Genetic diversity and structure of seed pools in an old planted *Pinus thunbergii* population and seed collection strategy for gene preservation

Masakazu G. Iwaizumi[1,*], Aziz Akbar Mukasyaf[2], Ichiro Tamaki[3], Jin'ya Nasu[4], Naoko Miyamoto[4], Miho Tamura[2] and Atsushi Watanabe[2]

[1] Kyushu Regional Breeding Office, Forest Tree Breeding Center, Forestry and Forest Products Research Institute, 2320-5, Suya, Koshi, Kumamoto 861-1102, Japan

[2] Department of Forest Environmental Science, Faculty of Agriculture, Kyushu University, Kuwabara, Nishi-ku, Fukuoka 819-0382, Japan

[3] Gifu Academy of Forest Science and Culture, 88 Sodai, Mino, Gifu 501‐3714, Japan

[4] Tohoku Regional Breeding Office, Forest Tree Breeding Center, Forestry and Forest Products Research Institute, 95, Osaki, Takizawa, Iwate 020-0621, Japan

***Corresponding author: Masakazu G. Iwaizumi**

Tel: +81-96-242-3151

Fax: +81-96-242-3150

E-mail: ganchan@ffpri.affrc.go.jp

Abstract

To achieve gene preservation in tree populations when planting seedlings to regenerate forests, specific and practical guidelines and criteria for seed collection are needed to ensure reliable coverage and effective capture of the current genetic variation. We examined the genetic variation of adult trees (524 trees) and seed pools (1,618 seeds) collected from 70 mother trees in an old planted population of *Pinus thunbergii* in Japan using seven nuclear microsatellite markers. To consider a suitable seed collection strategy, we monitored changes in allelic diversity of seed pools with an increasing number of mother trees and examined the required number of mother trees. We found significantly higher allelic diversity statistics in seed pools of mother trees with smaller diameter at breast height. Spatial genetic autocorrelation analysis showed significantly positive kinship coefficients between both pairs of adult trees and seeds collected from mother trees for up to 100 m distance classes. Based on the rarefaction curve, seed pools obtained from approximately up to 30 mother trees could cover and saturate most of the genetic diversity and composition of the adult tree population and the potential overall seed pools, particularly for statistics that are less likely affected by rare alleles. Indications from the mother tree number, together with considerations for the size or spatial distribution of mother trees, are expected to contribute not only to guidelines for *in situ* genetic management of local planted populations, but also to the development of strategies for *ex situ* gene preservation of populations as genetic resources.

Keywords: gene preservation; genetic diversity; mother tree; *Pinus thunbergii*; rarefaction curve; seed pool

Introduction

Preservation of regional genetic variation in forest tree populations is important not only to maintain the potential for adaptation to local environments, but also to sustain public functions as well as the historical, economic, social, and cultural values from which local peoples have benefited (Hosius et al. 2006; Iwasaki et al. 2019; see also Iwaizumi et al. 2018). Many gene flow studies, primarily in natural populations, have been based on parentage analysis or indirect estimators to monitor levels of genetic connectivity within or among (sub-) populations (e.g., Dow and Ashley 1996; Robledo-Arnuncio et al. 2004; Goto et al. 2006; Iwaizumi et al. 2013, 2022). These insights have been used to consider guidelines for achieving self-sustainable genetic maintenance of populations, such as appropriate population sizes, numbers, or necessary areas of preservation stand needed to cover the local genetic variation. However, in some major species, natural populations have become almost extinct, and the main reserve of genetic variation is in planted forests. In these cases, gene preservation within the populations cannot be achieved without artificial maintenance, such as through seed collection and replanting of seedlings. These efforts cannot be guided solely by monitoring genetic dynamics as for natural populations, but also require the development of new strategies for preserving current genetic variation in managed populations.

The conifer species *Pinus thunbergii* Parl. (Japanese black pine) is widely distributed throughout Honshu, Shikoku, and Kyushu (including Yaku Island) in Japan and the southern edge of the Korean Peninsula. Natural forests of the species were previously found in coastal areas, and the species has also been planted in coastal areas to prevent land erosion and to protect these areas from wind-blown sand and tidal waves, which contributed to the expansion of agriculture on inland plains. *P. thunbergii* forests also have socially and culturally important functions; old planted forests in many areas of Japan, which have been maintained and replanted by local residents, are places with traditional picturesque or relaxing scenery and

constitute an important part of Japanese historical coastal landscapes. However, arrival of the invasive pine wood nematode (*Bursaphelenchus xylophilus*) from North America resulted in a marked irreversible decline of most *P. thunbergii* forests in Japan over the past 40–50 years (Mamiya 1988), and, currently, *P. thunbergii* resources are limited primarily to the planted populations that have been strictly protected. Recent examination of genetic variation of these old planted populations of *P. thunbergii* throughout Japan using nuclear microsatellite (also known as simple sequence repeat or SSR) markers showed a general geographical cline in the genetic structure from southwestern to northeastern Japan (Iwaizumi et al. 2018), which is similar to the variation in historical natural populations previously analyzed using allozyme markers (Miyata and Ubukata 1993). The managed planting of *P. thunbergii* will continue in the future owing to its functions described above and consequently due to the great demand by local residents to preserve local pines. There is a need to consider strategies to conserve the genetic variation at the individual population level using seedlings collected from current populations in which the original genetic variation of the species is thought to be preserved.

To achieve gene preservation when replanting seedlings to regenerate populations, it is necessary to collect seeds that cover the genetic variation of the currently established adult tree population. Previous studies have evaluated genetic variation of seed pools collected from different mother trees or seed traps and, in particular, seed pools derived from smaller numbers of mother or paternal trees tended to show lower allelic diversities and also skewed genetic compositions (e.g., Irwin et al. 2003; Nakanishi et al. 2005; Iwaizumi et al. 2013). To enrich the genetic diversity of seed pools, it is important to select different mother trees that have seed pools with high genetic diversity or different genetic compositions, and also to collect seed pools from as many mother trees as possible. An empirical and conventional guideline, regarding the number of mother trees that should be targeted for seed collection when the forest is cut and replanted, has been set by a forestry agency in Japan (see Furukoshi 1991). However, specific and practical guidelines and criteria from a conservation genetics

perspective that are based on molecular marker tools are needed to examine whether previous guidelines are appropriate in order to ensure reliable coverage and effective capture of genetic variation in adult tree populations.

Therefore, in the present study, we considered a seed collection strategy that achieves gene preservation necessary to conserve genetic variation, with a focus on an old planted population of *P. thunbergii* in Japan. We examined genetic diversity and structure in the currently established adult tree population and seed pools collected from the adult trees using nuclear SSR markers. We then examined differences in the genetic diversity of seed pools among the mother trees and possible factors causing differences that affect this diversity. We also evaluated the increase in genetic diversity of seed pools and the decrease in genetic differentiation between seed pools and adult trees as a function of increasing the number of mother trees. Thus, we propose revised seed collection guidelines for the appropriate selection of mother trees or the required number of mother trees, based on previous seed collection guidelines and these study results.

