 polymorphic SNP loci. When
244 we used the UPGMA method, in which the distance between two groups is defined as the average
245 distance between each of their members, a very similar dendrogram to the one given by Hirsch et al.
246 (2013) was obtained. Instead, we used the Ward method for clustering, in which the distance between
247 two clusters is the ANOVA sum of squares between the two clusters summed over all of the variables.
248 That is, at each step, the two clusters that result in the least increase in the pooled within-group sum of
249 squares are fused. Thus, the Ward method tends to produce a large number of clusters with a small
250 number of members. We also treated the ordinal variables 0, 1, and 2 as the categorical variables and
251 conducted cluster analyses. Approximately similar dendrograms were obtained when using either the
252 UPGMA or Ward method. As shown in Fig. 2, the wild species and a subset of the genetic stocks were
253 distantly clustered as an outgroup. The second most distantly related group contained *S. phureja*-
254 derived genetic stocks and Inca Gold (a US-bred tetraploid variety). The French fry processing and
255 table russet group, the yellow and round white table group, the pigmented group, and the chip processing

group were classified into distinct clusters. The chip processing varieties Lenape, ND860-2, and S440 formed distinct subclusters. In addition, we identified another subcluster within the chip processing group that contained old chip processing varieties (Belchip, Atlantic and Wauseon).

Evaluation of the population 'All'

To understand the genetic relationships between Japanese potatoes and other potato varieties, all 432 genotypes were evaluated using 5972 polymorphic SNP loci with a diploid genotype model (Fig. 3). Similar to the clustering results obtained with the other populations, the group of wild species and some germplasms clustered most distantly. The second most distant cluster contained *S. phureja*-derived genetic stocks, US-bred Inca Gold and four Japanese diploid varieties, including Japanese-bred Inca Gold (a diploid with 37.5−62.5% *S. phureja* germplasm, bred from W822229-1 × W872209-5). The other genotypes were classified into three large clusters.

Cluster 1 was a mixture of all market classes. Out of 89 Japanese genotypes other than 2*x* varieties and CP clones, 80 genotypes were present in this cluster. The subcluster 1a is a group of old varieties such as pre-1930's US varieties Garnet Chili, Early Rose, Irish Cobbler, Russet Burbank, Green Mountain and pre-1950's Japanese varieties Norin 1 and Norin 2. One of the Japanese landraces, Kobo-imo, closely clustered with a group comprising Early Rose, Green Mountain, and Russet Burbank. Subclusters 1b and 1e contained Japanese and foreign genotypes of all market classes. Within the subcluster 1b, the Japanese landraces Nemuromurasaki and Kintoki-imo were clustered together. Subclusters 1c and 1f contained mostly the CP clones and their parents Andover, C44 and MN12828. Subcluster 1d contained one of Japanese landraces (Seinaiji-ki-imo), two CP clones, and 34 foreign genotypes, of which half were of the market class 'yellow'. Seinaiji-ki-imo clustered with the old European varieties May Queen and Bintje. Subcluster 1g contained exclusively Japanese genotypes, within which Nagasaki-bred varieties formed a distinct cluster, reflecting the close parentage relationships. The others within the subcluster 1g were Touya (released in 1992) and varieties released after 1992.

Cluster 2 was composed of the French fry processing and table russet subcluster and the pigmented subcluster, similarly identified in the population 'US'. Only one Japanese genotype in each subcluster was found.

285 Cluster 3 contained mostly chip processing varieties. Pike, NY115, Kitahime, and Monticello
286 formed respective subclusters with their progeny CP clones. However, three major sources for US chip
287 processing varieties (Lenape, ND860-2, and S440) did not form clusters with any Japanese genotypes or
288 CP clones with the exception of North Chip, bred from Hokkaikogane × ND860-2, which was closely
289 clustered with ND860-2. Furthermore, the old US chip processing varieties Belchip, Atlantic and
290 Wauseon were not clustered immediately with any Japanese genotypes. Although the Japanese
291 varieties, Snow March, Fugenmaru, Star Queen, and Okhotsk-chip and one of CP clones, C9132, were all
292 bred from Atlantic as the maternal parent, only Snow March and C9132 exhibited close relationships to
293 Atlantic. There was no Japanese genotype in the subcluster of Wauseon.
294 In the Fig. 3, parental clones currently used in the CP breeding program are indicated by stars.
295 Two stars indicate important parents frequently used in the past and present, whereas one star indicates a
296 parent rarely used. CP clones and their parents were clearly represented as subclusters. The parents
297 Andover, C44, C55, MN12828, Pike, NY115, Kitahime and Monticello were the most frequent
298 contributing parents to the CP clones. However, interestingly, however, six frequently used European
299 genotypes (C53, C54, C64, C65, C67 and Sassy) were members of the cluster 1d, in which no CP clones
300 were present.
301
302 Genetic diversity
303 The extent of genetic diversity within population and between populations was measured using mean
304 Euclidean distances (Fig. 4). All distances were significantly different from each other by *t* test
305 ($P<0.0005$) except for the distance of 48.4, obtained between Japanese 4*x* genotypes and foreign 4*x*
306 genotypes, between Japanese 4*x* genotypes and US 4*x* varieties, and among all 4*x* genotypes. The
307 largest within-populational distance of 51.4 and the largest between-populational distance of 57.6 were,
308 as expected, obtained from the population of wild species and genetic stocks and between this population
309 and the population of all 4*x* genotypes, verifying that the wild species and genetic stocks are most
310 distantly related from each other and from varieties and breeding clones. The Japanese 4*x* genotypes (86
311 named varieties, 2 breeding clones and 70 CP clones) showed a populational mean of 46.3, among which
312 named varieties (mean distance of 45.3) were less diverse than were CP clones (46.2) but more diverse
313 than Nagasaki-bred varieties (41.9). Inter-populational distances between Japanese 4*x* genotypes and

foreign 4*x* genotypes or US 4*x* varieties were 48.2–48.7, which were lower than the within-populational distances of US 4*x* varieties (49.1) or foreign 4*x* genotypes (49.3). These results indicate that the Japanese genotypes are less diverse and do not form a unique group but are included within the gene pool of US and foreign genotypes.

Heterozygosity

The percentage of heterozygous SNP loci (percent heterozygosity) in foreign genotypes including wild species and genetic stocks with all ploidy levels is shown in Fig. 5. Most of the data for foreign varieties were obtained from Hirsch et al. (2013). Thus, a very similar graph to Fig. 3A of Hirsch et al. (2013) was obtained. Early Rose was created from open-pollination of Garnet Chili; the seeds could have resulted from selfing or natural hybridization. Of the 5083 comparable SNP loci between the two varieties, 714 (14.0%) were homozygous in Garnet Chili and heterozygous in Early Rose. Thus, the average percent heterozygosity of Garnet Chili was 55.2%, whereas that of Early Rose was 61.8%. Apparently, Early Rose was derived by hybridization from Garnet Chili, as suggested by isozyme analysis (Douches et al. 1991). Irish Cobbler was formerly believed to be a sport of Early Rose (Plaisted and Hoopes 1989) but was later suggested to be a hybrid by isozyme analysis (Douches et al. 1991) and RAPD analysis (Hosaka et al. 1994). In the present study, the percent heterozygosity does not support a hybrid origin of Irish Cobbler because it is lower in Irish Cobbler (56.6%). However, of the 5149 comparable SNP loci between the two varieties, 397 (7.7%) were homozygous in Early Rose but heterozygous in Irish Cobbler. A likely explanation is that Early Rose was hybridized with a very closely related variety and originated Irish Cobbler.

The highest percent heterozygosity was observed in Atlantic (80.5%), followed by Lenape (74.3%), Snowden (69.6%), Pike (68.7%), MegaChip (68.3%), J138K6A22 (68.3%), Belchip (68.2%), Marcy (66.9%), Lehigh (66.6%), Yankee Chipper (66.4%), and Andover (66.3%). All of these except J138K6A22 (genetic stock) are chip processing varieties.

Using the 5972 SNP loci evaluated for the population ‘All’, or using 10810 SNP loci, 163 Japanese genotypes were analyzed (Table 3, Fig. 5). The average percent heterozygosity in the foreign 4*x* genotypes (56.7%) was not significantly different from that of the Japanese 4*x* varieties and breeding clones (57.3% using 5972 SNP loci, or 56.8% using 10810 SNP loci). However, the CP clones

exhibited a significantly higher percent heterozygosity (60.1% or 60.6%) than did the foreign and Japanese 4*x* genotypes by *t*-test ($P<0.001$). The average percent heterozygosity of the 2*x* varieties (23.1 or 23.8%) was much lower than that of the 4*x* genotypes. One 3*x* variety showed 51.0% or 51.5% heterozygosity, which was among the lowest heterozygosity of the 4*x* genotypes.

