Spatial genetic structure of the invasive tree *Robinia pseudoacacia* to determine migration patterns to inform best practices for riparian restoration

Sakiko Yaegashi¹, Tatsuo Omura², and Kozo Watanabe³,⁴

¹ Department of Civil and Environmental Engineering, University of Yamanashi, Kofu, Yamanashi prefecture, 400-8511, Japan

² New Industry Creation Hatchery Center (NIChe), Tohoku University, Aoba-yama 6-6-04, Sendai, 980-8579, Japan

³ Department of Civil and Environmental Engineering, Ehime University, Bunkyo-cho 3, Matsuyama, 790-8577, Japan

⁴ Department of Civil and Environmental Engineering, Ehime University, Bunkyo-cho 3, Matsuyama, Ehime 790-8577, Japan

**Corresponding author:** Kozo Watanabe, e-mail: watanabe_kozo@cee.ehime-u.ac.jp
Abstract

The black locust *Robinia pseudoacacia* (Robinieae, Fabaceae) is a common invasive riparian tree in Japan. There are less effective management strategies to remove the tree from the riparian area because of its quickly established high population. We investigated the expansion patterns of *R. pseudoacacia* through sympatric (i.e., between high and low water channel (HWC/LWC) within a study site) and allopatric (i.e., along river corridor) dispersal in the Tama River (Tokyo, Japan). Four microsatellites were used to examine the effects of gene flow on six populations in three sites. These subpopulations showed small genetic distance (i.e., no barrier or slightly limited) and genetically mixed population structure. It indicated that both sympatric and allopatric dispersal was active. Many migrants were younger individuals (i.e., less than 5 years old) and were found in the LWC area. Thus, the LWC could receive more migrants than the HWC through both types of dispersals. In addition, our age and genetic structure analyses reveal that recruited individuals likely settled immediately after the clearing project of *R. pseudoacacia* through sympatric dispersal. It appears that the migration by allopatric dispersal occurred following this. For the effective management of *R. pseudoacacia*, migrants should be removed regularly following initial removal of invaders during site restoration.

**Keywords:** invasive species, river management, gene flow, false acacia, elimination
Introduction

A fundamental challenge in conservation biology is to gain insights into the ecological process of invasions, such as the patterns of expansion and immigration strategies of alien species (Sakai et al. 2001, Washitani 2001). The black locust (Robinia pseudoacacia, Robinieae, Fabaceae) is a common invasive tree in riparian areas in Japan (Miyawaki and Washitani 2004). The black locust is native to the southern Appalachian Mountains in North America (Keresztesi 1980) but was introduced to the world for forestation and apiculture purposes (Keresztesi 1980, Cierjacks et al. 2013, Vítková et al. 2018). Its first introduction to Japan was reported in 1873 (Sakio 2009), and it has spread quickly throughout riparian areas (Maekawa and Nakagoshi 1997). According to a recent National Census on River Environments, R. pseudoacacia is now present in 84% of river basins in Japan (i.e., 97/115 basins) (Ministry of Land, Infrastructure, Transport and Tourism (MLIT), accessed in 2014).

The expansion of R. pseudoacacia has led to two primary problems in riparian areas. The first is the loss of native plant diversity due to invasion. For example, this has greatly affected the chrysanthemum (Aster kantoensis) in the Tama River in Tokyo (Washitani et al. 1997, reviewed in Cierjacks et al. 2013, Vítková et al. 2018). The second problem is the reduction of the flood-flow capacity within river channels. When flood events occur, R. pseudoacacia tends to be washed out, and its trunks prevent downstream water flow, resulting in high water levels during floods (Asaeda et al. 2010, Asaeda et al. 2011). Therefore, because of these issues related to ecological conservation and flood control, in 2002, the MLIT tried to remove trees of R. pseudoacacia in the area of S-R (Fig. 1) and the surrounding soils including their roots from the riparian area in the Tama River (Ogawa et al. 2011). However, the R. pseudoacacia population recovered after clearing.

Understanding the means for expansion can tell us about more effective methods to clear it from the riparian area. Propagation of R. pseudoacacia varies across the different phases of its life cycle, such as seeds, underground stems, basal shoots, epicormic branches, and damaged trunks. Several previous studies reported that allopatric dispersal (i.e., migrants carried by river water along riverine corridors, Fig. 2), occurred primarily via seed dispersal (Säumel and Kowarik 2013), whereas robust sympatric dispersal (i.e., reproduction occurred within small area, Fig. 2) was reported to be because of the extension of underground stems (Jung et al. 2009, Kurokochi et al. 2010, Asaeda et al. 2011). Seed dispersal occurs primarily in nearby areas because of gravity and wind (Robinson and Handel 1993, Morimot et al. 2010). Sprout outbreaks and the extension of underground stems have been studied using GIS and field digging observations (Jung et al. 2009, reviewed in Cierjacks et al. 2013). Several genetic analyses revealed the distribution of clones (Chang et al. 1998, Lian et al. 2009) and strong gene flow in basins (Lian et al. 2009, Kurokochi et al. 2010). Seeds showed inbreeding depression (Cierjacks et al. 2013, Yuan et al. 2013).
We predicted that allopatric dispersal from outside populations was more active in low-water channel (LWC), where is at lower elevation, near river channel, and frequently inundated (Fig. 2), than in high-water channel (HWC), which is at a considerably higher elevation and less inundated (Fig. 2). In contrast, sympatric dispersal in a native population was more common in HWC. Previous studies reported that the HWC showed higher densities of *R. pseudacacia*'s re-shootings than did the LWC (Asaeda *et al.* 2010, 2011, and 2014). Sapling density was reduced at middle elevations (i.e., 2–3 m) but was higher at low (i.e., 0–2 m) and higher elevations (i.e., >3m) (Asaeda *et al.* 2011). Although these phenomena suggest that the LWC, compared with HWC, has a different expansion process, this has not been tested genetically.