Materials and methods

Study site, field survey, and sample collection

The study site was an old planted coastal population of *Pinus thunbergii* managed by the experimental station of Kyushu University in Fukuoka, Kyushu, southwestern Japan (33°58´N, 130°30´E; Fig. 1). This population has been the subject of our previous studies on genetic variation based on nuclear SSR markers (48 trees) and morphological variation based on spherical fruit characteristics (20 of the 48 trees) in this species throughout Japan (Iwaizumi et al. 2018, 2021b, respectively). The forested study site was about 1.9 km long and a maximum of 0.2 km wide for an area of about 25 ha; the coastal side was *P. thunbergii* only, whereas the percentage of evergreen broad-leaved climax tree species was higher on the inland

side. More than 90% of *P. thunbergii* trees in this forest had a diameter at breast height (DBH) of less than 50 cm, and some larger trees with DBH > 60 cm were distributed randomly. In our recent study (Mukasyaf et al., 2021), tree age in this population was estimated to be 12-36 years for the DBH class of 10-30 cm, 32-190 years for the DBH class of 30-60 cm, and 170-195 years for the DBH class of 60-90 cm.

To investigate genetic diversity and structure of the adult tree population, from 2017 to 2020, we randomly selected 524 trees with DBH > 5 cm over the entire spatial distribution of this *P. thunbergii* population, keeping at least 10 m between trees. We measured the spatial location of each tree using a GPSMAP 64 (Garmin Inc.) and sampled fresh needles for DNA extraction. In October 2017, in which year there was a moderate cone production in the population, we selected 70 mother trees of various size ranges measured by height (range, 6.3 to 23.1 m; mean, 13.4 m) and DBH (range, 18.5 to 87.5 cm; mean, 44.0 cm) from the 524 adult trees to investigate the genetic diversity and structure of seed pools within the population. We sampled 5-10 mature cones from the upper crown of each mother tree. The sampled cones from each mother tree were dried separately in the drying room, and mature seeds were extracted and discriminated from empty and immature seeds. Needle and seed samples were stored at –25 °C prior to DNA extraction.

DNA analysis

In accordance with previous studies in *Pinus* and other tree populations (e.g., Robledo-Arnuncio et al. 2004; Miyamoto et al. 2008; Kim et al. 2015; Iwaizumi et al. 2021c), a maximum of 24 seeds per mother tree were randomly selected in order to evaluate genetic variation of seed pools at the family level as well as to compare among mother trees. Embryo tissue was extracted from each seed. Total DNA were extracted from needle samples of adult trees and tissue samples of seed pools using a modified cetyl trimethyl ammonium bromide (CTAB) method (Shiraishi and Watanabe 1995). Genotypes of the trees and seeds were

determined by seven polymorphic nuclear SSR markers: bcpt834 (Iwaizumi et al. 2013) and bcpt1075, bcpt1549, bcpt1671, bcpt1823, and bcpt2532 (Iwaizumi et al. 2018) of *P. thunbergii*, and bcpd119 (Iwaizumi et al. 2018) of the related Japanese pine species *Pinus densiflora*. This set of the seven markers was the same as the set used in our previous study of genetic variation of the species throughout Japan (Iwaizumi et al., 2018). PCR was performed using the GeneAmp PCR System 9700 version 3.05 (Applied Biosystems Inc.) with a Multiplex PCR kit (Qiagen Inc.) following the procedures described by Iwaizumi et al. (2010). The amplified fragment size was determined by the automated DNA sequencer ABI PRISM 3730 Genetic Analyzer (Applied Biosystems Inc.). The sizes of the amplified fragments were standardized with a GeneScan 500 LIZ Size Standard (Applied Biosystems Inc.). These seven loci were then genotyped using GeneMapper 5.0 software (Applied Biosystems Inc.). The number of seeds analyzed was planned to be 1,680 based on 24 seeds from 70 mother trees, but because fewer than 24 seeds were collected from some mother trees and PCR amplification failed for DNA extracted from some seeds due to immaturity or damage, multilocus genotypes of the seed pools were obtained from a total of 1,618 seeds (11-24 seeds per mother tree).

Data analysis

Genetic diversity and structure of adult tree population and seed pools

To assess allelic diversity within the adult tree population and seed pools, the number of detected alleles (n_a), the effective number of alleles (n_e), allelic richness (*AR*; El Mousadik and Petit 1996), and gene diversity (*HS*) were calculated at each locus in each generation using FSTAT version 2.9.3.2 software (Goudet 2002). Statistics of seed pools were calculated individually for the 70 mother trees and for all analyzed seeds. *AR* was calculated based on the smallest set of genotype data for 11 individuals (22 gene copies (‘$AR_{[22]}$’); El Mousadik and Petit 1996). To indirectly estimate the diversity of paternal parents contributing the seed pools for each of the mother trees, the effective number of pollen donors (N_{ep}) was calculated using

POLDISP version 1.0 software (Robledo-Arnuncio et al. 2007). To assess the level of outcrossing as a factor related to the genetic diversity of seed pools of mother trees, the multi-locus outcrossing rate (t_m) was estimated using MLTR version 3.4 software (Ritland 2002). To evaluate the levels of genetic differentiation of seed pools among mother trees, F_{ST} based on Weir and Cockerham (1984) was calculated over all loci using FSTAT. The significance of deviation of F_{ST} from zero was evaluated over 1,000 permutations. Confidence intervals (95% and 99%) of this value were determined based on 1,000 bootstrapping replications. In addition, to estimate the levels of genetic differentiation of paternal gamete pools [n] among mother trees, Φ_{FT} (Smouse et al. 2001; Austerlitz and Smouse 2001) also was calculated using POLDISP.

To examine the relationships between tree size of mother trees and genetic statistics of their seed pools, Kendall's rank-correlation coefficients (τ) between the height and DBH and the mean n_a, n_e, *AR*, *HS*, N_{ep}, and t_m values of the seven loci were calculated using R version 3.5.0 (R Development Core Team 2011) with EZR version 1.40, R-commander software (Kanda 2013).

To assess the spatial genetic structure within the adult tree population and among seed pools of mother trees, the S_p statistic (Vekemans and Hardy 2004) was calculated because it allows for comparisons of the magnitude of the genetic structure of a species among different years or life stages (Troupin et al. 2006; Iwaizumi et al. 2021a). The S_p statistic was manually computed using the following equation:

$$S_p = -b / [1 - f_{(1)}]$$

where $f_{(1)}$ is the individual-based kinship coefficient (f_{ij}: Loicelle et al. 1995) for the first distance class, and b is the regression slope based on the natural logarithm of spatial distance. The $f_{(1)}$ and b values were calculated using SPAGeDi version 1.2 software (Hardy and Vekemans 2002). The distance class for $f_{(1)}$ was set to 0–50 m for both adult trees and seed

pools. The significance of spatial genetic structure was estimated based on deviation of the *b* value from zero over 1000 permutations. To compare the degree of spatial autocorrelation in the genetic composition between adult trees and seed pools, f_{ij} was calculated using SPAGeDi software. The pairwise distance class between individual adult trees (for adults) and mother trees (for seed pools) was set up to 400 m at 50 m intervals and the significance of the deviation of f_{ij} values from zero was estimated over 1000 permutations.

Relationship between the number of mother trees and genetic diversity of seed pools

Using genotype data of the adult tree population and the seed pools of 70 mother trees, changes in the allelic diversity statistics of seed pools with an increasing number of mother trees were examined and compared with statistics of adult trees. Programs that (i) extract genotype data of seed pools of the specified numbers of mother trees randomly and then (ii) calculate the statistics (n_a, n_e, *AR*, *HS*) were constructed using R, and (iii) the calculation trials of (i) and (ii) were performed 100 times per number of mother trees specified. Similarly, to examine the change in the level of genetic differentiation between adult trees and seed pools with an increasing number of mother trees, similar calculation trials were performed by constructing a program to calculate F_{ST} between adult trees and the extracted pools of seed genotypes. At that time, extractions of seed pools were performed at the mother-tree level (fixing genotype data of seed pools within mother trees). We calculated the mean values and their 95% confidence intervals of the statistics of 100 trials and evaluated the changes with an increasing number of mother trees by drawing 'rarefaction curves', a biostatistics tool for diversity assessments (e.g., Li et al. 2015).