Discussion

Among currently available molecular markers, SSR markers have been widely utilized for variety identification and the assessment of genetic diversity in potato (Spooner et al. 2007; Ghislain et al. 2009; de Galarreta 2011; Salimi et al. 2016). Over 1000 European varieties have been characterized with unique SSR profiles (Reid et al. 2011). When the same system was transferred to the Canadian Food Inspection Agency to differentiate 217 varieties, occasional discrepancies between the two laboratories occurred due to their different interpretations of the presence/absence of the alleles (Côté et al. 2013). In contrast, identical results at all comparable SNP loci were obtained for the four sets of duplicates between Hirsch et al. (2013) and the present study. A high concordance of 99.98% between 39 replicated tetraploid samples was reported in a survey of 14530 SNP markers (Vos et al. 2015). These observations demonstrate that SNP array markers are highly reproducible. Hirsch et al. (2013) used 3763 SNP loci with a tetraploid genotype model (AAAA, AAAB, AABB, ABBB, and BBBB) and the UPGMA method for clustering 250 US genotypes. Although most of the genotypes used in this study were tetraploid, we used a diploid genotype model (AA, AB, and BB) for 5972 SNP loci and the Ward method for clustering, resulting in sufficient resolution for 432 genotypes. We treated the genotype data as ordinal variables (e.g., AA<AB<BB) for cluster analysis. However, treating these genotype data as categorical variables (e.g., AA≠AB≠BB) yielded similar clustering results. These findings indicate that if a large number of SNP data are used, the distinctiveness of observed clusters becomes robust regardless of the genotype model, clustering method, or variable type used. Therefore, the SNP array can readily generate a large number of reproducible data and is cost effective for analyzing the genetic relationships among potato varieties.

When the 8K SolCAP SNP array was applied to 250 diverse North American varieties, diploid genotypes were clearly separated from tetraploid genotypes, and the chip processing market class was the

most distinct group, clearly separated from all other markets classes (Hamilton et al. 2011; Hirsch et al. 2013). This is in good agreement with our results. Simko et al. (2006) analyzed 47 potato genotypes with 1088 SNPs and observed no subgroups despite the inclusion of 1 monoploid and 17 diploid genotypes. The number of SNPs analyzed by Simko et al. (2006) might have been insufficient to separate diploid genotypes from tetraploid ones. Using the same 8K SolCAP SNP array for European varieties, Stich et al. (2013) reported a clear separation of tetraploid from diploid genotypes but no distinct subgroups among 36 tetraploid genotypes. This lack of distinct subgroups may have been partly due to the method for detecting subgroups, which might have been inappropriate, and the lower differentiation of European potatoes than of US chip processing potatoes. However, Uidewilligen et al. (2013) analyzed the population structure of 84, mostly European, cultivars, using 43K sequence variants and identified three divergent groups, consisting of heirloom cultivars, frying cultivars, and cultivars and germplasm used in the starch industry. Thus, it is clear that a sufficient number of SNPs is needed to reveal the distinctiveness of genotypes.

Although we frequently used foreign varieties as parents in the past, most of the Japanese varieties in this study descended from Irish Cobbler (Supplementary Table 1), suggesting a narrow diversity of Japanese potatoes. This was verified by the significantly lower mean Euclidean distance among the Japanese varieties than that among the foreign genotypes (Fig. 4). However, we found a unique group of Japanese varieties (subcluster 1g in population 'All'), in which no foreign varieties were included. Within this subcluster, varieties bred in Nagasaki (all for table stocks) formed a distinct cluster. The Nagasaki potato breeding program is responsible for breeding varieties for double cropping (spring-season product is used as seed for fall-season cropping). Upon double cropping, the varieties are not in fact early-maturing but are late-maturing, and the tubers are bulked and harvested well before natural senescence and sold as "fresh new potatoes" at market. Consequently, fast emergence, rapid growth with low or no sensitivity to day length, early tuber bulking, and short tuber dormancy are required for double-cropping varieties. Apparently, high selection pressure was applied to achieve double cropping, resulting in a group of closely related varieties. European varieties for starch processing were identified as a distinct group; they were so identified because germplasm of *S. vernei* was incorporated to confer resistance to *G. pallida* (Uitdewilligen et al. 2013). In Japan, most starch processing varieties were initially selected as high-yielding clones for table stocks. However, their

401 inferior tuber quality lead to their use for starch production. Thus, these starch-processing varieties did
402 not form a distinct cluster. Only a few Japanese varieties were found within clusters of the yellow, the
403 pigmented, or the French fry and table russet groups of the population 'US'. There are yellow flesh
404 potatoes of recently released Japanese varieties (e.g., Kita-akari, Touya, Inca-no-mezame, Harimaru, and
405 Nagasaki Kogane). However, these were not intentionally selected for their flesh color; thus, they did
406 not cluster with the US yellow group. However, there are several varieties selected for their flesh color:
407 Inca Purple, Kitamurasaki, and Shadow Queen for purple flesh; and Inca Red and Northern Ruby for red
408 flesh. None of these varieties clustered with the US pigmented group. Therefore, we conclude that
409 although we found a distinct group of Japanese potatoes, Japanese varieties have not been genetically
410 differentiated for market types except for chip processing varieties. The first and still major chip
411 processing varieties, Waseshiro (released in 1974) and Toyoshiro (released in 1976), are a great-
412 grandchild and a second great-grandchild of Irish Cobbler, respectively, and were grouped with other
413 table stocks. However, the chip processing varieties released after 2000 were bred using the US chip
414 processing varieties Andover, Pike, Atlantic, Norking Russet and ND860-2 to improve chip color after
415 cold storage. For this reason, White Fryer (released in 2000), Kitahime (2001), North Chip (2002),
416 Okhotsk-chip (2005), and Poroshiri (2017) were grouped with the US chip processing varieties to form
417 Cluster 3 in population 'All' (Fig. 3).
418 Four landraces are of unknown origin and have been grown locally in very limited areas for
419 home consumption for many years since at least the early 1900s. There are many synonyms for each of
420 the four genotypes, but their genetic identities have been confirmed by RAPD and AFLP analyses
421 (Hosaka, unpublished). Nemuromurasaki has Andigena-type chloroplast DNA (Hosaka 1993). Its
422 close relationship with Andigena was demonstrated by RAPD analysis (Hosaka et al. 1994), providing
423 strong evidence that Nemuromurasaki is a relic potato of early European introduction (Hosaka et al.
424 1994). However, the present study does not support this hypothesis. Nemuromurasaki (purple tuber
425 skin) was most closely related to the other landrace in this study, Kintoki-imo (white tuber skin, T-type
426 cytoplasm) and was not separated from the other Japanese varieties (Figs. 1 and 3). Kobo-imo clustered
427 closely with Early Rose (Figs. 1 and 3), suggesting that the former is an immediate derivative of the
428 latter. Seinaiji-ki-imo was found within the cluster of European varieties, suggesting a European origin.
429 As Japan has systematically introduced foreign varieties since 1873 (Tsukikawa 1990) and tested for

local adaptability, some of those old foreign varieties might have escaped and spread without their correct names. Further investigation is needed to correctly identify these genotypes by comparing them with old foreign varieties.

Frequently used parents for the CP clones were Andover, Pike, NY115, and Monticello (Supplementary Table 1), all released from the Cornell University breeding program. These Cornell varieties formed large subclusters with CP clones (Fig. 3), suggesting that the CP clones are Japanese derivatives of Cornell chip varieties. Calbee Potato Inc. initially attempted to introduce superior US chip processing varieties and cultivate them in Japan. A threat against golden cyst nematodes (*G. rostochiensis*) forced the company to introduce varieties possessing the *H1* gene, which were available only from the Cornell breeding program at that time. Unfortunately, most of the US varieties did not perform well in Japanese environments. Thus, the CP breeders started to use the introduced Cornell varieties as parents for crossbreeding. Additional input of the other US and European varieties further modified the gene pool of CP clones. However, with these new introductions, the frequencies of D and W/γ cytoplasms increased, worrying the CP breeders with male sterility problems. One of the most important characteristics required for chip processing varieties is a low reducing sugar content after cold storage (cold-sweetening resistance). Three sources for cold-sweetening resistance have been identified among the US chip processing varieties (Hirsch et al. 2013): Lenape, the resistant variety originally derived from *S. chacoense* (Akeley et al. 1968); ND860-2, originally from *S. phureja* (Ehlenfeldt et al. 1990); and S440, originally from *S. tarijense* (Hermundstad 1986). Of the three sources, ND860-2 was used to incorporate resistance into Kitahime (a grandchild of ND860-2) and used frequently in Japan and for CP clones as well. To achieve more efficient breeding of chip processing varieties, the utilization of alternative sources for cold-sweetening resistance and crossing designed to avoid increasing the frequencies of undesired cytoplasms among the parental clones are suggested.

