Author for correspondence: Yuzu Sakata

Mailing address: Center for Ecological Research, Kyoto University, Otsu 520-2113 Japan

Tel: +81 077-549-8019

Fax: +81 077-549-8262

Email: sakata@ecology.kyoto-u.ac.jp

Membership of the Botanical Society of Japan: Yuji Isagi

Subject area: (1) Taxonomy/ Phylogenetics/ Evolutionary Biology

Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range

Author: Yuzu Sakata[1], Joanne Itami[2], Yuji Isagi[3], Takayuki Ohgushi[1]

Address: [1]*Center for Ecological Research, Kyoto University, Otsu 520-2113 Japan*

[2]*Department of Biology, University of Minnesota Duluth, Minnesota 55812 USA*

[3]*Laboratory of Forest Biology, Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502 Japan*

Abstract

Understanding the origins and diversity of invasive species can reveal introduction and invasion pathways, and inform an effective management of invasive species. Tall goldenrod, *Solidago altissima*, is a herbaceous perennial plant native to North America and it has become a widespread invasive weed in East Asian countries. We used microsatellite and chloroplast DNA markers to obtain information on neutral processes and on genetic diversity in native and invaded populations of *S. altissima* and to infer how it invaded and spread in Japan. We found that introduced (n = 12) and native (n = 21) populations had similar levels of genetic diversity at nuclear SSR loci. Genetic structure analysis indicated that at least two independent colonization events gave rise to current *S. altissima* populations in Japan. The majority (68%) of the Japanese *S. altissima* were genetically similar and likely shared a common origin from a single or a small number of populations from the southern USA populations, while the populations in Hokkaido were suggested to arise from a different source. Our results suggest that multiple and mass introductions have contributed to the persistence and rapid adaptation of *S. altissima* promoting its widespread establishment throughout Japan.

Keywords

Genetic diversity · Invasion history · Microsatellite · Phylogeography · Population genetics · *Solidago altissima*

Introduction

Determining which factors enable exotic plants to proliferate in new environments is a fundamental challenge worldwide (Thuiller et al. 2005; Mitchell et al. 2006). The genetic variation in founder populations is a crucial factor in determining whether an invasive species will successfully adapt to their new locations (Lee 2002). Genetic variation can be introduced into a new geographic area either through multiple introductions of genotypes (Ellstrand and Schierenbeck 2000; Bossdorf et al. 2005; Kelager et al. 2013), or through introduction of a few pre-adapted genotypes with a broad range of physiological tolerance and phenotypic plasticity (Lee 2002; Dlugosch and Parker 2008; Yu et al. 2014).

Phylogeographical studies of invasive species using neutral genetic markers can retrace the possible routes of introduction, and determine the number of introductions, genetic variability of invasive and source populations, and the degree of hybridization among source populations of invasive species (Estoup and Guillemaud 2010; Handley et al. 2011; Fitzpatrick et al. 2012). This information will help determine the relative importance of stochastic and deterministic forces in determining the success of an invasive species (Lee 2002; Lavergne and Molofsky 2007; Keller and Taylor 2008). The level of genetic diversity of exotic plants plays a key role in their ability to invade new areas (Lambrinos 2004; Dlugosch and Parker 2008; Keller and Taylor 2008; Vallejo-Marin and Lye 2013). Knowing the origins of invasive plants and how they spread is critical in designing strategies to control the colonization and spread of invasive plants (Roderick and Navajas 2003; Estoup and Guillemaud 2010).

Solidago altissima L. (Asteraceae), is a perennial herbaceous plant belonging to the *Solidago* subsect. Triplinervae. It is a dominant plant in the early stages of secondary succession in prairies, woodland edges, and old fields throughout North America in a broad geographic range (Semple and Cook 2006). It is among the most invasive introduced plants

61 in the East Asian countries of Japan, China, Korea, and Taiwan (Li 1978; Shimizu 2003;
62 Huang and Guo 2004; Kil et al. 2004). In its native range, *S. altissima* occurs as diploid,
63 tetraploid, and hexaploid (2n = 18, 36, 54; Halverson et al. 2008). However, in Japan only
64 hexaploids have been found (Sakata et al. 2013a). Semple et al. (2015) recognized three
65 varieties of *S. altissima*. These varieties are associated with cytotypic variation, with *S.*
66 *altissima* var. *gilvocanescens* reported as diploid and tetraploid, and *S. altissima* var.
67 *altissima* and *S. altissima* var. *pluricephala* primarily as hexaploid with a few tetraploids
68 reported at the western edge of their distribution and across the southeastern USA. *S.*
69 *altissima* var. *altissima* and var. *pluricephala* occur from the eastern edge of the Great Plains
70 to the Atlantic coast, while var. *altissima* are found predominately from 35° to 50° N latitude,
71 *S. a. pluricephala* primarily south of 35° N latitude. *Solidago altissima* var. *gilvocanescens* is
72 found in the Great Plains (Semple et al. 2015).
73 *Solidago altissima* was first introduced to Japan as an ornamental plant in 1897, but it
74 was not until the 1980's that became naturalized throughout the country (Fukuda 1982), and
75 it is still rare on the island of Hokkaido. *Solidago altissima* reproduces through obligate
76 outcrossing via pollination by a wide range of insects (Melville and Morton 1982), and it
77 produces large numbers of wind-dispersed seeds. Once established, it spreads almost
78 exclusively through vegetative reproduction (Meyer and Schmid 1999). Two closely related
79 species, *S. canadensis* L. and *S. gigantea* Aiton, have also become invasive throughout
80 Europe (Weber 1997; Weber and Jakobs 2005). *Solidago altissima* has been studied for its
81 phenotypic traits such as growth, allelopathic, and defensive traits in its invaded range in
82 Japan (Ito et al. 1998; Sakata et al. 2014). Despite *S. altissima*'s status as a globally important
83 invasive species, little is known about its invasion history, and how its genetic diversity has
84 changed through the invasion process. Information of the invasion history would allow
85 comparisons of phenotypic traits between native ancestors of the invasive populations and

allow inferences to be made about the evolution of plant traits in the invasive populations (Kellor and Taylor 2008).

In order to obtain information on genetic diversity in populations of *S. altissima* in both its native and invaded range, and to identify the source populations of lineages invasive in Japan to infer how it was introduced and spread in Japan, we studied the genetic diversity and structure of multiple *S. altissima* populations in their native and invaded ranges. We used both chloroplast DNA (cpDNA) markers and recently-developed nuclear simple sequence repeat (nSSR) markers, which differ in their mode of inheritance (maternal only vs. biparental) and mutation rate (higher at nuclear markers), and therefore give complementary insights into the invasion history and population dynamics (i.e., changes in size and age of populations) of the species.

Materials and Methods

Population sampling

Samples were collected from 20 populations of *S. altissima* from its native range in North America, and 11 populations from its invaded range in Japan, and one population in Korea during June to August in 2011-2013 (Table S1). In each population, samples of 10-24 (22.3 on average) leaves of *S. altissima* were collected from plants that were at least 5m apart. Because *S. altissima* in Japan are all hexaploid, we collected throughout the hexaploid range in the USA. Because of the possibility that hexaploid plants were derived from the hybridization between diploids and tetraploids plants after their invasion of Japan, diploid (CA: California, USA) and tetraploid plants (EF: Oklahoma, USA) were also collected to examine the genetic relationship among plants of different ploidy levels (Table S1). We determined the ploidy level of plants in the Midwestern USA since other ploidy level individuals occur sympatrically. We collected and cultivated rhizomes of 5 to 10 individuals

111 that we had used for leaf samples in the Midwest populations (HB, FB, CL, PE, CG, EF, I20,
112 and HIL). Fresh leaf samples were collected from these cultivars to determine the ploidy
113 levels by using chromosome counts, flow cytometry as described in Sakata et al. (2013a). In
114 addition, allele numbers with nSSR genotyping were also used to determine the ploidy level
115 of individuals without rhizome samples and plants in Korean populations. We also collected
116 two individuals of *Solidago virgaurea* (subsection Solidago; from northern Honshu, Japan),
117 24 individuals of *Solidago gigantea* (subsection Triplinerviae; from Hokkaido, Japan), and
118 six individuals of *Solidago canadensis* (subsection Triplinervae; from Jena, Germany) as
119 outgroups. We collected larger number of samples of *S. gigantea* compared to other outgroup
120 species to determine whether they have hybridized with *S. altissima* since they co-occur with
121 *S. altissima* in the Hokkaido region in Japan.
122
123 DNA extraction and nSSR genotyping
124 Total genomic DNA was extracted from 1.0 cm^2 of plant tissue using a modified
125 CTAB (cetyltrimethyl ammonium bromide) method (Milligan 1992). Fifteen nSSR markers
126 out of 16 markers developed for *S. altissima* except for Salt 9 (Sakata et al. 2013b) were
127 scored in all samples (N = 713; Table S1). The nSSR loci were polymerase chain reaction
128 (PCR) amplified following Sakata et al. (2013b) and loaded on an ABI Prism 3130 Genetic
129 Analyzer (Applied Biosystems Foster City, California, USA), and scored using GeneMapper
130 (Applied Biosystems). Of the 15 nSSR primer pairs, two loci (Salt 6 and 7) did not amplify
131 well on the diploid population (CA) and *S. gigantea*. Therefore, the two loci were excluded
132 from the analysis of outgroups and different ploidy level populations.
133
134 Chloroplast DNA sequencing
135 First, we screened cpDNA variation in intergenic spacers with 20 sets of PCR

primers from which six: psbH-psbB, rps12-rpl20 (Shaw et al. 2005), psbJ-petA, rpl-trnL, rps16-trnK, and trnQ-rps16 (Shaw et al. 2007) were selected due to amplification and variable sites and were sequenced for 142 samples (i.e., four samples per population and two samples per outgroup species). PCRs were performed on 20 μL samples containing 50 ng of template DNA, 2 μL of 10×PCR buffer, 1.6 μL of 2.5 mM dNTP, 0.1 μL of 50 μM each primer and 0.5 unit of TaKaRa Ex Taq™ (TaKaRa, Shiga, Japan). The PCR cycle for all six fragments was as follows: template denaturation at 94 °C for 3 min, followed by 35 cycles of denaturation at 95 °C for 1 min, annealing at 50 °C for 1.5 min and extension at 72 °C for 1.5 min; followed by a final extension of 72 °C for 7 min. PCR products were sequenced using reverse primers with ABI Prism BIGDYE Terminator Cycle Sequencing Ready Reaction kit v. 3.1 (Applied Biosystems), and electrophoresed on an ABI Prism 3130 Genetic Analyzer (Applied Biosystems).

Genetic diversity, differentiation, and demographic analysis

For nSSR analysis, it was not possible to determine genotypes and allele frequencies from peak heights due to the high ploidy level. Therefore, only genetic diversity statistics that are not affected by genotype ambiguity were calculated. The number of maximum alleles per locus (*A'*), Shannon diversity index (*Shannon*), and genetic differentiation statistics G'_{ST} (Hedrick 2005) were calculated at each nSSR locus using the software GenoDive 2.0b19 (Meirmans and Van Tienderen 2004) and POLYSAT v. 1.3-2 (Clark and Jasieniuk 2011) in R v. 3.0.1 (R Development Core Team 2013). To measure the level of within-population genetic diversity, we calculated the following statistics: number of alleles per locus (N_a), gametic heterozygosity (H_o) (Moody et al. 1993), and Shannon diversity index (*Shannon*). Genetic differentiation statistics: G_{ST} (Nei 1973), G'_{ST} (Hedrick 2005), and D_{EST} (Jost 2008) were calculated per population per loci using GenoDive 2.0b19.

For cpDNA sequence, the number of haplotypes (*H*), haplotype richness (H_R) and G_{ST} were calculated using CONTRIB v. 1.02 (Petit et al. 1998).