Based on these examinations, for each diversity statistic (n_a, n_e, *AR*, *HS*), we evaluated the number of mother trees at which their seed pools can cover genetic variation of the adult tree population. We calculated the percentages of the diversities of seed pools from two viewpoints: namely, the percentages of statistical values calculated for the seed pools at the

specified number of mother trees against (1) statistical values of adult trees (524 trees), and (2) those of all sampled seed pools (a total of 1,618 seeds from 70 mother trees). We considered (1) as a 'coverage-based' viewpoint, or the number of mother trees at which seed pools exceed the particular percentages of statistical values of adult trees; and (2) as a 'saturation-based' viewpoint, or the number of mother trees at which the collected seed pools cover the particular percentages of the potential gene pool of the next generation within the population (i.e., the statistical values of seed pools at which the diversity statistics would not be increased any further).

Results

Genetic diversity and structure of adult trees and seed pools

Among the seven loci analyzed, the mean values of n_a, n_e, $AR_{[22]}$, and *HS* of the adult tree population (524 trees) were 17.57, 5.04, 6.35, and 0.771, respectively (see also Supplementary Table 1). The corresponding values for all seed pools (of 1,618 seeds from 70 mother trees) were 18.57, 5.46, 6.56, and 0.786, respectively, and these values were all higher than the corresponding values for adult trees.

The values of allelic diversity and mating pattern statistics of seed pools among loci at the mother tree level ranged 3.86 to 7.71 (mean 6.38) for n_a, 1.81 to 3.86 (mean 2.98) for n_e, 2.68 to 5.46 (mean 4.65) for $AR_{[22]}$, 0.386 to 0.741 (mean 0.640) for *HS*, 1.06 to 177.31 (mean 25.39) for N_{ep}, and 0.727 to 1.000 (mean 0.991) for t_m. Kendall's rank-correlation tests showed that none of these six statistics correlated significantly with the tree height of the mother trees (Table 1). For mother tree DBH, n_e and *HS* values showed significantly negative relationships with DBH (see also Supplementary Fig. 1 for DBH vs. n_e); that is, seed pools collected from mother trees with smaller diameter had higher effective numbers of alleles and gene diversities.

The S_p statistic for seed pools (0.0019), calculated from the *b* value, was higher than that for adult trees (0.0013). These values for both generations were low but deviated significantly from zero (permutation test, $P < 0.05$). Spatial genetic autocorrelation analysis showed significantly positive kinship coefficient (f_{ij}) between pairs of adult trees at the 50 m distance class (permutation test, $P < 0.05$; Supplementary Fig. 2). The f_{ij} values between pairs of seeds collected from mother trees were significantly positive up to the 100 m distance classes. Values for both generations decreased with an increase in distance class, and those in longer distance classes did not deviate significantly from zero.

The overall value of the statistics for genetic differentiation (F_{ST}) between adult trees and all seed pools was 0.0026, which was significantly higher than zero at the 95% confidence interval ($P < 0.05$) but not at the 99% confidence interval ($P > 0.01$). The overall F_{ST} value of seed pools among the 70 mother trees was 0.190, which was significantly higher than zero at the 99% confidence interval ($P < 0.01$). The overall Φ_{FT} value for the 70 seed pools was estimated to be 0.063.

Relationship between the number of mother trees and genetic diversity of seed pools

The percentages of the allelic diversity statistics of all seed pools against those of adult trees were 105.7%, 108.4%, 103.2%, and 102.0% for n_a, n_e, $AR_{[22]}$, and *HS*, respectively. The mean coverage-based percentage (vs. adult trees) of n_a of seed pools for 100 trials exceeded 95% and 100% at 39 and 51 mother trees, respectively, and mother tree numbers required were the largest for n_a among the four statistics (Table 2 and Fig. 2). The corresponding numbers for *AR* were the second largest (exceeding 95% and 100% at 6 and 17 mother trees, respectively), and the numbers were smallest for *HS* (exceeding 95% and 100% at 3 and 9 mother trees, respectively). The mean saturation-based percentage (vs. all seed pools) of n_a of seed pools for 100 trials exceeded 94% and 98% at 50 and 62 mother trees, respectively, and the increase in the percentage was greater with increasing number of mother trees to 70 mother trees. The

corresponding mother tree numbers for n_e were the second largest at 11 and 28 mother trees (exceeding 94% and 98%, respectively), and at around 28 mother trees, the increase of the percentage with increasing number of mother trees was almost saturated. The corresponding numbers for *HS* were the smallest among the four statistics (at 3 and 9 mother trees exceeding 94% and 98%, respectively).

The mean value of the 100 trials of the multi-locus genetic differentiation (F_{ST}) between adult trees and seed pools decreased below 0.01 and 0.005 at 12 and 25 mother trees, respectively (Fig. 3). At around more than 25 mother trees, the decrease of F_{ST} with an increasing number of mother trees almost converged.

Discussion

Differences in genetic diversities of seed pools among mother trees

We found variations in allelic diversity statistics of seed pools among mother trees in the old planted population of *Pinus thunbergii*. We found significantly negative relationships between DBH of mother trees and n_e and *HS* (Table 2 and Supplementary Fig. 1), indicating that mother trees with smaller DBHs tend to be pollinated by the genetically diverse paternal gamete pools. Although the present study examined seed pools of one moderate year, the tendency based on the analysis of more than 1,600 seedlings from as many as 70 mother trees would be relatively similar over years, considering the low genetic differentiation in paternal gametes among years in *Pinus* (Iwaizumi et al. 2010; 2013) due to the small year-to-year fluctuation in male reproductive amount (Nakayama and Kobayashi 1981).

The lower genetic diversity of seed pools in the larger mother trees in our study population could firstly be considered to be the result of a higher percentage of self-pollinated seeds. Some conifer studies of mating parameters in natural populations and seed orchards have shown differences in the outcrossing rate of seed pools among mother trees, for example,

Pinus (Robledo-Arnuncio et al., 2004), *Abies* (Arista and Talavera 1997), *Picea* (Morgante et al. 1991), and *Tsuga* (El-Kassaby et al. 2003). However, the multi-locus outcrossing rate (t_m) of seed pools were estimated to be almost 1.0 for many mother trees and did not correlate significantly with DBH (Table 1). The more possible reason would be that larger mother trees tend to be pollinated by lower percentages or narrow variety of outcrossing male gametes, because of their higher relative reproductive effort put into male flowers compared to female ones (e.g., in *Abies*: Kohyama 1982; Iwaizumi et al. 2021c). Arista and Talavera (1997) reported that a positive size-dependency in reproductive investment of *A. pinsapo* trees in a natural population was more apparent in male reproduction than in female reproduction.