The maximum heterozygosity hypothesis is a well-known hypothesis that states that maximum heterosis is achieved by maximizing intra-locus diversity in autopolyploids (Chase 1963). However, Hirsch et al. (2013) demonstrated that after ~150 years of breeding, including intensive breeding in the chip and French fry processing market classes, substantial changes in heterozygosity and allele dosage have not occurred. The relationship between genome-wide percent heterozygosity and tuber yield has not been supported universally (Bonierbale et al. 1993). However, we recognized that highly

heterozygous US varieties tend to grow relatively in larger areas. Atlantic had the highest percent heterozygosity of the studied varieties and is the most popular variety grown worldwide. We introduced many foreign varieties in the past. However, only a few varieties, such as Atlantic, Snowden, and Andover, are grown in Japan at present. Thus, the genome-wide percent heterozygosity may be associated with the genetic plasticity: The higher the percent heterozygosity, the wider adaptability or more consistent performance the variety is expected to exhibit. An alternative explanation for the higher heterozygosities of Atlantic and Snowden is that the 8K and 12K SolCAP SNPs were originally developed as a collection of polymorphic nucleotides among these varieties (Hamilton et al. 2011).

The wild species were all clustered together in a single cluster, as shown in Figs. 1–3. Some of these, such as *S. bulbocastanum* and *S. pinnatisectum*, are reproductively isolated Mexican diploid species and the most distantly related potatoes from the cultivated potatoes in this study (Hawkes 1990). According to whole-genome re-sequencing of 20 South American landraces, 23 North American cultivars, 20 wild progenitor species, and four outgroup species (Hardigan et al. 2017), 68.9 million SNPs were found. Nucleotide divergence from the reference genome sequence (= *S. phureja* DM in this study) averaged 2.2% for Andigena, 2.4% for cultivars, 3.4% for wild progenitor species, and 4.2% for Mexican diploid species (Hardigan et al. 2017). Simko et al. (2006) found one SNP in every 23 bp in 66 DNA fragments from 47 potato genotypes. Thus, there must exist many SNPs in wild species that differ from the SolCAP SNPs detected among only six cultivars (Hamilton et al. 2011). This could result in significant bias when the SolCAP SNP array is applied to much wider or different germplasm (Moragues et al. 2010; Thomson et al. 2012; Vos et al. 2015; Torkamaneh et al. 2018) and likely caused contracted clustering of all of the wild species (as seen in Figs. 1–3), the unexpected positioning of the relic cultivar Nemuromurasaki, and biased heterozygosity.

Conclusion

This study demonstrated the usefulness of SNP array technology for performing a cost-effective, robust and direct comparison of genetic diversity among different gene pools. We conclude that Japanese varieties have not been genetically differentiated for market types except for recently released chip processing varieties. Compared with foreign genotypes, Japanese potatoes are not uniquely

differentiated but are included within the genetic diversity of foreign genotypes. In fact, we identified unutilized genetic variation in US French fry processing varieties and pigmented varieties (Cluster 2 in Fig. 3). This study was made possible by direct comparison using publicly available SNP data produced by Hirsch et al. (2013). Although the ascertainment bias is unavoidable in a SNP array technology, accumulation of the SNP data generated from the same SNP array could make it possible to compare the genetic diversity and relationships of different gene pools or breeding programs worldwide, which would be effective for understanding our potatoes and for developing breeding strategies, including determinations of what should be introduced next and what mating combination should be used.

Acknowledgments This research was supported by Calbee Inc., Hokkaido Potato Growers Association, Kewpie Corp., KENKO Mayonnaise Co., Ltd., and Japan Snack Cereal Foods Association.

Author contributions TI, MT and KH conceived and designed the experiments. TI, KO, EK, and RS performed the experiments and analyzed the data. TI, RS, and KH wrote the paper. All authors have read and approved the manuscript.

Conflict of interest The authors declare that they have no competing interests.

References

Akeley R, Mills W, Cunningham C, Watts J (1968) Lenape: a new potato variety high in solids and chipping quality. Am Potato J 45:142–145

Asama K (1978) Jagaimo 43 wa (43 stories on potato). Hokkaido-Shinbunsha, Sapporo

Bethke PC, Nassar AMK, Kubow S, Leclerc YN, Li XQ, Haroon M, Molen T, Bamberg J, Martin M, Donnelly DJ (2014) History and origin of Russet Burbank (Netted Gem) a sport of Burbank. Amer J Potato Res 91:594−609

Bonierbale MW, Plaisted RL, Tanksley SD (1993) A test of the maximum heterozygosity hypothesis using molecular markers in tetraploid potatoes. Theor Appl Genet 86:481−491

517 Chase SS (1963) Analytic breeding in *Solanum tuberosum* L.—a scheme utilizing parthenotes and
518 other diploid stocks. Can J Genet Cytol 5:359–363
519 Côté MJ, Leduc L, Reid A (2013) Evaluation of simple sequence repeat (SSR) markers established in
520 Europe as a method for the identification of potato varieties grown in Canada. Amer J Potato
521 Res 90:340−350
522 de Galarreta JIR, Barandalla L, Rios DJ, Lopez R, Ritter E (2011) Genetic relationships among local
523 potato cultivars from Spain using SSR markers. Genet Resour Crop Evol 58:383−395
524 Dionne LA (1961) Cytoplasmic sterility in derivatives of *Solanum demissum*. Am Potato J
525 38:117−120
526 Douches DS, Ludlam K, Freyre R (1991) Isozyme and plastid DNA assessment of pedigrees of
527 nineteenth century potato cultivars. Theor Appl Genet 82:195−200
528 Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue.
529 Phytochem Bull 19:11−15
530 Ehlenfeldt M, Lopez-Portilla D, Boe A, Johansen R (1990) Reducing sugar accumulation in progeny
531 families of cold chipping potato clones. Am Potato J 67:83–91
532 Felcher KJ, Coombs JJ, Massa AN, Hansey CN, Hamilton JP, Veilleux RE, Buell CR, Douches DS
533 (2012) Integration of two diploid potato linkage maps with the potato genome sequence.
534 PLoS ONE 7:e36347
535 Ghislain M, Núñez J, del Rosario Herrera M, Pignataro J, Guzman F, Bonierbale M, Spooner DM
536 (2009) Robust and highly informative microsatellite-based genetic identity kit for potato.
537 Mol Breed 23:377−388
538 Hamilton JP, Hansey CN, Whitty BR, Stoffel K, Massa AN, Van Deynze A, De Jong WS, Douches DS,
539 Buell CR (2011) Single nucleotide polymorphism discovery in elite North American potato
540 germplasm. BMC Genomics 12:302
541 Hardigan MA, Laimbeer FPE, Newton L, Crisovan E, Hamilton JP, Vaillancourt B, Wiegert-Rininger K,
542 Wood JC, Douches DS, Farré EM, Veilleux RE, Buell, CR (2017) Genome diversity of
543 tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in
544 the cultivated potato. Proc Natl Acad Sci USA 114:E9999-E10008

545 Hawkes JG (1990) The potato - evolution, biodiversity and genetic resources. Belhaven Press,
546 London
547 Hermundstad SA (1986) Haploid-wild species hybrids in potato breeding, genetics, and germplasm
548 enhancement. Ph.D. Thesis, University of Wisconsin-Madison
549 Hirsch CN, Hirsch CD, Felcher K, Coombs J, Zarka D, Van Deynze A, De Jong W, Veilleux RE, Jansky
550 S, Bethke P, Douches DS, Buell CR (2013) Retrospective view of North American potato
551 (*Solanum tuberosum* L.) breeding in the 20th and 21st centuries. G3 3:1003−1013
552 Hosaka K (1993) Similar introduction and incorporation of potato chloroplast DNA in Japan and
553 Europe. Jpn J Genet 68:55−61
554 Hosaka K, Hanneman REJr (1998) Genetics of self-compatibility in a self-incompatible wild diploid
555 potato species *Solanum chacoense*. 2. Localization of an S locus inhibitor (*Sli*) gene on the
556 potato genome using DNA markers. Euphytica 103:265−271
557 Hosaka K, Sanetomo R (2009) Comparative differentiation in mitochondrial and chloroplast DNA
558 among cultivated potatoes and closely related wild species. Genes Genet Syst 84:371−378
559 Hosaka K, Sanetomo R (2012) Development of a rapid identification method for potato cytoplasm
560 and its use for evaluating Japanese collections. Theor Appl Genet 125:1237−1251
561 Hosaka K, Mori M, Ogawa K (1994) Genetic relationships of Japanese potato cultivars assessed by
562 RAPD analysis. Am Potato J 71:535−546
563 Kawakami K (1948) Bareisho-Tsuron (Textbook on potato). Yokendo, Tokyo
564 Kolech SA, Halseth D, Perry K, Wolfe D, Douches DS, Coombs J, De Jong W (2016) Genetic
565 diversity and relationship of Ethiopian potato varieties to germplasm from North America,
566 Europe and the International Potato Center. Amer J Potato Res 93:609−619
567 Laufer B (1938) The American plant migration. Part 1: The potato. The potato in Japan and Korea.
568 Anthropological Series, Field Museum of Natural History 28:80−83
569 Lössl A, Götz M, Braun A, Wenzel G (2000) Molecular markers for cytoplasm in potato: male
570 sterility and contribution of different plastid-mitochondrial configurations to starch production.
571 Euphytica 116:221−230
572 Mendoza HA, Haynes FL (1974) Genetic relationship among potato cultivars grown in the United
573 States. HortScience 9:328−330