To assess whether populations in the native and invaded regions experienced past population expansions, mismatch distribution analysis (Rogers and Harpending 1992) was performed using Arlequin v. 3.11 (Excoffier et al. 2005). This analysis detects past population expansions and declines by measuring the signature of molecular changes (i.e. the frequency distribution of pairwise nucleotide or restriction site differences) that follow population fluctuations. The sum of square deviations between the observed and expected mismatch distributions and the raggedness index of the observed distribution were used as statistics to validate fit of the models (Rogers and Harpending 1992; Harpending 1994).

We calculated pairwise genetic differences between individuals averaged over loci with POLYSAT (Clark and Jasieniuk 2011) using two distance measures appropriate for polyploids: a band-sharing dissimilarity index (one minus the similarity index in eqn 1 of Lynch (1990)) and a measure taking into account mutational distance between microsatellite alleles (Bruvo et al. 2004). Because the hexaploid *S. altissima* is considered an autopolyploid (J. Semple personal communication), we calculated both Bruvo and Lynch distance matrices. We assessed the hierarchical partitioning of genetic variance within and among populations and between the two regions (native and invaded regions) by performing analysis of molecular variance (AMOVA) (Excoffier et al. 1992) on the basis of the two distance matrices using GenAlEx 6.4 (Peakall and Smouse 2006). The same distance matrices were used to estimate pairwise Φpt, an analogue of *F*st (Weir and Cockerham 1984), between populations. Pairwise Φpt were estimated using GenAlEx 6.4 (Peakall and Smouse 2006). We also estimated pairwise Φpt using the cpDNA sequence data. We tested for correlations between geographical and genetic distance [Φpt/(1-Φpt)] following a classical isolation by a distance method in each region using Mantel tests performed with GenAlEx 6.4 (Peakall and

186 Smouse 2006).

187

188 Population genetic structure

189 Both flow cytometry analysis and nSSR genotyping indicated that the outgroup
190 population CA (California, N°39.43, W°120.24) consists of diploid plants and population
191 EF (Oklahoma) consists of tetraploid plants, and that all other populations were hexaploid. To
192 clarify the genetic relationship with other taxa, we conducted a PCoA analysis based on
193 Bruvo genetic distance matrices among outgroups (*S. gigantea* 4X and *S. canadensis* 2X) and
194 the different ploidy populations. In the PCoA analysis, we included five hexaploid
195 populations (SN, KRF, CG, I20, PAa), including the same geographic areas as outgroups.
196 Because our focus was on the invasion history of *S. altissima* in Japan, we only used the
197 hexaploid populations for the rest of the population analyses with nSSR.

198 Population structure based on the nSSR markers was examined by both distance
199 and model-based methods. We increased the probability of obtaining correct phylogenetic
200 tree topology (Takezaki and Nei 1996) using pairwise genetic distances of Nei's D_A (Nei et al.
201 1983), calculated among the 32 populations to construct a neighbor-joining tree (Saitou and
202 Nei 1987) using package APE (Paradis et al. 2004) in R. The application of a simple
203 branching tree model to data, however, can be problematic when systematic errors, such as
204 inappropriate assumptions in the evolutionary model and sampling errors resulting from
205 small numbers of observed loci, exists (Huson and Bryant 2006). Thus, to explore and
206 graphically present these ambiguities, we also constructed a split network (Bryant and
207 Moulton 2004) on the basis of a distance matrix of Nei's D_A (Nei et al. 1983) using
208 SplitsTree4 v. 4.10 (Huson and Bryant 2006). We also used the recently developed
209 Discriminant Analysis of Principal Components (DAPC), which is a multivariate analysis
210 that describes clusters of genetically related individuals (Jombart et al. 2010) using adegenet

v. 1.3-4 (Jombart and Ahmed 2011) in R using presence–absence (binary) genetic data. To find an optimal number of clusters in our data, we used k-means clustering of the principal components and calculated the statistical fit of the data for a given k, using the function find.clusters in adegenet. The optimal number of clusters in the data was determined using the diffNgroup option, which identifies sharp changes in the fit of models (measured using Bayesian Information Criterion) with different numbers of clusters. We used 10^6 iterations of the model to search for convergence and obtained the likelihood associated with each value of k between 1 and 20. Finally, PCoA analysis using a pairwise *Φ*pt matrix obtained from two distance matrices (Bruvo and Lynch) was conducted with GenAlEx 6.4 (Peakall and Smouse 2006). This method produces a few axes containing most of the genetic variation in the data set and separates the populations.

Genetic structure in native and invaded populations was investigated using two model-based Bayesian algorithms implemented in STRUCTURE 2.3 and TESS 2.3.1 (Pritchard et al. 2000; Durand et al. 2009). The STRUCTURE analysis aims to cluster individuals in *K* genetic groups, using the multilocus genotypes of individuals. We performed ten independent runs with different proposals for *K*, testing each possible *K* from 2 to 10 using 100,000 iterations after a burn-in period of 50, 000 iterations. All runs were conducted with the admixture model, assuming correlated allele frequencies (Pritchard et al., 2000; Falush et al., 2003) with prior information on the sampling location (Hubisz et al. 2009). To ensure convergence of the Markov Chain Monte Carlo estimates, the consistency of results was checked for the ten replicates performed for each value of *K*. The most probable number of clusters (*K*) was then determined using the change in log likelihood of the data between successive values of *K*, as described in Evanno et al. (2005). Parameters in the method of Evanno et al. (2005) were calculated using the program Structure Harvester v. 6.0 (Earl and vonHoldt 2012). Population genetic structure was also estimated using a spatial hierarchical

Bayesian algorithm implemented in TESS 2.3.1, which includes spatial prior distributions on the individual admixture proportions (Durand et al. 2009). We estimated the population structure within the samples using this algorithm by incorporating individual geographic covariates in the prior distributions on the admixture memberships. Ten individual simulations were run for each *K*max (*K*max 2–10), with 10,000 burn-in steps followed by 20,000 Markov Chain Monte Carlo steps.

To identify genetic groups, for cpDNA sequence data, a spatial analysis of molecular variance (SAMOVA) was performed using SAMOVA 1.0 (Dupanloup et al. 2002). Based on a simulated annealing procedure, SAMOVA algorithm iteratively seeks the composition of a user-defined number of groups (*K*) of geographically adjacent populations that maximizes the proportion of total genetic variance (F_{CT}) as a result of the differences between groups of populations. We set the number of initial condition to 100 with $K = 2–10$.

Phylogenetic relationships

The chloroplast sequence data were edited and aligned using BIOEDIT v. 7.0.8.0 (Hall 1999) and GENEIOUS PRO, version 5.4.6 (Drummond et al. 2011). Variable sites are listed in Table S2. Phylogenetic relationships between cpDNA haplotypes were assessed using a median-joining network with *Solidago virgaurea*, *Solidago gigantea*, and *Solidago canadensis* as outgroups using NETWORK v. 4.6.0.0 (Bandelt et al. 1999). In addition, phylogenetic relationships between cpDNA haplotypes were assessed with MEGA version 3.1 (Kumar et al. 2004), with conducting 1,000 bootstrap replicates to test the robustness of clades in maximum-parsimony trees. We ran analyses excluding and including indels and mononucleotide repeat length variations and found that though the resolution did not change the inclusion of indels resulted in complicated haplotype networks and the phylogenetic trees that were difficult to interpret. We will report data from analyses excluding indels.

Results

Genetic diversity and differentiation

The number of alleles observed in the native and invaded range per locus varied 9 to 29 (18.3 on average) and 7 to 31 (15 on average), respectively (Table 1). Out of the 306 different alleles found across the whole data set, 69 were unique to the native and 19 to the invaded range (Table 2). In the population genetic analysis, we used only the values averaged for all loci (Table S1). The invaded populations had lower mean number of alleles per locus within populations and a lower Shannon diversity index, whereas the observed heterozygosity was higher than the native populations. The genetic differentiation was higher in the invaded range (G_{ST} = 0.03) than the native range (G_{ST} = 0.011, which was also true with alternative estimators of genetic differentiation G'_{ST} and D_{EST} (Table 2).

The concatenated cpDNA sequence had a length of 4417 bp with 21 substitutions (Table S2). Among the 32 haplotypes found in the total data set, seven haplotypes were found in outgroups (H6 and H10: *S. virgaurea*, H7: *S. gigantea*, H19 and H20: *S. canadensis*, H31: diploid, H12: tetraploid). There were five haplotypes shared with both native and introduced ranges (H9, H24, H25, H26, and H30), two haplotypes unique to the introduced range (H5 and H23), and 20 haplotypes unique to the native range (Table 2, Fig. 1). Overall genetic differentiation was low in both ranges: G_{ST} = 0.24 and 0.27 in the native and invaded range, respectively (Table 2). Haplotype richness was smaller in the invaded range (Table 2: 3.15 and 2.54 in native and invaded range, respectively).

Analysis of mismatch distributions, which shows a large increase between 0_0 and 0_1 (Table 2) revealed support for recent expansion in both ranges. Non-significant sum of square deviations (SSD) and the raggedness index (HRI) of the observed distribution indices were obtained for both ranges ($P > 0.05$), while both indices and observed value of time since

286 divergence (tau) in the invaded region showed lower values than those in the native region
287 (Table 2).

288 The AMOVA analysis showed little genetic differentiation among individuals
289 between the two ranges (4 and 5%), while there was a large genetic differentiation among
290 individuals within populations (85 and 82%). The genetic divergence between the invasive
291 and native groups was significant (Table 3).

292 We found a weak but significant association between genetic and geographic
293 distance with the cpDNA markers in the native range (Fig. S1a: $R^2 = 0.02$, $P < 0.001$), but a
294 non-significant association in the invaded range (Fig. S1b: $R^2 = 0.002$, $P = 0.42$). With the
295 two genetic distance matrices obtained with the nSSR analysis, significant association
296 between genetic and geographic distance was found for both ranges (Fig. S1c-f; Native: $R^2 =$
297 0.15 for Bruvo, and 0.16 for Lynch, $P < 0.001$ for both; Invaded: $R^2 = 0.17$ for Bruvo, and
298 0.26 for Lynch, $P < 0.001$ for both), indicating genetic isolation by geographic distance.

299

300 Population genetic structure and phylogenetic relationships

301 The PCoA analysis, including other taxa that are outgroups and the different ploidy
302 populations showed that all the taxa were clustered separately from each other, except for the
303 tetraploid *S. altissima*, which partly overlapped with the cluster of the hexaploid *S. altissima*
304 (Fig. S2). The neighbor-joining tree based on the genetic distance of Nei's D_A (Nei et al.
305 1983) showed distinct differentiation between the native and invaded populations, with the
306 exception of the Hokkaido (SN) population that was rather closely related to native
307 populations (Fig. 2). In the native range, the most closely related populations to the invaded
308 range were the southern populations (LSa, LSb, FLa, FLb). The Midwest USA populations
309 (CL, PE, CG, I20, HIL) were genetically differentiated from other USA populations (Fig. 2).
310 Korean population (KY) was strongly genetically differentiated from the Japanese

311 populations as indicated by length of this terminal clade estimated to be much longer than
312 that of other populations. The population network included many boxes where more than one
313 split and was distinctly non-tree-like (Fig. S3), meaning ambiguity in applying a simple tree
314 model to the data, although the genetic relationship between populations was almost the same
315 as those shown in the neighbor-joining tree. The PCoA analysis appeared to reflect a native–
316 invaded differentiation pattern on the first axis, and a north–south differentiation in the
317 invaded range in the second axis. The southern populations (LSa, LSb, FLa, FLb) in the
318 native range were most closely associated with the invaded populations (Fig. S4), in
319 agreement with the neighbor-joining tree and the population network analyses.
320 The STRUCTURE analysis showed a clear peak in the modal value of ΔK at $K = 5$
321 as determined by the method of Evanno et al. (2005), indicating that the most likely number
322 of ancestral gene pools is $K = 5$. A spatial hierarchical Bayesian clustering TESS analysis and
323 the DAPC analysis produced five genetic clusters similar to the genetic clustering pattern of
324 populations with STRUCTURE analysis, although the extent of genetic admixture between
325 populations appeared to be slightly larger than estimated by STRUCTURE analysis (Fig. S5).
326 It suggests that it is unlikely that there was any spurious clustering in the STRUCTURE
327 analysis caused by spatial autocorrelation. The red cluster was dominant in all Japanese
328 populations, except for Hokkaido (SN) and the two other northern Japanese populations.
329 These three populations had very different compositions that characterized by an admixture
330 of more than three gene pools (Fig 3b). The population in Korea (SY) was a distinct cluster.
331 Native USA populations varied with the green cluster being dominant in the central Midwest
332 populations (CL, PE, CG), while the northern Midwest (HB, FB) and Eastern populations
333 exhibited a highly admixed composition dominated by the purple cluster (Fig. 3b). Only the
334 southern USA populations (LSa, LSb, FLa) and populations in the Atlantic coast (PAa, PAb,
335 MD, VS) had a significant proportion of the red cluster.

