# Shared Loci Correspondence Analysis: A New Method for Full-sib Reconstruction with Genetic Polymorphic Data

January 2009

**Nozomu KOKUVO**

# Shared Loci Correspondence Analysis: A New Method for Full-sib Reconstruction with Genetic Polymorphic Data

A Dissertation Submitted to

the Graduate School of Life and Environmental Sciences,

the University of Tsukuba

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy in Science

( Doctoral Program in Integrative Environmental Sciences )

**Nozomu KOKUVO**

# Contents

# Abstract

I proposed a simple visualization method, Shared Loci Correspondence Analysis (SLCA) to reconstruct full-sib patterns of haplodiploid females with co-dominant genetic marker data. The method begins with constructing a similarity matrix based on pairwise comparison of allele-sharing patterns among sampled individuals, followed by sorting with the correspondence analysis for displaying full-sib patterns among sampled individuals. In particular, SLCA does not require the estimation of allele frequency in a population. Therefore, SLCA would solve various problems in estimating social structure, mating behavior, and the dispersal range of wild insect populations, for which the collection of large amounts of data is impossible.

In Chapter 1, I first introduced SLCA algorithm and applied it to two types of actual microsatellite data. The results indicated that SLCA correctly distinguished multiple full-sib families of native bumble bees, *Bombus ardens* and *B. diversus*. In Chapter 2, I conducted a brief survey to estimate the population size of alien bumble bees, *B. terrestris*, in Biratori, Hokkaido, Japan. Using SLCA algorithm, I found that *B. terrestris* queens were foraging from many colonies in the field. In Chapter 3, I proposed a modified version of SLCA. Various types of simulation was conducted for comparing the accuracy of mSLCA with the original SLCA. Results indicated that mSLCA outperformed SLCA for various data settings, and it produced fairly accurate estimates with eight or more loci with eight or more alleles. Finally in chapter 4, I applied mSLCA to estimate the number of fathering males in native bumble bee colonies. I found no evidence for polyandory in native colonies of *B. ardens*, *B. diversus*, and *B. honshuensis*.

**Key words:** *Bombus* · full-sib reconstruction · haplodiploid species · microsatellites ·

monandrous · Shared Loci Correspondence Analysis · mSLCA

# List of Tables

# List of Figures

# General Introduction

There are a number of ecological areas of study that require methods to reconstruct the sibship structure from genetic data (Blouin 2003). In particular, knowledge of genetic relationships among haplodiploid individuals is necessary for studying the concept of inclusive fitness and the evolutional aspects of social behavior, as well as mating systems, parental care, and dispersal range of social insects. Recent progress in molecular ecology has made it possible to estimate sibship from molecular marker data even in the absence of parental information. With such information, the estimation of population size and resource sharing by pollinators, for example, can help lead to effective plant-pollinator conservation (Chapman et al. 2003; Knight et al. 2005).

How best to reconstruct a single-generation sample of individuals into full-sib families is an active area of research. Strict full-sib reconstruction is Non-deterministic polynomial time complete (NP-complete) (Beger-Wolf et al. 2005; Beger-Wolf et al. 2007), and heuristic methods are required in order to solve practical problems. Recently, several heuristic methods have been proposed (Almudevar and Field 1999; Beyer and May 2003; Blouin et al. 1996; Painter 1997; Smith et al. 2001; Thomas and Hill 2000; Thomas and Hill 2002), and several surveys have been conducted to test their accuracy and robustness of these methods (Butler et al. 2004; Konovalov et al. 2005; Beger-Wolf et al. 2007). In many cases, the accuracy of these methods largely depends on an accurate estimation of allele frequency in the sample population. Therefore, large-scale sampling or prior knowledge of the distribution of allele frequency is necessary. For the study of wild populations, however, there are many situations where we cannot fulfill these conditions: Not only is it impossible to have prior knowledge of allele frequency,

but large-scale sampling is also extremely difficult. An alternative method that does not require the accurate estimation of allele frequency is required for such cases.

In this study, I propose a simple visualization method for partitioning single-generation haplodiploid females into full-sib families with co-dominant marker data. I tested the validity of the algorithms using simulation data sets and real data sets of several bumble bee (Hymenoptera: Apidae) colonies, and then estimated the nest-mate patterns of Japanese native bumble bees. Bumble bees are important pollinators of crops as well as wild flowers in temperate regions. Colonies of most species are generally founded by a single once-mated queen, and various microsatellite DNA markers have already been proposed (Estoup et al. 1995).

In Chapter 1, I first introduce the Shared Loci Correspondence Analysis (SLCA) to reconstruct full-sib females, and apply it to two types of actual microsatellite data. In Chapter 2, I discuss a brief survey that was conducted, using SLCA algorithm, to estimate the number of colonies of alien bumble bees, *Bombus terrestris*. Then I discuss the impact of the intrusion of *B. terrestris* in Hokkaido, Japan. In Chapter 3, I propose a modified version of the SLCA (mSLCA) to reconstruct full-sib patterns of haplodiploid species. A large-scale simulation was conducted to compare the accuracy of the mSLCA with that of the original SLCA. Results indicated that mSLCA outperformed SLCA for various data settings, and it produced fairly accurate estimates with eight or more loci with eight or more alleles. Even if sample size is large, mSLCA produced accurate estimations with eight or more loci with eight or more alleles. I also investigate how these algorithms can produce accurate partitions with several types of real data sets. Accuracy levels and the utility of the algorithm are discussed. In Chapter 4, I use the mSLCA to estimate fathering males of native bumble bee colonies: there was no evidence for polyandry in any of the investigated colonies. Further extensions and applications of our algorithm are discussed.

# Chapter 1

# A simple visualization method to reconstruct nest-mate patterns among bumble bees (Hymenoptera, Apidae) using genetic data

## 1.1 Introduction

Recent developments in molecular ecology have made it possible to estimate kinship from molecular marker data (Blouin 2003; Jones and Ardren 2003). Sibling reconstruction is an effective approach to detect the overall family structure of a population, and is widely used for the study of conservation, evolution, and population ecology. Examples include the identification and subsequent management of threatened populations (Darvill et al. 2004; Knight et al. 2005), and it could also become an important tool for estimating social structure, mating behavior, and dispersal range of animal populations.

Bumble bees (Hymenoptera: Apidae) are annual eusocial insects where only the fertilized queens over-winter and start colonies in spring (Goulson 2003). Reproduction occurs at the end of the seasonal cycle, with the production of sons (drones) and daughters (young queens) that leave the colony and mate. They are important pollinators of crops as well as wild flowers in temperate regions. Therefore, knowledge of their resource sharing and space use is required for effective plant-pollinator conservation (Chapman et al. 2003). In particular, because most bumble bees nest

under ground, locating nests in the field is often difficult. There is little knowledge about their foraging behavior in the wild. They are also known as social insects, and each of their nests is founded by a single once-mated queen (Estoup et al. 1995; Schmid-Hempel and Schmid-Hempel 2000). Thus, all the workers in a nest are full siblings, and one can estimate nest mates, or sibling structure, using their genetic data.

Several statistical methods have been proposed to partition single-generation individuals into sibling groups with molecular marker data (Almudevar and Field 1999; Beyer and May 2003; Blouin et al. 1996; Painter 1997; Smith et al. 2001; Thomas and Hill 2000; Thomas and Hill 2002). For example, Painter (1997) used the Bayesian approach to estimate full sibling groups, which was applied to reconstruct the partition of nine falcons. Thomas and Hill (2000) used the maximum-likelihood approach to estimate quantitative genetic parameters. In many cases, the accuracy of these methods largely depends on the accurate estimation of allele frequency in a sample population. Therefore, large-scale sampling or prior knowledge of the distribution of allele frequency is necessary. For the study of wild populations, however, there are many situations where I cannot fulfill these conditions. It is impossible to have the prior knowledge of allele frequencies in a wild population, and it is difficult to estimate them. Alternative and practical method that does not require the accurate estimation of allele frequency is essential for these studies.

In this paper, I proposed a simple visualization method, Shared Loci Correspondence Analysis (SLCA), for partitioning one-generation females (i.e. workers and new queens) of haplodiploid species into full sibling families with co-dominant marker data. It does not require complex calculations, and visually shows inferred pedigree relationships in an instant. This allows us to easily understand the sibling relationships of a social insect population. In particular, since the algorithm does not require the accurate estimation of allele frequencies in a population, it can be applied to relatively small populations. Using data collected from wild bumble bees, I applied SLCA to two types of actual microsatellite data. First, I estimated the nest-mate patterns of 11 bumble bees

collected from three *Bombus ardens* colonies. Second, I estimated the sibling structure of 87 *Bombus diversus* workers collected from one colony. Further extensions and applications of SLCA will be discussed.

## 1.2 Materials and Methods

### 1.2.1 SLCA Algorithm

A list of haplodiploid females and their genotypes at each locus is input into the SLCA (Fig. 1 (a)). Individuals are assumed to belong to a single generation, and genotypes are assumed to be from neutral, unlinked, and co-dominant. In haplodiploid species, female individuals (workers and queens) are diploid, developing from fertilized eggs, and males are haploid, developing from unfertilized eggs. A male transmits a single copy of his genome to his daughter. Thus, a full-sibling pair of the haplodiploid species should share the same copy of genome inherited from their common father.

Considering haplodiploid genetics, SLCA first calculates a score for each sample pair and, using Correspondence Analysis (CA), it partitions individuals into estimated groups based on these scores. At this point, a shared-loci score that counts the number of loci where at least one allele is shared by a pair is calculated (Fig. 1 (b)). Once shared-loci score is assigned to every sample pair, a scoring matrix comprising pairwise shared-loci scores is created (Fig. 1 (c)). The resulting matrix is a similarity matrix, to which I applied CA (Fig. 1 (d)). CA is a method commonly used for categorical data, collected, for instance, in social surveys or linguistics (Benzecri 1992). Recently, it has also been applied for the study of ecology, especially for studying community structures and plant-pollinator interactions (Lewinsohn et al. 2006). I used CA to obtain clusters along the diagonal axis in the similarity matrix to visualize sibling structure of individuals.

With CA, higher-score pairs are ordered along the diagonal in the matrix (Fig. 1 (d)). An initial scoring matrix may appear to have no structure (Fig. 1 (c)), but by rearranging this matrix using CA, a pattern emerges along the diagonal (see Fig. 1 (e)). As the order of individuals is the same along rows and columns, the matrix is always

symmetrical along the diagonal. Moreover, the diagonal is always highest in the matrix. I finally used the number of the highest score groups as an estimate of colony number by detecting the boundaries of these groups along the diagonal, where $a_{i,i+1}$ elements of the matrix are not the highest (Fig. 1 (e)). Ideally, full sisters share at least one allele for every observed locus (Evans 1998). Thus, I determined that a pair sharing at least one allele for every locus is a full-sibling pair, and that they inherited one copy of genes from their common father.

I implemented the SLCA algorithm in Ruby script, EstNest 1.0, that called the correspondence analysis module in the MASS library provided by R (R Development Core Team 2004). The entire process was fully automated and the Ruby source code is available from the corresponding author.

### 1.2.2 Data Collection

To evaluate my algorithm, I used two types of actual data sets. The first was a sample of microsatellite genotypes for 11 *B. ardens ardens* Smith workers and new queens collected from three wild colonies. The second was a sample of 87 *B. diversus diversus* Smith workers collected from a wild colony.

The three colonies of *B. ardens* were sampled in July, 2005 from a small field in Tsukuba city (36° 06′ N, 140° 06′ E). One, two and eight bees were collected from the three colonies, respectively. Although these colonies were less than 5m apart, genotypes of individuals were genetically separate (Table 1 A). A large *B. diversus* colony was also sampled in October, 2004 on the campus of the University of Tsukuba, and 87 bees were collected. The bees were chilled immediately, and frozen for DNA extraction.