574 Ministry of Agriculture, Forestry and Fisheries (2018)
575 http://www.maff.go.jp/j/seisan/tokusan/imo/siryou.html. Accessed 15 February 2018
576 Moragues M, Comadran J, Waugh R, Milne I, Flavell A, Russell JR (2010) Effects of ascertainment
577 bias and marker number on estimations of barley diversity from high-throughput SNP genotype
578 data. Theor Appl Genet. 120:1525–1534
579 Mori K, Mukojima N, Nakao T, Tamiya S, Sakamoto Y, Sohbaru N, Hayashi K, Watanuki H, Nara K,
580 Yamazaki K, Ishii T, Hosaka K (2012) Germplasm release: Saikai 35, a male and female
581 fertile breeding line carrying *Solanum phureja*-derived cytoplasm and potato cyst nematode
582 resistance (*H1*) and *Potato virus Y* resistance (Ry_{chc}) genes. Amer J Potato Res 89:63−72
583 Mori M, Tsuda S, Mukojima N, Kobayashi A, Matsuura-Endo C, Ohara-Takada A, Zaidul ISM (2007)
584 Breeding of potato cyst nematode resistant varieties in Japan. In: Haverkort AJ, Anisimov BV
585 (eds) Potato production and innovative technologies. Wageningen Academic Publishers, The
586 Netherlands, pp 328−339
587 Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D (2010) Rapid genomic
588 characterization of the genus *Vitis*. PLoS One 5:e8219
589 Plaisted RL, Hoopes RW (1989) The past record and future prospects for the use of exotic potato
590 germplasm. Am Potato J 66:603−627
591 Reid A, Hof L, Felix G, Rücker B, Tams S, Milczynska E, Esselink D, Uenk G, Vosman B, Weitz A
592 (2011) Construction of an integrated microsatellite and key morphological characteristic
593 database of potato varieties on the EU common catalogue. Euphytica 182:239−249
594 Ross H (1986) Potato breeding-problems and perspectives. Verlag Paul Parey, Berlin
595 Salimi H, Bahar M, Mirlohi A, Talebi M (2016) Assessment of the genetic diversity among potato
596 cultivars from different geographical areas using the genomic and EST microsatellites. Iran J
597 Biotech 14:e1280
598 Sanetomo R, Hosaka K (2011) A maternally inherited DNA marker, descended from *Solanum*
599 *demissum* (2*n*=6*x*=72) to *S. tuberosum* (2*n*=4*x*=48). Breed Sci 61:426−434
600 Simko I, Haynes KG, Jones RW (2006) Assessment of linkage disequilibrium in potato genome with
601 single nucleotide polymorphism markers. Genetics 173:2237–2245
602 Spooner DM, Núñez J, Trujillo G, del Rosario Herrera M, Guzmán F, Ghislain M (2007) Extensive

603 simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene
604 pool structure and classification. Proc Natl Acad Sci USA 104:19398−19403
605 Stich B, Urbany C, Hoffmann P, Gebhardt C (2013) Population structure and linkage disequilibrium
606 in diploid and tetraploid potato revealed by genome-wide high-density genotyping using the
607 SolCAP SNP array. Plant Breed 132:718–724
608 Song Q, Hyten DL, Jia G, Quigley CV, Fickus EW, Nelson RL, Cregan PB (2013) Development and
609 evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS ONE 8:e54985
610 Sukhotu T, Kamijima O, Hosaka K (2004) Nuclear and chloroplast DNA differentiation in Andean
611 potatoes. Genome 47:46−56
612 Takase N (1977) Breeding and variety. In: Taguchi K, Murayama D (eds) Potatoes in Japan.
613 Green Disen Fukyukai, Sapporo, pp 37−72
614 Thomson M, Zhao K, Wright M, McNally K, Rey J, Tung C-W, Reynolds A, Scheffler B, Eizenga G,
615 McClung A, Kim H, Ismail A, de Ocampo M, Mojica C, Reveche MY, Dilla-Ermita C, Mauleon
616 R, Leung H, Bustamante C, McCouch S (2012) High-throughput single nucleotide
617 polymorphism genotyping for breeding applications in rice using the BeadXpress platform.
618 Mol Breed 29:875–886
619 Torkamaneh D, Boyle B, Belzile F (2018) Efficient genome-wide genotyping strategies and data
620 integration in crop plants. Theor Appl Genet 131:499–511
621 Tsukikawa M (1990) Nagasaki jaga-imo hattatsushi (Potato history in Nagasaki). Nagasaki Seed
622 Potato Association, Nagasaki
623 Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, Borm TJA, Visser RGF, van Eck HJ (2013) A
624 next-generation sequencing method for genotyping-by-sequencing of highly heterozygous
625 autotetraploid potato. PLoS ONE 8:e62355
626 Vos PG, Uitdewilligen JGAML, Voorrips RE, Visser RGF, van Eck HJ (2015) Development and
627 analysis of a 20K SNP array for potato (*Solanum tuberosum*): an insight into the breeding
628 history. Theor Appl Genet 128:2387−2401
629 Ward JH Jr (1963) Hierarchical grouping to optimize an objective function. J Am Stat Assoc
630 58:236−244
631

632 **Table 1** The numbers of genotypes in each population

Population and category	Market class	n
1) Population ‘Japan’		164
Japanese variety	Table (4*x*)	50
	Table (3*x*)	1
	Table (2*x*)	5
	Starch	19
	Chip	9
	French fry	4
Japanese landrace	Table	4
Japanese advanced breeding clone	Table	2
Calbee Potato breeding clone	Chip	70
2) Population ‘Japan Plus’		205
Population ‘Japan’		163
US variety	Table	5*
	Chip	3+6*
Canadian varieties	French fry	1
US advanced breeding clones	Chip	8+1*
European varieties	Table	1
	Chip	8
European advanced breeding clones	Chip	5
Wild species as an outgroup		4*
3) Population ‘US’ (Hirsch et al. 2013)		247

		Round white table	38
		Table russet	11
		Pigmented	31
		Yellow	28
		Chip	69
		French fry	33
		Genetic stock	25
		Wild species	12
	4) Population ‘All’		432

633 The SNP data of the genotypes shown by asterisks used in the population ‘Japan
634 Plus’ were cited from Hirsch et al. (2013)
635

636 **Table 2** The number of SNPs used for cluster analyses and heterozygosity calculation

Population	No. of clones	Initial no. of SNPs	Monomorphic			>10% missing value	Final no. of SNPs
			AA	AB	BB		
US	247	6373	56	0	56	970	5306
Japan	163	12720	250	177	278	1269	10810
Japan Plus	205	6373	59	0	64	84	6172
All	432	6373	54	0	54	304	5972

637
638

639

640 **Table 3** The mean heterozygosity and standard deviation in parenthesis

641 calculated using 5972 (upper column) or 10810 (lower column) SNP loci

Category	n	Heterozygosity (%)
Foreign 4*x* genotypes	232	56.7 (4.70)
		- (-)
Japanese 4*x* varieties and	88	57.3 (3.63)
breeding clones		56.8 (3.37)
Japanese 3*x* variety	1	51.0
		51.5
Japanese 2*x* varieties	4	23.1 (3.45)
		23.8 (3.43)
Calbee Potato breeding	70	60.1 (2.79)
clones		60.6 (1.83)