336 The phylogenetic structure among *Solidago* species investigated was shallow, and
337 haplotypes were not grouped to species monophyletically (Figs. 1 and S6). The most
338 frequently observed haplotype (H25) was the most common haplotype in both the invaded
339 range and the native range. All haplotypes observed in the invaded range were closely related
340 to H25 except for H5 (Figs. 1 and S6). In the SAMOVA analysis, the F_{CT} value was highest
341 when the number of population groups was defined as three, and included the diploid
342 population (CA) and the tetraploid population (EF) which were detected as a separate group.
343 When these two populations were excluded, the F_{CT} value was highest when the number of
344 population groups was defined as five, and continued to decrease as the number of groups
345 increased. The five groups included two single-population groups and two groups containing
346 two populations, and a large group containing the rest of the populations. However, these
347 groups had no biological consequence.
348
349 **Discussion**
350 Relationships among species and ploidy levels
351 The genus *Solidago* is known for complex taxonomy and clear differentiation of the species
352 is further complicated, not only by the occurrence of interspecific hybrids, but also because
353 many species are polytopic with overlapping ranges, and several species have multiple
354 cytotypes (Semple and Cook 2006). A recent phylogenetic study of *Solidago* species has
355 revealed a single origin involving reticulate evolution and introgression in an allohexaploid *S.*
356 *houghtonii* (Laureto and Barkman 2011). While Laureto and Barkman (2011) resolved
357 relationships among Triplinerviae species using chloroplast DNA sequence data, Schlaepfer
358 et al. (2008) could not distinguish between *S. gigantea* and *S. canadensis*, and found that
359 some haplotypes were shared with more than one species. Similarly, our results of the
360 cpDNA sequence data could not resolve relationships among *S. altissima*, *S. gigantea*, and *S.*

canadensis. This was consistent even when we included the indels in the DNA sequence for the analysis. We cannot deny the possibility of either incomplete lineage sorting or chloroplast capture by secondary contact, both of which are consistent with recent speciation (Twyford and Ennos 2012). On the other hand, using microsatellite markers, three species were clustered separately and were shown to be genetically distinct. These high-resolution markers can be used to distinguish among polyploid taxa, except for tetraploids and hexaploids.

Genetic diversity and structure in the native range

In the native range, we found high genetic diversity in nSSR markers within populations (Table S1). The genetic structure of nSSR data analyses showed that Central Midwest populations (CL, PE, CG, I20, HIL) were genetically differentiated from the rest of the populations (Figs. 2 and 3b). This could be explained by one of the most commonly observed phylogenetic breaks in Eastern North America, which is between the two sides of the Mississippi Valley (Soltis et al. 2006; Jaramillo-Correa et al. 2009). In addition, the genetic differentiation between HB, FB and the other Midwest populations is likely to reflect the southern limit of the Laurentide Ice Sheet. However, since most of our samples are hexaploid plants and include only a few other ploidy plants, we need more diploid and tetraploid samples from other geographic range, to examine the evolutionary history of *S. altissima* in North America. Although the closely related species *S. gigantea* showed the possibility of glacial survival in different refugial areas and separate migration routes on opposite sides of the Appalachian Mountains (Schlaepfer et al. 2008), we did not find any genetic discontinuity in the Appalachian populations. The phylogeographic study of the *Solidago* subsect., *Humiles* suggested Holocene polyploid speciation supported by restriction of endemic polyploid taxa to post-glacial habitats (Peirson et al. 2013). In line with these

findings, the cpDNA sequence data showed no pattern in genetic structure among hexaploid populations and the mismatch distribution analysis suggested rapid expansion of the populations (Table 2, Fig. 3a). The significant isolation by distance in both nSSR and cpDNA markers (Fig. S1) shows that genetic differentiation is influenced by geography in North America, and that most subsequent gene flow seems mediated through pollen, as indicated by the much stronger among-populational differentiation at cpDNA (only dispersed by seeds) than nDNA (dispersed by both seeds and pollen) markers (Table 2). Alternatively, this could simply reflect the difference in the effective population size between cpDNA and nDNA markers. From the results of the mismatch distribution analysis and the fact that the hexaploid *S. altissima* is likely to have multiple origins from different diploid lineages (Halverson et al. 2008), the hexaploid *S. altissima* in North America would have expanded its range recently and the genetic structure been largely shaped through post-glacial migration and gene flow.

Multiple and mass introductions in the invaded range

Many studies have documented that the founding population reduces genetic variation within population, relative to the source population, because of a reduction in population size during colonization (Henry et al. 2009; Okada et al. 2007; Dybdahl and Drown 2011). However, there is increasing evidence for similar or even higher levels of within-population genetic diversity in exotic populations. This is typically explained by high propagule pressure, multiple introductions or admixture in the introduced range of individuals from different sources (Bossdorf et al. 2005; Dlugosch and Parker 2008; Kelagar et al. 2013).

The low haplotype diversity and a few private alleles indicate that introduced populations came from a single or a few independent origins (Table 2 and Fig. 1). Moreover, the introduced populations with the red cluster representing 68% of the introduced sites in

411 Japan are genetically similar and likely share a common origin, indicating that a single or a
412 few populations were particularly successful in colonizing the invaded range. This can also
413 be suggested as a result of post-invasion selection of genotypes adapted in invaded regions.
414 This scenario has been confirmed in many successful invasive species (Lee 2002; Lombaert
415 et al. 2010). Despite the small numbers of private alleles in the invaded range, the difference
416 between introduced and native populations was significant in the genetic structure analysis
417 (Figs. 2, S2, and S4). This may be caused by dissimilarities in allele frequencies rather than
418 allelic identities. Hence, it is likely that high propagule pressure at primary establishment (i.e.,
419 mass introductions from few sources) and substantial geneflow through obligate outcrossing
420 and production of large amounts of wind dispersed seeds have co-founded the widely
421 distributed red cluster. This is also indicated in the Korean SY population with the high F_{ST}
422 value of the yellow cluster in STRUCTURE analysis. Although the SY population has
423 genetically differentiated from the other invaded populations, it was most closely related to
424 the northern Honshu populations (Figs. 2 and 3). We speculate that the SY population
425 originated from a source in the northern region of Japan, rather than a separate colonization
426 event from the native source.

427 Our study infers multiple colonization with at least two primary colonization events
428 of *S. altisssima* into Japan. The Hokkaido population (SN) was clustered with the native
429 populations (Fig. 2) and thus was genetically differentiated from the other introduced
430 populations in the population structure analyses (Fig. S5). Fukuda (1982) notes that the
431 Hokkaido populations have been established more recently in Japan compared to other
432 regions. The populations in Hokkaido are more likely derived from an independent
433 colonization event rather than a range expansion from the Honshu populations. In addition,
434 the northern Honshu region showed higher admixture and higher genetic diversity than
435 southern Japan populations (Table S1 and Fig. 3). These results indicate that secondary

colonization and gene flow (shown in Honshu and Hokkaido clusters) have contributed to the maintenance of high genetic diversity within populations in the invaded range.

The results of the genetic structure analyses in nSSR showed congruent patterns by different methods (neighbor-joining tree, population network analysis and PCoA analysis), which showed that the introduced populations dominated by the red cluster were genetically most related to the South region in North America (LSa, LSb, FLa, FLb), where the highest genetic diversity was observed (Table S1, S2). However, populations in the Atlantic coast region (PAa, PAb, MD, VS) are also related to populations in the invaded range. Moreover, the cpDNA analysis showed that only two haplotypes (H9 and H25) were shared between the invaded region and South region in North America, while four haplotypes (H9, H24, H25 and H26) were shared with the central Midwest populations. Larger sample size for cpDNA analysis, and analysis with higher resolution markers are necessary to identify the origin of the Japanese populations.

Spatial genetic structure in the invaded range and invasive spread during invasion

While the isolation by distance was significant for both cpDNA and nSSR data in the native range, it was only significant for nSSR data in the invaded range. The population differentiation was higher in the invaded range than in the native range for nSSR data (Table 2), but when the Korean (SY) and the Hokkaido (SN) populations were excluded they showed similar low values in both ranges (G_{ST}, 0.011 vs. 0.014; G'_{ST}, 0.052 vs. 0.061; D_{EST}, 0.041 vs. 0.048 in native and invaded range, respectively). There are two possible explanations. First, the multiple introductions from distinct native sources to different regions in Japan (i.e., Hokkaido) may account for the larger population differentiation in the invaded region than in the native range. Second, substantial gene flow has been present and/or genetic drift has been weak in the invaded range. High admixture in the northern Honshu region also

461 suggests this. Gene flow is promoted by the fact that *S. altissima* invaded along railways and
462 roads, which form a highly connected habitat. In addition, the high ploidal nature of the
463 species may have mitigated the effect of genetic drift (Gaudeul et al. 2011). Asai (1970)
464 reported that Japanese bee-keepers have transplanted the naturalized *S. altissima* as
465 honey-bee food plants. It implies that range expansion in Japan occurred by a series of
466 long-distance human-mediated dispersal commonly found in invasive species (Okada et al.
467 2007; Gaudeul et al. 2011; Kelager et al. 2013).

468 Our reconstruction of *S. altissima*'s invasion history based on historical records and the
469 results of genetic diversity and structure in both range, suggests that the invaded populations
470 arose from introductions from a few sources potentially in the southern part of North
471 America in early 1900s by ornamental planting and commercial exchange. Later, the
472 independent introduction from different native sources to Hokkaido generated admixture
473 between genetic clusters leading to high genetic diversity in the northern region in Japan.
474 Soon after, human-mediated dispersal and the high ploidal nature of the species contributed
475 to the rapid and successful invasion. We summarize that multiple and mass introductions may
476 have contributed to the persistence and rapid adaptation of the species enabling it to be spread
477 throughout Japan. The high genetic diversity maintained within populations in the invaded
478 range is often found in other polyploid plants (Schlaepfer et al. 2008; Hornoy et al. 2013;
479 Vallejo-Marin and Lye 2013). High levels of heterozygosity (and nearly double the effective
480 population size) compared with diploid plants are characterized by autopolyploids as a result
481 of polysomic inheritance (te Beest et al. 2012). Furthermore, the fact that *S. altissima* is
482 perennial may have also increased the effective population size as discussed in other systems
483 (Hornoy et al 2013). Together with the growing number of studies of invasive polyploid
484 plants, our results also emphasize the significance of considering polyploid as an important

trait in invasion models, as polyploids benefit from reduced genetic impact of bottlenecks, and a high evolutionary potential (Lee 2002; te Beest et al. 2012).

Future directions

Determining the mechanisms underlying successful range expansions and rapid adaptive processes including genetic factors, environmental factors, or their interactions is a critical challenge. Closely related *Solidago* species exhibit latitudinal variation in phenology (Weber and Schmid 1998) and enhanced competitive ability (Yuan et al. 2012) in the invaded range. Our previous study on *S. altissima* in Japan showed that plant resistance was rapidly selected when it was re-associated with a recently invaded herbivorous insect, *Corythucha marmorata* from North America (Sakata et al. 2014). The present results support the argument that large genetic variability was introduced and local adaptation is ongoing in the invaded range. Direct ancestor-descendent comparisons of phenotypic traits between native and introduced populations (Kellor and Taylor 2008) will clarify whether the adaptation is a result of historical factors or of evolutionary changes involved in the invasion process (i.e., separating the role of stochastic and demographic events from that of selection in shaping the evolution of phenotypes), which is an important focus in future research.