Workers and new queens were genotyped by microsatellite DNA loci. DNA was extracted from an entire middle leg. The leg was homogenized in a tube containing 30 $\mu$l lysis buffer (1 $mg/ml$ proteinase K, 0.01 M NaCl, 0.1 M EDTA, 0.01 M Tris-HCl (pH 8.0), 0.5% Nonidet P-40). The homogenate was incubated at 50°$C$ for 60 min and then at 94°$C$ for 10 min. After incubation, the homogenate was diluted with 270 $\mu$l TE buffer

(0.001 M EDTA, 0.01 M Tris-HCl (pH 8.0)), and used as a DNA template in polymerase chain reaction (PCR). Microsatellite DNA was amplified by PCR using locus-specific pairs of flanking oligonucleotide primers that were developed by Estoup et al. (1996). Polymorph patterns of five microsatellite loci (B11, B96, B100, B121, and B132) and six microsatellite loci (B11, B96, B100, B118, B121, and B 132) were analyzed for *B.ardens* and *B.diversus*, respectively.

One of each primer pair was fluorescence-labeled beforehand. PCR assays were conducted with 2 $\mu$l of each template DNA in a total reaction volume of 50 $\mu$l. The PCR reaction mix contained 0.2 ml each of dNTP, 2 mM $MgCl_2$, 1.25 units of Taq DNA polymerase (Amplitaq Gold, Applied Biosystems), and 0.4 $\mu$M each primer. All PCR reagents were purchased from Perkin Elmer Applied Biosystems. PCR conditions were an initial 9 min at 95°C, followed by 30 s at 94°C, 30 s at 50°C, and 30 s at 72°C for 30 cycles, and a final 7 min at 72°C. Each PCR product was added in an equal volume of formamide loading buffer (95% formamide, 20 mM EDTA, 0.05% xylene cyanol, 0.05% bromophenol blue), then denatured at 95°C for 3 min. About 3 $\mu$l of denatured products were electrophoresed with an external fluorescent-labeled size marker (Sizer 50-500, Pharmacia) in 6% denatured polyacrylamide gel, using an ALF DNA Sequencer-II (Pharmacia). Raw data gained from the sequencer were analyzed by Fragment Manager (Pharmacia), which gave the genotypes of microsatellite loci for each bee. Data gained from successfully amplified diploid females were used for further analysis. All the observed genotypes are described in Table 1. I estimated the nest-mate pattern of *B.ardens* and *B.diversus*, respectively by SLCA algorithm.

## 1.3 Results

The observed genotypes showed that all nest-mates shared at least one allele for every locus (Table 1). My method correctly identified nest mates for each of the three *B. ardens* colonies (Fig. 2).

My method also reconstructed the true sibling structure of a large *B. diversus*

*diversus* colony. As all individuals in a colony share at least one allele for every six loci (Table 1), all elements of the resulting score matrix were filled with the maximum number of shared loci, or 6.

## 1.4 Discussion

I have presented the SLCA algorithm, which clusters one-generation haplodiploid females into full sibling groups. It successfully reconstructed three *B. ardens* colonies using a small amount of data. It also correctly identified *B. diversus* data as a single family.

My method can be applied to all haplodiploid species, which share at least one allele for every loci, to cluster sample individuals into full-sibling groups. In particular, it does not require the accurate estimation of allele frequency. My algorithm, therefore, is suitable for estimating the sibling structure of a wild population, where I do not have detailed genetic information. Moreover, my algorithm can be applied to a small sample, where there is difficulty in estimating population allele frequency. For small samples, it is often difficult to estimate some ecological parameters, such as inbreeding coefficients. By altering the scoring method, SLCA can also be applied to other types of studies, such as the reconstruction of half-sib families.

To illustrate my algorithm, I used two types of actual microsatellite data, and considered further applications to wild insect populations. Compared with data generated by simulation experiments, it can be difficult to estimate the sibling structure from data collected from wild organism, because of possible natural errors or mutations. Actually, two individuals from colony D of *B.diversus* had a unique genotype (164,183) in B132 locus. This may have resulted from errors in genotyping data or mutations. Although these errors or mutations did not affect the total results in this case, the accuracy of my method will be affected if errors or mutations occur in paternal alleles. From a brief survey of simulation experiments, the variance of estimate slightly tends to increase in data consisting of large number of individuals (Kokuvo, unpublished data).

Further exploration is needed to test the validity against the number of colonies, number of loci, or other conditions, and comparisons with other methods.

How best to partition single-generation individuals into sibling families is an active area of research (Blouin 2003). Although several methods have been proposed, there is a problem in applying these methods to various types of data (Butler et al. 2004). In particular, most of these methods require accurate estimation or prior knowledge of population allele frequency. However, it is often difficult to estimate population allele frequency in wild populations, since, in many case, large-scale sampling cannot be conducted, and knowledge about their genetics is limited. Therefore, accurate reconstruction without the necessity of estimating allele frequency would enable us to study the genetics and ecology of wild populations. My method could play an important role in estimating sibling structures or conducting an instant survey of genetic relationships for wild populations.

# Chapter 2

# Estimating colony number of *Bombus terrestris* (Hymenoptera, Apidae) queens foraging at Biratori, Hokkaido, Japan

## 2.1 Introduction

*Bombus terrestris* is a widespread European species of bumble bees. Colonies of this species are relatively large, and workers visit a wide range of flowers. This efficient pollinator species has been commercialized in Europe, and commercial colonies of *B. terrestris* have been imported into many countries for pollination of agricultural crops in greenhouses. Since 1991, this species has been also imported into Japan for pollination of tomato plants in greenhouses (Ono 1998). More than 70,000 colonies of *B. terrestris* are imported from Europe each year, because they help to produce less labor intensive and pesticide free tomatoes (Kunitake and Goka 2006).

Japanese ecologists, however, have warned of the ecological risks caused by the biological invasion of this species (Goka et al. 2001; Washitani 1998). This exotic bumble bees has the potential to affect native bumble bees or the native pollination system in three ways (Goka et al. 2001): transmission of exotic parasites and pathogens (Goka et al. 2001), genetic disturbance caused by hybridyzation with native bees (Goka 1998), and competition for ecological niches, such as food or nesting sites (Hingston and Mcquillan 1999). In 1996, a feral colony of *B. terrestris* was found at last in the vicinity

of a greenhouse in Monbetsu in the Hidaka region of Hokkaido (Washitani and Matsumura 1998). Since then, several studies have been done to investigate the state of invasion and its ecological impacts on the native species mainly in Hokkaido (Matsumura et al. 2004; Inari et al. 2005; Goka et al. 2006).

Estimation of the population size of naturalized *B. terrestris* is needed as basic data for assessing ecological risks and eradicating the naturalized alien bees. Each bumble bee colony is usually founded by a single queen, and queens of *B. terrestris* are monoandrous (Estoup et al. 1995; Schmid-Hempel and Schmid-Hempel 2000). The number of reproductive individuals released varies greatly between colonies (Goulson 2003); therefore, the number of colonies, rather than the number of individuals, represents the effective population size. The number of colonies also represents propagule pressure of the species, because bumble bees are social an colonial insects. To prevent more expansion and establishment in the field, we should decrease propagule pressure on native ecosystems; however, locating bumble bee colonies in the field is often difficult because *B. terrestris* queens usually nest underground using pre-existing holes made by mice and voles (Goulson 2003).

Using genotypes of foraging worker bees, we can estimate the number of *B. terrestris* colonies with so-called full-sib reconstruction methods (Konovalov et al. 2005) because workers within a colony forms full siblings. Several surveys have been conducted to estimate the number of bumble bee colonies using genetic markers such as microsatellite DNA (Chapman et al. 2003; Darvill et al. 2004; Knight et al. 2005). Chapman et al. (2003) used a likelihood approach, which maximizes the likelihood of overall full-sib configurations of sample individuals (Thomas and Hill 2000; Wang 2004). Darvill et al. (2004) and Knight et al. (2005) used the pairwise approach, which infer the potential relationships of a pair of individuals, and then cluster them into sibling families. Maximum likelihood approaches require a large number of individuals because they need an accurate estimation of population allele frequencies. Most of these methods are computationally intensive and, in many cases, do not work successfully on large

challenging data set (Butler et al. 2004). Darvill et al. (2004) and Knight et al. (2005) used a pairwise likelihood approach, which infers the potential relationship of a pair of individuals, and then clusters them into sibling families. To study the initial phase of the intrusion, the pairwise approach would be one of a suitable method for estimating the number of foraging bumble bees with DNA data.

In chapter 1, I propose a simple pairwise scoring method, Shared Loci Correspondence Analysis (SLCA), to reconstruct nest-mate patterns of haplodiploid species. SLCA does not require the knowledge of population allele frequencies. Using haplodiploid genetics and Mendelian exclusion principal, SLCA enables the estimation of full siblings for haplodiploid species even with a small data set.

In this study, I conducted a brief survey for investigating the initial phase of the intrusion by *B. terrestris* in the field. In May 2005, I collected 39 *B. terrestris* queens foraging at Biratori, Hokkaido, Japan, and I estimated the population size of this alien bees using the SLCA algorithm and a pairwise method using a software KINSHIP (Darvill et al. 2004; Knight et al. 2005). The impact of the intrusion of *B. terrestris* on natural habitats will be discussed.

## 2.2 Materials and Methods

### 2.2.1 Sample Collection.

On four different days in May of 2005, 39 *Bombus terrestris* queens foraging on *Taraxcum* spp. and *Corydalis ambigua* were captured with a net over about two hours a day at six sites in Biratori, Hokkaido, Japan (Fig. 3).

In general, *Bombus* species have an annual life cycle (Goulson 2003). Queens emerge from hibernation in late winter or spring, and at this time of year they can often be seen searching for suitable nest sites. In 2005, commercial colonies were also used from March to August in the Biratori region (Yoneda et al. 2007); thus, our sample was possibly from commercial colonies in greenhouses or wild colonies.

In these regions, the use of commercial colonies of *B. terrestris* has rapidly increased

since 1997 (Inari et al. 2005), and about one thousand colonies were shipped into Biratori in 2005 (M. Yoneda, pers. comm.). Thus, this area is suitable for studying the initial phase of the intrusion process of *B. terrestris*. Each sample was kept in a plastic tube, chilled immediately, and later frozen for DNA extraction at - 80°$C$.

### 2.2.2 DNA Extraction and Microsatellite Genotyping.

Bees were genotyped by six microsatellite DNA loci. DNA was extracted from an entire middle leg of each sample. The leg was homogenized in a tube containing 30$\mu$l of lysis buffer (1 $mg/ml$ proteinase K, 0.01 M NaCl, 0.1 M EDTA, 0.01 M Tris-HCl (pH 8.0) 0.5% Nonidet P-40). The homogenate was incubated at 50°$C$ for 60 min and then at 94°$C$ for 10 min. After incubation, the homogenate was diluted with 270$\mu$l of TE buffer (0.001 M EDTA, 0.01 M Tris-HCl (pH 8.0)) and used as a DNA template in polymerase chain reaction (PCR). Six microsatellite loci (B11, B96, B100, B118, B121, and B132) were amplified by PCR using locus-specific pairs of flanking oligonucleotide primers that were developed by Estoup et al. (1996). PCRs were performed and polymorphic patterns of each microsatellite locus were analyzed, following protocols described in Chapter 1. The mean number of allele per locus and heterozygosity was 6.83 (range, 4 - 10) and 61.5%, respectively (Table 2).

### 2.2.3 Estimating the Number of Colonies

I estimated the number of colonies in a sample using Shared-Loci Correspondence analysis (SLCA) algorithms. As sample queens might be from commercial or wild colonies, the estimated number of colonies is the number of commercial colonies in 2005 and feral or commercial colonies in the previous year. SLCA is an algorithm for reconstructing nest-mate patterns among haplodiploid females. The confidence intervals and standard errors of the estimate were calculated by Jackknife resampling (Efron 1982).

I also used another pairwise method to estimate the number of colonies following the algorithm described in Knight et al. (2005). Sister relationships among individuals were

established using the likelihood function of KINSHIP 1.3.1 (Queller and Goodnight 1989) where $Rm = 0.5$ and $Rp = 1.0$. Confidence in the sister pair assignment was calculated from 1,000,000 simulations. To minimize type I errors, only sisters designated at $P \leq 0.001$ (the most stringent value that KINSHIP will return) were used for further analysis. The total number of colonies was calculated as the total number of sister groups found by KINSHIP . I did not estimate the number of colonies that were not sampled (the 'zero' bee category) because I was interested in the number of colonies from which the sampled bees came.