642
643

644
645 Figure Legends
646

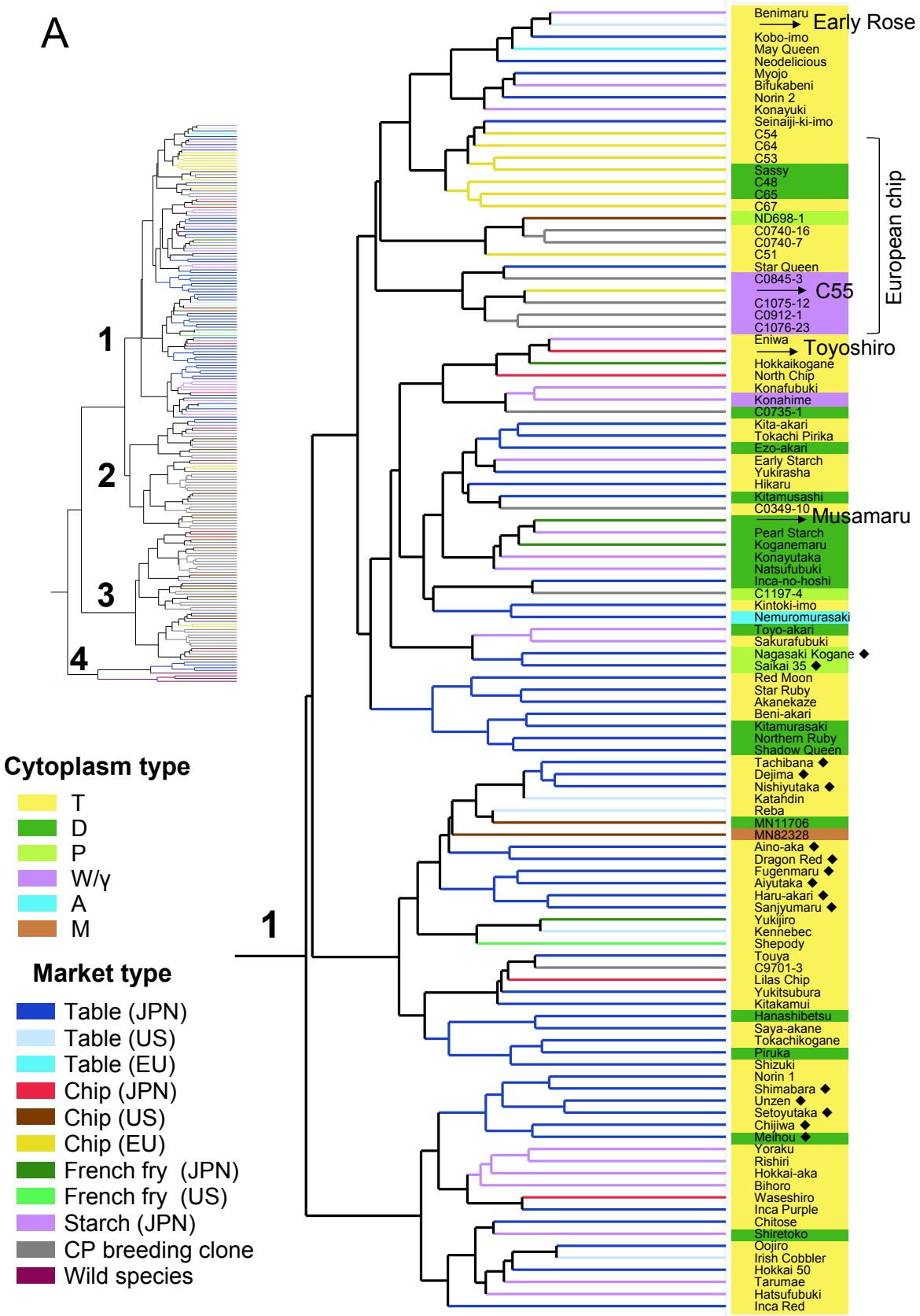
647 **Fig. 1** A Ward dendrogram, the upper (A) and the lower (B) parts shown separately, exhibiting
648 relationships among genotypes in the population “Japan Plus”. Cytoplasm types and market classes of
649 the genotypes are indicated by different colors in the box and on the branches, respectively. Genotypes
650 with diamond symbols are Nagasaki-bred varieties.
651

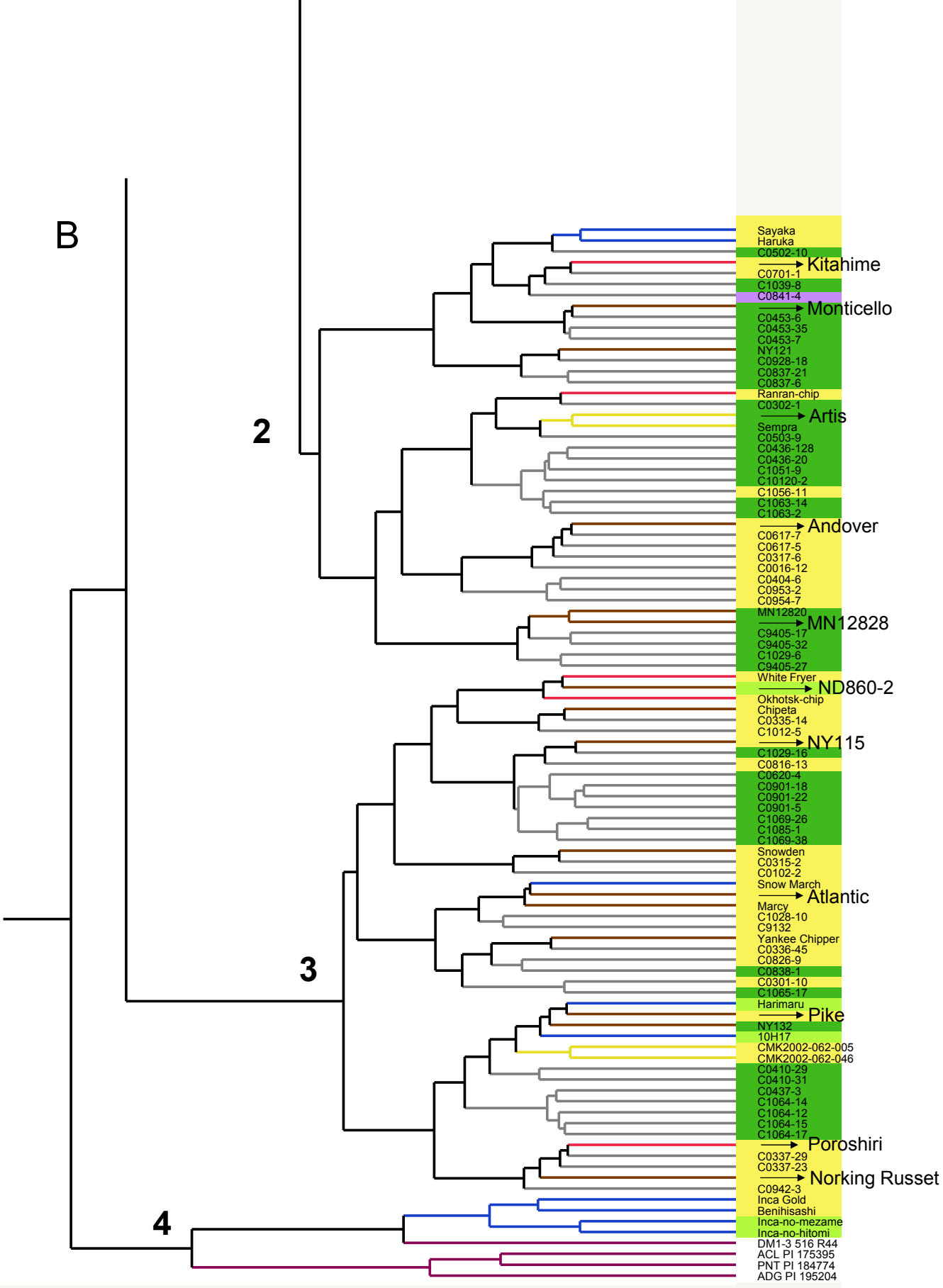
652 **Fig. 2** A Ward dendrogram exhibiting relationships among genotypes in the population “US”. Market
653 types of the genotypes are indicated by different colors on the branches.
654

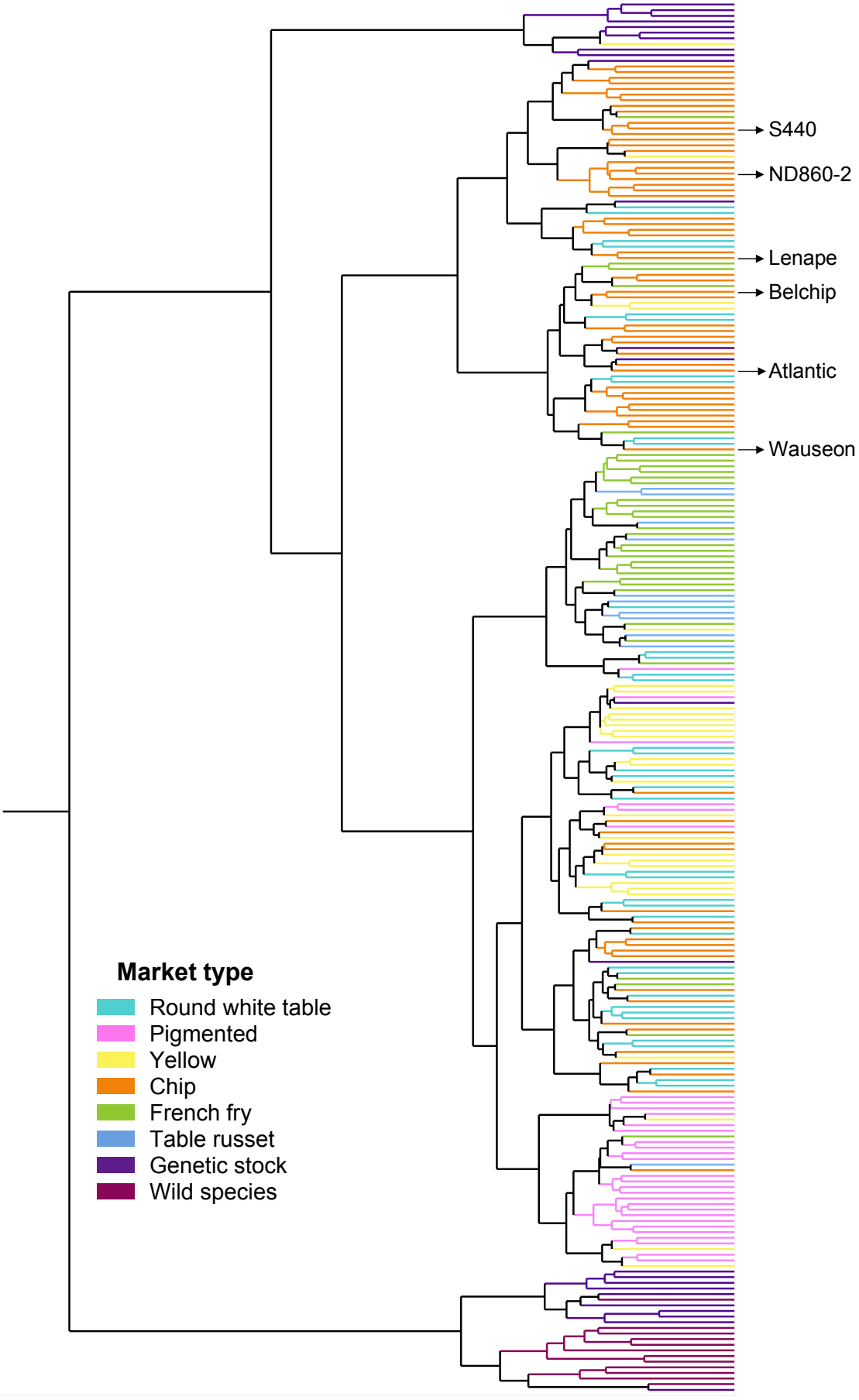
655 **Fig. 3** A Ward dendrogram, the upper (A), the middle (B) and the lower (C) parts shown separately,
656 exhibiting relationships among genotypes in the population “All”. Japanese genotypes and market types
657 for foreign genotypes are indicated by different colors on the branches. Parental clones frequently used
658 in the Calbee Potato breeding program are indicated by stars (see text). Genotypes with diamond
659 symbols are Nagasaki-bred varieties.
660