Knowledge of the genetic variation in *S. altissima* in its native and invaded ranges can aid in the development of effective strategies to manage its spread. Information on the amount and distribution of genetic diversity can help to predict its response to chemical and biological control by considering the impacts of genetic variability on the interactions of the invasive species and the biological agents (Garcia-Rossi et al. 2003).

Acknowledgment

We thank H. Choi, M. Ikemoto, K. Shiojiri, A. Uesugi, W. Licht, C. Hafdahl, C. Sacchi, J. Cronin, and K. Dixon for their great help in collecting plant materials. We are indebted to T. Craig for help with site selection for plant sampling and manuscript revision. We are grateful to M. Yamasaki for assistance with figure preparation and S. Sakaguchi for insightful comments on data analysis and manuscript improvement. We thank the two reviewers for invaluable comments and suggestions. This work was partly supported by JSPS Core-to-Core Program (No. 20004) from Japan Society for the Promotion of Science and the National Science Foundation grant (DEB 0949280).

References

Asai Y (1970) On the *Solidago canadensis* group widely spread by bee-keepers in Japan. J Jap Bot 45:82-83 (in Japanese)

Bandelt H, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37-48

Bossdorf O, Auge H, Lafuma L, Rogers W, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. Oecologia 144:1-11

Bruvo R, Michiels NK, D'Souza TG, Schulenburg H (2004) A simple method for the calculation of microsatellite genotype distances irrespective of ploidy level. Mol Ecol 13:2101-2106

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255-265

Clark LV, Jasieniuk M (2011) POLYSAT: an R package for polyploid microsatellite analysis. Mol Ecol Resour 11:562-566

533 Dlugosch K, Parker I (2008) Founding events in species invasions: genetic variation,
534 adaptive evolution, and the role of multiple introductions. Mol Ecol 17:431-449
535 Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J,
536 Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A
537 (2011) *Geneious*, Version 5.4. New Zealand: Biomatters Ltd. Available at:
538 http://www.geneious.com/
539 Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the
540 genetic structure of populations. Mol Ecol 11:2571-2581
541 Durand E, Jay F, Gaggiotti O, Francois O (2009) Spatial inference of admixture proportions
542 and secondary contact zones. Mol Biol Evol 26:1963-1973
543 Dybdahl M, Drown D (2011) The absence of genotypic diversity in a successful
544 parthenogenetic invader. Biol Invasions 13:1663-1672
545 Earl DA, vonHoldt BM. (2012) STRUCTURE HARVESTER: a website and program for
546 visualizing STRUCTURE output and implementing the Evanno method. Conserv
547 Genet Resour 4:359-361
548 Ellstrand N, Schierenbeck K (2000) Hybridization as a stimulus for the evolution of
549 invasiveness in plants? Proc Natl Acad Sci USA 97:7043-7050
550 Estoup A, Guillemaud T (2010) Reconstructing routes of invasion using genetic data: why,
551 how and so what? Mol Ecol 19:4113-4130
552 Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using
553 the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620
554 Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software
555 package for population genetics data analysis. Evol Bioinform 1:47-50

556 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from
557 metric distances among DNA haplotypes - application to human mitochondrial-DNA
558 restriction data. Genetics 131:479-491
559 Falush D, Stephens M, Pritchard J (2003) Inference of population structure using multilocus
560 genotype data: linked loci and correlated allele frequencies. Genetics 164:1567-1587
561 Fitzpatrick B, Fordyce J, Niemiller M, Reynolds RG (2012) What can DNA tell us about
562 biological invasions? Biol Invasions 14:245-253
563 Fukuda I (1982) Distributions and population structures of North American plants, *Solidago*
564 *altissima* L. and *S. gigantea* AIT., introduced into the Japanese islands. Science
565 reports of Tokyo Woman's Christian University 32: 675-690 (in Japanese)
566 Garcia-Rossi D, Rank N, Strong D (2003) Potential for self-defeating biological control?
567 Variation in herbivore vulnerability among invasive *Spartina* genotypes. Ecol Appl
568 13:1640-1649
569 Gaudeul M, Giraud T, Kiss L, Shykoff J (2011) Nuclear and chloroplast microsatellites show
570 multiple introductions in the worldwide invasion history of common ragweed,
571 *Ambrosia artemisiifolia*. PLoS ONE 6:e17658
572 Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis
573 program for Windows 95/98/NT. Nucleic Acids Symposium Series, 41:95–98.
574 Halverson K, Heard SB, Nason JD, Stireman JO (2008) Origins, distribution, and local
575 co-occurrence of polyploid cytotypes in *Solidago altissima* (Asteraceae). Am J Bot
576 95:50-58
577 Handley L, Estoup A, Evans D, Thomas, CE, Lombaert E, Facon B, Aebi A, Roy HE (2011)
578 Ecological genetics of invasive alien species. Biocontrol 56:409-428
579 Harpending H (1994) Signature of ancient population-growth in a low-resolution
580 mitocondrial-DNA mismatch distribution. Hum Biol 66:591-600

581 Hedrick PW (2005) A standardized genetic differentiation measure. Evolution 59:1633-1638
582 Henry P, Le Lay G, Goudet J, Guisan A, Jahodová S, Besnard G (2009) Reduced genetic
583 diversity, increased isolation and multiple introductions of invasive giant hogweed in
584 the western Swiss Alps. Mol Ecol 18:2819-2831
585 Hornoy B, Atlan A, Roussel V, Buckley YM, Tarayre M (2013) Two colonisation stages
586 generate two different patterns of genetic diversity within native and invasive ranges
587 of *Ulex europaeus*. Heredity 111:355-363
588 Huang H. Guo. S. (2004). Review on ecological studies on three invasive species of
589 European genus *Solidago*. Guangxi Sciences 11:69-74
590 Hubisz M, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure
591 with the assistance of sample group information. Mol Ecol Resour 9:1322-1332
592 Huson D, Bryant D (2006) Application of phylogenetic networks in evolutionary studies.
593 Mol Biol Evol 23:254-267
594 Ito I, Kobayashi K, Yoneyama T (1998) Fate of dehydromatricaria ester added to soil and its
595 implications for the allelopathic effect of *Solidago altissima* L. Ann Bot-London
596 82:625-630
597 Jaramillo-Correa JP, Beaulieu J, Khasa DP, Bousquet J (2009) Inferring the past from the
598 present phylogeographic structure of North American forest trees: seeing the forest
599 for the genes. Can J Forest Res 39:286–307
600 Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP
601 data. Bioinformatics 27:3070-3071
602 Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a
603 new method for the analysis of genetically structured populations. BMC Genet 11:94
604 Jost, L (2008). G_{ST} and its relatives do not measure differentiation. Mol Ecol 17:4015–4026

605 Kelager A, Pedersen J, Bruun H (2013) Multiple introductions and no loss of genetic
606 diversity: invasion history of Japanese rose, *Rosa rugosa*, in Europe. Biol Invasions
607 15:1125-1141
608 Keller SR, Taylor DR (2008) History, chance and adaptation during biological invasion:
609 separating stochastic phenotypic evolution from response to selection. Ecol Lett
610 11:852-866
611 Kil JH, Shim KC, Park SH, Koh KS, Suh MH, Ku YB, Suh SU, Oh HK, Kong HY
612 (2004) Distributions of naturalized alien plants in South Korea. Weed Technol
613 18:1493-1495
614 Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary
615 genetics analysis and sequence alignment. Brief Bioinform 5:150-163
616 Lambrinos JG (2004) How interactions between ecology and evolution influence
617 contemporary invasion dynamics. Ecology 85:2061-2070
618 Laureto P, Barkman T (2011) Nuclear and chloroplast DNA suggest a complex single origin
619 for the threatened allopolyploid *Solidago houghtonii* (Asteraceae) involving reticulate
620 evolution and introgression. Syst Bot 36:209-226
621 Lavergne S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive
622 the success of an invasive grass. Proc Natl Acad Sci USA 104:3883-3888
623 Lee C (2002) Evolutionary genetics of invasive species. Trends Ecol Evol 17:386-391
624 Li HL (1978). Compositae. Flora of Taiwan, Epoch Publishing Co, Taipei, Taiwan, 4: 768–
625 965
626 Lombaert E, Guillemaud T, Cornuet J, Malausa T, Facon B, Estoup A (2010) Bridgehead
627 effect in the worldwide invasion of the biocontrol harlequin ladybird. PLoS ONE
628 5:e9743
629 Lynch M (1990) The similarity index and DNA fingerprinting. Mol Biol Evol 7:478-484

630 Meirmans PG and Van Tienderen PH (2004) GENOTYPE and GENODIVE: two programs
631 for the analysis of genetic diversity of asexual organisms. Mol Ecol Notes 4:792-794.
632 Melville M, Morton J (1982) A biosystematic study of the *Solidago canadensis* (Compositae)
633 complex. 1. The Ontario populations. Can J Bot 60:976-997
634 Meyer AH, Schmid B (1999) Experimental demography of rhizome populations of
635 establishing clones of *Solidago altissima*. J Ecol 87:42-54
636 Milligan B (1992) Plant DNA isolation. In: Hoelzel AR (ed) Molecular genetic analysis of
637 populations: a practical approach. IRL Press, Oxford, pp 59–88
638 Mitchell CE, Agrawal AA, Bever JD, Gilbert GS, Hufbauer RA, Klironomos JN, Maron JL,
639 Morris WF, Parker IM, Power AG, Seabloom EW, Torchin ME, Vázquez DP (2006)
640 Biotic interactions and plant invasions. Ecol Lett 9:726-740
641 Moody ME, Mueller LD, Soltis DE (1993) Genetic variation and random drift in
642 autotetraploid populations. Genetics 134:649-657
643 Nei M (1973) Analysis of Gene diversity in subdivided populations. Proc Natl Acad Sci USA
644 70:3321-3323
645 Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from
646 molecular-data 2. Gene frequency data. J Mol Evol 19:153-170
647 Okada M, Ahmad R, Jasieniuk M (2007) Microsatellite variation points to local landscape
648 plantings as sources of invasive pampas grass (*Cortaderia selloana*) in California.
649 Mol Ecol 16:4956-4971
650 Paradis E, Claude J, Strimmer K (2004) APE: analyses of phylogenetics and evolution in R
651 language. Bioinformatics 20:289–290
652 Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic
653 software for teaching and research. Mol Ecol Notes 6:288-295

654 Peirson JA, Dick CW, Reznicek AA (2013) Phylogeography and polyploidy evolution of
655 North American goldenrods (*Solidago* subsect. *Humiles*, Asteraceae). J Biogeogr
656 40:1887-1898
657 Petit R, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis
658 of genetic markers. Conserv Biol 12:844-855
659 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
660 multilocus genotype data. Genetics 155:945–959
661 R Development Core Team (2013) R 3.0.1. R project for statistical computing. Vienna,
662 Austria. www.r-project.org
663 Roderick G, Navajas M (2003) Genes in new environments: genetics and evolution in
664 biological control. Nat Rev Genet 4:889-899
665 Rogers A, Harpending H (1992) Population growth makes waves in the distribution of
666 pairwise genetic differences. Mol Biol Evol 9:552-569
667 Saitou N, Nei M (1987) The neighbor joining method - A new method for reconstructing
668 phylogenetic trees. Mol Biol Evol 4:406-425
669 Sakata Y, Ohgushi T, Isagi Y 2013a. Geographic variations in phenotypic traits of the exotic
670 herb *Solidago altissima* and abundance of recent established exotic herbivorous
671 insects. J Plant Interact 8:216-218
672 Sakata Y, Kaneko S, Hayano A, Inoue-Murayama M, Ohgushi T, Isagi Y. 2013b. Isolation
673 and characterization of microsatellite loci in the invasive herb *Solidago altissima*
674 (Asteraceae). Appl Plant Sci 1:1200313
675 Sakata Y, Yamasaki M, Isagi Y, Ohgushi T (2014) An exotic herbivorous insect drives the
676 evolution of resistance in the exotic perennial herb *Solidago altissima*. Ecology
677 95:2569-2578