## 2.3 Results and Discussion

I conducted a genetic analysis to estimate the number of colonies of 39 *Bombus terrestris* queens foraging in the field of Biratori, Hokkaido, Japan. I found that queens were foraging from many colonies, suggesting that most bees found in the field come from different colonies. A total of 28.82 ± 1.62 families (mean ± $t_{\alpha}^{n-1}$SD, range 22-32, n=39) were detected by Jackknife resampling with SLCA. Estimated sister pairs for 39 individuals by KINSHIP and SLCA are shown in Fig. 4. Using whole individuals, 30 colonies were estimated by SLCA. Twenty-three were assigned to one individuals, 8 to two individuals, and 6 to three individuals. On the other hand, 31 colonies were estimated by KINSHIP. Twenty-three were assigned to one individuals, and the other 8 colonies were assigned to two individuals.

Accumulation of knowledge and quantitative data is necessary for a better understanding of the intrusion and the effective management of this efficient pollinator. Unfortunately, we cannot know the actual number of colonies in the wild, and we should be careful when evaluating this estimate.

Propagule pressure is a key factor of alien species to expand their distributions in a new circumstance (Williamson 1996). In order to calculate the propagule pressure of commercial *B. terrestris*, I estimated the effective population size, that is, the number of colonies, of the species. My results indicate that a large number of colonies is releasing

the bees into the wild, suggesting that the propagule pressure of the species is high in my study area. My estimate and the method used here would provide valuable information to discuss the impacts of commercial bee populations on native bee species. This method could also be used to estimate the number of colonies of foraging workers in the field.

In 2006, *B. terrestris* was listed as an invasive alien species by the Japanese government, but their importation and rearing is permitted on the condition that the greenhouse where they will be used is fully netted and colonies are burned or otherwise destroyed at the end of their useful life (Kunitake and Goka 2006; Goka 2006b; Goka 2006a). Proper usage of this efficient pollinator is essential both for Japanese agriculture and Japanese ecosystems. My present study shows, however, that a large number of colonies may be releasing bees into the wild. In order to allow the sustainable and continued use of this efficient pollinator, steps should be taken as soon as possible to prevent additional commercial bees from being released into the wild, and I should develop more efficient methods to control and eradicate the escaped alien bumble bees. Long-term monitoring studies and conducted vaster activities for the extermination of *B. terrestris* in the field have been conducted by the Japanese Society of Conservation Biology, and the results of annual "sighting" surveys are published on their web site (http://www003.upp.so-net.ne.jp/consecol/). Very valuable samples and data have been accumulated by the research team. The collaboration of various fields of scientists is required for more detailed understanding of this introduced species of bees.

# Chapter 3

# Modified SLCA algorithm for full-sib reconstruction of haplodiploid species

## 3.1 Introduction

A number of ecological areas require methods to reconstruct sibship structure from genetic data (Blouin 2003). In particular, knowledge of the genetic relationships among haplodiploid individuals is necessary for studying the concept of inclusive fitness and evolutional aspects of social behavior (Hamilton 1964), as well as mating systems, parental care, and dispersal range of social insects. In natural situations, however, actual sibling information is lacking to a greater or lesser degree (Herbinger et al. 2006). Recent developments in molecular ecology have made it possible to estimate the sibships from molecular marker data even in the absence of parental information (Blouin 2003; Konovalov et al. 2005; Herbinger et al. 2006). Using such information to estimate population size and resource sharing by pollinator bees, for example, can support effective plant-pollinator conservation (Chapman et al. 2003; Darvill et al. 2004; Knight et al. 2005).

Several methods have been proposed to assign single-generation individuals into full-sib families based on their genetic data (Painter 1997; Almudevar and Field 1999; Thomas and Hill 2000; Smith et al. 2001; Beyer and May 2003). Various approaches developed thus far can be characterized by how all the possible partitions are searched

and how these partitions are ranked to determine the best-fit partition (Konovalov et al. 2005). One simple approach is to estimate pairwise relatedness between all individuals and graphically cluster the pairs of individuals (Blouin et al. 1996). The approach of Almudevar and Field (1999) (denoted as AF), based on the Mendelian exclusion principal, constructs and evaluates sets of feasible sibling groups. The method of Butler et al. (2004) (denoted as Simpson), based on a simple exclusion method, chooses the best partition according to the Simpson index. Some methods are based on maximizing some form of the likelihood of proposed full-sib configurations and use a Markov Chain Monte Carlo (MCMC) method to select the best partition (Thomas and Hill 2000; Smith et al. 2001; Wang 2004). In particular, Smith et al. (2001) used the MCMC method for pairwise scoring (denoted as SC) and for estimating likelihood (denoted as Likelihood). Other methods are based on the likelihood ratio of null and hypothesized kin relationships (Goodnight and Queller 1999) and cluster the relationships by various methods (Beyer and May 2003; Darvill et al. 2004; Konovalov et al. 2004; Knight et al. 2005). For example, Beyer and May (2003) used a graph-based algorithm (denoted as Graph), and Konovalov et al. (2004) used 'Descending Ratio' (denoted as DR) search algorithms.

Butler et al. (2004) conducted a large-scale simulation study and compared the AF, Simpson, SC and Likelihood approaches at various parameter settings. Their results indicated that, in the absence of genotype errors, the exclusion based methods (AF, Simpson, and SC) could be more accurate than the likelihood based approach, especially when a small number of loci are considered. The Likelihood method produces accurate estimates only when sufficient alleles and loci were available. SC requires accurate population allele frequencies for estimation. The calculation speed of AF depends primarily on the family structure of the sample, and in some cases, AF failed to produce answers within a reasonable time. Konovalov et al. (2005) compared DR, Graph, Simpson, and a modified version of Simpson (denoted as MS) and found that MS outperformed the other three. But in the presence of unrelated individuals, MS was less

accurate than DR. I addition, MS requires a heuristic parameter, and there are theoretical limits of MS for highly unrelated data sets (Konovalov et al. 2005). Alternative and practical method that neither require heuristic parameter settings nor family structure-dependent accuracy are essential for the study of full-sib reconstruction.

In Chapter 1, I proposed a simple full-sib reconstruction method, Shared Loci Correspondence Analysis (SLCA), to estimate the nest-mate patterns of bumble bees. SLCA is based on a pairwise scoring method using the degree of shared alleles between individuals, and this method can be used with applied all haplodiploid species to reconstruct full-sib structure. In general, pairwise scoring approaches are simple to implement but lose valuable information (Thomas and Hill 2000; Herbinger et al. 2006), and they often require population allele frequencies (Thomas and Hill 2000). SLCA, however, solves these problems by focusing on haplodiploid genetics (haplodiploid full-sibs share at least one genome (a set of alleles across multiple loci) from their haploid father, and thus no valuable information is lost). In addition, it does not require population allele frequencies. SLCA could correctly assign real DNA data collected from natural bumble bee colonies into true configurations. I conducted additional surveys to study the accuracy of the estimates, using microsatellite genotypes of the commercial bumble bees (Kokuvo et al., unpublished data) and found that SLCA tended to overestimate the number of colonies as the number increased. In addition, SLCA tended to give large variances, and the accuracy of the algorithm depended much on the distribution of the data.

In this chapter, I propose a modified version of SLCA (mSLCA) to reconstruct full-sib patterns of haplodiploid species. Instead of the multi-level pairwise scoring used in SLCA, the approach taken here is a binary pairwise scoring based primarily on the Mendelian exclusion principal. Paradoxically, the blunting of sensitivity by pairwise comparisons reduces the variance of the estimates. The results of a large-scale simulation study to compare the accuracy of mSLCA with the original SLCA indicated that mSLCA outperforms the original SLCA for at almost all the parameter settings. I

also studied sensitivity to the number of loci, number of alleles per loci, and family size distribution. The accuracy and use of the algorithm are discussed.

## 3.2 Material and Methods

### 3.2.1 SLCA and mSLCA Algorithm

SLCA is a simple visualization method to reconstruct full-sib families of haplodiploid species. The algorithm fist assigns a pairwise score, the number of loci that share at least one allele, for all the sample pair. Then a similarity matrix is produced, which is clustered by the correspondence analysis. SLCA produces an estimate of the number of families by detecting the highest scoring groups. The details of the algorithm are described in Chapter 1.

The new algorithm replaces the scoring and the estimation methods. mSLCA first assigns a binary score to all sample pairs (Fig. 5 (a)). Distinguishing the pairs that share at least one allele at every locus from the other pairs, mSLCA gives the higher score to the former and the lower core to the latter (Fig. 5 (b)). Once a shared-loci binary score is assigned to every sample pair, a scoring similarity matrix is created (Fig. 5 (c)), to which I apply correspondence analysis (CA, Fig. 5 (d)). The algorithm clusters individuals using CA, gathering the higher score pairs together in a scoring matrix (Fig. 5 (e)). When the higher scoring pairs are gathered into the same group according to the order of individuals in the clustered matrix, the number of higher scoring groups become the number of families. I implemented the mSLCA in R (R Development Core Team 2004) using the correspondence analysis module in the MASS library. The entire process was fully automated, and the R source code is available from the author.

### 3.2.2 Simulation Data Sets

Three types of simulations were conducted to test the algorithm's ability to estimate the number of families for haplodiploid species. All data sets were designed to be similar to one of the simulation sets used by Beyer and May (2003) except that individuals were

assumed as haplodiploid. Parental genotypes, that is, pairs of haploid and diploid, were first generated for each combination of the number of loci and and the number of alleles per locus. According to Mendelian inheritance, I generated diploid genotypes for the descendants of each pair. For simplicity, I assumed that alleles are equally distributed at each locus and that there are no mutations, crossing over, genotyping errors, or null alleles. One hundred simulated data sets were run with each combination of parameters. Finally, boxplots of deviations from the actual number of colonies, and accuracy levels (percentages of correctly estimated data) were plotted using the R (version 2.0.1) software package (R Development Core Team 2004).

Simulation set A contains 50 haplodiploid females from five full-sib families. Four family distributions were used, with family sizes for each set at (10, 10, 10, 10, 10), (20, 10, 10, 5, 5), (30, 5, 5, 5, 5) and (40, 5, 2, 2, 1), respectively. Each family distribution was tested with 4, 6, 8, 10, and 12 loci, each with 4, 6, 8, 10 and 12 alleles per locus.

Simulation set B was designed to test the performance of mSLCA with large and small data sets. All parameters were identical to those used in simulation set A. I prepared equally distributed families with 25, 50, and 100 individuals, that is, (5,5,5,5,5), (10,10,10,10,10) and (20,20,20,20,20), respectively, were prepared.

Simulation set C was used to evaluate the effects of the number of families per data set on the algorithm's ability to estimate a correct number of families. The number of families was varied between 5 and 45, using 50 individuals divided randomly among the families. Tests were conducted with similar parameters to those used in simulation set A.

Finally, simulation set D was prepared to test the performance of mSLCA for larger data set. The number of families was varied from 50 to 150, with 200 individuals. I compared the estimated numbers of families with the true number of families with eight loci with eight alleles.

### 3.2.3 Real Data sets

In addition to the simulated data sets, I tested the algorithms on real data sets. I used 20 commercial *Bombus terrestris* colonies, all supplied by Api Co. (Gifu, Japan). Because most bumblebees are known to be monandrous and monogynous (Estoup et al. 1996; Schmid-Hempel and Schmid-Hempel 2000; Goulson 2003), workers collected from the same colony formed a full-sib group. Five workers from each colony were caught, chilled immediately, and later frozen for DNA extraction.

Bees were genotyped by six microsatellite DNA loci. DNA was extracted from an entire middle leg of each sample. The leg was homogenized in a tube containing 30$\mu$l of lysis buffer (1 $mg/ml$ proteinase K, 0.01 M NaCl, 0.1 M EDTA, 0.01 M Tris-HCl (pH 8.0) 0.5% Nonidet P-40). The homogenate was incubated at 50°$C$ for 60 min and then at 94°$C$ for 10 min. It was diluted with 270$\mu$l of TE buffer (0.001 M EDTA, 0.01 M Tris-HCl (pH 8.0)) and used as DNA template in a polymerase chain reaction (PCR). Microsatellite DNA was amplified by PCR using locus-specific pairs of flanking oligonucleotide primers that were developed by Estoup et al. (1996). Polymorphic patterns of six microsatellite loci (B11, B96, B100, B118, B121, and B132) were analyzed, following protocols described in Chapter 1. Data gained from successfully amplified bees were used for further analysis. The mean number of allele per locus and heterozygosity was 6.5 (range, 3 - 12) and 65.7%, respectively (Table 3).