In this study, we examined the spatial genetic structure of *R. pseudoacacia* populations to infer its spatial expansion patterns focused on allopatric and sympatric dispersal. Our main objectives were: 1) to investigate the existence of two patterns of dispersal (i.e., sympatric and allopatric dispersal) by its gene flow, 2) to reveal how these dispersals were related to population establishment in HWC and LWC area, and 3) to estimate the process of population establishment after the restoration project. To approach these objectives, we examined spatial-genetic variation using highly polymorphic microsatellites for *R. pseudoacacia* sampled from three study sites, including a restoration site. Because microsatellites are able to reveal the genetic structure and dispersal pattern at a narrow-geographic scale (i.e., within a basin, plant community) (Lian *et al.* 2009, Kurokochi *et al.* 2010, and Jung *et al.* 2009, Love *et al.* 2013 Hevroy *et al.* 2018), we employed the microsatellites to determine migration pattern of the invasive locust tree *R. pseudoacacia* in Tama River (Tokyo, Japan) to consider efficient riparian restoration.

**Method**

**Study sites and sampling**

Field surveys were conducted in two subpopulations (LWC and HWC, Fig. 2) at three sites (six subpopulations in total) along the Tama River (Tokyo, Japan). The three sites are located at 53.5, 52.0, and 41.5 km (named site upstream (S-U), site restoration (S-R), site downstream (S-D), respectively; Fig. 1) from the river mouth. A restoration project had been conducted at the S-R km in 2002. Trees including *R. pseudoacacia* were removed from the LWC at the S-R to restore gravel bars. Simultaneously, surface soil, including tree roots, was also removed to prevent reproduction of *R. pseudoacacia* (Ogawa *et al.* 2011). The other two study sites were located upstream and downstream of the S-R. There is a weir that leads water into the Tamagawa aqueduct upstream of the S-U. The S-D is located downstream of the confluence of the Tama River and the other two tributaries. According to vegetation maps (Iketani 2001), *R. pseudoacacia* showed rapid expansion in all three sites. That is,
at the S-U, the expansion was from 1.2 ha in 1979 to 7.0 ha in 1995, and at the S-R, it was from 1.0 ha in 1979 to 8.55 ha in 1995). In particular, the *R. pseudoacacia* habitat at the S-D in 1995 expanded 12.5 times compared to that in 1979 (i.e., from 0.55 ha in 1979 to 6.85 ha in 1995). Takahashi and Minagawa (2007) reported that the population of *R. pseudoacacia* at S-R increased exponentially (i.e., from 12 individuals in 1,982 to 2,435 individuals in 1993). According to the reported water levels from January 1989 to December 2009 at 1.5 km downstream for site S-D (Fig. 1) (Water Information System by MLIT, accessed in 2014), the LWC area was inundated for 806 days (40.3 days/year on average). In contrast, the HWC was covered by water only for 3 days (0.15 days/year on average).

Between August 28 and September 18, 2009, we collected sprout leaves from each tree on the right-side shore at S-U and S-R and on the left-side shore at S-D. The range of other sampling sites was adequate to be nearly as large as that in S-R. The leaves were stored at -20°C until genetic analysis. We also recorded latitude, longitude, locations (i.e., HWC or LWC), and the circumference of the trunk at 120 cm height ($C_t$) from each tree that we sampled. A previous study reported that $C_t$ was correlated with the age of each tree (Takahashi and Minagawa 2007). Distance from the riverside of the main stream ($D_r$) was measured using the Google Earth tool on the basis of the most recent aerial photos of our field sampling.

DNA extraction and microsatellite genotyping

We genotyped 86 different trees from the three sites (Table 1). Each ~50 mg leaf was frozen in liquid nitrogen and crushed with a pestle. The crushed leaves were incubated at 60°C for 30 min with 500 µl of cetyltrimethyl ammonium bromide buffer. DNA was isolated with chloroform extraction and isopropanol precipitation. After the extraction, DNA was stored in TE buffer at -20°C. DNA concentrations were measured using a NanoDrop-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, UK).

We genotyped all samples at four microsatellites (Rops02, Rops04, Rops05, and Rops08; Lian and Hougetsu 2002). PCR products showed clear peaks during subsequent fragment analysis, even when our budget was limited. We ran PCR with a Veriti 96-well thermal cycler (Applied Biosystems, Foster City, California) in 10 µl reaction volumes with 10 ng template DNA, 0.25 U Taq DNA polymerase (TaKaRa, Tokyo, Japan), 1 × PCR buffer (TaKaRa), 2.5 mM MgCl$_2$, and 0.4 mM dNTP (TaKaRa). PCR cycling conditions were 2 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at annealing temperature for each primer (Lian and Hougetsu 2002), 1 min at 72°C, and a final extension step of 5 min at 72°C. We analyzed fragment sizes on an ABI 310 automated sequencer (Applied Biosystems) with GeneScan (Applied Biosystems). We estimated fragment sizes with a GeneScan 500 ROX size standard (Applied Biosystems). Although we employed fewer
microsatellites than standard-microsatellite studies, these markers reliably amplified our DNA. As several previous studies that focused on the gene flow of trees also employed four microsatellite loci (Marzouki et al. 2009, Pandey and Geburek 2010, Nielsen and Kjær 2010, Pandey and Geburek 2011), studying the evaluation of gene flow is possible.