661 **Fig. 4** **W**ithin and between populational mean Euclidean distances $\pm$ standard deviation calculated
662 using 5972 SNP loci. The numbers of genotypes are shown in parentheses.
663

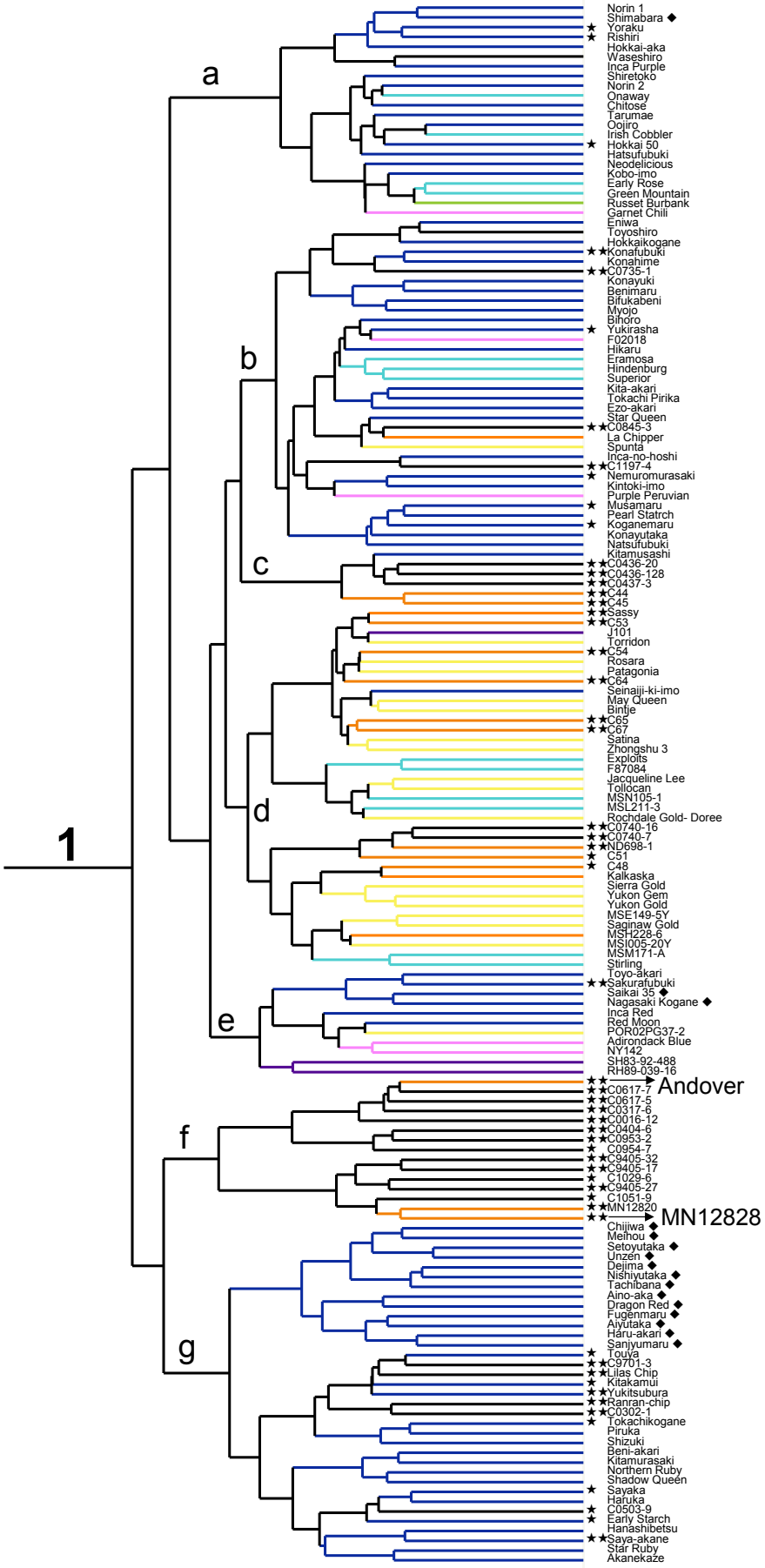
664 **Fig. 5** Heterozygosity (%) in foreign and Japanese varieties. 1*x* (red), 2*x* (green), 3*x* (purple), 4*x* (light
665 blue or blue), and Calbee Potato breeding clones (black) are represented.


A
1
2
3
4
Cytoplasm type
T
D
P
W/γ
A
M
Market type
Table (JPN)
Table (US)
Table (EU)
Chip (JPN)
Chip (US)
Chip (EU)
French fry (JPN)
French fry (US)
Starch (JPN)
CP breeding clone
Wild species
1
Benimaru
Early Rose
Kobo-imo
May Queen
Neodelicious
Myojo
Bifukabeni
Norin 2
Konayuki
Seinaiji-ki-imo
C54
C64
C53
Sassy
C48
C65
C67
ND698-1
C0740-16
C0740-7
C51
Star Queen
C0845-3
C55
C1075-12
C0912-1
C1076-23
European chip
Eniwa
Toyoshiro
Hokkaikogane
North Chip
Konafubuki
Konahime
C0735-1
Kita-akari
Tokachi Pirika
Ezo-akari
Early Starch
Yukirasha
Hikaru
Kitamusashi
C0349-10
Musamaru
Pearl Starch
Koganemaru
Konayutaka
Natsufubuki
Inca-no-hoshi
C1197-4
Kintoki-imo
Nemuromurasaki
Toyo-akari
Sakurafubuki
Nagasaki Kogane ◆
Saikai 35 ◆
Red Moon
Star Ruby
Akanekaze
Beni-akari
Kitamurasaki
Northern Ruby
Shadow Queen
Tachibana ◆
Dejima ◆
Nishiyutaka ◆
Katahdin
Reba
MN11706
MN82328
Aino-aka ◆
Dragon Red ◆
Fugenmaru ◆
Aiyutaka ◆
Haru-akari ◆
Sanjyumaru ◆
Yukijiro
Kennebec
Shepody
Touya
C9701-3
Lilas Chip
Yukitsubura
Kitakamui
Hanashibetsu
Saya-akane
Tokachikogane
Piruka
Shizuki
Norin 1
Shimabara ◆
Unzen ◆
Setoyutaka ◆
Chijiwa ◆
Meihou ◆
Yoraku
Rishiri
Hokkai-aka
Bihoro
Waseshiro
Inca Purple
Chitose
Shiretoko
Oojiro
Irish Cobbler
Hokkai 50
Tarumae
Hatsufubuki
Inca Red