678 Schlaepfer DR, Edwards PJ, Widmer A, Billeter R (2008) Phylogeography of native ploidy
679 levels and invasive tetraploids of *Solidago gigantea*. Mol Ecol 17:5245-5256
680 Semple JC, Cook RE (2006) In Flora North America Editorial Committee (ed) Flora of North
681 America, vol. 20. Oxford University Press, Oxford, UK, pp 107-166
682 Semple JC, Rahman R, Sbovski S, Sorour MK, Kornobis K, Laphitz RL, Tong L (2015) A
683 multivariate morphometric study of the *Solidago altissima* complex and *S. canadensis*
684 (Asteraceae: Astereae). Phytoneuron 2014-10:1-31
685 Shaw J, Lickey EB, Beck JT, Farmer SB, Liu W, Miller J, Siripun KC, Winder CT, Schilling
686 EE, Small RL (2005) The tortoise and the hare II: relative utility of 21 noncoding
687 chloroplast DNA sequences for phylogenetic analysis. Am J Bot 92:142-166
688 Shaw J, Lickey E, Schilling E, Small RL (2007) Comparison of whole chloroplast genome
689 sequences to choose noncoding regions for phylogenetic studies in angiosperms: the
690 tortoise and the hare III. Am J Bot 94:275-288
691 Shimizu T (2003) Naturalized plants of Japan. In Heibonsha Tokyo, Japan. (in Japanese)
692 Soltis D, Morris A, McLachlan J, Manos PS, Soltis PS (2006) Comparative phylogeography
693 of unglaciated eastern North America. Mol Ecol 15:4261-4293
694 Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from
695 microsatellite DNA. Genetics 144:389-399
696 te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubešová M, Pyšek P (2012)
697 The more the better? The role of polyploidy in facilitating plant invasions. Ann
698 Bot-London 109:19-45
699 Thuiller W, Richardson D, Pysek P, Midgley GF, Hughes GO, Rouget M (2005) Niche-based
700 modelling as a tool for predicting the risk of alien plant invasions at a global scale.
701 Glob Change Biol 11:2234-2250

Twyford A, Ennos R (2012) Next-generation hybridization and introgression. Heredity 108:179-189

Vallejo-Marin M, Lye G (2013) Hybridisation and genetic diversity in introduced *Mimulus* (Phrymaceae). Heredity 110:111-122

Weber E (1997) Phenotypic variation of the introduced perennial *Solidago gigantea* in Europe. Nord J Bot 17:631-638

Weber E, Jakobs G (2005) Biological flora of central Europe: *Solidago gigantea* Aiton. Flora 200:109-118

Weber E, Schmid B (1998) Latitudinal population differentiation in two species of *Solidago* (Asteraceae) introduced into Europe. Am J Bot 85:1110-1121

Weir BS, Cockerham CC (1984) Estimating F statistics for the analysis of population structure. Evolution 38:1358-1370

Yu X, He T, Zhao J, Li Q (2014) Invasion genetics of *Chromolaena odorata* (Asteraceae) extremely low diversity across Asia. Biol Invasions DOI 10.1007/s10530-014-0669-2

Yuan Y, Wang B, Zhang S, Tang J, Tu C, Hu S, Yong JWH, Chen X (2013) Enhanced allelopathy and competitive ability of invasive plant *Solidago canadensis* in its introduced range. J Plant Ecol 6:253-263

Table 1. Characteristics of the 15 microsatellite loci examined in this study.

Locus	Native (*N* = 443)			Invaded (*N* = 270)		
	A'	*Shannon*	*G'*$_{ST}$	*A'*	*Shannon*	*G'*$_{ST}$
Salt1	26	2.66	0.146	21	2.31	0.357
Salt2	19	1.86	0.057	10	1.29	0.085
Salt3	28	2.83	0.128	31	2.49	0.307
Salt4	16	1.83	0.063	12	1.57	0.132
Salt5	29	2.46	0.076	21	2.09	0.164
Salt6	24	2.57	0.074	14	1.98	0.265
Salt7	13	1.29	0.018	13	1.28	0.191
Salt8	16	2.28	0.064	17	2.09	0.084
Salt13	13	2.01	0.025	7	1.78	0.068
Salt14	9	1.61	0.029	9	1.11	0.121
Salt16	15	1.49	0.021	17	1.08	0.083
Salt17	18	1.78	0.05	18	1.58	0.086
Salt18	18	1.95	0.079	12	1.59	0.066
Salt19	12	1.73	0.095	8	1.44	0.053
Salt21	19	2.01	0.048	15	1.72	0.123
Overall	275	2.88	0.052	225	2.67	0.126

A', number of maximum alleles per locus; *Shannon*, Shannon diversity (mean across population); *G'*$_{ST}$, standardized genetic differentiation index (mean across population) (Hedrick 2005).

Table 2.

Genetic diversity of cpDNA (upper half) and nSSR (lower half) analysis, and the results of the mismatch distribution analyses of the native and invaded populations of *S. altissima*.

	Native	Invaded
cpDNA		
H	25 (2X: H31, 4X: H12)	7
H_R	3.15	2.54
G_{ST}	0.238	0.272
SSD	0.087 (0.45)	0.064 (0.26)
HRI	0.401 (0.62)	0.319 (0.34)
tau	1.71	0.94
θ_0	4.2×10^{-3}	2.9×10^{-4}
θ_1	5.8×10^{4}	7.0×10^{4}
nSSR		
N_a	9.68	7.92
A	275	225
*H*o	0.863	0.873
P_R	69	19
Shannon	2.88	2.67
G_{ST}	0.011	0.03
G'_{ST}	0.052	0.13
D_{EST}	0.041	0.099

H, number of haplotypes; $\boldsymbol{H_R}$, haplotypic richness; $\boldsymbol{G_{ST}}$, genetic differentiation index (Nei 1973); **SSD**, the sum of square deviations between the observed and expected distributions (Rogers and Harpending 1992); **HRI**, the raggedness index of the observed distribution

(Harpending 1994); **tau**, the time parameter to the population expansion; $\boldsymbol{\theta_0}$, θ before population expansion; $\boldsymbol{\theta_1}$, θ after population expansion, pairwise difference of population after growth; $\boldsymbol{N_a}$, number of alleles per locus per population; ***A***, total number of alleles, $\boldsymbol{H_o}$, gametic heterozygosity (Moody 1993); $\boldsymbol{P_R}$, number of private alleles, **Shannon**, Shannon diversity; $\boldsymbol{G_{ST}}$, genetic differentiation index (Nei 1973); $\boldsymbol{G'_{ST}}$, standardized genetic differentiation index (Hedrick 2005); $\boldsymbol{D_{EST}}$, standardized genetic differentiation index (Jost 2008).

N_a, H_o, Shannon, G_{ST}, G'_{ST}, and D_{EST} are mean values across loci and population.

Values in parentheses in *H* represent haplotypes of diploid (2X) and tetraploid (4X) samples.

Values in parentheses in SSD and HRI represent *P* values.

777 Table 3. Results of the analysis of molecular variance for both ranges.

Distance measure	Source	d.f.	Variance Components	%	Fixation indices
Bruvo	Among range	1	0.010	4	Φ_{RT} = 0.039*
	Among population	30	0.027	11	Φ_{PR} = 0.113*
	Within population	681	0.212	85	Φ_{PT} = 0.148*
Lynch	Among range	1	0.012	5	Φ_{RT} = 0.047*
	Among population	30	0.035	13	Φ_{PR} = 0.138*
	Within population	681	0.218	82	Φ_{PT} = 0.179*

778

779

780 *P < 0.0001, the probability of obtaining by chance an fixation indices equal to or greater
781 than the observed value, estimated from 999 permutations.

782

Figure legend

Fig. 1

The haplotype network of *Solidago altissima* based on chloroplast DNA sequence variation. Circle size is proportional to haplotype frequency. Each line between haplotypes corresponds to one mutational change. Small black squares indicate missing haplotypes. Haplotypes found in both native and invaded populations are shown in dotted outlines, and haplotypes only found in invaded populations are shown in diamonds, and haplotypes of outgroup species are shown in triangles: *S. virgaurea*, H6 and H10; *S. gigantea*, H7; *S. canadensis*, H19 and H20.

Fig. 2

A neighbor-joining tree summarizing the relationships among hexaploid *S. altissima* populations based on Nei's genetic distance (Nei et al. 1983). * Populations in the invaded range. Abbreviations refer to Table S1.

Fig. 3

(a) Geographic distribution of the 27 chloroplast haplotypes found in *S. altissima* populations. H12 is observed in the tetraploid population EF. Haplotype names correspond to Fig. 1. (b) Geographic distributions of five nSSR gene pools estimated by STRUCTURE analysis (Pritchard et al. 2000). A neighbor-joining tree showing the relationships of each gene pool is depicted. The number of samples analyzed per population is proportional to the circle's size.

Supplementary material

Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range

Yuzu Sakata, Joanne Itami, Yuji Isagi, Takayuki Ohgushi

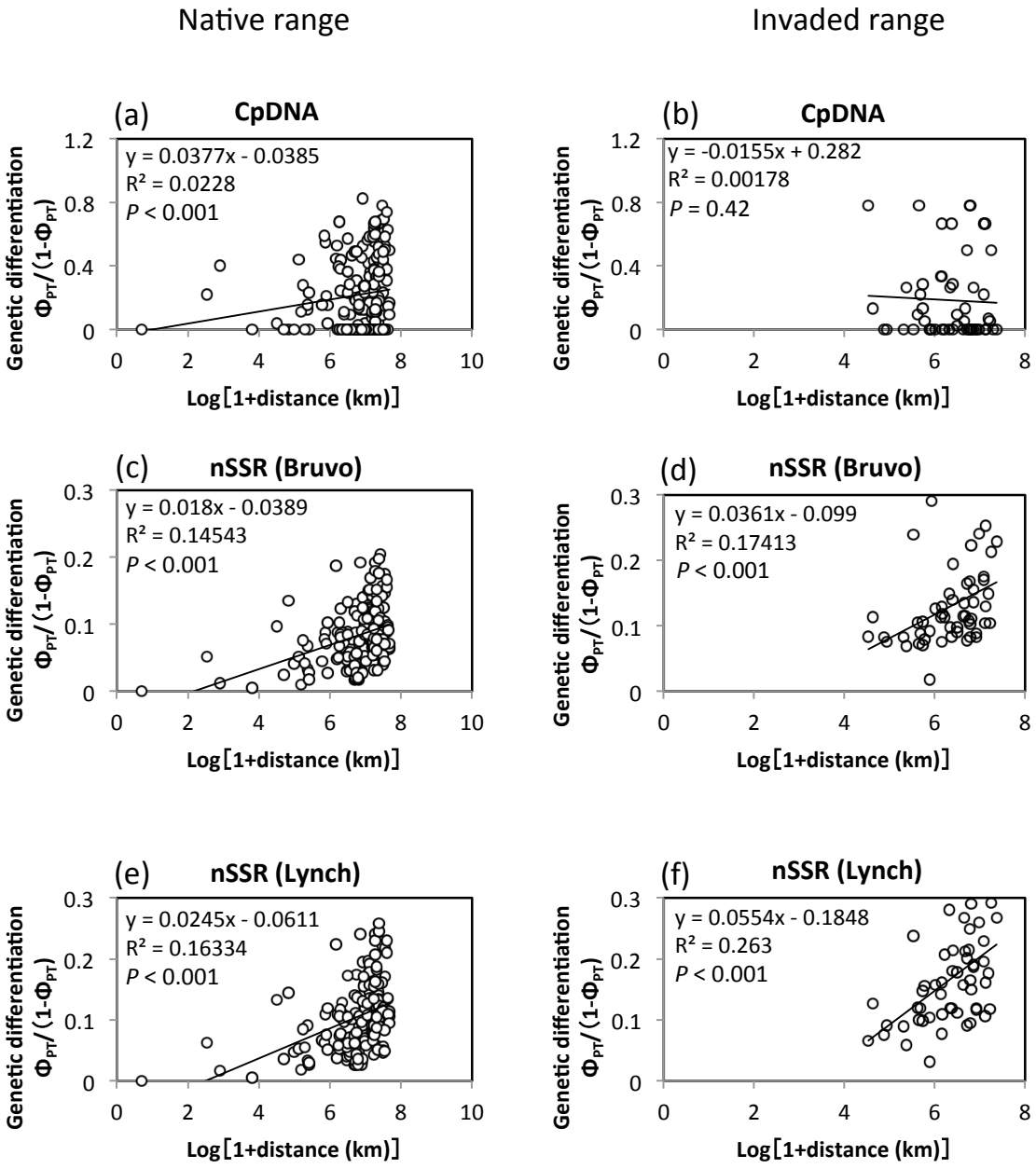

Fig. S1

Relationships between the pairwise genetic and geographic distance matrices obtained from (a), (b) CpDNA; (c), (d) Bruvo distance for nSSR; (e), (f) Lynch distance for nSSR.