I tested the robustness of SLCA and mSLCA with these actual DNA data to verify the validity of the results with an increasing number of colonies. Random numbers of individuals, ranging from 1 to 5, were withdrawn from each of 1 -10 colonies, randomly selected out of the 20, for analysis. A hundred hypothetical data sets were generated for each number of colonies, ranging from 1 to 10. The number of colonies estimated were compared with the actual numbers.

## 3.3 Results

### 3.3.1 Simulation Set A: Effect of Family Distributions

Boxplots of the deviations from the actual number of families are presented in Fig. 6. The distribution of family sizes had little effects on deviations in either SLCA or mSLCA. mSLCA correctly estimated the number of colonies at almost all parameter settings (Fig. 6b), whereas SLCA tended to overestimate it at all parameters (Fig. 6 (a)). Even with four loci and four alleles per loci, mSLCA can produced accurate and stable estimates. However, it slightly underestimated the number of colonies at fewer loci or fewer alleles per locus (Fig. 6 (b)).

### 3.3.2 Simulation Set B: Effect of Sample Size for Various Family Distributions of the Data

Accuracies of mSLCA at different sample sizes were shown in Fig. 7. The total sample size did not affect on the accuracy. Accuracies of 90% or higher were obtained at 12 loci with 4 alleles per locus, 6 or more loci with 8 alleles per locus, and in 4 or more loci with 12 alleles per locus.

### 3.3.3 Simulation Set C: Effect of the Number of Families

The effects of the number of families on the accuracies of SLCA and mSLCA is shown in (Fig. 8). When using eight or more loci with eight or more alleles per locus, mSLCA correctly estimated the number of families, but with fewer loci or fewer alleles, it tended to underestimate the number (Fig. 8 (b)). However, at almost all parameter settings, SLCA overestimated the number (Fig. 8 (a)). The number of families had a large effect on the deviation of SLCA but no effect on mSLCA at eight or mo loci with eight or more alleles.

### 3.3.4 Simulation Set D: Effect of Sample Size for Various Numbers of Families

The accuracy of the mSLCA was not largely differed from the result of simulation set C with 50 individuals, although the estimated numbers of families were slightly less than the actual number of families in 150 families (Fig. 9).

### 3.3.5 Real Data Sets

mSLCA successfully predicted the actual number of families at smaller numbers of families, but slightly underestimated the number at larger numbers (Fig. 10 (b)). SLCA had larger variances and overestimated the number with increasing number of families (Fig. 10 (a)).

## 3.4 Discussion

Modified Shared Loci Correspondence Analysis (mSLCA) is a simple full-sib reconstruction method that strictly obeys the Mendelian exclusion principal. It starts by identifying pairs that share at least one allele at every locus and clusters them into full-sib families using correspondence analysis. Although some previous methods required the choice of an appropriate prior distributions for the population allele frequencies (Painter 1997), family size (Thomas and Hill 2000), or other parameters (Konovalov et al. 2004; Konovalov et al. 2005), our algorithm does not require those parameters. The accuracy of the estimates is not affected much by the family size distribution, number of families, or sample size of the data. mSLCA also gives relatively accurate estimates of real data sets, even with as few as six loci. The algorithm seems to be stable with sufficient parameter settings, although the data used here are based on haplodiploid genetics, and mutations and null alleles were not considered in the simulation sets.

The simulation showed that mSLCA generated robust estimates for various types of distributions, even using only four loci with four alleles per locus. In addition, the

accuracy did not vary much from small to large data sets. Moreover, the algorithm can produce accurate and precise estimates using eight or more loci with eight or more alleles in all family structures. In actual situations, we have no prior knowledge of the family size or the number of families in a sample. mSLCA, therefore, can be a good tool for practical applications in the absence of mutations, null alleles, or other factors causing genotyping errors.

With four loci or four alleles, however, mSLCA tended to produce less accurate estimates and to underestimate the number of families. This was probably because the amount of information was not sufficient to distinguish unrelated families. However, the estimates of SLCA depend greatly on the data distribution: SLCA seems to be accurate when the numbers of families is extremely small or extremely large (Fig. 8 (a)). In Chapter 1, I showed that SLCA could correctly estimate the number of families from data collected in the field. This might be because the data were collected from 11 bumble bees from three colonies.

The software for several other algorithms for reconstructing full-sib families are now available on the internet. However, one should take care when using them. For example, Kinship (Goodnight and Queller 1999), which was previously used for estimating family numbers of bee species (Paxton et al. 2002; Cameron et al. 2004; Darvill et al. 2004; Knight et al. 2005; Ellis et al. 2006), had difficulty estimating the true full-sib families for the small data set examined in Chapter 1. An exception is KinGroup (Konovalov et al. 2005) in which a couple of modules correctly estimated the true number of full-fib families, even in the cases where other likelihood algorithms could not be applied. However, it is sometimes very difficult to decide which module one should use because the algorithms are in a black box of user-friendly but researcher-unfriendly coding style. My approach consists of two simple steps: 1) binary scoring of pairwise comparison and 2) clustering the scores with the corresponding analysis. I used R-language for implementing the two steps, but one can use any programming language that is capable of coding the above two steps.

Because mSLCA is based on a simple scoring and clustering procedure, the approach taken here allows for many possible extensions. For example, an alternative scoring system might be possible for diploid species or to detect the other types of sibling relationships. It is interesting that this simple algorithm achieved accurate estimates. This might be possible because the aim of my study was to reconstruct full-sib families for haplodiploid species: by focusing only on haplodiploid species, I avoided the theoretical limitation of the exclusion principal (Thomas and Hill 2000).

I propose the mSLCA algorithm for reconstructing full-sib families of haplodiploid species. Simulation study shows that the accuracy of mSLCA does not depend on the number of families, the distribution of family size, or the number of sampled individuals if we analyze eight or more loci with eight or more alleles. mSLCA could be used to estimate population size or dispersal ranges and to understand mating systems of social insects. Its accurate partitioning in haplodiploid species allows for many extensions to the fields of conservation, behavioral ecology, and evolutionary genetics.

# Chapter 4

# Effective paternity in natural colonies of Japanese native bumble bees

## 4.1 Introduction

Hamilton (1964) introduced the hypothesis of kin selection to explain the frequent evolution of eusociality in Hymenoptera compared with other insect groups. The initial hypothesis relies upon a high coefficient of relatedness among siblings and haplodiploidy of Hymenopteran species (Hamilton 1972). Afterward, conflicts between queen and workers over male parentage (Starr 1984) or sex ratio (Trivers and Hare 1976) are pointed out. In a monogynous colony, mating frequency of the queen is the sole factor that affects genetic structure among the siblings. Mating frequency of females in social insects is therefore of particularly interest for investigating the evolution of eusociality or genetic conflicts among nestmates.

In a monandrous colony, workers are more closely related to their sisters (Coefficient of relatedness (Hamilton 1972), $r$= 3/4) than to their own sons (1/2). Moreover, they are more closely related to their nephews (3/8) than to their brothers (1/4). Although workers cannot lay fertilized (diploid) eggs, they can lay haploid male eggs to get direct inclusive fitness benefits. Therefore, kin selection predicts that relatedness leads to conflicts between queens and workers among male production (Starr 1984; Woyciechowski and Lomnicki 1987).

If a queen has mated with multiple males, the situation is more complicated as the colony membership is divided into several patrilines. In these situations, workers within a patriline are still related to each other by 3/4, but only by 1/4 to their half sib sisters in other patrilines. They are similarly related to their brothers (1/4), but more than to sons of their half sib sisters (1/8) if their fathers are unrelated. Male production by workers is therefore to be expected in monandrous colonies but not in polyandrous colonies because of mutual policing of worker reproduction by other workers (Paxton et al. 2001).

Bumble bees are social hymenoptera where only fertilized queens overwinter and start colonies in the spring. Each colony is founded by a single queen, and lasts for a single breeding season. They usually nest underground using pre-existing holes, very often abandoned burrows of rodents or artificial cavities (Goulson 2003). In general, the queens of most bumble bee species are monandrous (Estoup et al. 1995; Schmid-Hempel and Schmid-Hempel 2000; Takahashi et al. 2008). However, *Bombus hypnorum* and some other species of bumble bees are known to be polyandrous, and males in *B. hypnorum* colonies are produced by a foundress queen (Paxton et al. 2001; Payne et al. 2003).

Developments in molecular markers make it easier to access the genetic relatedness among nestmates and to determine the parentage of social insects (Blouin 2003). By analyzing the DNA patterns of nestmates, we can estimate paternity or genetic relatedness among members in social colonies. In the Eastern Asia region, Takahashi et al. (2008) investigated paternity in *B. ignitus* using microsatellite markers. They found that *B. ignitus* queens from seven colonies that were reared in laboratory each mated only once. However, parentage of other native bumble bees has not been investigated because of the difficulties to collect several colonies to study in the field.

Here, I investigate female mating frequency of the natural colonies of native bumble bees, *B. ardens*, *B. diversus*, and *B. honshuensis*, using microsatellite DNA data. Genetic conflict between nestmates will be discussed.

## 4.2 Material and Methods

### 4.2.1 Sample Collection

Five colonies of *Bombus ardens*, three colonies of *B. diversus*, and one colony of *B. honshuensis* were collected in 2004, 2005, and 2006, from three localities, Tsukuba, Kasama, and Kuzuu (Fig. 11). Each colony was assigned a name for greater ease in discussing them (for details on the name, species, collection location, and the other information, please refer to Table 4). All of these colonies were feral colonies, typically found in holes or artificial cavities. They were collected in the field at the end of the breeding season or at intermediate stages. For colony Kuri, as only one new queen was collected, the new queen was released in the field after her left leg was cut for DNA sample (Holehouse et al. 2003). Colonies Kuzuu and Miyama were collected at the intermediate stages, but later reared in the laboratory until the end of their breeding activities. All the sample individuals were kept in plastic tubes, chilled immediately, and later frozen for DNA extraction at - 80°C.

### 4.2.2 DNA Extraction and Microsatellite Genotyping

Bees were genotyped by six microsatellite loci. DNA was extracted from an entire middle leg of each sample. The leg was homogenized in a tube containing 30 $\mu$l of lysis buffer (1 mg/ml proteinase K, 0.01 M NaCl, 0.1 M EDTA, 0.01 M Tris-HCl (pH 8.0), 0.5% Nonidet P-40). The homogenate was incubated at 50 °C for 60 min and then at 94°C for 10 min. After incubation, the homogenate was diluted with 270 $\mu$l of TE buffer (0.001 M EDTA, 0.01 M Tris-HCl (pH 8.0)) and used as a DNA template in a polymerase chain reaction (PCR). Six microsatellite loci (B11, B96, B100, B118, B121, and B132) were amplified by PCR using primers and the PCR conditions that were developed by Estoup et al. (1996). The polymorphic patterns of each microsatellite loci were analyzed, following protocols described in Chapter 1. Alleles at successfully amplified loci were used for further analysis.

### 4.2.3 Genetic Data Analysis

Identification of paternal alleles was conducted according to Paxton's method (Paxton et al. 2001). If queens had not been sampled (colonies MagJr, Kuzuu, and Miyama), her genotypes were determined indirectly from those of nestmates. Following the identification of maternal alleles, it was straightforward to infer the paternal allele for each worker. Workers that shared no allele with the queen at a given locus were regarded as alien workers (see colony Hyaku in Table 4).

The minimum number of fathering males of queen-produced nestmates was also estimated using mSLCA. mSLCA is software designed to estimate the number of colonies of sampled haplodiploid females (workers and new queens) without using parental information.

The algorithm uses a simple Mendelian exclusion rule and correspondence analysis to detect a minimum number of full-sib families, estimating the number of male patrilines. If this algorithm is used for single-generation nestmate females in a monogynous colony, we can estimate the minimum number of fathering males in that colony.