Data analyses

We calculated five genetic indices: the number of alleles ($N_a$), the number of effective alleles ($N_e$), observed heterozygosity ($H_o$), expected heterozygosity ($H_e$), and the fixation index ($F_{is}$) within each population using GenAlEx (version 6.4;Peakall and Smouse 2006). We tested for linkage disequilibrium (LD) and deviation from the Hardy–Weinberg equilibrium (HWE) for each population and each locus using GENEPOP (version 4.2; Raymond and Rousset 1995) with a likelihood test (settings: dememorization, 1000; batches, 100; iteration per batch, 1000). We adjusted the significance levels using the Bonferroni correction. We estimated null allele (ENA) frequencies and global $F_{ST}$ with and without the ENA correction using FreeNA (Chapuis and Estoup 2007) and the EM algorithm (Dempster et al. 1977) with 1000 permutations.

We delineated populations using a model-based Bayesian clustering analysis of all 86 genotyped individuals in STRUCTURE (version 2.3.2; Pritchard et al. 2000). The correlated allele-frequency model (Falush et al. 2003) infers population structure where there are K genetic groups (K is unknown a priori). The model estimates the log likelihood (lnP[K]) for each K (named No. K-pop) and an optimal K can be chosen on the basis of the highest standardized second-order rate of change ($\Delta K$) of lnP(K) (Evanno et al. 2005). For comparing the relationship between subpopulations and those among sites, we also chose two and three as number of K-pop. We performed 20 runs of 100,000 iterations with a burn-in of 50,000 for each K ranging from 1 to 20 using the admixture model. We used a uniform prior for $\alpha$, the parameter representing the degree of admixture, with a maximum of 10.0, and set Alphapropsd to 0.05. Lambda, the parameter representing the correlation in the parental allele frequencies, was estimated in a preliminary run using $K = 1$. The prior $F_{ST}$ was set to default values (mean ± SE, 0.01 ± 0.05). We employed STRUCTURE HARVESTER (Earl and von Holdt 2012) to calculate $\Delta K$ and CLUMPP version 1.1.2 (Jakobsson and Rosenberg 2007) to find optimal alignment clusters across multiple runs with following approaches: LargeKGreedy method and the $G'$ pairwise matrix similarity statistics. Each individual was assigned to the genetic population having the highest $q$ value (i.e., assignment proportion to each K-pop).

Correlation analyses were conducted among each individual’s geographic distance from the riverside of the main stream, the assignment proportion to each K-pop with the most assigned individuals in each site ($Q_k$), and its trunk circumference ($C_t$) using the test for association/correlation
between paired samples in R (Best and Roberts 1975, Hollander and Wolfe 1973). In addition, seven measurements (i.e., \( N_e, N_a, H_o, H_e, F_{is}, \) mean geographic distance from the riverside, the mean 120 cm high circumference of the trunk) were compared between HWC and LWC with Wilcoxon rank sum and signed-rank tests in R (Bauer 1972, Hollander and Wolfe 1973).

A frequency-based assignment test (Peatkau et al. 2004) was performed to find putative migrants GenAlEx (version 6.4; Peakall and Smouse 2006). All individuals were divided into three prior-populations based on sampling sites. Each individual was assigned to the population with the highest log likelihood.

We calculated pairwise \( F_{ST} \) (Wright 1978) among the six subpopulations or two populations & two subpopulations (i.e., S-U, HWC and LWC in S-R and S-D) using Arlequin (version 3.1; Excoffier et al. 2005). The pairwise FST value of 0.14 was used as a criterion for genetic differentiation and as 0.26 for slight differentiation because Hamrick et al. 1993 reported that \( F_{st} \) was lower than ~0.14 in species with wind-promoted (= likely no barrier) gene flow and was ~0.26 in species with animal-promoted gene flow. The analysis of molecular variance (AMOVA) provided estimates of genetic differentiation at three hierarchical spatial levels: 1) among sites, 2) between HWC and LWC within a site, and 3) among individuals within a subpopulation, with 10,000 permutations.

Individual-based genetic distances \( (GD_i, Smouse and Peakall 1999) \) were calculated by GenAlEx (version 6.4; Peakall and Smouse 2006) to compare genetic differentiation among the following categories: individuals from 1) within a site or between sites, 2) HWC or LWC subpopulations within the same site, and 3) HWC or LWC between other sites (e.g., between HWC in S-U and in S-R) and LWC between other sites. The difference in average values in each category was tested by the Tukey–Kramer test using R v 3.5.3. The correlation between \( GD_i \) and \( F_{is} \) was tested by Spearman’s rank correlation test using R v 3.5.3.

Result

All four microsatellite loci were polymorphic (\( Na \) ranged from 8 to 18; see Supporting Information Table S1). Significant LD was detected between Rops02 and Rops05 at the HWC in S-R, and deviations from HWE were shown for Rops02, Rops05, and Rops04 (see Supporting Information Table S1). Null alleles were detected for all loci (Appendices S1 and S2). However, global \( F_{ST} \) among the six subpopulations with the ENA correction (= 0.070, range 0.009–0.129; see Supporting Information Table S1) was nearly identical to the global \( F_{ST} \) without the ENA correction (= 0.071,
range 0.005–0.122; see Supporting Information Table S1). Therefore, we did not exclude any loci from further analyses.