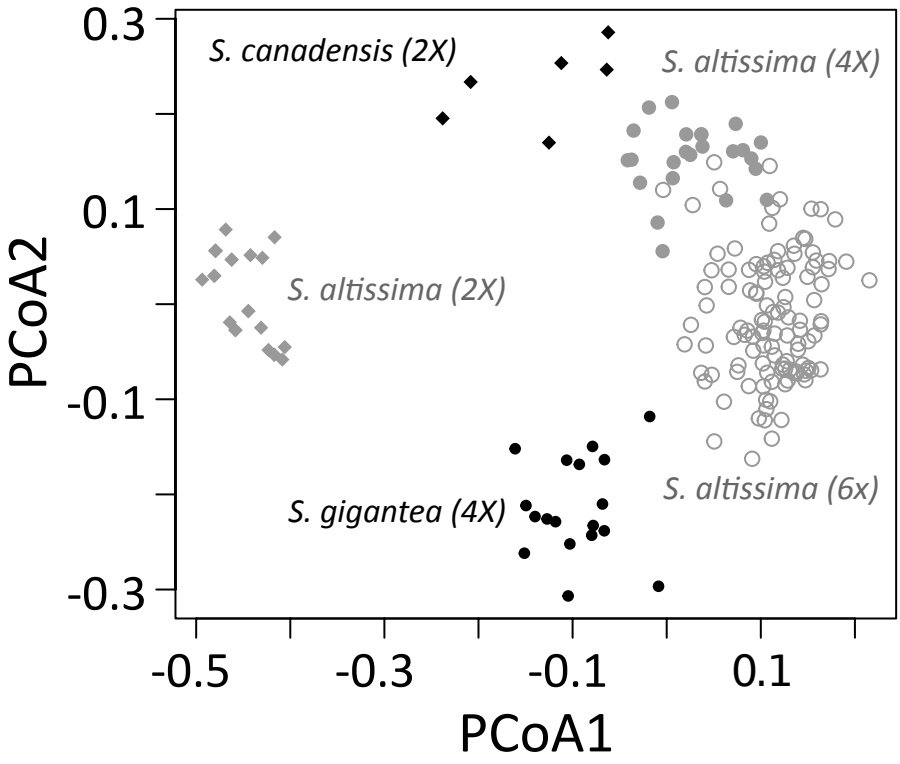

Fig. S2

Principal coordinate analyses of three different ploidy taxa of *S. altissima* (2X, 4X, and 6X) and two closely-related species (*S. gigantea* and *S. canadensis*). The first two principal coordinate axes are shown. Individuals of *S. altissima* are shown in grey, and other species are shown in black. Symbols represent different ploidy levels.

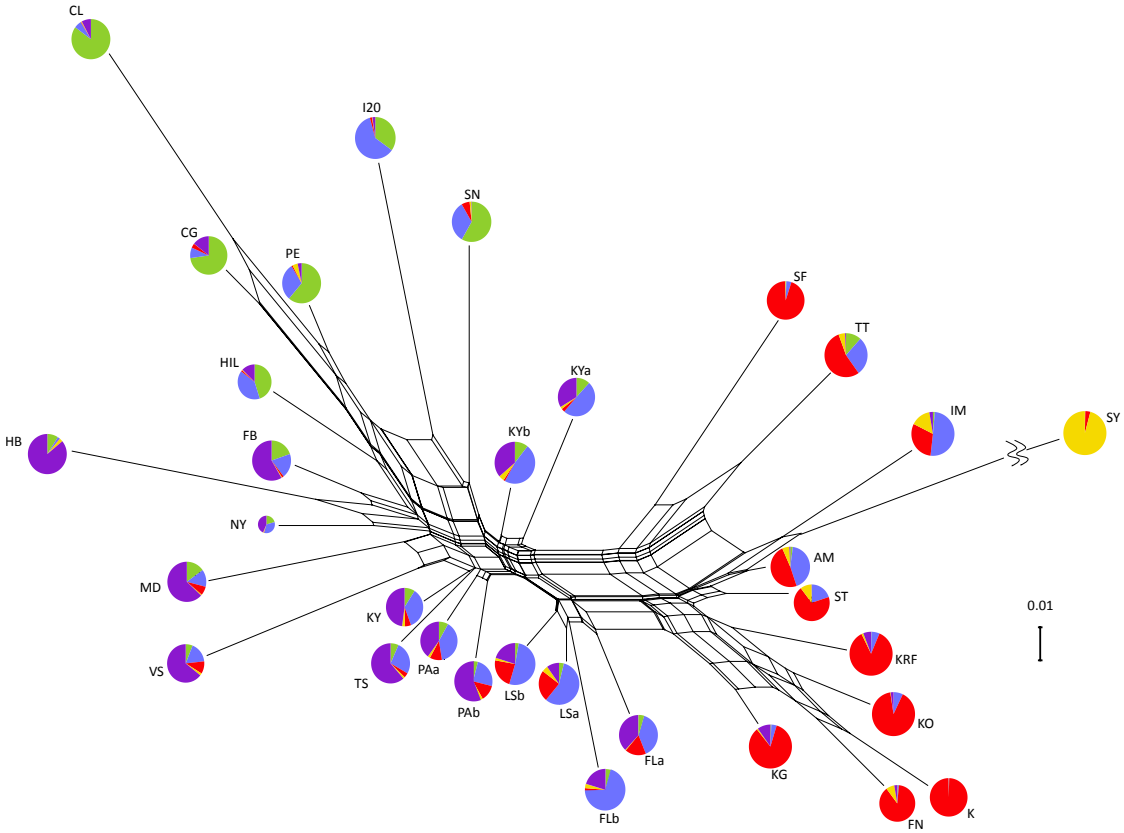

Fig. S3

Relationships among 32 *S. altissima* populations based on Nei's D_A (Nei et al. 1983), using a split network constructed by the neighbour-net method (Bryant and Moulton 2004) with the proportions of ancestry of each population superimposed at the node tips. Colors correspond with the result of the STRUCTURE analysis shown in Fig. 3b. Abbreviations refer to Table S1.

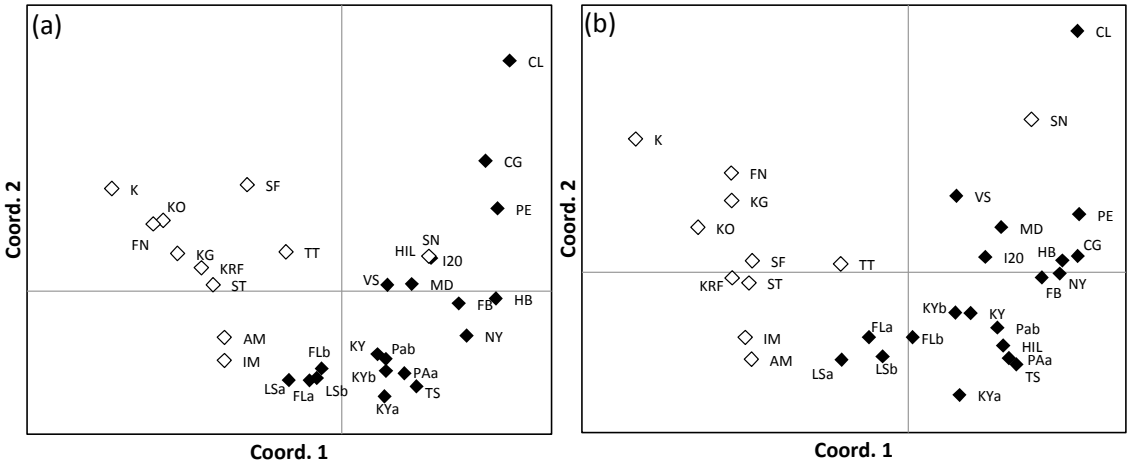

Fig. S4

PCoA based on pairwaise Φ_{RT} between the sampled populations of *S. altissima* using genetic distance of (a) Lynch and (b) Bruvo. Native and invaded populations are shown in closed and open symbols, respectively. Population abbreviations refer to Table S1.

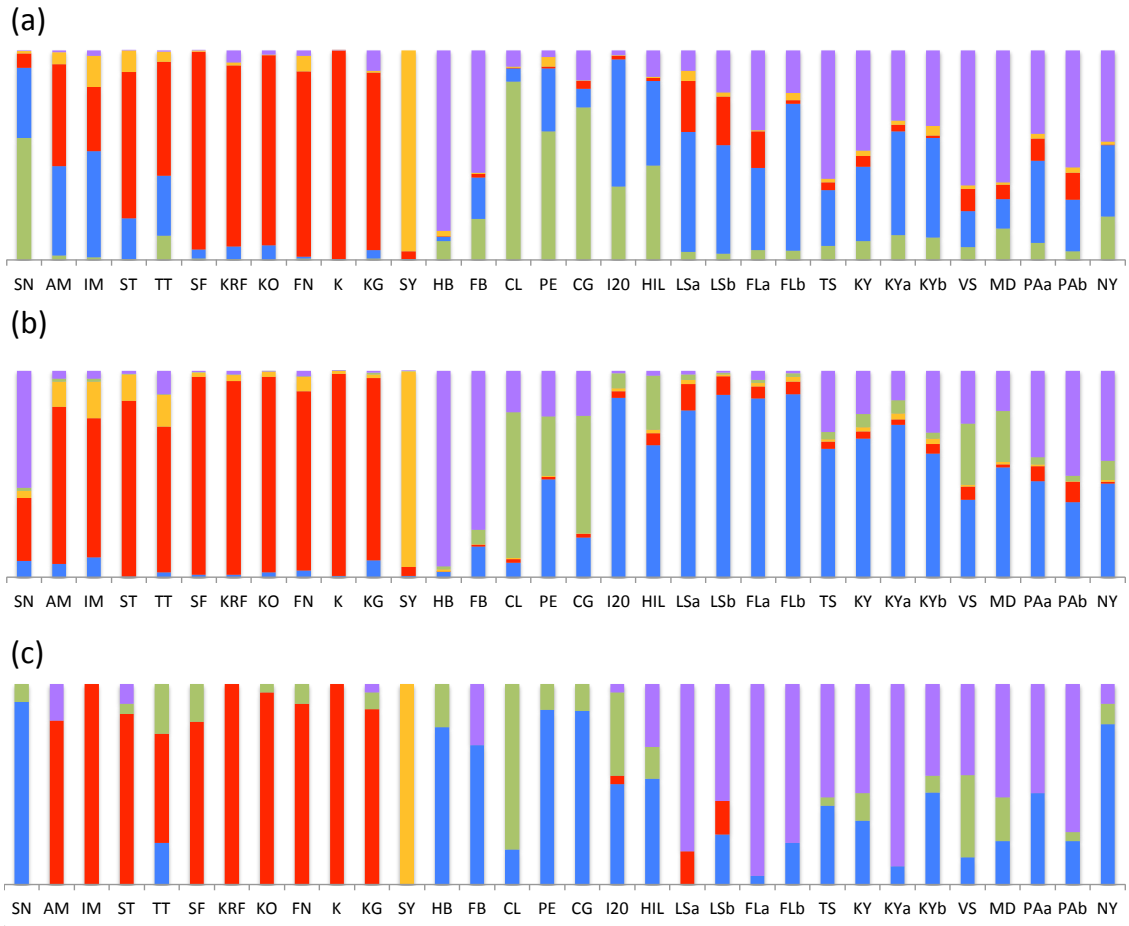

Fig. S5

Probabilities of membership for each population to each of the five genetic clusters (K = 5) estimated by (a) the model-based clustering method STRUCTURE (Pritchard et al. 2000), (b) the spatial model-based clustering TESS analysis (Durand et al. 2009), and (c) the multivariate analysis that identifies and describes clusters of genetically related individuals DAPC (Jombart et al. 2010). Population abbreviations refer to Table S1.