The idea of effective paternity number ($k_e$) was proposed by Starr (1984) as follows:

$$k_e = 1/\sum_{i=1}^{k} p_i^2$$

where $p_i$ is the proportion of offspring sired by $i$th fathering male and $k$ is the number of fathering males in each colony. Because of small sample size, I used bias-corrected calculation by Nielsen et al. (2003) as follows:

$$k_e = (n-1)^2/\sum_{i=1}^{k} p_i^2(n+1)(n-2) + 3 - n.$$

I employed six highly variable DNA microsatellite loci, and collected individuals at the most active phase of colony development so as to minimize the chance that we would not detect offspring sired by uncommon males. I also estimated pairwise relatedness among workers and new queens in a colony using *Relatedness 5.0.8* software (Queller and Goodnight 1989). I calculated pairwise relatedness of every pairs of individuals in a

colony for every species. Then, the average and the standard-deviation of pairwise relatedness for each colony were calculated. These values were then averaged over every species (Table 6).

## 4.3 Results

### 4.3.1 Microsatellite Variability

All six loci were highly variable, with observed heterozygosities ($H_o$) ranging from 0.11 to 0.99. The mean number of allele per locus and heterozygosity for each species were 5.80 (range, 3 - 9) and 56.9% for *B. ardens*, 6.17 (range 4 - 10) and 82.4% for *B. diversus*, and 1.83 (range, 1 - 3) and 48.6% for *B. honshuensis*, respectively. The observed and estimated genotypes of queens and fathering males for each colony are shown in Table 5. In addition, all genotypes are shown in the electronic supplementary materials (ESM).

### 4.3.2 Colony Kin Structure

The estimated number of fathering males, and thus the effective paternity, was one for all the analyzed colonies (Table 6). Mean pairwise relatedness (Queller and Goodnight 1989) of each colony was 0.76 for *Bombus ardens* and 0.78 for *B. diversus*, and 1.00 for *B. honshuensis* (Table 6).

## 4.4 Discussion

I conducted DNA analysis for natural colonies of three native bumble species: *Bombus ardens*, *B. diversus*, and *B. honshuensis*. Using six highly variable DNA microsatellite loci, I estimated the paternity of natural bumble bee colonies. All the analyzed colonies were monandrous, and thus within-colony genetic relatedness was very high. Although I were not able to collect enough samples to draw solid conclusions, I found strong evidence that nestmates within a colony share fathers and that they are highly related to each other.

Earlier studies in Europe show that bumble bee queens, with the exception of *B.*

*hypnorum*, mate with a single male, which maximizes within-colony relatedness (Paxton et al. 2001). Data for the three native species explored here is consistent with these previous findings. Although there have been no systematic studies to examine the mating behaviors of these species, Katayama (1964) observed the mating behavior of *B. ardens*: He found that *B. ardens* queens resist approaches by males, but once she has accepted him, she mates with the male for a considerably long period of time (about 50 minutes). These observations may suggest that *B. ardens* queens avoid multiple mating, or that there exist mating plugs as with *B. terrestris* (Brown et al. 2002).

Strassmann (2001) claimed that single mating is predominant in social Hymenoptera groups. Recently, Hughes et al. (2008) showed that monandry is the ancestral state of social hymenoptera, and that worker reproductive totipotency is associated with monandry. In monogynous and monandrous colonies, the inclusive fitness benefits to the queen and workers are contradicted in male production (Starr 1984). However, male production by workers appears to be an integral part of the reproductive effort of bumble bee colonies even though it may arise from worker-queen conflicts of (Owen and Plowright 1982; Paxton et al. 2001; Takahashi et al. 2008). Thus in bumble bee colonies, high relatedness among workers and the balanced tension between the queen and workers may maintain social activity in this monandrous species.

Here, I have shown that the effective mating frequency of native bumble bees is one for all the investigated colonies. However, much of their ecology and mating behaviors remain unknown, because of difficulties in finding their nests in the field. Population size or other ecological aspects have not been fully understood for native bumble bees as well. Additional investigation is needed immediately.

# General Discussion

I proposed two simple visualization methods, the Shared Loci Correspondence Analysis (SLCA) and its modified version (mSLCA), to reconstruct full-sib patterns of haplodiploid females. I conducted a simulation study for investigating the accuracy levels of these methods. Results indicate that the mSLCA outperforms the SLCA, and that it can produce accurate estimates using eight or more loci with eight or more alleles. I also conducted a brief survey to estimate the population size of *Bombus terrestris* in Biratori, Hokkaido, Japan and to estimate the number of fathering males in Japanese native colonies.

The mSLCA uses the Mendelian exclusion principal for clustering procedures. Although the theoretical limitation of the exclusion principle has been cited as a limitation in previous research (Thomas and Hill 2000), I overcame this by focusing only on haplodiploid species. Using haplodiploid genetics and the Mendelian exclusion principle, mSLCA can quickly estimate full siblings for Haplodiploid species even with small samples. In addition, the mSLCA does not require the accurate estimation of allele frequencies and it does not assume genotyping errors. It is thus suitable for estimating the sibling structure with small sample sizes, such as a sample from a wild population for which we do not have detailed genetic information. mSLCA can produce accurate estimates with large sample size as well.

How best to partition single-generation individuals into sibling families is an active area of research (Blouin 2003). Although several methods have been proposed, there is a problem in applying these methods to various types of data (Butler et al. 2004). It is interesting that the simpler and blunter mSLCA method outperforms the SLCA for

almost all data settings. The mSLCA uses a binary scoring system, while the SLCA gives the number of loci at which at least one allele is shared.

An accurate partition of haplodiploid species allows for many extensions into the fields of conservation, evolution, and behavioral ecology: It can be applied to the estimation of population size or dispersal ranges, and to mating systems of various organisms, as well as social insects. A simple scoring approach based on the Mendelian exclusion principle could be modified to estimate full sibling structure of diploid species by changing the scoring procedures. For example, it could estimate the number of colonies even in the polyandrous colonies by adjusting multi-level scoring rather than applying binary scoring. The algorithms could also be modified to estimate full siblings nested among half siblings (Thomas and Hill 2000; Thomas and Hill 2002) by altering the scoring or clustering process.

The population decline of pollinator species is a serious problems in many parts of the world (Darvill et al. 2004; Knight et al. 2005). Key ecological parameters, such as dispersal range or effective population size, are vital for planning a strategy for the effective management of plant-pollinator populations. The rarity of multiple mating in social hymenoptera would be a very interesting topics for the study of inclusive fitness and social organization of insect populations. Using DNA data of larva in a colony, we would be able to detect the effective paternity of the natural colonies in the strict sense. The mSLCA will contribute to these areas of research because of its effectiveness for detecting the number of mating patterns of monandrous social insects such as wasps, ants, and bees (Strassmann 2001; Kikuta and Tsuji 1999; Evans 1998). It can be also applied for estimating pollen dispersal, by estimating sibling structure of trees with pollen DNA (Streiff et al. 1999). Investigating the route and origin of the introduced populations of hymenoptera species such as the Argentina ant (*Linepithema*), is another application of this method (Tsutsui et al. 2001). It is my hope that these algorithms will be widely used to investigate the problems in ecology, genetics, and other exciting field of research.

# Acknowledgments

I am deeply indebted to Associate Professor Yukihiko Toquenaga for his support and guidance throughout the course of my study. I am also grateful to Assistant Professor Kazuharu Ohashi for his support and valuable suggestions on my research. Thanks are also due to Dr. Koichi Goka of the National Institute for Environmental Studies for his experimental guidance and valuable comments throughout my research. My special thanks go to Professor Mamoru Watanabe and Associate Professor Yoichi Kaino for his valuable comments and supports on the thesis.

I wish to thank Mr. Masahiro Yoneda, Dr. Yoko Kunitake, Dr. Natsuko Kondo, Dr. Jun Yokoyama, and Mr. Yosuke Saito for providing me with the *Bombus terrestris* samples in Hokkaido. I also thank to the member of MARUHANA team in Population Ecology Laboratory, especially to Ms. Lina Kawaguchi, Ms. Dulee Munidasa, Ms. Yaya Hoshino, and Chikako Ishida for providing me with the valuable samples of wild colonies of Japanese bumble bees.

Thanks are also due to Dr. Hiroyuki Mano and Shinya Yoshimura for their valuable comments and fruitful discussions. Chapter 3 would not be possible without their technical and educational supports.

My thanks are also to the member of the Population Ecology laboratory and Invasive Alien Species Team, especially to Dr. Koichi Fuji for his valuable comments and educational advice, and the members of MARUHANA team for their valuable comments and consistent encouragements.

Finally, I would like to thank my families for their spiritual and financial support throughout my study. Whole support of colleagues in my work place is also necessary

for my education in Doctoral program. My special thanks are also due to all my friends for their support and encouragements.

# Appendix

All the analyzed genotypes. In column Caste, "Q" represents queen, "w" represents worker, "m" represents male, and "nq" new queen.

Colony: Kuri (*Bombus ardens*)

| No. | Caste | No. of individuals | B11 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 173 | 187 | 236 | 262 | 152 | 158 | - | - | 160 | 162 |
| 2 | nq,w | 4 | 159 | 173 | 236 | 236 | 150 | 158 | - | - | 162 | 162 |
| 3 | w | 2 | 159 | 173 | 236 | 236 | 150 | 152 | - | - | 160 | 162 |
| 4 | w | 3 | 159 | 173 | 236 | 262 | 150 | 152 | - | - | 162 | 162 |
| 5 | w | 3 | 159 | 187 | 236 | 262 | 150 | 158 | - | - | 160 | 162 |
| 6 | w | 3 | 159 | 187 | 236 | 236 | 150 | 152 | - | - | 162 | 162 |
| 7 | w | 1 | 159 | 173 | 236 | 262 | 150 | 152 | - | - | 160 | 162 |
| 8 | w | 1 | 159 | 187 | 236 | 236 | 150 | 152 | - | - | 160 | 162 |
| 9 | w | 1 | 159 | 187 | 236 | 262 | 150 | 158 | - | - | 162 | 162 |
| 10 | w | 3 | 159 | 187 | 236 | 262 | 150 | 152 | - | - | 162 | 162 |
| 11 | w | 3 | 159 | 173 | 236 | 236 | 150 | 158 | - | - | 160 | 162 |
| 12 | w | 1 | 159 | 187 | 236 | 236 | 150 | 158 | - | - | 160 | 162 |
| 13 | w | 1 | 159 | 187 | 236 | 236 | 150 | 158 | - | - | 162 | 162 |
| 14 | w | 1 | 159 | 187 | 236 | 262 | 150 | 152 | - | - | 160 | 162 |

Colony: MagES (*Bombus ardens*)

| No. | Caste | No. of individuals | B11 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 177 | 177 | 250 | 254 | 156 | 158 | 147 | 147 | 160 | 160 |
| 2 | w | 3 | 177 | 179 | 250 | 254 | 150 | 158 | 147 | 147 | 160 | 160 |
| 3 | w | 2 | 177 | 179 | 254 | 254 | 150 | 156 | 147 | 147 | 160 | 160 |
| 4 | w | 2 | 177 | 179 | 254 | 254 | 150 | 158 | 147 | 147 | 160 | 160 |
| 5 | w | 1 | 177 | 179 | 254 | 254 | 150 | 156 | 147 | 174 | 160 | 160 |
| 6 | m | 1 | 177 | - | 250 | - | 156 | - | 147 | - | - | - |

Colony: MagJr (*Bombus ardens*)

| No. | Caste | No. of individuals | B11 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | w | 1 | 179 | 169 | 246 | 252 | 162 | 164 | 147 | 179 | 160 | 160 |
| 2 | w | 1 | 167 | 169 | 251 | 252 | 162 | 164 | 147 | 179 | 160 | 160 |
| 3 | w | 1 | 167 | 169 | 246 | 252 | 162 | 164 | 147 | 179 | 160 | 160 |
| 4 | w | 1 | 167 | 169 | 251 | 252 | 162 | 164 | 147 | 147 | 160 | 183 |
| 5 | w | 1 | 179 | 169 | 246 | 252 | 162 | 164 | 147 | 147 | 160 | 183 |