The significantly negative relationship between the mother tree size and genetic diversity of seed pools could also be a result of differences in stand structures around mother trees such as size distribution or tree density. In conifers, studies have shown that mother trees with higher tree density have higher genetic diversity of seed pools (e.g., Morgante et al. 1991; El-Kassaby et al. 2003; Robledo-Arnuncio et al. 2004). Regarding this possibility, a plot-based census of stand structure, including some mother trees, has been conducted in some areas of the study population (A.A. Mukasyaf, unpublished data). However, parameters of stand structure around the mother trees (i.e., the number of trees (# trees), total basal area (BA) with more than 5 cm in DBH, and the mean DBH, within the 25 m × 25 m-square plot of each mother tree) did not show any significant relationships with genetic diversity of the seed pools ($N = 19$; $\tau = 0.135$ for # trees vs. n_e; $\tau = 0.035$ for BA vs. n_e; $\tau = -0.246$ for mean DBH vs. n_e; all ns). Although the slightly negative correlation between mean DBH vs. n_e may also imply low allelic diversity of seed pools in an area of large surrounding trees, due to the size-dependency in reproductive investment supposed as above, the level of influence of stand structure would differ among different species or populations.

Relationship between number of mother trees and genetic diversity of seed pools

By constructing calculation programs with randomly extracted seed pools at the mother tree level, we examined the percentage of genetic diversity of seed pools at specific numbers of mother trees against that of adult trees and the overall seed pool based on several diversity statistics. Based on n_e, *AR*, *HS*, and F_{ST}, seed pools obtained from approximately up to 30 mother trees could account for most of the genetic diversity and composition of adult tree populations (coverage-based viewpoint) and also the potential overall seed pools (saturation-based viewpoint; Table 2, Figs. 2 and 3).

Among the four allelic diversity statistics (n_a, n_e, *AR*, and *HS*), there were differences in the increasing patterns of the diversity of seed pools as an increasing numbers of mother trees. Of these, *HS* covered most of the genetic diversity of both adult trees and the overall seed pool while requiring the smallest number of mother trees (i.e., over 100% and 98% at 9 mother trees in the coverage-based and saturation-based viewpoints, respectively), and n_e and *AR* tended to cover the second smallest number of mother trees (Table 2; Figs. 2 and 3). Theoretical population genetics studies suggest that reduction in the genetic diversity of populations accompanied by reduction in population size appeared later for statistics based on the skewness of allele frequencies (such as n_e and *HS*) than those based on allele counts (like n_a: Cornuet et al. 1996). Thus, findings of previous studies do not contradict our study results for the smallest number of mother trees required for the *HS*-diversity coverage. On the other hand, genetic diversity of seed pools based on n_a did not saturate with an increasing number of mother trees, probably due to continuous capture of new rare alleles with increasing sample numbers. Further discussion based on theoretical or practical examinations may be required to understand to what extent we should consider covering such rare alleles. However, considering the theoretical mutation rates suggested for nuclear SSR markers (10^{-3} to 10^{-4}: Guichoux et al. 2011) and the long-distance immigrant gene flow from outside of the population assumed to be frequent in *Pinus* (e.g., Nathan et al. 2002; Iwaizumi et al. 2010; 2013), rare alleles may be

more likely to be supplied. Thus, coverage based on n_e and *HS* would be rather more important from the viewpoint of maintaining the number of major alleles and/or allele frequencies that constitute the main variation within populations. Marshall and Brown (1975) recommended that population sizes be maintained so that 95% of the major alleles (the frequencies of more than 0.05) can be preserved.

Previous population genetics studies have considered the effective sample size per population required to evaluate the genetic variation of species, and the numbers required were reported as 25-30 (Hale et al. 2012; Hoban et al. 2014), 30 (Petit et al. 1998; Sjogren and Wyoni 1994) and 50 (Nybom 2004). Volk et al. (2005) reported that, in the core collection trees from the next-generation population of *Malus sieversii* wild cultivars, extraction of 35 trees covered more than 92.5% of the effective number of alleles of the potential whole next generation (all 278 trees). According to the replantation manual of gene preservation forests by The Forestry Ministry of Japan, seed pools are recommended to be derived from more than 30 mother trees collected in the mast year(s) and used for replantation of the next-generation forest when the original or the current forest is cut (see Furukoshi 1991). This number of mother trees is also similar for our results, even though the technical molecular tools of population genetics are not yet efficient. In addition, we evaluated the appropriate number of mother trees from the viewpoint of preservation of genetic composition (allele frequencies) of seed pools in terms of minimizing genetic differentiation (F_{ST}) with adult trees; to the best of our knowledge, the present study represents the first application of this approach in a major forestry tree species.

Evaluation of the number of mother trees for which genetic diversity of seed pools is covered by or saturated using the rarefaction curve can be mostly applied to other tree populations. However, the number may vary according to the level of spatial genetic structure within the population and the relative proportions of genetic variation within and among mother trees within the populations, to some extent. Significant spatial genetic structures up to

100 m were found for both generations (adult trees and seed pools; Supplementary Fig. 2). In fact, it is difficult to consider that replantation of this large-scale *P. thunbergii* forest (about 25 ha) has been conducted entirely by the same seed lot at the same time over the entire area, and it is more likely to have been replanted incrementally and from a different seed pool (Mukasyaf et al. 2021). Seed pools with somewhat different genetic compositions might be used for replantation in different areas within the population, at the scale of the spatial genetic structure. Thus, it may be more appropriate to select mother trees that are distant (as far as possible in scale of genetic structure) from each other in order to reduce the risk of collecting seed pools that are not sufficiently genetically variable and/or to avoid inbreeding. In addition, the level of genetic differentiation of seed pools among mother trees in our study (F_{ST} = 0.190; Φ_{FT} = 0.063; see also Results) seems to be somewhat high compared with those of studies of other major *Pinus* populations which analyzed similar numbers of seeds per mother tree (12-28 seeds in average; Φ_{FT} = 0.006–0.007 for *Pinus sylvestris*: Robledo-Arnuncio et al. 2004; Φ_{FT} = 0.021 for *Pinus densiflora*: Kim et al. 2015). In general, genetic differentiation of the next generation among cohorts within populations tends to be lower for the mainly conifer species with substantial gene flow than for endemic or rare species with small population size or tree density (Hamrick et al. 1992). Practical studies involving consideration of sampling strategy (Miyamoto et al. 2008 for *Fraxinus excelsior*) suggest that increasing the number of samples (groups) at the hierarchical level with larger genetic variation (e.g., increasing the mother trees, in the case of large F_{ST} of seed pools among mother trees) is more efficient for covering genetic diversity of a species or population.

Implications for seed collection strategies for gene preservation of* P. thunbergii *populations

In the studied population of *Pinus thunbergii*, we obtained significant practical indications for revising seed collection strategies for gene preservation of forest tree

populations that build on previous knowledge. First, the number of mother trees required for gene preservation in the present study (approximately more than 30 mother trees; Table 2 and Figs. 2 and 3) was generally similar to numbers indicated or recommended in previous studies or empirical manuals. More effective evaluation of coverage and saturation of genetic diversity and composition could be achieved using statistical parameters that are less likely to be affected by rare alleles such as n_e, *HS*, and F_{ST} (rather than n_a). Second, according to the newest results, considerations of tree sizes of mother trees or spatial genetic structure within the population are recommended for effective coverage or saturation of the genetic diversity of seed pools. For example, considering the significantly higher n_e and *HS* of seed pools collected from the smaller mother trees based on DBH (Table 1 and Supplementary Fig. 1), the seed collection strategy might, in fact, be differentiated depending on whether the objective is (A) the individual-level conservation of particular superior (larger) trees using their F1 progenies as genetic resources or (B) the population-level conservation of gene pools to cover the parental population. In addition, a sampling strategy that takes information on replantation history or processes within populations into consideration would be more reliably and effectively to avoid collection of genetically related seeds, in light of the range of significant spatial genetic autocorrelations (as in Supplementary Fig. 2). Although the replication of investigation in other populations would be better to generalize the strategy, results and indications obtained from the present study are expected to contribute not only to the techniques and guidelines for *in situ* genetic management of planted forests, but also to the development of strategies for *ex situ* gene preservation of populations as genetic resources.