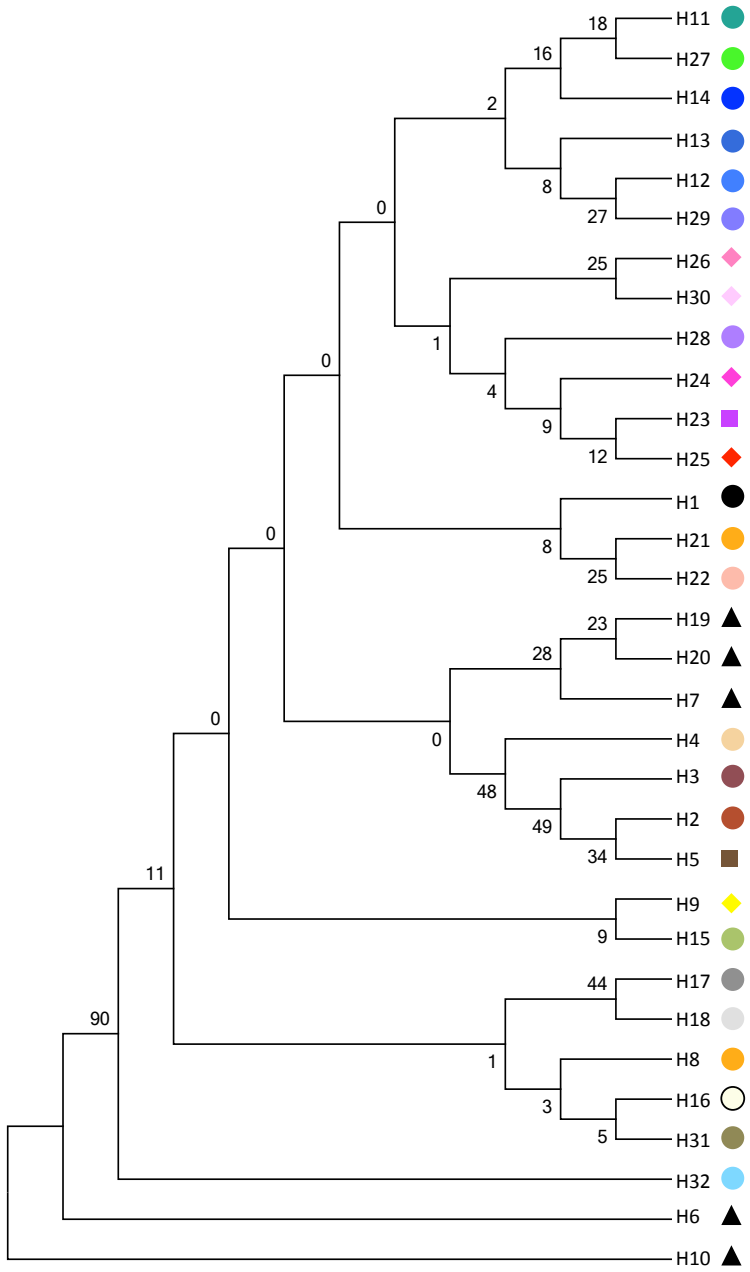

Fig. S6

The inferred phylogenetic relationships among *S. altissima* chloroplast haplotypes and their relationship to outgroups reconstructed by maximum parsimony analysis based on the cpDNA sequence data. Bootstrap values are shown above branches. Haplotypes found in both native and invaded populations are shown in diamonds, and haplotypes only found in invaded populations are shown in squares, and haplotypes of other species are shown in triangles: *S. virgaurea*, H6 and H10; *S. gigantea*, H7; *S. canadensis*, H19 and H20.

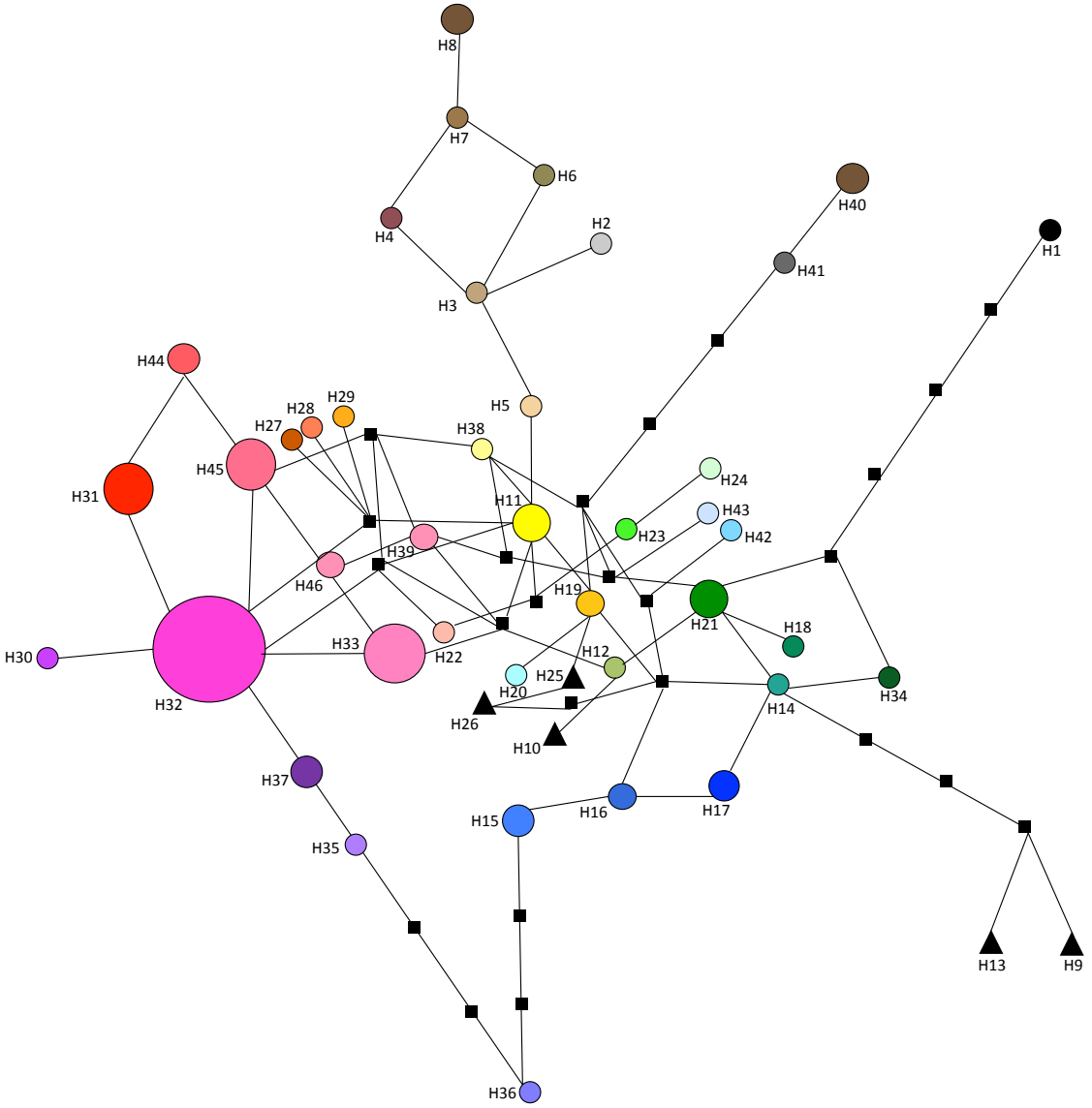

Fig. S7

The haplotype network of *Solidago altissima* based on chloroplast DNA sequence variation including indels. Circle size is proportional to haplotype frequency. Each line between haplotypes corresponds to one mutational change. Small black squares indicate missing haplotypes. Haplotypes of outgroup species are shown in triangles: *S. virgaurea*, H9 and H13; *S. gigantea*, H10; *S. canadensis*, H25 and H26.

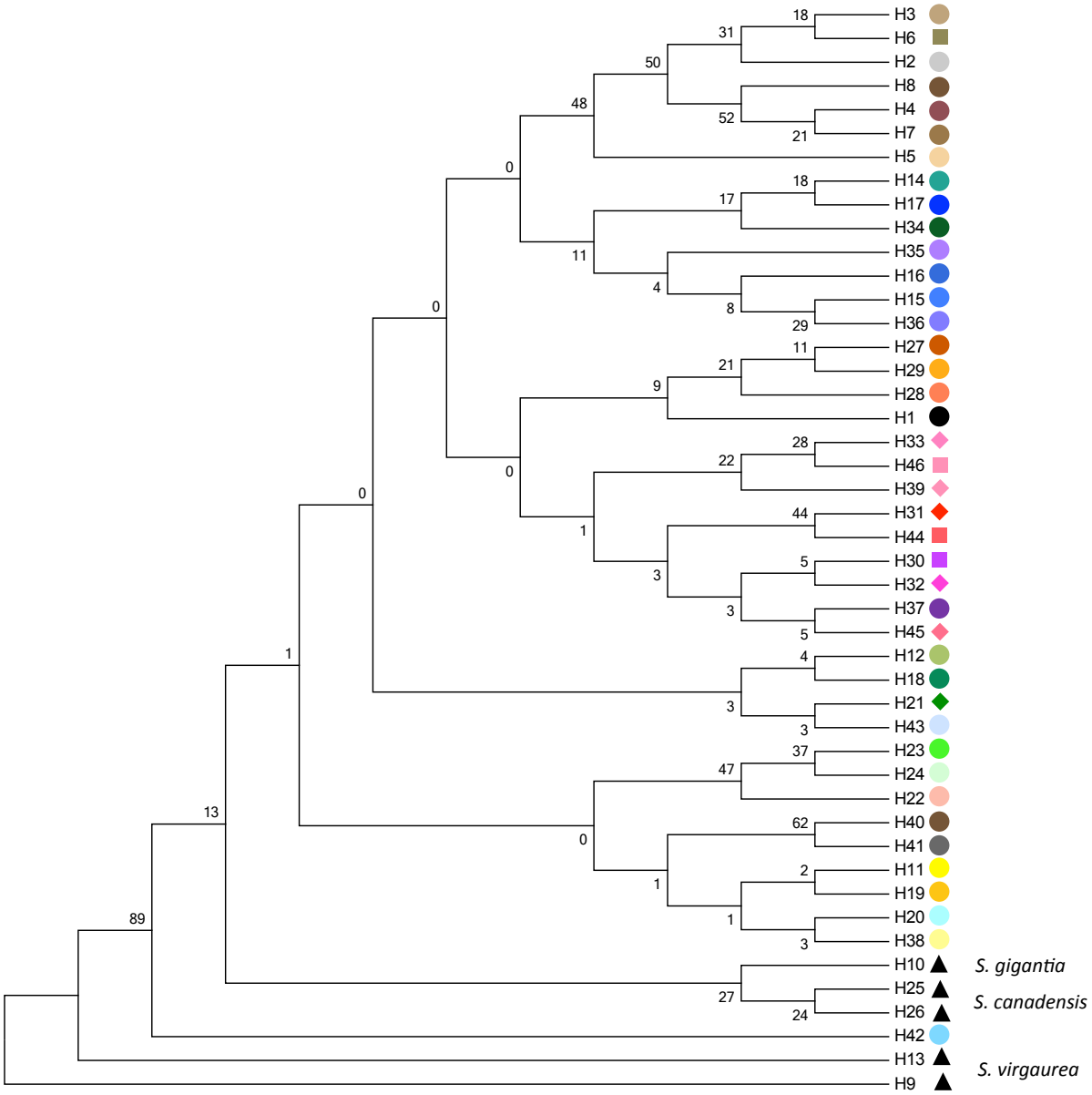

Fig. S8

The inferred phylogenetic relationships among *S. altissima* chloroplast haplotypes and their relationship to outgroups reconstructed by maximum parsimony analysis based on the cpDNA sequence data including indels. Bootstrap values are shown above branches. Haplotypes found in both native and invaded populations are shown in diamonds, and haplotypes only found in invaded populations are shown in squares, and haplotypes of other species are shown in triangles: *S. virgaurea*, H9 and H13; *S. gigantea*, H10; *S. canadensis*, H25 and H26.