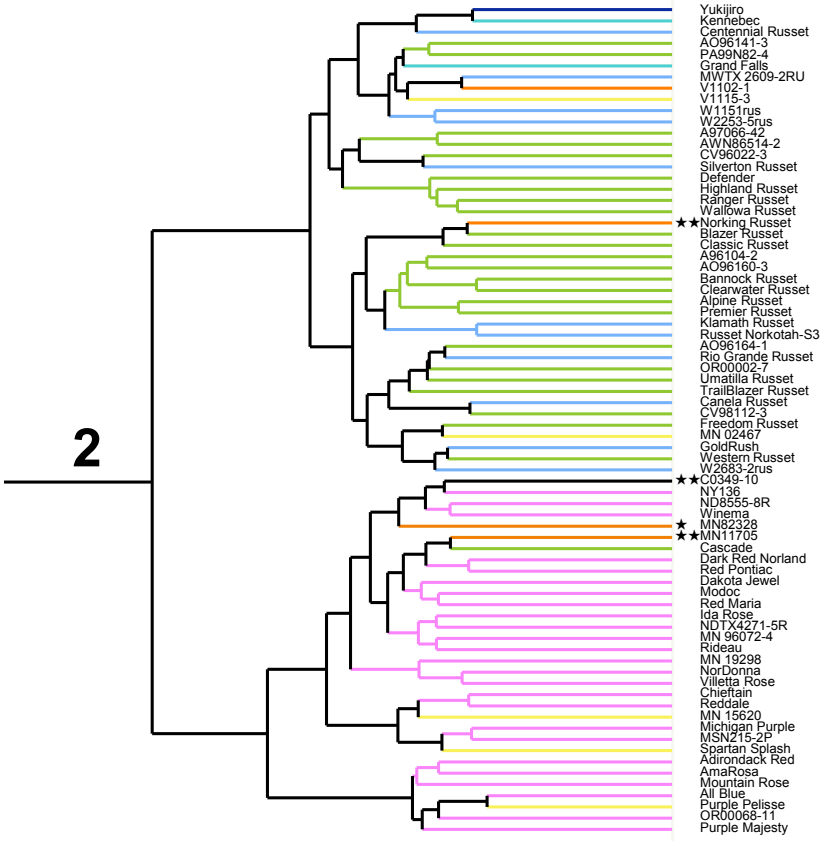

B
2
3
4
Sayaka
Haruka
C0502-10
Kitahime
C0701-1
C1039-8
C0841-4
Monticello
C0453-6
C0453-35
C0453-7
NY121
C0928-18
C0837-21
C0837-6
Ranran-chip
C0302-1
Artis
Sempra
C0503-9
C0436-128
C0436-20
C1051-9
C10120-2
C1056-11
C1063-14
C1063-2
Andover
C0617-7
C0617-5
C0317-6
C0016-12
C0404-6
C0953-2
C0954-7
MN12820
MN12828
C9405-17
C9405-32
C1029-6
C9405-27
White Fryer
ND860-2
Okhotsk-chip
Chipeta
C0335-14
C1012-5
NY115
C1029-16
C0816-13
C0620-4
C0901-18
C0901-22
C0901-5
C1069-26
C1085-1
C1069-38
Snowden
C0315-2
C0102-2
Snow March
Atlantic
Marcy
C1028-10
C9132
Yankee Chipper
C0336-45
C0826-9
C0838-1
C0301-10
C1065-17
Harimaru
Pike
NY132
10H17
CMK2002-062-005
CMK2002-062-046
C0410-29
C0410-31
C0437-3
C1064-14
C1064-12
C1064-15
C1064-17
Poroshiri
C0337-29
C0337-23
Norking Russet
C0942-3
Inca Gold
Benihisashi
Inca-no-mezame
Inca-no-hitomi
DM1-3 516 R44
ACL PI 175395
PNT PI 184774
ADG PI 195204

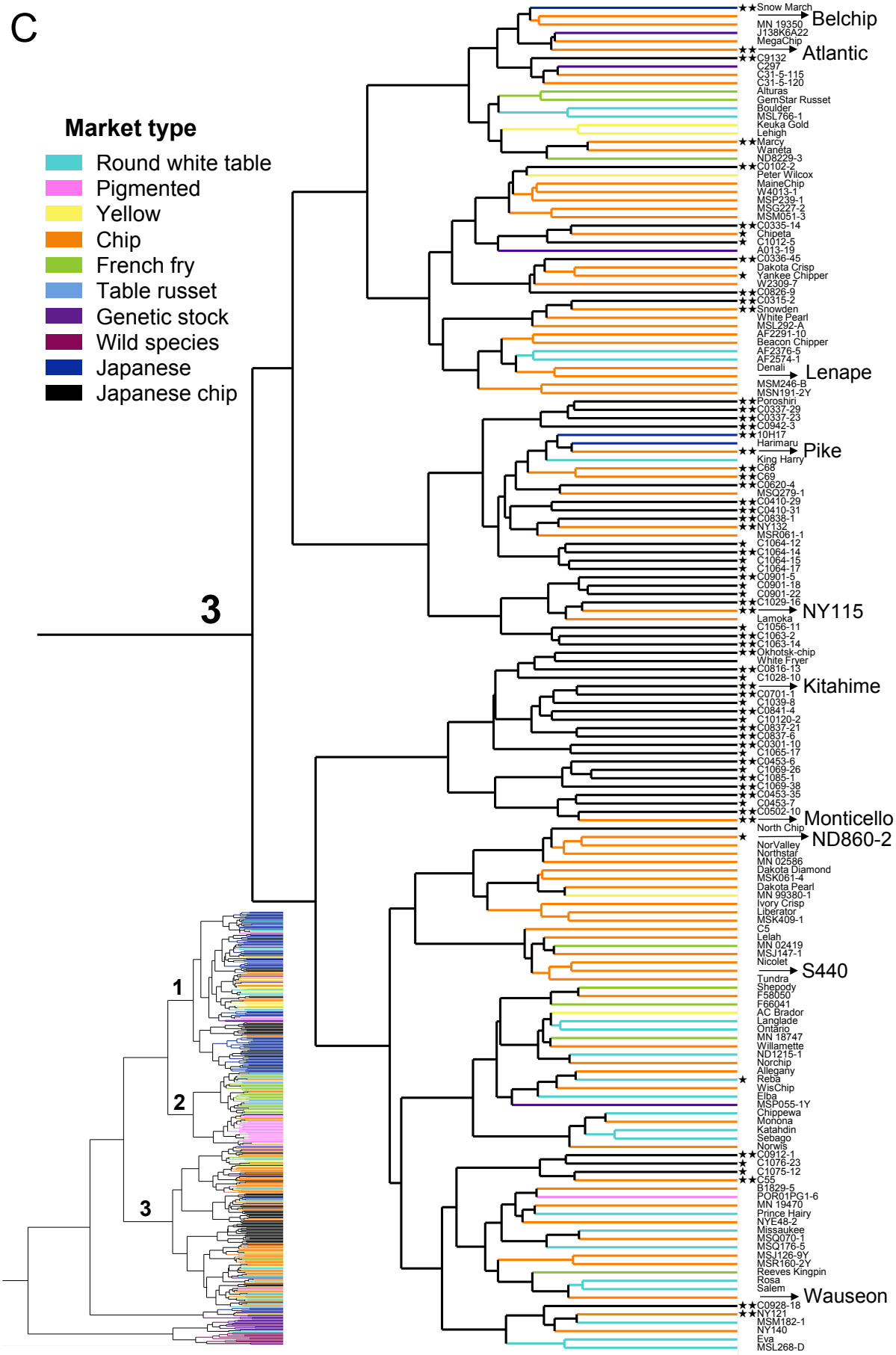

S440
ND860-2
Lenape
Belchip
Atlantic
Wauseon
Market type
Round white table
Pigmented
Yellow
Chip
French fry
Table russet
Genetic stock
Wild species