Mean \( H_o \) among the three subpopulations in LWC was significantly higher than that among subpopulations in HWC (d.f. = 11, \( V = 11, p < 0.05 \)). Mean \( F_{is} \) among HWC subpopulations was significantly higher than that among LWC subpopulations (d.f. = 11, \( V = 69, p < 0.05 \)). The trend of the other indices (i.e., \( N_n, N_e, \) and \( H_e \)) at S-U (i.e., the LWC is higher than the HWC) was different from that at the other reaches (i.e., the LWC are less than the HWC).

Bayesian clustering analysis delineated four populations among the 86 genotyped individuals (Fig. 3). The \( \Delta K \) was found in 4 populations. The dominant K-pops were complementary (i.e., \( K = 4; K4-p2 \) in S-U, \( K4-p1 \) in S-R, and \( K4-p4 \) in S-D). Half of the individuals in the LWC subpopulation in S-R were assigned to \( K4-p3 \). However, all K-pops were found in small numbers at all sites. In the case of \( K = 2 \), \( K2-p2 \) was found frequently at HWC in S-U and S-D, whereas \( K2-p1 \) was found in S-R. Comparing the relationship between \( K = 2 \) and \( K = 4 \), \( K2-p1 \) corresponded to \( K4-p1 \) and \( K4-p3 \), whereas \( K2-p1 \) corresponded to \( K4-p1 \) and \( K4-p3 \).

The assignment test showed that the proportion of self-assignment in individuals at HWC was higher than that in individuals at LWC in the S-U and S-R (Table 2). The overall self-assignment probability was 70.9%. The direction of the assignment was two-directional: from upstream into downstream (e.g., from S-U to S-R) and from downstream to upstream (e.g., from S-D to S-R). Most migrants (i.e., individuals assigned to non-sample sites) were less than 5 years old and located in the LWC area in S-U and S-R and in the HWC area in S-D (Fig. 1 and Table 2). In particular, migrants in S-R were more prevalent in the LWC area.

All \( F_{st} \) paths between HWC and LWC within a site did not show any barrier (i.e., <0.14, Table S3). The HWC in S-D showed strong connectivity in all subpopulations, whereas most links connected to the LWC in S-D meant genetic differentiation. On average, \( F_{st} \) in HWC, LWC, between HWC and LWC within a site, and between sites was 0.098, 0.122, 0.046, and 0.111, respectively. In the case of four populations (two sites and two subpopulations in S-R), all \( F_{st} \) s were <0.12 (i.e., no differentiation among site/subpopulations). The \( F_{st} \) values related to S-D were higher than those between S-U and S-R.

Of note, the average \( GD_i \) within a site was significantly lower than that between sites (\( p < 0.01 \), Fig.4). The average \( GD_i \) in the LWC subpopulation within the same site was also significantly lower than that in the HWC subpopulation within the same site (\( p < 0.01 \), Fig. 4a). In the comparison of \( GD_i \) between sites, the \( GD_i \) at the LWC was significantly lower than that at the HWC (\( p < 0.01 \), Fig.4b). In addition, the number at LWC between other sites was also significantly lower than between LWC and between LWC and HWC among other sites (\( p < 0.01 \), Fig. 4c). On average, \( GD_i \) s
were 8.84 at HWC within a site, 9.97 between sites, 7.41 at LWC within a site, 9.00 between sites, 8.57 between HWC and LWC within a site, and 9.46 between sites. Individuals having the same genotype (i.e., $GD_i = 0$) were found within the same HWC subpopulations, within the same LWC subpopulation, and among the HWC and LWC subpopulations in the same site. These genotypes were found in the range of 3–42 m. Although the $GD_i$ was not correlated to $Fst$ in the case using all subpopulation pairs ($r = 0.39$, $p > 0.05$), it showed significant positive correlation in the case of pairs excluding $Fst$ related to HWC in S-D ($r = 0.83$, $p > 0.01$). Thus, the pairwise $Fst$ values related to the HWC appeared to be higher due to the effect of sample size.

The AMOVA detected significant genetic variation at all three hierarchical levels, with 5.19% of the total variation partitioned to the among-sites level (d.f. = 2, $p < 0.01$), 4.22% between HWC and LWC subpopulations within a site (d.f. = 3, $p < 0.01$), and 90.59% among individuals within a subpopulation (d.f. = 166, $p < 0.01$).

The $Cs$ were correlated to several $Q_k$ ($Q_{k2.p2}$ in S-U, $r = 0.43$, $p < 0.05$; $Q_{k1.p2}$ in S-U, $r = 0.47$, $p < 0.05$) and the $Ds$ were also correlated to several $Q_k$ ($Q_{k1.p1}$ in S-R, $r = 0.39$, $p < 0.05$). The correlations between $D_i$ and $C_i$ were not significant at any of the sites. The mean $C_i$ in the LWC at S-U ($= 9.3$ cm) and S-R ($= 6.1$ cm) were significantly smaller than in the HWC (S-U $= 26.0$ cm and S-R $= 26.8$ cm) ($p < 0.01$), whereas no significant difference was found at S-D ($p > 0.05$). In S-R, migrants were more prevalent at the LWC, whereas the oldest individuals at the LWC were found at the border with HWC. In S-R, the oldest individuals at the LWC were found on the border with HWC.