Supplementary material

Multiple and mass introductions from limited origins: genetic diversity and structure of *Solidago altissima* in the native and invaded range

Yuzu Sakata, Joanne Itami, Yuji Isagi, Takayuki Ohgushi

Table S1. Distribution of sampling populations and indices of genetic diversity within populations of *S. altissima*.

Population	Location	Latitude	Longitude	*N*	H	N_a	*Ho*	*Shannon*
Invaded range								
SN	Hokkaido, JP	43.03N	141.30E	22	H:24,25,26	6.267	0.802	2.878
AM	Aomori, JP	40.82N	140.70E	22	H:23,24,25	8.467	0.932	3.028
IM	Iwate, JP	39.69N	141.12E	24	H:25,26	7.733	0.883	3.041
ST	Sado, JP	38.04N	138.44E	20	H:24,25	8.000	0.888	2.926
TT	Tokyo, JP	35.64N	139.58E	24	H:26	8.467	0.876	2.705
SF	Sizuoka, JP	35.13N	138.64E	21	H:9,24,25,26	9.000	0.900	2.912
KRF	Hyogo, JP	34.69N	135.26E	24	H:25	9.067	0.896	3.120
KO	Kochi, JP	33.56N	133.55E	24	H:5,25,26	8.133	0.895	2.694
FN	Fukuoka, JP	33.59N	130.21E	20	H:26	7.467	0.819	2.995
K	Kumamoto, JP	32.89N	130.73E	21	H:24,25	7.0000	0.880	2.306
KG	Kagoshima, JP	31.65N	130.42E	24	H:24,25,26,30	7.467	0.840	3.041
SY	Suncheon, KR	34.53N	127.30E	24	H:25	4.333	0.954	0.344
Native range								
HB	Minnesota, USA	47.39N	92.81W	23	H:2,3,9	7.933	0.904	2.368
FB	Minnesota, USA	44.23N	93.23W	23	H:18,25	9.533	0.896	2.894
CL	Iowa, USA	43.10N	93.45W	23	H:8,17,18,25	8.933	0.709	3.015
PE	Missouri, USA	39.78N	94.38W	23	H:8,9,24,26	8.933	0.829	2.645
CG	Kansas, USA	38.73N	96.48W	22	H:8,21,22	9.067	0.900	2.689
EF	Okulahoma, USA	35.45N	-95.88W	24	H:12	7.286	0.713	3.120
I20	Texas, USA	32.70N	96.58W	24	H:24,25	8.933	0.822	3.005
HIL	Texas, USA	32.00N	97.08W	19	H:25,28,29	8.267	0.890	2.552
LSa	Louisiana, USA	30.36N	91.15W	24	H:8,25	10.73	0.903	3.178
LSb	Louisiana, USA	30.34N	91.14W	24	H:4,25	11.13	0.899	3.178
FLa	Florida, USA	30.60N	84.38W	23	H:9,16,25	10.73	0.879	3.015
FLb	Florida, USA	30.50N	84.32W	24	H:25	10.13	0.876	2.925
TS	Tennessee, USA	35.23N	83.45W	23	H:9,25,26,27	10.26	0.882	2.955
KY	Kentucky, USA	36.82N	83.80W	22	H:13,25,32	10.46	0.874	3.028
KYa	Kentucky, USA	38.01N	84.51W	22	H:8,25	9.600	0.904	2.965

KYb	Kentucky, USA	37.05N	84.21W	24	H:8,25	10.20	0.868	3.178
VS	Virginia, USA	38.13N	78.45W	22	H:3,30	8.867	0.781	2.752
MD	Maryland, USA	39.07N	76.95W	23	H:11,13,14,25	9.933	0.796	2.789
PAa	Pennsylvania, USA	40.51N	75.79W	22	H:9,25	9.200	0.904	3.028
PAb	Pennsylvania, USA	40.54N	75.58W	23	H:3,8,14	11.00	0.894	3.135
NY	New York, USA	42.45N	76.48W	10	H:1,15,25	8.800	0.859	2.303
CA	California, USA	39.43N	-120.24W	24	H:31	2.500	0.528	2.658

N, number of individuals; N_a, number of alleles per locus (mean across loci); H_o, gamatic heterozygosity (mean across loci) (Moody 1993); *Shannon*, Shannon diversity (mean across loci). EF is tetraploid and CA is diploid, and the rest of the population is hexaploid.

Table S2

Chloroplast DNA sequence polymorphisms detected in three non-coding regions of *S. altissima* and outgroup species identifying 32 haplotypes (H1–32). Accession numbers for the DNA sequences of outgroup species: Abbreviations: *n*, the number of samples assigned to particular haplotype.

Haplotype	n	Nucleotide Position																				GeneBank accession number					
		psbH-psbB		*psbJ-petA*				*rpl-trnL*				*rps12-rpl20*		*rps16-trnK*					*trnQ-rps16*								
							1	1	1	1	2	2	2	3	3	3	3	3	3	3	3	4						
		1	2	7	9	9	1	5	6	7	0	3	7	5	5	5	5	6	7	7	8	0						
		1	9	1	4	7	3	7	7	9	5	0	3	4	8	9	9	2	6	9	4	3	*psbH-psbB*	*psbJ-petA*	*rpl-trnL*	*rps12-rpl20*	*rps16-trnK*	*trnQ-rps16*
		4	4	5	8	0	4	8	1	6	7	3	3	3	3	2	4	0	8	4	6	3						
H1	1	G	T	C	G	C	C	T	G	T	A	C	G	C	A	C	T	G	C	T	G	C	AB907859	AB907891	AB907923	AB907955	AB907987	AB908019
H2	2	T	C	A	G	A	C	T	G	T	A	A	G	C	A	C	T	G	A	T	A	C	AB907860	AB907892	AB907924	AB907956	AB907988	AB908020
H3	5	T	C	A	G	A	C	T	G	T	A	A	G	C	A	C	T	G	C	T	G	C	AB907861	AB907893	AB907925	AB907957	AB907989	AB908021
H4	1	T	C	A	G	A	C	T	G	T	A	C	G	C	A	C	T	G	A	T	G	C	AB907862	AB907894	AB907926	AB907958	AB907990	AB908022
H5	1	T	C	A	G	A	C	T	G	T	A	A	G	C	A	C	T	G	A	T	G	C	AB907863	AB907895	AB907927	AB907959	AB907991	AB908023
H6	1	T	C	C	G	A	A	T	G	A	C	C	G	C	A	A	T	G	A	T	G	C	AB907864	AB907896	AB907928	AB907960	AB907992	AB908024
H7	1	T	C	C	G	A	C	T	G	T	A	C	G	T	A	C	T	G	C	T	G	C	AB907865	AB907897	AB907929	AB907961	AB907993	AB908025
H8	7	T	C	C	G	A	C	T	G	T	A	C	G	C	A	C	T	G	A	T	G	C	AB907866	AB907898	AB907930	AB907962	AB907994	AB908026
H9	6	T	C	C	G	A	C	T	G	T	A	C	G	C	A	C	T	G	C	T	G	C	AB907867	AB907899	AB907931	AB907963	AB907995	AB908027
H10	1	T	C	C	G	A	C	T	G	A	C	C	G	C	C	A	T	G	A	T	G	C	AB907868	AB907900	AB907932	AB907964	AB907996	AB908028

H11	1	T	C	C	G	A	C	C	G	T	A	C	G	C	A	C	T	G	C	T	G	C	AB907869	AB907901	AB907933	AB907965	AB907997	AB908029
H12	4	T	C	C	G	A	C	C	G	T	C	C	G	C	A	C	T	G	A	T	A	C	AB907870	AB907902	AB907934	AB907966	AB907998	AB908030
H13	2	T	C	C	G	A	C	C	G	T	C	C	G	C	A	C	T	G	A	T	G	C	AB907871	AB907903	AB907935	AB907967	AB907999	AB908031
H14	3	T	C	C	G	A	C	C	G	T	C	C	G	C	A	C	T	G	C	T	G	C	AB907872	AB907904	AB907936	AB907968	AB908000	AB908032
H15	1	T	C	C	G	A	C	T	A	T	A	C	G	C	A	C	T	G	C	T	G	C	AB907873	AB907905	AB907937	AB907969	AB908001	AB908033
H16	1	T	C	C	G	A	C	T	G	T	A	C	G	C	A	C	T	G	A	T	G	T	AB907874	AB907906	AB907938	AB907970	AB908002	AB908034
H17	1	T	C	C	T	A	C	T	G	T	A	C	G	C	A	C	T	A	A	T	G	C	AB907875	AB907907	AB907939	AB907971	AB908003	AB908035
H18	2	T	C	C	T	A	C	T	G	T	A	C	G	C	A	C	T	G	A	T	G	C	AB907876	AB907908	AB907940	AB907972	AB908004	AB908036
H19	1	T	C	C	G	A	C	T	G	T	A	C	G	T	A	C	T	G	A	T	G	C	AB907877	AB907909	AB907941	AB907973	AB908005	AB908037
H20	1	T	C	C	G	A	C	T	G	T	C	C	G	T	A	C	T	G	A	T	G	C	AB907878	AB907910	AB907942	AB907974	AB908006	AB908038
H21	2	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	T	G	A	T	G	C	AB907879	AB907911	AB907943	AB907975	AB908007	AB908039
H22	1	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	G	G	A	T	G	C	AB907880	AB907912	AB907944	AB907976	AB908008	AB908040
H23	1	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	T	A	A	G	G	C	AB907881	AB907913	AB907945	AB907977	AB908009	AB908041
H24	11	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	T	A	A	T	A	C	AB907882	AB907914	AB907946	AB907978	AB908010	AB908042
H25	58	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	T	A	A	T	G	C	AB907883	AB907915	AB907947	AB907979	AB908011	AB908043
H26	15	T	T	C	G	A	C	T	G	T	A	C	G	C	A	C	T	A	C	T	G	C	AB907884	AB907916	AB907948	AB907980	AB908012	AB908044
H27	1	T	T	C	G	A	C	C	G	T	A	C	G	C	A	C	T	G	C	T	G	C	AB907885	AB907917	AB907949	AB907981	AB908013	AB908045
H28	1	T	T	C	G	A	C	C	G	T	A	C	G	C	A	C	T	A	A	T	G	C	AB907886	AB907918	AB907950	AB907982	AB908014	AB908046
H29	1	T	T	C	G	A	C	C	G	T	C	C	G	C	A	C	T	A	A	T	A	C	AB907887	AB907919	AB907951	AB907983	AB908015	AB908047
H30	2	T	C	C	G	A	C	T	G	T	A	C	G	C	A	C	T	A	C	T	G	C	AB907888	AB907920	AB907952	AB907984	AB908016	AB908048
H31	4	T	C	C	G	A	C	T	G	T	A	C	T	C	A	C	T	G	A	T	G	C	AB907889	AB907921	AB907953	AB907985	AB908017	AB908049

H32		1		T	C			C	G	A	C			T	G	T	C			C	G			C	A	C	T	G			A	T	G	C		AB907890	AB907922	AB907954	AB907986		AB908018	AB908050