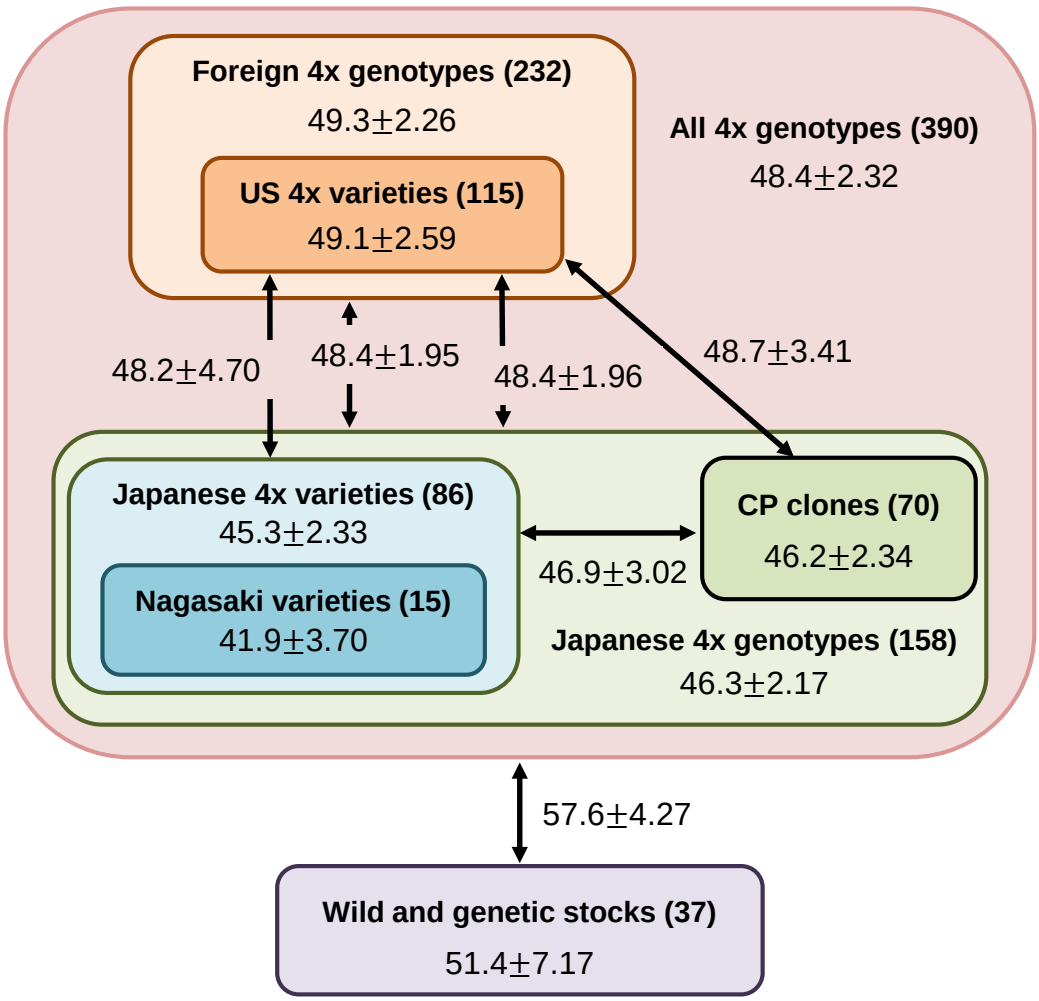
A


1
a
b
c
d
e
f
g
Norin 1
Shimabara ◆
★ Yoraku
★ Rishiri
Hokkai-aka
Waseshiro
Inca Purple
Shiretoko
Norin 2
Onaway
Chitose
Tarumae
Oojiro
Irish Cobbler
★ Hokkai 50
Hatsufubuki
Neodelicious
Kobo-imo
Early Rose
Green Mountain
Russet Burbank
Garnet Chili
Eniwa
Toyoshiro
Hokkaikogane
★★Konafubuki
Konahime
★★C0735-1
Konayuki
Benimaru
Bifukabeni
Myojo
Bihoro
★ Yukirasha
F02018
Hikaru
Eramosa
Hindenburg
Superior
Kita-akari
Tokachi Pirika
Ezo-akari
Star Queen
★★C0845-3
La Chipper
Spunta
Inca-no-hoshi
★★C1197-4
★ Nemuromurasaki
Kintoki-imo
Purple Peruvian
★ Musamaru
Pearl Statrch
★ Koganemaru
Konayutaka
Natsufubuki
Kitamusashi
★★C0436-20
★★C0436-128
★★C0437-3
★★C44
★★C45
★★Sassy
★★C53
J101
Torridon
★★C54
Rosara
Patagonia
★★C64
Seinaiji-ki-imo
May Queen
Bintje
★★C65
★★C67
Satina
Zhongshu 3
Exploits
F87084
Jacqueline Lee
Tollocan
MSN105-1
MSL211-3
Rochdale Gold- Doree
★★C0740-16
★★C0740-7
★★ND698-1
★ C51
★ C48
Kalkaska
Sierra Gold
Yukon Gem
Yukon Gold
MSE149-5Y
Saginaw Gold
MSH228-6
MSI005-20Y
MSM171-A
Stirling
Toyo-akari
★★Sakurafubuki
Saikai 35 ◆
Nagasaki Kogane ◆
Inca Red
Red Moon
POR02PG37-2
Adirondack Blue
NY142
SH83-92-488
RH89-039-16
★★ Andover
★★C0617-7
★★C0617-5
★★C0317-6
★★C0016-12
★★C0404-6
★★C0953-2
★ C0954-7
★★C9405-32
★★C9405-17
★ C1029-6
★★C9405-27
★ C1051-9
★★MN12820
★★ MN12828
Chijiwa ◆
Meihou ◆
Setoyutaka ◆
Unzen ◆
Dejima ◆
Nishiyutaka ◆
Tachibana ◆
Aino-aka ◆
Dragon Red ◆
Fugenmaru ◆
Aiyutaka ◆
Haru-akari ◆
Sanjyumaru ◆
★ Touya
★★C9701-3
★★Lilas Chip
★ Kitakamui
★★Yukitsubura
★★Ranran-chip
★★C0302-1
★ Tokachikogane
Piruka
Shizuki
Beni-akari
Kitamurasaki
Northern Ruby
Shadow Queen
★ Sayaka
Haruka
★ C0503-9
★ Early Starch
Hanashibetsu
★★Saya-akane
Star Ruby
Akanekaze

B


2
Yukijiro
Kennebec
Centennial Russet
AO96141-3
PA99N82-4
Grand Falls
MWTX 2609-2RU
V1102-1
V1115-3
W1151rus
W2253-5rus
A97066-42
AWN86514-2
CV96022-3
Silverton Russet
Defender
Highland Russet
Ranger Russet
Wallowa Russet
★★Norking Russet
Blazer Russet
Classic Russet
A96104-2
AO96160-3
Bannock Russet
Clearwater Russet
Alpine Russet
Premier Russet
Klamath Russet
Russet Norkotah-S3
AO96164-1
Rio Grande Russet
OR00002-7
Umatilla Russet
TrailBlazer Russet
Canela Russet
CV98112-3
Freedom Russet
MN 02467
GoldRush
Western Russet
W2683-2rus
★★C0349-10
NY136
ND8555-8R
Winema
★MN82328
★★MN11705
Cascade
Dark Red Norland
Red Pontiac
Dakota Jewel
Modoc
Red Maria
Ida Rose
NDTX4271-5R
MN 96072-4
Rideau
MN 19298
NorDonna
Villetta Rose
Chieftain
Reddale
MN 15620
Michigan Purple
MSN215-2P
Spartan Splash
Adirondack Red
AmaRosa
Mountain Rose
All Blue
Purple Pelisse
OR00068-11
Purple Majesty


C
Market type
Round white table
Pigmented
Yellow
Chip
French fry
Table russet
Genetic stock
Wild species
Japanese
Japanese chip
3
1
2
3
Belchip
Atlantic
Lenape
Pike
NY115
Kitahime
Monticello
ND860-2
S440
Wauseon
Snow March
MN 19350
J138K6A22
MegaChip
C9132
C297
C31-5-115
C31-5-120
Alturas
GemStar Russet
Boulder
MSL766-1
Keuka Gold
Lehigh
Marcy
Wanéta
ND8229-3
C0102-2
Peter Wilcox
MaineChip
W4013-1
MSP239-1
MSG227-2
MSM051-3
C0335-14
Chipeta
C1012-5
A013-19
C0336-45
Dakota Crisp
Yankee Chipper
W2309-7
C0826-9
C0315-2
Snowden
White Pearl
MSL292-A
AF2291-10
Beacon Chipper
AF2376-5
AF2574-1
Denali
MSM246-B
MSN191-2Y
Poroshiri
C0337-29
C0337-23
C0942-3
10H117
Harimaru
King Harry
C68
C69
C0620-4
MSQ279-1
C0410-29
C0410-31
C0838-1
NY132
MSR061-1
C1064-12
C1064-14
C1064-15
C1064-17
C0901-5
C0901-18
C0901-22
C1029-16
Lamoka
C1056-11
C1063-2
C1063-14
Okhotsk-chip
White Fryer
C0816-13
C1028-10
C0701-1
C1039-8
C0841-4
C10120-2
C0837-21
C0837-6
C0301-10
C1065-17
C0453-6
C1089-26
C1085-1
C1069-38
C0453-35
C0453-7
C0502-10
North Chip
NorValley
Northstar
MN 02586
Dakota Diamond
MSK061-4
Dakota Pearl
MN 99380-1
Ivory Crisp
Liberator
MSK409-1
C5
Lelah
MN 02419
MSJ147-1
Nicolet
Tundra
Shepody
F58050
F66041
AC Brador
Langlade
Ontario
MN 18747
Willamette
ND1215-1
Norchip
Allegany
Reba
WisChip
Elba
MSP055-1Y
Chippewa
Monona
Katahdin
Sebago
Norwis
C0912-1
C1076-23
C1075-12
C55
B18229-5
POR01PG1-6
MN 19470
Prince Hairy
NYE48-2
Missaukee
MSQ070-1
MSQ176-5
MSJ126-9Y
MSR160-2Y
Reeves Kingpin
Rosa
Salem
C0928-18
NY121
MSM182-1
NY140
Eva
MSL268-D


Foreign 4x genotypes (232)
49.3±2.26
US 4x varieties (115)
49.1±2.59
All 4x genotypes (390)
48.4±2.32
48.2±4.70
48.4±1.95
48.4±1.96
48.7±3.41
Japanese 4x varieties (86)
45.3±2.33
Nagasaki varieties (15)
41.9±3.70
CP clones (70)
46.2±2.34
46.9±3.02
Japanese 4x genotypes (158)
46.3±2.17
57.6±4.27
Wild and genetic stocks (37)
51.4±7.17