Acknowledgements

We thank M. Kubota, H. Yamada, H. Miyashita, M. Miura, Y. Kawai-Munehara and other members of the Forest Tree Breeding Center and Kyushu University, for kind collaboration in field surveys and cone samplings and also the continued advices and encouragements during the present study. We express our special thanks to M. Shimoyama for her assistance in the DNA analyzes. The present study was funded by the Ministry of Education, Culture, Sports, Science, and Technology of Japan (Grant-in-Aid for Scientific Research, No. 17K07853).

Conflicts of Interest

The authors declare no conflict of interest.

Data Archiving Statement

Location, size and genotype data of adult tree population and seed pools are available via online as the Supplementary Materials.

References

Arista M, Talavera S (1997) Gender expression in *Abies pinsapo* Boiss., a Mediterranean fir. Ann Bot 79:337–342.

Austerlitz F, Smouse PE (2001) Two-generation analysis of pollen flow across a landscape. II. Relation between Φ_{FT}, pollen dispersal, and inter-female distance. Genetics 157:851–857.

Cornuet JM, Luikart G (1996) Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. Genetics 144:2001–2014.

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5:615–627.

El-Kassaby YA (2003) Clonal-row vs. random seed orchard designs: mating pattern and seed yield of western hemlock (*Tsuga heterophylla* (Raf.) Sarg.). For Genet 10:121–127.

El Mousadik A, Petit RJ (1996) High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. Theor Appl Genet 92:832–839.

Finkeldey R (1995) Homogeneity of pollen allele frequencies of single seed trees in *Picea abies* (L.) Karst. plantations. Heredity 74:451–463.

Furukoshi T (1991) A gene bank project for preservation of forest tree genetic resources. In: Ohba K, Katsuta M (eds.) Forest tree breeding. Bun-eido Shuppan, Tokyo, pp 224–241.

Goto S, Shimatani K, Yoshimaru H, Takahashi Y (2006) Fat-tailed gene flow in the dioecious canopy tree species *Fraxinus mandshurica* var. japonica revealed by microsatellites. Mol Ecol 15:2985–2996.

Goudet J (2002) FSTAT version 2.9.3, a program to estimate and test gene diversities and fixation indices [computer program]. Available from: http://www.unil.ch/izea/softwares/fstat.html.

Guichoux E, Lagache L, Wagne, S, Chaumeil P, Leger P, Lepais O, Lepoittevi C, Malausa T, Revardel E, Salin F, Petit RJ (2011) Current trends in microsatellite genotyping. Mol Ecol Resour 11:591–611.

Hale ML, Burg TM, Steeves TE (2012) Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS ONE 7:e45170.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. New Forests 6:95–124.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620.

Hoban S, Schlarbaum S (2014) Optimal sampling of seeds from plant populations for ex-situ conservation of genetic biodiversity, considering realistic population structure. Biol Cons 177:90–99.

Hosius B, Leinemann L, Konnert M, Bergmann F (2006) Genetic aspects of forestry in the Central Europe. Eur J Forest Res 125:407–417.

Irwin AJ, Hamrick JL, Godt MJW, Smouse PE (2003) Significance multiyear estimate of the effective pollen donor pool for *Albizia julibrissin*. Heredity 90:187–194.

Iwaizumi MG, Takahashi M, Watanabe A, Ubukata M (2010) Simultaneous evaluation of paternal and maternal immigrant gene flow and the implications for the overall genetic composition of *Pinus densiflora* dispersed seeds. J Hered 101:144–153.

Iwaizumi MG, Takahashi M, Isoda K, Austerlitz F (2013) Consecutive five-year analysis of paternal and maternal gene flow and contributions of the gametic heterogeneities to overall genetic composition of *Pinus densiflora* dispersed seeds. Am J Bot 100:1896–1904.

Iwaizumi MG, Miyata S, Hirao T, Tamura M, Watanabe A (2018) Historical seed use and transfer affects geographic specificity in genetic diversity and structure of old planted *Pinus thunbergii* populations. For Ecol Manag 408:211–219.

Iwaizumi MG, Takahashi M, Yano K (2021a) Temporal variation in regeneration events affecting population structure in different size- and life-stages contributes to overall genetic diversity of natural *Zelkova serrata* population. J For Res 26:32–42.

Iwaizumi MG, Matsunaga K, Iki T, Yamanobe T, Hirao T, Watanabe A (2021b) Geographical cline and inter-seaside difference in cone characteristics related to climatic conditions of old planted *Pinus thunbergii* populations throughout Japan. Plant Species Biol 36:218–229.

Iwaizumi MG, Nasu J, Miyamoto N, Isoda K (2021c) Genetic variation of seed pools in two mast years and a genetic preservation strategy for a population of the endemic conifer *Abies veitchii* var. *Shikokiana*. J Jpn For Soc 103:172–179 (in Japanese with English summary).

Iwaizumi MG, Ohtani M, Yano K, Nasu J, Miyamoto N, Takahashi M, Ubukata M (2022) Accurate paternal and maternal immigrant gene flow analysis over three mast years detects relative levels of gametic heterogeneity components in a natural population of *Abies firma*. J For Res 27 (in press).

Iwasaki H, Uchiyama K, Kimura MK, Saito Y, Hakamata T, Ide Y (2019) Impact of a tree improvement program on the genetic diversity of sugi (*Cryptomeria japonica* D Don) plantations. For Ecol Manag 448:466–473.

Kanda Y (2013) Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marr Transplant 48:452–458.

Kohyama T (1982) Studies on the *Abies* population of Mt. Shimagare II. Reproductive and life history traits. Bot Mag Tokyo 95:167–181.

Li G, Yang M, Zhou K, Zhang L, Tian L, Lv S, Jin Y, Qian W, Xiong H, Lin R, Fu Y, Hou X (2015) Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. J Microbiol Biotechnol 25:1136–1145.

Loicelle BA, Sork VL, Nason J, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinails* (Rubiaceae). Am J Bot 82:1420–1425.

Marshall DR, Brown AHD (1975) The charge-state model of protein polymorphism in natural populations. J Mol Evol 6:149–163.

Mamiya Y (1988) History of pine wilt disease in Japan. J Nematol 20:219–226.

Miyamoto N, Fernandez-Manjarres JF, Morand-prieur M-E, Bertolino P, Frascaria-Lacoste N (2008) What sampling is needed for reliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)?. Ann For Sci 65:403.

Miyata M, Ubukata M (1994) Genetic variations of allozymes in natural stands of Japanese black pine. J Jpn For Soc 76:445–455 (In Japanese with English sumamary).

Morgante M, Vendramin GG, Rossi P (1991) Effects of stand density on outcrossing rate in two Norway spruce (*Picea abies*) populations. Can J Bot 69:2704–2708.

Mukasyaf AA, Matsunaga K, Tamura M, Iki T, Watanabe A, Iwaizumi MG (2021) Reforestation or genetic disturbance: a case study of *Pinus thunbergii* in the Iki-no-Matsubara coastal forest (Japan). Forests 12:72.