Discussion

Gene flow of *Robinia pseudoacacia* in riparian areas

To address the invasion strategies of black locust *R. pseudoacacia* in the Tama River, we examined its genetic structure and gene flow using four microsatellites. The *R. pseudoacacia* in our study showed both sympatric and allopatric gene flow actively. We initially focused on the allopatric dispersal. The subpopulation-based genetic distance (i.e., pairwise $Fst$) within a site indicated no barrier, and the individual-based genetic distance (i.e., $GD_i$) also showed the closest relationship in other $GD_i$s. The individuals with the same genotype, who could be clones or seeds from the same origin, were found only within the same site. In addition, dominant K-pops were found in both subpopulations within a site. A previous genetic study also reported the occurrence of close relatives within approximately 50 m$^2$ (Kurokochi and Hougetsu 2014). These results indicated that active gene flow occurred in a narrow area. This active gene flow could be led by seed and pollen dispersal. The dispersal of *R. pseudoacacia* was conducted by wind in their native habitat (Robinson and Handel
Partial seed dispersal occurred close to the parents by gravity and wind (Morimoto et al. 2010, Cierjacks et al. 2013). Most pollen could be transported to nearby individuals by animal pollinators (Stacy et al. 1996) or by wind (Geng et al. 2008, Isagi et al. 2017). At our study sites, strong sympatric dispersal was believed to have occurred due to seed and pollen dispersal. However, it is difficult to determine its direction. HWC subpopulations could provide seeds to those of LWC as there were older individuals (i.e., more than 8 years old) at HWC in S-U and S-R. Reverse direction of gene flow led could also have possibly occurred by wind and animal. This needs to be further studied in the future.

The allopatric dispersal was also active. Furthermore, it may have occurred in two directions: from upstream to downstream (e.g., from S-U to S-D) and from downstream to upstream (e.g., from S-D to S-R), although we predicted the gene flow from upstream to downstream conducted by river water. Most pairwise $F_{st}$s indicated no barrier, and three of them related to the LWC subpopulation in S-D showed a little genetic differentiation (i.e., higher $F_{st}$ than wind promoted one). All individuals in each K-pop were found at all sites, although dominant K-pop was complicated. The migrant analysis also supported the exchange of individuals among study sites. The *R. pseudoacacia* was known to disperse at a long-distance from forested mountains or upland (Lian et al. 2009, Kurokochi et al. 2009, Love et al. 2013, Säumel and Kowarik 2013 Hevroy et al. 2018). Its pollen could disperse several kilometers (Lian et al. 2009, Kurokochi et al. 2014). Its seed could be carried by animals such as birds and humans (reviewed in Cierjacks et al. 1992, Vítková et al. 2017). Not only water flow but also these biological factors could contribute to the allopatric gene flow in two directions.

It was possible that the recruitments occurred in a site and migrants coming from other site grew and established their populations easier in the LWC area than in the HWC one. The $GD_S$ for LWC within a site or between sites were significantly lower than those for HWC. The fixation indices in the LWC showed lower than the HWC. These facts meant that the LWC subpopulations were more homogenized than the HWC subpopulations. In S-U and S-R, most migrants found in the LWC area and, the average age of LWC subpopulation was younger than that in the HWC subpopulation. It supported the fact that the LWC received more recruitments and migrants. In previous studies, the *R. pseudoacacia* required bright conditions for seed establishment (Groninger et al. 1997), and tends not to propagate new ramets in areas already occupied by former residents (Jung et al. 2009). These mean that open space is needed for their growth. Since *R. pseudoacacia* is vulnerable to flooding related to its height and the length of underground stems (Rechinger 1992, Asaeda et al. 2010), the old population can be washed away easily. As a result, a suitable environment to grow new populations had been created. Migrants were also present in the HWC which was likely to carry by past large flood or animals.
Establishment *R. pseudoacacia* population after clearing project

The restored site the S-R provides an opportunity to observe the recovery of *R. pseudoacacia* populations after clearing. The restoration project was conducted about 7 years ago before this study and removed not only tree bodies and also surface soils, including tree roots (Ogawa et al. 2011) to prevent population from re-establishing. Therefore, it is likely that the re-established population was derived from migration from outside of restoration area or remaining seed rather than from remaining roots in this area. However, the oldest seven individuals in LWC (named OL) were almost 6-7 years old, corresponding to the year of project completion.

The establishment of the new population at the LWC seemed to be started from the boundary of the HWC. The Bayesian clustering analysis indicated that different K-pops (i.e., K2-p1, K4-p1, and K4-p3) with other sites establish their populations in both HWC and LWC area. Most OL individuals also were assigned to these K-pops and located in the boundary between HWC and LWC. Moreover, the OL individuals distributed in the boundary between the HWC and the LWC areas. The migrants from HWC subpopulation could contribute to the population establishment at first rather than soil seed bank or remaining root. Next, the migration by allopatric dispersal occurred after the first establishment. Migrants from S-U found in the place close to the river and were quite younger (i.e., less than five years old). They might be carried by river water. The allopatric migration would be delayed than the sympatric one even though migrants came from close areas (i.e., about 1-1.5 km depart from). In conclusion, it is needed for effective clearing of *R. pseudoacacia* in riparian areas to remove not only its tree body but also surface soil. Subsequently, it is necessary to deal with individuals carried by allopatric dispersal.

Conclusion

We investigated the gene flow and genetic structure of the invasive tree *R. pseudoacacia* using four polymorphic markers in the three riparian sites and six subpopulations of the Tama River based on 86 individuals.

1. Both sympatric and allopatric dispersals were active in our study area. The genetic distance based on subpopulations indicated that no barrier level and the dispersals of pollen and seed would have occurred within a site by wind or animals. The allopatric dispersal also shown to be active (i.e., no barrier or a little limited) and two-directional (i.e., from upstream to downstream, from downstream and upstream). It could be conducted by seed dispersal not only river flow but also animal and pollen dispersal.
2. The LWC area seemed to have the potential to receive migrants from both other sites and the same site rather than the HWC. Most migrants were young (i.e., < 5) and distributed at the LWC area in S-R and S-U. The *R. pseudoacacia* needs bright and open environments such as the LWC area where old populations are washed away by frequent floods.