Colony: Kuzuu (*Bombus ardens*)

| No. | Caste | No. of individuals | B11 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | nq,w | 5 | 167 | 177 | 250 | 252 | 156 | 156 | 147 | 147 | 160 | 168 |
| 2 | nq,w | 3 | 167 | 177 | 252 | 252 | 150 | 156 | 147 | 147 | 160 | 160 |
| 3 | nq,w | 7 | 167 | 177 | 250 | 252 | 150 | 156 | 147 | 147 | 160 | 160 |
| 4 | nq,w | 9 | 167 | 177 | 252 | 252 | 156 | 156 | 147 | 147 | 160 | 160 |
| 5 | nq,w | 9 | 167 | 177 | 250 | 252 | 156 | 156 | 147 | 147 | 160 | 160 |
| 6 | nq,w | 8 | 167 | 177 | 250 | 252 | 150 | 156 | 147 | 147 | 160 | 168 |
| 7 | nq,w | 14 | 167 | 177 | 252 | 252 | 150 | 156 | 147 | 147 | 160 | 168 |
| 8 | nq,w | 8 | 167 | 177 | 252 | 252 | 156 | 156 | 147 | 147 | 160 | 168 |
| 9 | nq | 1 | 167 | 167 | 250 | 252 | 150 | 156 | 147 | 147 | 168 | 168 |
| 10 | nq | 1 | 167 | 177 | - | - | 156 | 156 | 147 | 147 | 160 | 168 |
| 11 | m | 2 | 167 | - | 250 | - | 150 | - | - | - | 168 | - |
| 12 | m | 1 | 167 | - | 250 | - | 156 | - | 147 | - | 168 | - |
| 13 | m | 1 | 167 | - | 252 | - | 156 | - | - | - | 168 | - |
| 14 | m | 1 | 167 | - | 252 | - | 150 | - | - | - | - | - |
| 15 | m | 2 | 167 | - | 252 | - | 150 | - | 147 | - | 168 | - |

Colony: Legend (*Bombus ardens*)

| No. | Caste | No. of individuals | B11 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 171 | 181 | 252 | 252 | 156 | 164 | 147 | 147 | 164 | 168 |
| 2 | w | 11 | 169 | 181 | 246 | 252 | 152 | 164 | 147 | 147 | - | - |
| 3 | w | 2 | 169 | 181 | 246 | 252 | 152 | 156 | 147 | 147 | - | - |
| 4 | w | 6 | 169 | 171 | 246 | 252 | 152 | 164 | 147 | 147 | - | - |
| 5 | w | 2 | 169 | 171 | 246 | 252 | 152 | 156 | 147 | 147 | - | - |

Colony: MagDG (*Bombus diversus*)

| No. | Caste | No. of individuals | B11 | | B118 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 182 | 182 | - | - | 283 | 283 | 213 | 217 | 155 | 155 | 170 | 179 |
| 2 | w | 1 | 182 | 182 | - | - | - | - | 213 | 223 | 155 | 163 | 179 | 181 |
| 3 | w | 1 | 182 | 182 | - | - | 283 | 299 | - | - | 155 | 163 | 170 | 181 |
| 4 | w | 2 | 182 | 182 | - | - | 283 | 299 | 213 | 223 | 155 | 163 | 179 | 181 |
| 5 | w | 2 | 182 | 182 | - | - | 283 | 299 | 217 | 223 | 155 | 163 | 170 | 181 |
| 6 | w | 1 | 182 | 182 | - | - | 283 | 299 | 217 | 223 | - | - | 179 | 181 |
| 7 | w | 1 | 182 | 182 | - | - | 283 | 299 | 217 | 223 | - | - | 171 | 181 |
| 8 | w | 1 | 182 | 182 | - | - | 283 | 299 | 213 | 223 | - | - | 179 | 181 |

Colony: Back (*Bombus diversus*)

| No. | Caste | No. of individuals | B11 | | B118 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 182 | 182 | 234 | 240 | 273 | 275 | 209 | 209 | 155 | 163 | 173 | 179 |
| 2 | w | 1 | 182 | 187 | 234 | 234 | 273 | 283 | 209 | 209 | 157 | 163 | 173 | 183 |
| 3 | w | 1 | 182 | 187 | 234 | 240 | 275 | 283 | 209 | 209 | 157 | 163 | 179 | 183 |
| 4 | w | 1 | 182 | 187 | 234 | 240 | 275 | 283 | 209 | 209 | 157 | 155 | 173 | 183 |
| 5 | w | 1 | 182 | 187 | 234 | 240 | 273 | 283 | 209 | 209 | 157 | 163 | 173 | 183 |
| 6 | w | 2 | 182 | 187 | 234 | 240 | 275 | 283 | 209 | 209 | 157 | 163 | 173 | 183 |
| 7 | w | 1 | 182 | 187 | 234 | 234 | 273 | 283 | 209 | 209 | 157 | 155 | 179 | 183 |
| 8 | w | 1 | 182 | 187 | 234 | 234 | - | - | 209 | 209 | 157 | 163 | 173 | 183 |
| 9 | w | 2 | 182 | 187 | 234 | 234 | 273 | 283 | 209 | 209 | 157 | 163 | 179 | 183 |
| 10 | w | 1 | 182 | 187 | 234 | 240 | 273 | 283 | 209 | 209 | 157 | 155 | 179 | 183 |

Colony: Hyaku (*Bombus diversus*)

| No. | Caste | No. of individuals | B11 | | B118 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Q | 1 | 182 | 182 | - | - | 275 | 275 | 213 | 219 | - | - | - | - |
| 2 | w | 3 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 213 | 155 | 163 | 164 | 185 |
| 3 | w | 4 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 213 | 155 | 163 | 164 | 179 |
| 4 | w | 6 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 213 | 155 | 163 | 164 | 185 |
| 5 | w | 7 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 213 | 161 | 163 | 164 | 179 |
| 6 | w | 4 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | 161 | 163 | 164 | 185 |
| 7 | w | 7 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | 155 | 163 | 164 | 185 |
| 8 | w | 6 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 219 | 161 | 163 | 164 | 185 |
| 9 | w | 13 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 213 | 155 | 163 | 164 | 179 |
| 10 | w | 3 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | 155 | 163 | 164 | 179 |
| 11 | w | 6 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 219 | 155 | 163 | 164 | 179 |
| 12 | w | 3 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 219 | 155 | 163 | 164 | 185 |
| 13 | w | 2 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 213 | 161 | 163 | 164 | 185 |
| 14 | w | 5 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 213 | 161 | 163 | 164 | 179 |
| 15 | w | 1 | 182 | 182 | - | - | 275 | 285 | 209 | 213 | 161 | 163 | 164 | 179 |
| 16 | w | 6 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 219 | 161 | 163 | 164 | 179 |
| 17 | w | 9 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | 161 | 163 | 164 | 179 |
| 18 | w | 1 | 182 | 182 | 232 | 236 | - | - | 209 | 213 | 161 | 163 | 164 | 179 |
| 19 | w | 1 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 213 | 161 | 163 | - | - |
| 20 | w | 1 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | 155 | 163 | - | - |
| 21 | w | 1 | 182 | 182 | 232 | 240 | 275 | 285 | 209 | 219 | - | - | 164 | 185 |
| 22 | w | 3 | 182 | 182 | 232 | 236 | 275 | 285 | 209 | 213 | 161 | 163 | 164 | 185 |
| 23 | unrelated | 1 | 147 | 149 | 215 | 219 | 240 | 240 | 142 | 142 | 151 | 155 | 156 | 160 |

Colony: Miyama (*Bombus honshuensis*)

| No. | Caste | No. of individuals | B11 | | B118 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | w | 1 | 147 | 149 | - | - | - | - | 142 | 142 | 151 | 155 | 152 | 156 |
| 2 | w | 2 | 147 | 149 | - | - | - | - | 142 | 142 | 155 | 155 | 152 | 156 |
| 3 | w | 1 | 147 | 149 | 216 | 220 | 240 | 240 | 142 | 142 | 151 | 155 | 152 | 156 |
| 4 | w | 1 | 147 | 149 | 216 | 216 | 240 | 240 | 142 | 142 | 151 | 155 | - | - |
| 5 | w | 2 | 147 | 149 | 216 | 216 | 240 | 240 | 142 | 142 | 151 | 155 | 160 | 156 |
| 6 | w | 2 | 147 | 149 | 216 | 220 | 240 | 240 | 142 | 142 | 155 | 155 | 152 | 156 |
| 7 | w | 1 | 147 | 149 | 216 | 216 | 240 | 240 | 142 | 142 | 155 | 155 | 160 | 156 |
| 8 | w | 2 | 147 | 149 | 216 | 216 | 240 | 240 | 142 | 142 | 151 | 155 | 152 | 156 |

# References


Almudevar, A. and C. Field (1999). Estimation of single-generation sibling relationships based on DNA markers. *Journal of Agricultural, Biological, and Environmental Statistics 4*, 136–165.

Beger-Wolf, T. Y., B. DasGupta, W. Chaovalitwongse, and M. V. Ashley (2005). Combinatorial reconstruction of sibling relationships. *In Proceedings of the 6th International Symposium on Computational Biology and Genome Information, Utah*, 1252–1255.

Beger-Wolf, T. Y., S. Sheikh, B. DasGupta, M. V. Ashley, I. Caballero, W. Chaovalitwongse, and L. Putrevu (2007). Reconstructing sibling relationships in wild populations. *Bioinformatics 23*, 148–156.

Benzecri, J. P. (1992). *Correspondence Analysis Handbook: Statistics, Text Books and Monographs*. Vol.125. Marcel Dekker, New York. 665 pp.

Beyer, J. and B. May (2003). A graph-theoretic approach to the partition of individuals into full-sib families. *Molecular Ecology 12*, 2243–2250.

Blouin, M. (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. *Trends in Ecology and Evolution 18*, 503–511.

Blouin, M., M. Parsons, V. Lacaille, and S. Lotz (1996). Use of microsatellite loci to classify individuals by relatedness. *Molecular Ecology 5*, 394–401.

Brown, M., B. Baer, R. Schmid-Hempel, and P. Schmid-Hempel (2002). Dynamics of multiple mating in the bumble bee *Bombus hypnorum*. *Insectes Sociaux 49*, 315–319.

Butler, K., C. Field, C. Herbinger, and B. Smith (2004). Accuracy, efficiency and robustness of four algorithms allowing full sibship reconstruction from DNA marker data. *Molecular Ecology 13*, 1589–1600.

Cameron, E., P. Franck, and P. Oldroyd (2004). Genetic structure of nest aggregations and drone configurations of the southeast Asian stingless bee *Trigona collina*. *Molecular Ecology 13*, 2357–2364.

Chapman, R., J. Wang, and A. Bourke (2003). Genetic analysis of spatial foraging patterns and resource sharing in bumble bee pollinators. *Molecular Ecology 12*, 2801–2808.

Darvill, B., M. Knight, and D. Goulson (2004). Use of genetic markers to quantify bumblebee foraging range and nest density. *Oikos 107*, 471–478.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, 92pp.

Ellis, J., M. Knight, B. Darvill, and D. Goulson (2006). Extremely low effective population sizes, genetic structuring and reduced genetic diversity in a threatened bumblebee species, *Bombus sylvarum* (Hymenoptera: Apidae). *Molecular Ecology 15*, 4375 – 4386.

Estoup, A., A. Scholl, A. Pouvreau, and M. Solignac (1995). Monandry and polyandry in bumble bees (Hymenoptera: Bombinae) as evidenced by highly variable microsatellites. *Molecular Ecology 4*, 89 – 93.

Estoup, A., M. Solignac, J. Cornuet, J. Goudet, and A. Scholl (1996). Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. *Molecular Ecology 5*, 19 – 31.

Evans, J. (1998). Parentage and sex allocation in the facultatively polygynous ant *Myrmica tahoensis*. *Behavioral Ecology and Sociobiology 44*, 35–42.

Goka, K. (1998). Influence of invasive species on native species-will the European bumble bee, *b*ombus terrestris, bring genetic pollution into the Japanes native

species? *Bulletin of Biography Society in Japan 53*, 91–101.