Nakanishi A, Tomaru N, Yoshimaru H, Manabe T, Yamamoto S (2005) Interannual genetic heterogeneity of pollen pools accepted by *Quercus salicina* individuals. Mol Ecol 14:4469–4478.

Nakayama M, Kobayashi Y (1981) *Pinus* Linn. In Asakawa S, Katsuta M, Yokoyama T (eds.) Seeds of woody plants in Japan: Gymnospermae. Japan Forest Tree Breeding Association, Tokyo, pp. 65-76 (in Japanese).

Nathan R, Katul CG, Horn HS, Thomas SM, Oren R, Avissar R, Pacala SW, Levin SA (2002) Mechanisms of long-distance dispersal of seeds by wind. Nature 418:409–413.

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York.

Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. Mol Ecol 13:1143–1155.

Petit RJ, El Mousadik A, Pons O (1998) Identifying population for conservation on the basis of genetic markers. Conserv Biol 12:844–855.

R Development Core Team (2011) R: A language and environment for statistical computing. http://www.R-project.org.

Ritland K (2002) Extensions of models for the estimation of mating systems using *n* independent loci. Heredity 88:221–228.

Robledo-Arnuncio JJ, Smouse PE, Gil L, Alia R (2004) Pollen movement under alternative silvicultural practices in native populations of Scots pine (*Pinus sylvestris* L.) in central Spain. For Ecol Manag 197:243–253.

Robledo-Arnuncio JJ, Austerlitz F, Smouse PE (2007) POLDISP: a software package for indirect estimation of contemporary pollen dispersal. Mol Ecol Notes 7:763–766.

Shiraishi S, Watanabe A (1995) Identification of chloroplast genome between *Pinus densiflora* Sieb. et Zucc. and *Pinus thunbergii* Parl. based on the polymorphism in *rbc*L gene. J Jpn For Soc 77:429–436 (in Japanese with English summary).

Sjogren P, Wyoni PI (1994) Conservation genetics and detection of rare alleles in finite populations. Conserv Biol 8:267–270.

Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. Evolution 55:260–271.

Troupin D, Nathan R, Vendramin GG (2006) Analysis of spatial genetic structure in an expanding *Pinus halepensis* population reveals development of fine-scale genetic clustering over time. Mol Ecol 15:3617–3630.

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13:921–935.

Volk GM, Richards CM, Reilley AA, Henk AD, Forsline PL, Aldwinckle HS (2005) Ex-situ conservation of vegetatively propagated species: development of seed-based core collection for *Malus sieversii*. J Am Soc Hort Sci 130:203–210.

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

Figure captions

Fig. 1. Map of the spatial distributions of the 70 mother trees in an investigated *Pinus thunbergii* population (in Fukuoka, Japan).

Fig. 2. Relationships between the number of sampled mother trees and the mean values of genetic diversity statistics of seed pools ((a) number of alleles: n_a; (b) effective number of alleles: n_e; (c) allelic richness: *AR*; (d) gene diversity: *HS*) based on seven microsatellite loci using a rarefaction curve. Gray and chick bars indicate the corresponding statistics values of adult trees (also shown in parentheses). Error bars indicate the upper and lower 95% confidence intervals based on the 100 times of collection trials per the specified mother tree numbers. A gray arrow for each figure objects the mother tree number where the statistics value of seed pools exceeded that of adult trees (over 100%).

Fig. 3. Relationship between the number of sampled mother trees and the mean value of genetic differentiation statistics (Weir and Cockerham's F_{ST}) between seed pools and adult trees, based on seven microsatellite loci using a rarefaction curve. Error bars indicate the upper and lower 95% confidence intervals based on the 100 times of collection trials per the specified mother tree numbers. A gray arrow objects the mother tree number where the statistics value has become below 0.005 and almost converged.

Supplementary Fig. 1. Relationship between the tree size (diameter at breast height: DBH) of mother trees and the genetic diversity statistics of the corresponding seed pools (effective number of alleles: n_e) based on Kendall's rank correlation coefficient (τ).

Supplementary Fig. 2. The kinship coefficient (f_{ij}: solid line) estimated for (a) the adult trees and (b) seed pools for each pairwise distance class at consecutive 50 m intervals using SPAGeDi software. The significantly positive f_{ij} value was estimated using 1,000

permutations and based on comparisons with the upper 95% confidence levels (upper dashed line); *: $P < 0.05$.

Table 1 Relationships between tree size (height and diameter at breast height (DBH)) and mean values of allelic diversity statistics of seven loci of the seedlings collected by 70 mother trees based on Kendall's rank correlation coefficient (τ).

	n_a	n_e	HS	$AR_{[22]}$	N_{ep}	t_m
Height	-0.067 ns	-0.074 ns	-0.088 ns	-0.067 ns	-0.067 ns	0.034 ns
DBH	-0.077 ns	**-0.170** *	**-0.162** *	-0.113 ns	-0.113 ns	-0.033 ns

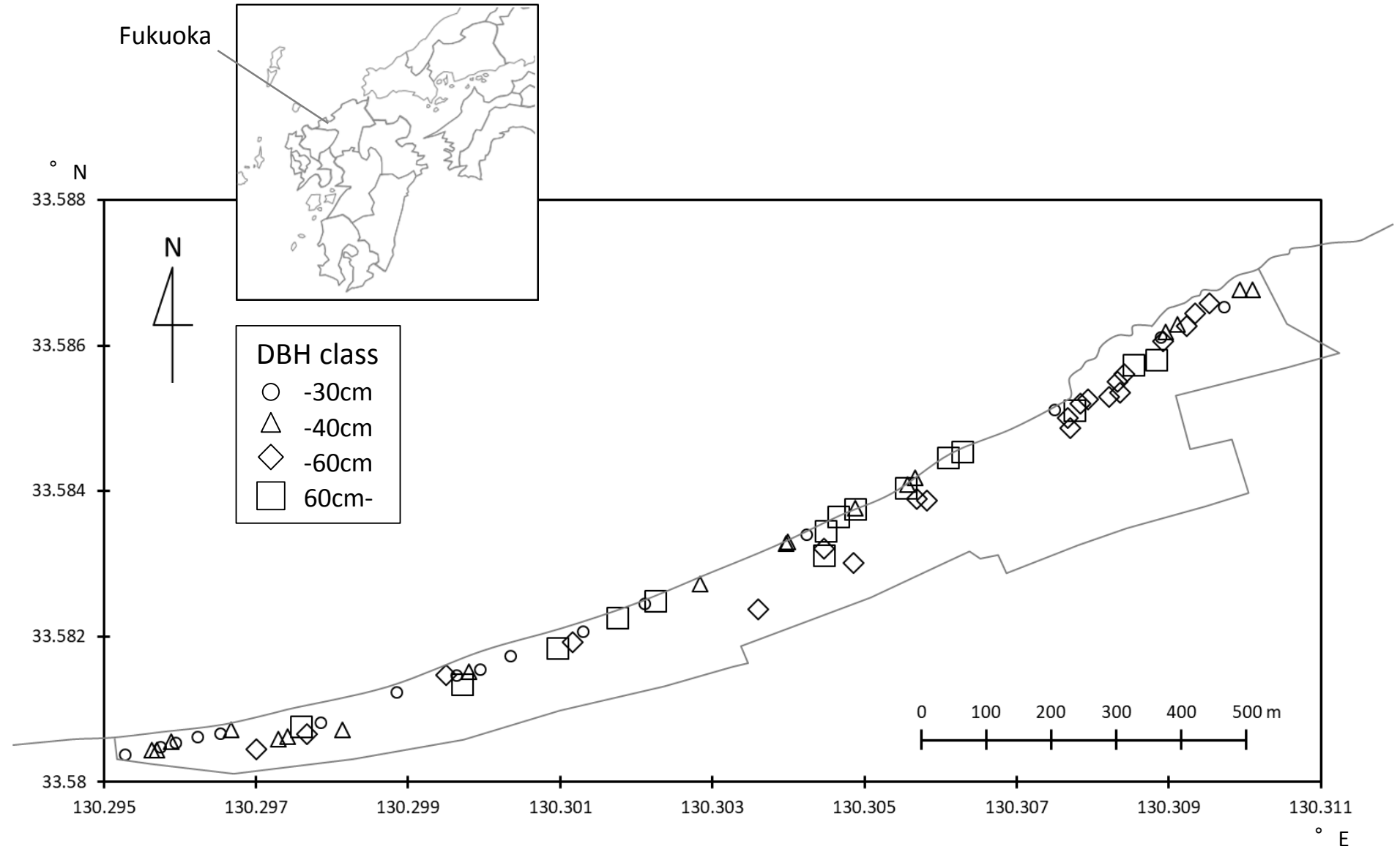
n_a: number of alleles; n_e: effective number of alleles; HS: gene diversity; AR: allelic richness; N_{ep}: effective number of pollen donors; t_m: multi-locus outcrossing rate. Marks with the coefficient values represent their significance levels; *: $P < 0.05$; ns: $P \geq 0.05$; and the underlined coefficient values indicate those with a significant relationship.