3. After the restoration project, the first migration seemed to occur at the boundary of HWC and LWC by sympatric dispersal. Subsequently, migrants could be carried by river flow.

4. To successfully clear *R. pseudoacacia* from riparian areas, it is necessary first to remove trees from whole area. As allopatric dispersal through the river corridor also provides migrants from outside, migrants should be removed regularly.
Acknowledgement

This research was supported by the Japan Society for the Promotion of Science (grant numbers 18K13858, 19H02276, and 19K21996) and the Japan River Front Research Center. K. Goto and Y. Mori helped with field work. The research was suggested by J. Sago. S. Suzuki helped with field and laboratory work. The anonymous reviewers gave us meaningful feedback. The English language in this manuscript was reviewed by Enago (www.enago.jp).
References

Asaeda, T., Gomes, P.I.A., Sakamoto, K. and Rashid, M.H. (2011), Tree colonization trends on a sediment bar after a major flood. *River Research and applications*, 27, 976-984.

Asaeda, T., Gomes, P.I.A. and Takeda, E. (2010) Spatial and temporal tree colonization in a midstream sediment bar and the mechanisms governing tree mortality during a flood event. *River Research and applications*, 26, 960-976.

Asaeda, T and Rashid, MDH. (2014) Modelling of nutrient dynamics and vegetation succession in midstream sediment bars of a regulated river, *International Journal of River Basin Management*, 12, 123-133.

Bauer, D. F. (1972) Constructing confidence sets using rank statistics. *Journal of the American Statistical Association*, 67, 687–690.

Best, D. J. and Roberts, D. E. (1975) Algorithm AS 89: The Upper Tail Probabilities of Spearman’s rho. *Applied Statistics*, 24, 377–379.

Chapuis, M.P. and Estoup, A. (2007) Microsatellite null alleles and estimation of population differentiation. Molecular Biology and Evolution 24, 621-631.

Chang, C., Bongarten, B. and Hamrick, J. (1998) Genetic structure of natural populations of black locust (*Robinia pseudoacacia* L.) at Coweeta, North Carolina. *Journal of Plant Research*. 111, 17–24.

Cierjacks, A., Kowarik, I., Joshi, J., Hempel, S., Ristow, M., von der, Lippe, M. and Weber, E. (2013) Biological Flora of the British Isles: *Robinia pseudoacacia*. Journal of Ecology, 101, 1623-1640.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B*, 39,1-38.

Earl, D. A., and von Holdt, B. M. (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361.

Evanno, G., S. Regnaut and J. Goudet. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology, 14, 2611–2620.

Excoffier, L. G. Laval and S. Schneider. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, 1, 47–50.
Falush D., Stephens, M. and Pritchard, J. K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics, 164, 1567–1587.

Geng, Q., Lian, C., Goto, S., Tao, J., Kimura, M., Islam, M.S. and Hogetsu, T. (2008), Mating system, pollen and propagule dispersal, and spatial genetic structure in a high-density population of the mangrove tree Kandelia candel. Molecular Ecology, 17, 4724-4739.

Guo, Q., Wang, J. X., Su, L. Z., Lv, W., Sun, Y. H., and Li, Y. (2017) Development and Evaluation of a Novel Set of EST-SSR Markers Based on Transcriptome Sequences of Black Locust (Robinia pseudoacacia L.). Genes, 8, 177.

Groninger, J. W., Zedaker, S. M., and Fredericksen, T. S. (1997) Stand characteristics of intercropped loblolly pine and black locust, Forest Ecology and Management, 91, 221-227.

Hamrick, J., Murawski, D., and Nason, J. (1993). The Influence of Seed Dispersal Mechanisms on the Genetic Structure of Tropical Tree Populations. Vegetatio, 107/108, 281-297.

Hevroy, T. H., Moody, M. L., and Krauss, S. L (2018) Population genetic analysis reveals barriers and corridors for gene flow within and among riparian populations of a rare plant, AoB PLANTS, 10, plx065.

Hollander, M. and Wolfe D. A. (1973) Nonparametric Statistical Methods. New York: John Wiley & Sons. 185–194.

Hollander, M. and Wolfe, D. A. (1973) Nonparametric Statistical Methods. New York: John Wiley & Sons. 68–75.

Iketani, H. (2001). Riparian habitat of Robinia pseudoacacia on Tama River. The Tokyu Foundation for Better Environment. Tokyo (in Japanese). (Available from: http://www.tokyuenv.or.jp/wp/wp-content/uploads/2011/02/e977849cdd987e08226668d683757c757.pdf)

Isagi, Y., Saito, D., Kawagushi, H., Tateno, R. and Watanabe, S. (2007), Effective pollen dispersal is enhanced by the genetic structure of an Aesculus turbinata population. Journal of Ecology, 95: 983-990.

Jakobsson, M., and Rosenberg, N. A. (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23, 1801–1806.

Jung, S. C., Matsushita, N., Wu, B.-Y., Kondo, N., Shiraishi, A., and Hogetsu, T. (2009) Reproduction of a Robinia pseudoacacia population in a coastal Pinus thunbergii windbreak along the Kujukurihama Coast, Japan. Journal of Forest Research, 14(2), 101–110.
Keresztes, B. (1980) The black locust. Unasylva, 32, 23-33.