Goka, K. (2006a). The European bumblebee *Bombus terrestris* and Invasive Alien species Act. *Nature and Insects 41*, 22–26 (in Japanese).

Goka, K. (2006b). Perspective of invasive alien species problem in Japan suggested by exotic insects introduction. *Bulletin of Kansai Organization for Nature Conservation 28*, 37–47 (in Japanese).

Goka, K., K. Okabe, and M. Yoneda (2006). Worldwide migration of parasitic mites as a result of bumblebee commercialization. *Population Ecology 48*, 285–291.

Goka, K., K. Okabe, M. Yoneda, and S. Niwa (2001). Bumblebee commercialization will cause worldwide migration of parasitic mite. *Molecular Ecology 10*, 2095–2099.

Goodnight, K. and D. Queller (1999). Computer software for performing likelihood tests of pedigree relationship using genetic markers. *Molecular Ecology 8*, 1231–1234.

Goulson, D. (2003). *Bumblebees: Their Behaviour and Ecology*. Oxford University Press, Oxford. 235 pp.

Hamilton, W. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology 7*, 1–52.

Hamilton, W. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of Ecological Systems 3*, 193–232.

Herbinger, C., P. O'reilly, and E. Verspoor (2006). Unravelling first-generation pedigrees in wild endangered salmon populations using molecular genetic markers. *Molecular Ecology 15*, 2261–2275.

Hingston, A. and P. Mcquillan (1999). Displacement of tasmanian native megachilid bees by the recently introduced bumblebee *b*ombus terrestris (Linnaeus, 1758)(Hymenoptera: Apoidae). *Australian Journal of Zoology 47*, 59–65.

Holehouse, K., R. Hammond, and A. Bourke (2003). Non-lethal sampling of DNA from bumblebee for conservation genetics. *Insectes sociaux 50*, 277–285.

Hughes, W., B. Oldroyd, M. Beekman, and F. Ratnieks (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science 320*, 1213–1216.

Inari, N., T. Nagamitsu, T. Kenta, K. Goka, and T. Hiura (2005). Spatial and temporal pattern of introduced *Bombus terrestris* abundance in Hokkaido, Japan, and its potential impact on native bumblebees. *Population Ecology 47*, 77–82.

Jones, A. and W. Ardren (2003). Methods of parentage analysis in natural populations. *Molecular Ecology 12*, 2511–2523.

Katayama, E. (1964). Observation on the layer stage colonies of *Bombus ardens* Smith (Hymenoptera, Apidae). *Kontyu 32*, 393–402.

Kikuta, N. and K. Tsuji (1999). Queen and worker policing in the monogynous and monandrous ant, *Diacamma sp. Behavioral Ecology and Sociobiology 46*, 180–189.

Knight, M., A. Martin, S. Bishop, J. Osborne, R. Hale, A. Sanderson, and D. Goulson (2005). An interspecific comparison of foraging range and nest density of four bumblebee (*Bombus*) species. *Molecular Ecology 14*, 1811–1820.

Konovalov, D., N. Bajema, and B. Litow (2005). Modified SIMPSON O($n^3$) algorithm for the full sibship reconstruction problem. *Bioinformatics 21*, 3912–3917.

Konovalov, D., C. Manning, and M. Henshaw (2004). KINGROUP: a program for pedigree relationship reconstruction and kin group assignments using genetic markers. *Molecular Ecology Note 4*, 779–782.

Kunitake, Y. and K. Goka (2006). Environmental risk assessment and management decisions for introduced insects – Legal controls on *Bombus terrestris* by invasive alien species act. *Monthly Plant Protection Journal 4*, 196–198 (in Japanese).

Lewinsohn, T., P. Prado, P. Jordano, and J. Bascompte (2006). Structure in plant-animal interaction assemblages. *Oikos 13*, 174–184.

Matsumura, C., J. Yokoyama, and I. Washitani (2004). Invasion status and potential ecological impacts of an invasive alien bumblebee, *Bombus terrestris* L. (Hymenoptera: Apidae) naturalized in Southern Hokkaido, Japan. *Global Environmental Research 8*, 51–66.

Nielsen, R., D. Tarpy, and H. Reeve (2003). Estimating effective paternity number in social insects and the effective number of alleles in a population. *Molecular Ecology 12*, 3157–3164.

Ono, M. (1998). Why is now the bumblebees. *Nature and Insects 33*, 2–3 (in Japanese).

Owen, R. and R. Plowright (1982). Worker-queen conflict and male parentage in bumble bees. *Behavioral Ecology and Sociobiology 11*, 91–99.

Painter, I. (1997). Sibship reconstruction without parental information. *Journal of Agricultural, Biological, and Environmental Statistics 2*, 212–229.

Paxton, R., M. Ayasse, J. Field, and A. Soro (2002). Complex sociogenetic organization and reproductive skew in a primitively eusocial sweat bee, *Laioglossum malachurum*, as revealed by microsatellite. *Molecular Ecology 11*, 2405–2416.

Paxton, R., P. Thoren, A. Estoups, and J. Tengo (2001). Queen-worker conflict over male production and the sex ratio in a facultatively polyandrous bumble bee, *Bombus hypnorum*:the consequences of nest usurpation. *Molecular Ecology 10*, 2489–2498.

Payne, C., L. T.M., and L. M.A. (2003). The frequency of multiple paternity in bumble bee (*Bombus*) colonies based on microsatellite DNA at B10 locus. *Insectes Sociaux 50*, 375–378.

Queller, D. and K. Goodnight (1989). Estimating relatedness using genetic markers. *Evolution 43*, 258–275.

R Development Core Team (2004). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Schmid-Hempel, R. and P. Schmid-Hempel (2000). Female mating frequencies in *Bombus* spp. from Central Europe. *Insectes Sociaux 47*, 36–41.

Smith, B., C. Herbinger, and H. Merry (2001). Accurate partition of individuals into full-sib families from genetic data without parental information. *Genetics 158*, 1329–1338.

Starr, C. (1984). *Sperm competition, kinship, and sociality in the aculeate Hymenoptera. In: Smith RL (ed) Sperm competition and the evolution of animal mating systems.* Academic press Orlando, Florida, pp 427-464.

Strassmann, J. (2001). The rarity of multiple mating by females in the social Hymenoptera. *Insectes Sociaux 48*, 01–13.

Streiff, R. et al. (1999). Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur* L. and *Q. ptraea*(Matt.) Liebl. *Molecular Ecology 8*, 831–841.

Takahashi, J., M. Itoh, I. Shimizu, and M. Ono (2008). Male parentage and queen mating frequency in the bumblebee *Bombus ignitus* (Hymenoptera, Bombinae). *Ecological Research DOI 10.1007/s11284-007-0456-y*.

Thomas, S. and W. Hill (2000). Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics 155*, 1961–1972.

Thomas, S. C. and W. G. Hill (2002). Sibship reconstruction in hierarchical population structures using Markov chain Monte Carlo techniques. *Genetical Research 79*, 227–234.

Trivers, R. and H. Hare (1976). Haplodiploidy and the evolution of the social insect. *Science 191*, 249–263.

Tsutsui, N., A. Shuarez, D. Holwasy, and T. Case (2001). Relationships among native

and introduced populations of the Argentina ant (*Linepithema humile*) and the source of introduced populations. *Molecular Ecology 10*, 2151–2161.

Wang, J. (2004). Sibship reconstruction from genetic data with typing errors. *Genetics 166*, 1963–1979.

Washitani, I. (1998). Conservation-ecological issues of the recent invasion of *Bombus terrestris* into Japan. *Japanese Journal of Ecology 48*, 73–78 (in Japanese).

Washitani, I. and C. Matsumura (1998). European bumblebee sighting information. *Japanese Journal of Conservation Ecology 3*, 176–177 (in Japanese).

Williamson, M. (1996). *Biological Invasions - Population and Community Biology Series 15*. Chapman and Hall, London. 244 pp.

Woyciechowski, M. and A. Lomnicki (1987). Multiple mating of queen and the sterility of worker among eusocial hymenoptera. *Journal of Theoretical Biology 128*, 317–327.

Yoneda, M., J. Yokoyama, K. Tsuchida, S. Osaki, S. Itoya, and K. Goka (2007). Preventing *b*ombus terrestris from escaping with a net covering over a tomato greenhouse in Hokkaido. *Japanese Journal of Applied Entomology and Zoology 51*, 39–44.

Table 1 Genotype frequency for each colony and locus, and number of alleles for each locus.

| | Locus | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B11 | *n* | B96 | *n* | B100 | *n* | B118 | *n* | B121 | *n* | B132 | *n* |
| Colony | | | | | | | | | | | | |
| *B. ardens* | | | | | | | | | | | | |
| A (n=1) | (157,181) | 1 | (252,262) | 1 | (152,156) | 1 | | | (147,147) | 1 | (158,160) | 1 |
| B (n=2) | (167,169) | 1 | (246,251) | 1 | (156,156) | 1 | | | (177,177) | 1 | (160,166) | 1 |
| | (169,169) | 1 | (246,254) | 1 | (156,158) | 1 | | | (177,179) | 1 | (166,174) | 1 |
| C (n=8) | (177,179) | 8 | (250,254) | 3 | (150,156) | 3 | | | (147,147) | 7 | (160,160) | 8 |
| | | | (254,254) | 5 | (150,158) | 5 | | | (147,174) | 1 | | |
| *B. diversus* | | | | | | | | | | | | |
| D (n=87) | (182,182) | 87 | (275,285) | 87 | (209,213) | 43 | (232,236) | 50 | (155,163) | 45 | (164,179) | 53 |
| | | | | | (209,219) | 44 | (232,240) | 37 | (161,163) | 42 | (164,185) | 32 |
| | | | | | | | | | | | (164,183) | 2 |
| Number of alleles | 7 | | 8 | | 7 | | 3 | | 7 | | 8 | |

Table 2 Number of alleles ($n$), observed heterozygosities ($H_o$) and allele frequencies of six microsatellite loci from *Bombus terrestris* in Hokkaido.

| | | | Allele frequency | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | $n$ | $H_o$ | a | b | c | d | e | f | g | h | i | j |
| B11 | 10 | 0.615 | 0.282 | 0.256 | 0.218 | 0.115 | 0.038 | 0.026 | 0.026 | 0.013 | 0.013 | 0.013 |
| B96 | 4 | 0.462 | 0.513 | 0.385 | 0.077 | 0.026 | | | | | | |
| B100 | 8 | 0.744 | 0.398 | 0.180 | 0.167 | 0.103 | 0.103 | 0.026 | 0.013 | 0.013 | | |
| B118 | 5 | 0.641 | 0.385 | 0.372 | 0.103 | 0.077 | 0.026 | | | | | |
| B121 | 4 | 0.359 | 0.795 | 0.103 | 0.090 | 0.013 | | | | | | |
| B132 | 10 | 0.872 | 0.321 | 0.179 | 0.128 | 0.115 | 0.077 | 0.051 | 0.051 | 0.038 | 0.026 | 0.013 |

Table 3 Number of alleles ($n$), observed heterozygosities ($H_o$) and allele frequencies of six microsatellite loci from *Bombus terrestris* collected from 20 commercial colonies.

| | | | Allele frequency | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Locus | $n$ | $H_o$ | a | b | c | d | e | f | g | h | i | j | k | l |
| B11 | 7 | 0.73 | 0.310 | 0.300 | 0.145 | 0.110 | 0.075 | 0.045 | 0.015 | | | | | |
| B96 | 3 | 0.64 | 0.68 | 0.27 | 0.050 | | | | | | | | | |
| B100 | 8 | 0.55 | 0.515 | 0.250 | 0.095 | 0.065 | 0.030 | 0.020 | 0.015 | 0.010 | | | | |
| B118 | 6 | 0.56 | 0.010 | 0.015 | 0.295 | 0.020 | 0.555 | 0.105 | | | | | | |
| B121 | 3 | 0.54 | 0.730 | 0.225 | 0.045 | | | | | | | | | |
| B132 | 12 | 0.92 | 0.200 | 0.195 | 0.140 | 0.135 | 0.115 | 0.080 | 0.050 | 0.020 | 0.020 | 0.020 | 0.015 | 0.010 |