Table 2 The numbers of mother trees when allelic diversity of seed pools exceed the percentages of those of (1) adult tree population and (2) all sampled seed pools (70 mother trees).

Statistics	(1) Coverage-based (vs. Adults)						
	90%	92.5%	95%	97.5%	100%	102.5%	105%
n_a	27	33	39	44	51	59	69
n_e	5	5	6	7	10	11	20
AR	4	5	6	9	17	41	-
HS	2	2	3	5	9	-	-

Statistics	(2) Saturation-based (vs. All seed pools: 70 mother trees)					
	90%	92%	94%	96%	98%	100%
n_a	40	44	50	57	62	70
n_e	8	9	11	18	28	58
AR	5	6	9	12	21	70
HS	2	3	3	5	9	58

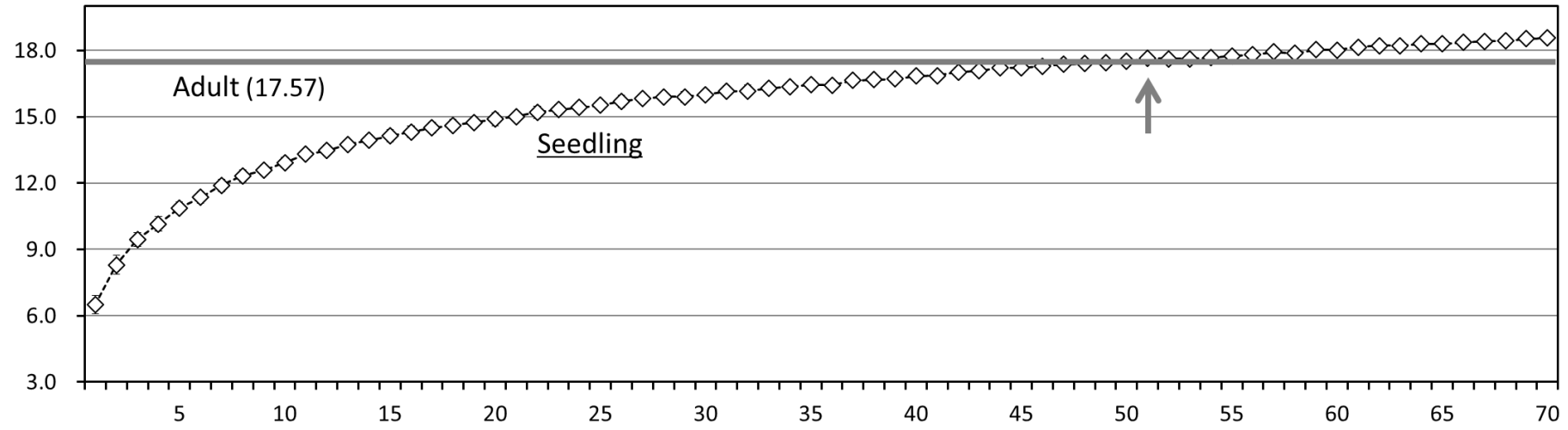
n_a: number of alleles; n_e: effective number of alleles; *AR*: allelic richness; *HS*: gene diversity.



Iwaizumi et al.