Kurokochi, H., Toyama, K., and Hogetsu, T. (2010) Regeneration of Robinia pseudoacacdia riparian forests after clear-cutting along the Chikumagawa River in Japan. Plant Ecology, 210, 31–41.

Kurokochi, H., and Hogetsu, T. (2014). Fine-scale initiation of non-native Robinia pseudoacacdia riparian forests along the Chikumagawa River in central Japan. Journal of Ecology and Environment, 37(1), 21–29.

Lian, C., Kimura, M., Sakio, H. and Hogetsu, T. (2009) Characteristic of propagation on Robinia pseudoacacdia using microsatellites. In Sakio, H. (Eds.), Ecology of Robinia pseudoacacdia – invasion history, utility, ecology and management-. Bun-ichiSogo Shuppan, Tokyo (in Japanese).

Lian, C., and Hogetsu, T. (2002) Development of microsatellite markers in black locust (Robinia pseudoacacdia) using a dual-supression-PCR technique, Molecular Ecology Notes, 2, 211-213.

Lian, C., Oishi, R., Miyashita, N. and Hougetsu. T. (2004) High somatic instability of a microsatellite locus in a clonal tree, Robinia pseudoacacdia. Theoretical and Applied Genetics, 108, 836–841.

Love, H. M., Maggs, C. A., Murray, T. E. and Provan, J. (2013) Genetic evidence for predominantly hydrochoric gene flow in the invasive riparian plant Impatiens glandulifera (Himalayan balsam), Annals of Botany, 112, 1743–1750.

Maekawa, M., and Nakagoshi, N. (1997) Riparian landscape changes over a period of 46 years, on the Azusa River in Central Japan. Landscape and Urban Planning, 37, 37–43.

Marzouki, H., Nasri, N., Jouaud, B., Bonnet, C., Khalidi, A., Bouzid, S., and Fady, B. (2009) Population Genetic Structure of Laurus nobilis L. Inferred From Transferred Nuclear Microsatellites, Silvae Genetica, 58, 270-276.

Mishima, K., Hirao, T., Urano, S., Watanabe, A. and Takata, K. (2009). Isolation and characterization of microsatellite markers from Robinia pseudoacacdia L. Molecular ecology resources, 9, 850-852.

Miyawaki, S., and Washitani, I. (2004) Invasive Alien Plant Species in Riparian Areas of Japan: The Contribution of Agricultural Weeds, Revegetation Species and Aquacultural Species. Global Environmental Research, 8, 89–101.

Morimoto, J., Kominami, R. and Koike, T. (2010) Distribution and characteristics of the soil seed bank of the black locust (Robinia pseudoacacdia) in a headwater basin in northern Japan. Landscape and Ecological Engineering, 6, 193–199.
Nielsen, L.R., and Kjær, E.D. (2010) Gene flow and mating patterns in individuals of wych elm (*Ulmus glabra*) in forest and open land after the influence of Dutch elm disease. *Conservation Genetics*, 11, 257–268.

Ogawa, G. Naito, M., and Yoshimura, M. (2011) Restoration project of gravel bars of the Tama River: study of the detailed implementation plan, Japan River Front research Center report, 22, 86-95 (in Japanese with English abstract).

Pandey, M., and Geburek, T. (2011). Fine-scale genetic structure and gene flow in a semi-isolated population of a tropical tree, *Shorea robusta* Gaertn. (Dipterocarpaceae). *Current Science*,101, 293-299.

Pandey, M., and Geburek, T. (2010) Genetic differences between continuous and disjunct populations: some insights from sal (*Shorea robusta* Roxb.) in Nepal. *Conservation Genetics*, 11, 977–984.

Peakall, R., and Smouse, P. E. (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6(1), 288–295.

Peatka, D., Slade, R., Burden, M. and Estoup, A. (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology*, 13, 55-65

Pritchard J. K, Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Raymond, M. and Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248-249.

Robinson, G, and Handel, S. (2002). Forest Restoration on a Closed Landfill: Rapid Addition of New Species by Bird Dispersal. *Conservation Biology*, 7, 271 - 278.

Sakai, A. K., Allendorf, F. W., Holt, J. S., Lodge, M., Molofsky, J., With, K. A., … Weller, S. G. (2001). The population biology of invasive species. *Annual review of ecology, evolution, and systematics*, 33, 305–332.

Sakio, H. (2009). Ecology of *Robinia pseudoacacia* – invasion history, utility, ecology and management-, Bun-ichiSogo Shuppan, Tokyo (in Japanese).

Säumel, I. and Kowarik, I. (2013) Propagule morphology and river characteristics shape secondary water dispersal in tree species. *Plant Ecology*, 214, 1257-1272.
Smouse, P., and Peakall, R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, 82, 561–573.

Stacy, E. A., Hamrick, J. L., Nason, J. D., Hubbell, S. P., Foster, R. B. and Condit, R., Pollen dispersal in low-density populations of three neotropical tree species, *The American Naturalist*, 148, 275-298.

Takahashi, T. and Minagawa, T. (2007) Vegetation change analysis of invasive alien species, *Robinia pseudoacacia*, using tree census data and multitemporal vegetation maps. Proceedings of hydraulic engineering, 51, 1261-1266 (in Japanese with English abstract).

Vítková, M., Müllerová, J., Sádlo, J., Pergl, J., and Pyšek, P. (2017) Black locust (*Robinia pseudoacacia*) beloved and despised: A story of an invasive tree in Central Europe. *Forest Ecology and Management*, 384, 287-302.

Washitani, I. (2001) Plant conservation ecology for management and restoration of riparian habitats of lowland Japan. *Population Ecology*, 43, 189–195.

Washitani, I., Takenaka, A., Kuramoto, N., and Inoue, (1997) K., *Aster kantoensis* Kitam., an endangered LWC endemic plant in Japan: Its ability to form persistent soil seed banks. *Biological Conservation*, 82, 67-72.

Wright S (1978) Evolution and the Genetics of Populations. Variability within and among natural populations. Vol. 4. The University of Chicago Press, Chicago.

Yuan, C. Q., Li, Y. F., Wang, L., Zhao, K. Q., Hu, R.Y., Sun, P., Sun, Y.H., Li, Y., Gu, W.X. and Zhou, Z.Y. (2013) Evidence for inbreeding depression in the tree *Robinia pseudoacacia* L. (Fabaceae). *Genetics and molecular research*: GMR, 12, 6249-6256.
Figure legends

Figure 1. Map of the study area indicating study sites (a: all sampling sites, b: S-R, c: S-U, d: S-D). Each circle indicated the sampling location of each individual. Circle color distinguish subpopulation (i.e., HWC or LWC) and size stated the age of individuals. Capital letters indicated individuals assigned to other population. U = assigned to S-U, R = assigned to S-R, D = assigned to S-D.

Figure 2. The image of migration flow in riparian area. Arrows indicated how migrants come.

Figure 3. The result of Bayesian clustering analysis (a, b) and the age of each individual (c). Samples are sorted on the basis of whether they come from HWC (left) or LWC (right) areas. Individuals are subsequently sorted by their the 120cm high circumference of trunk (C). Each upper rectangle, containing colored vertical lines, indicates each individual’s proportion of assignment to each of K populations (K = 2 and 4, in a, b) based on hierarchical Bayesian clustering analysis. Each lower rectangle indicates the majority of K-pop (in a, b).

Figure 4. Comparison individual based genetic distance ($G_D$). a) b) within the same site, c) between the other site (e.g., between HWC in S-U and HWC in S-R). HWC = comparison between individuals sampled from HWC, LWC = comparison between individuals sampled from LWC, and HWC/LWC = comparison between individuals sampled from HWC and LWC. Bold line is the median, the boxes means the range of 25-75% range, and outsiders were open circle. The small letters indicated that each combination showed $p < 0.01$. 
Table 1. Genetic diversity of *Robinia pseudoacacia* at each site measured at four polymorphic microsatellite loci. $N$ = number of individuals genotyped, $D_t$ = average distance from the riverside, $C_t$ = average of the 120cm high circumference of trunk (cm), $N_a$ = average number of alleles per locus, $N_e$ = number of effective alleles, $H_o$ = observed heterozygosity, $H_e$ = expected heterozygosity, and $F_{is}$ = fixation index.

| Site | Population | $N$ | $D_t$ | $C_t$ | $N_a$ | $N_e$ | $H_o$ | $H_e$ | $F_{is}$ |
|------|------------|-----|-------|-------|-------|-------|-------|-------|--------|
| S-U  | HWE        | 11  | 35.7  | 26.0  | 5.25  | 3.517 | 0.386 | 0.708 | 0.423  |
|      | LWC        | 12  | 58.3  | 9.3   | 8.25  | 6.169 | 0.521 | 0.849 | 0.337  |
|      | All        | 23  | 17.3  | 47.5  | 9.25  | 5.405 | 0.457 | 0.808 | 0.411  |
| S-R  | HWE        | 19  | 81.5  | 26.8  | 9.50  | 5.590 | 0.487 | 0.789 | 0.337  |
|      | LWC        | 18  | 83.7  | 6.1   | 8.50  | 4.637 | 0.583 | 0.771 | 0.241  |
|      | All        | 37  | 82.6  | 16.8  | 11.25 | 5.879 | 0.534 | 0.795 | 0.314  |
| S-D  | HWE        | 18  | 43.8  | 28.3  | 7.75  | 4.326 | 0.333 | 0.728 | 0.535  |
|      | LWC        | 8   | 35.9  | 37.6  | 4.50  | 2.674 | 0.438 | 0.515 | 0.028  |
|      | All        | 26  | 41.4  | 31.2  | 8.00  | 4.080 | 0.365 | 0.677 | 0.452  |
Table 2. The result of the assignment analysis.

| Sampled site | Assigned site | No. individuals | Proportion of | Self-assignment | Non-self-assignment |
|--------------|---------------|-----------------|---------------|-----------------|---------------------|
|              | S-U           |                 |               |                 |                     |
| HWC          | 10            | 0               | 1             | 0.909           | 0.091               |
| LWC          | 5             | 4               | 3             | 0.417           | 0.583               |
|              | S-R           |                 |               |                 |                     |
| HWC          | 2             | 16              | 1             | 0.842           | 0.158               |
| LWC          | 6             | 12              | 0             | 0.667           | 0.333               |
|              | S-D           |                 |               |                 |                     |
| HWC          | 3             | 4               | 11            | 0.611           | 0.389               |
| LWC          | 0             | 1               | 7             | 0.875           | 0.125               |
Figure 3

a) K=2

b) K=4

c) Age

- ≥ 8 years old (C ≥ 31 cm)
- 6 - 7 years old (C ≥ 31 - 13 cm)
- ≤ 5 years old (C < 13 cm)
Figure 4

a) All sites

b) Within the same site

c) Between the other sites

Within a site Between sites

HWC LWC

HWC LWC

G_{ij}