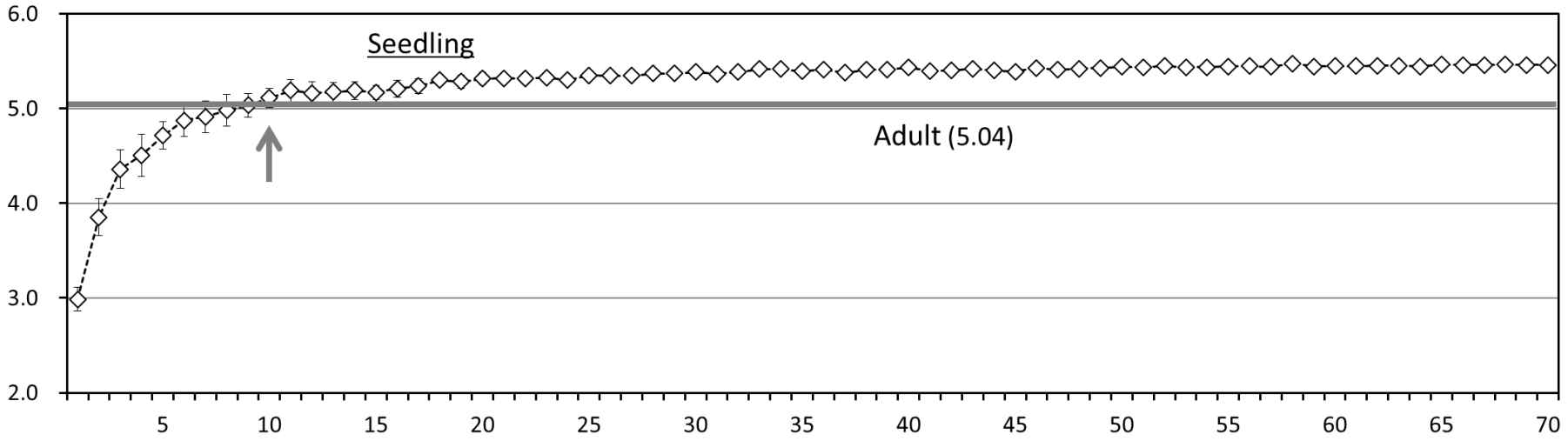
Fig. 1

(a) n_a



Mother trees

(b) n_e

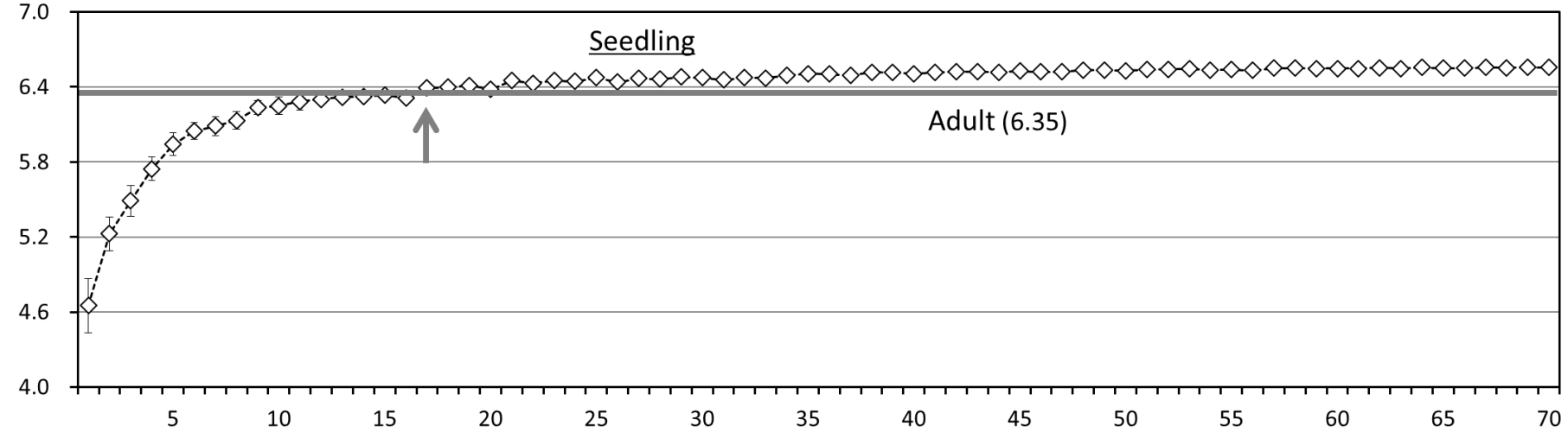


Mother trees

Iwaizumi et al.

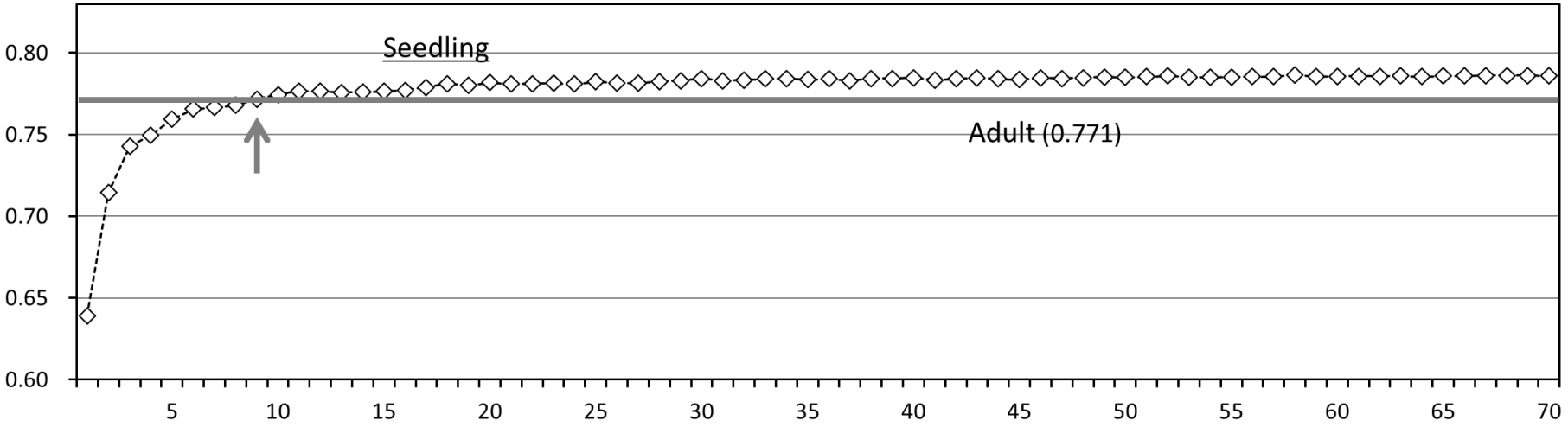
Fig. 2

(c) $AR_{[22]}$



Mother trees

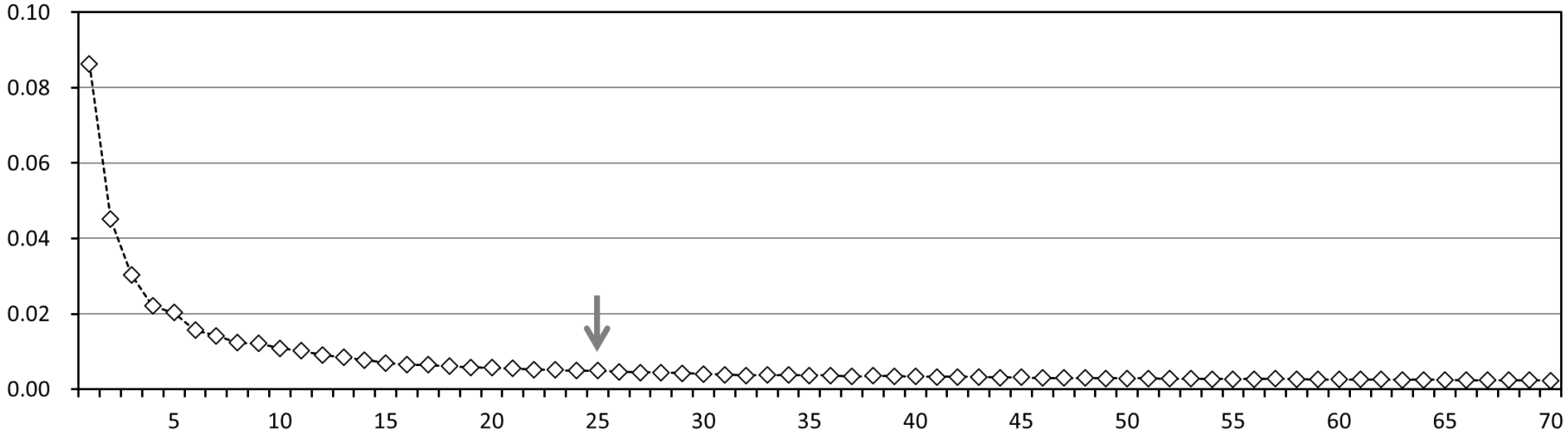
(d) *HS*



Mother trees

Iwaizumi et al.
Fig. 2 (continued)

F_{ST} (Adults vs. Seedlings)



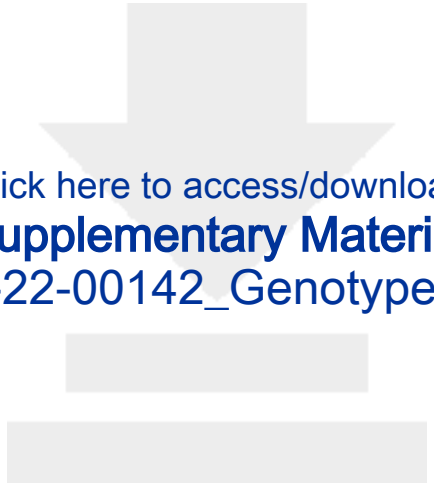
Iwaizumi et al.
Fig. 3

SuppTable


Click here to access/download
Supplementary Material
SuppTable1_220624.xls

SuppFigs

SuppData



Click here to access/download

Supplementary Material

TGGE-D-22-00142_GenotypeData.xlsx