Table 4 Number of genotyped individuals for each colony of *Bombus ardens*, *B. diversus*, and *B. honshuensis*.

| ID | Species | Place | Date | No. of Genotyped Individuals | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Queen | Worker | Male | Newqueen | Unrelated | Total |
| Kuri | *B.ardens* | Kasama | 2004. 6.11 | 1 | 26 | - | 1 | - | 28 |
| MagES | *B.ardens* | Tsukuba | 2005. 6.14 | 1 | 8 | 1 | - | - | 10 |
| MagJr | *B.ardens* | Tsukuba | 2006. 6. 2 | - | 5 | - | - | - | 5 |
| Kuzuu | *B.ardens* | Kuzuu | 2005. 6. 2 | - | 18 | 7 | 47 | - | 72 |
| Legend | *B.ardens* | Tsukuba | 2005. 6. 1 | 1 | 21 | - | - | - | 22 |
| MagDG | *B.diversus* | Tsukuba | 2006. 6. 2 | 1 | 9 | - | - | - | 10 |
| Back | *B.diversus* | Tsukuba | 2004. 9.24 | 1 | 11 | - | - | - | 12 |
| Hyaku | *B.diversus* | Tsukuba | 2004.10. 7 | 1 | 92 | - | - | 1 | 94 |
| Miyama | *B.honshuensis* | Tsukuba | 2006. 8. 4 | - | 12 | - | - | - | 12 |

Table 5 Observed or estimated genotypes of queens (Q) and fatherig males (F) for each colony.

| Colony | B11 | | B118 | | B96 | | B100 | | B121 | | B132 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Q | F | Q | F | Q | F | Q | F | Q | F | Q | F |
| Kuri | 173,187 | 159 | - | - | 236,262 | 236 | 152,158 | 150 | - | - | 160,162 | 162 |
| MagES | 177 | 179 | - | - | 250,245 | 254 | 156,158 | 150 | 147 | 147 | 160 | 160 |
| MagJr | 167,179 | 169 | - | - | 246,251 | 252 | 162or164 | 162or164 | 147,179 | 147 | 160,183 | 160 |
| Kuzuu | 167,177 | 177 | - | - | 250,252 | 252 | 150,156 | 156 | 147 | 147 | 160,168 | 160 |
| Legend | 171,181 | 169 | - | - | 252 | 246 | 156,164 | 152 | 147 | 147 | 164,168 | - |
| MagDG | 182 | 182 | - | - | 283 | 299 | 213,217 | 223 | 155 | 163 | 170,179 | 181 |
| Back | 182 | 187 | 234,240 | 234 | 273,275 | 283 | 209 | 209 | 155,163 | 157 | 173,179 | 183 |
| Hyaku | 182 | 182 | 236,240 | 232 | 275 | 285 | 213,219 | 209 | 155,161 | 163 | 179,185 | 164 |
| Miyama | 147or149 | 147or149 | 216,220 | 216 | 240 | 240 | 142 | 142 | 151,155 | 155 | 152,160 | 156 |

Table 6 Effective paternity and genetic relatedness among workers and new queens of each colony.

| ID | Species | State of the queen | number of females | Effective Paternity | Estimated number of fathering males | Average (SD) of pairwise relatedness |
|---|---|---|---|---|---|---|
| Kuri | *B.ardens* | Queen-present | 27 | 1 | 1 | 0.76 (0.12) |
| MagES | *B.ardens* | Queen-present | 8 | 1 | 1 | 0.84 (0.11) |
| MagJr | *B.ardens* | Queen-absent | 5 | 1 | 1 | 0.58 (0.18) |
| Kuzuu | *B.ardens* | Queen-absent | 65 | 1 | 1 | 0.80 (0.13) |
| Legend | *B.ardens* | Queen-present | 21 | 1 | 1 | 0.84 (0.14) |
| MEAN | | | | 1 | 1 | 0.76 (0.14) |
| MagDG | *B.diversus* | Queen-present | 9 | 1 | 1 | 0.81 (0.13) |
| Back | *B.diversus* | Queen-present | 11 | 1 | 1 | 0.79 (0.11) |
| Hyaku | *B.diversus* | Queen-present | 92 | 1 | 1 | 0.73 (0.19) |
| MEAN | | | | 1 | 1 | 0.78 (0.14) |
| Miyama | *B.honshuensis* | Queen-absent | 12 | 1 | 1 | 1.00 (0.00) |

# Figure Legends

Fig. 1

Flow of SLCA algorithm. See the algorithm section in Materials and Methods in Chapter 1 for a full explanation.

Fig. 2

Scoring matrix (a) and clustered matrix (b) for *B. ardens* data. Colonies are identified by letters, and individuals are identified by numbers. SLCA accurately reconstructed three sibling groups along the diagonal of the matrix.

Fig. 3

A map of Biratori indicating the location and sample sizes for each site.

Fig. 4

The resulted matrix for the number of families estimated by SLCA. Each block represents estimated full sibling families.

Fig. 5

Flow of mSLCA algorithm. See the algorithm section in the main text for a full explanation.

Fig. 6

Boxplots of deviations from the actual number of families with four distributions of 50 individuals by (a) SLCA and (b) mSLCA. Left-hand panels correspond to 4 alleles per locus, centre panels correspond to 8 alleles per locus, and right-hand panels to 12 alleles per locus. Upper-panels corresponds to 4 loci, middle to 8, and

lower to 12. In each panel, left-hand plots correspond to a family distribution of (10, 10, 10, 10, 10), centre-left plots to (20, 10, 10, 5, 5), centre-right plots to (30, 5, 5, 5, 5), and right-hand plots to (40, 5, 2, 2, 1). To simplify the output presentation, each family distribution is represented by the largest size of the family, i.e. 10, 20, 30, and 40, respectively.

Fig. 7

Effect of family size on accuracy. Results are shown for tests using 4, 6, 8, 10 or 12 loci, with 4, 8 or 12 alleles per locus. The key gives the distribution of the families that were tested. Equally frequent distributions with 25, 50 or 100 individuals were tested.

Fig. 8

Mean and standard deviations of the estimated number of families with 50 individuals, using (a) SLCA and (b) mSLCA. The number of loci and the number of alleles per locus are as in Fig. 6.

Fig. 9

Mean and standard deviations of the estimated number of families with 200 individuals. Results are shown for 8 loci with 8 alleles.

Fig. 10

Mean and standard deviations for the estimated number of families for *B. terrestris* data, using (a) SLCA and (b) mSLCA. Results are shown for one to ten colonies.

Fig. 11

Locations of the study area of Tsukuba, Kasama, and Kuzuu.

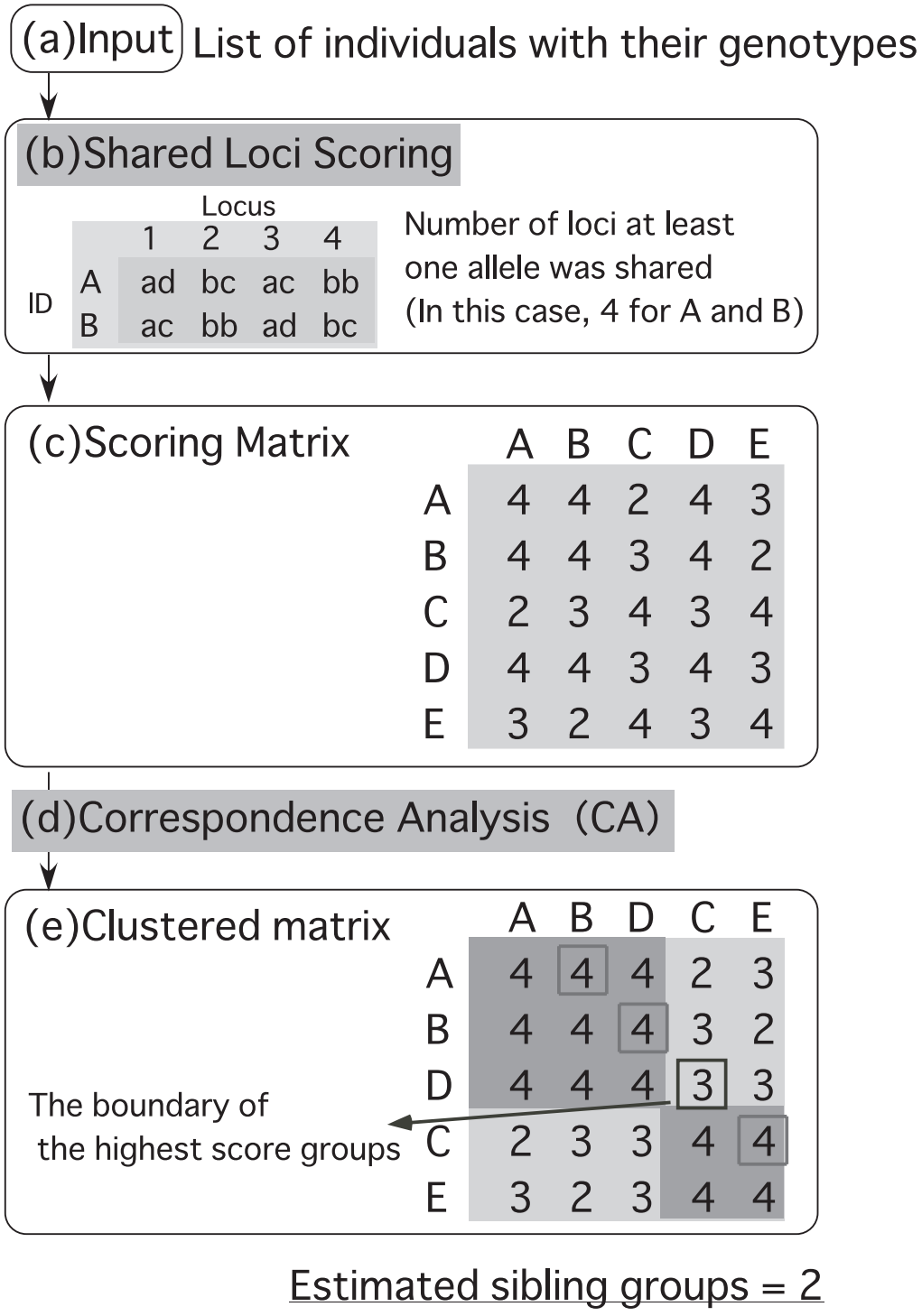


Figure 1 Flow of SLCA algorithm.

(a) Scoring matrix

| | C01 | C02 | A03 | B04 | C05 | C06 | C07 | B08 | C09 | C10 | C11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C01 | 5 | 5 | 2 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 |
| C02 | 5 | 5 | 3 | 5 | 3 | 5 | 5 | 1 | 5 | 5 | 5 |
| A03 | 2 | 3 | 5 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 2 |
| C04 | 5 | 5 | 2 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 |
| B05 | 2 | 3 | 2 | 2 | 5 | 3 | 2 | 5 | 2 | 3 | 2 |
| C06 | 5 | 5 | 3 | 5 | 3 | 5 | 5 | 1 | 5 | 5 | 5 |
| C07 | 5 | 5 | 2 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 |
| B08 | 1 | 1 | 1 | 1 | 5 | 1 | 1 | 5 | 1 | 1 | 1 |
| C09 | 5 | 5 | 2 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 |
| C10 | 5 | 5 | 3 | 5 | 3 | 5 | 5 | 1 | 5 | 5 | 5 |
| C11 | 5 | 5 | 2 | 5 | 2 | 5 | 5 | 1 | 5 | 5 | 5 |

(b) Clustered matrix

| | C01 | C09 | C11 | C05 | C07 | C02 | C10 | C06 | A03 | B04 | B08 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| C01 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 2 | 1 |
| C09 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 2 | 1 |
| C11 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 2 | 1 |
| C05 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 2 | 1 |
| C07 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 2 | 2 | 1 |
| C02 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 3 | 1 |
| C10 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 3 | 1 |
| C06 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 3 | 1 |
| A03 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 5 | 2 | 1 |
| B04 | 2 | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 2 | 5 | 5 |
| B08 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 |

Figure 2 Scoring matrix and clustered matrix for *B. ardens* data.

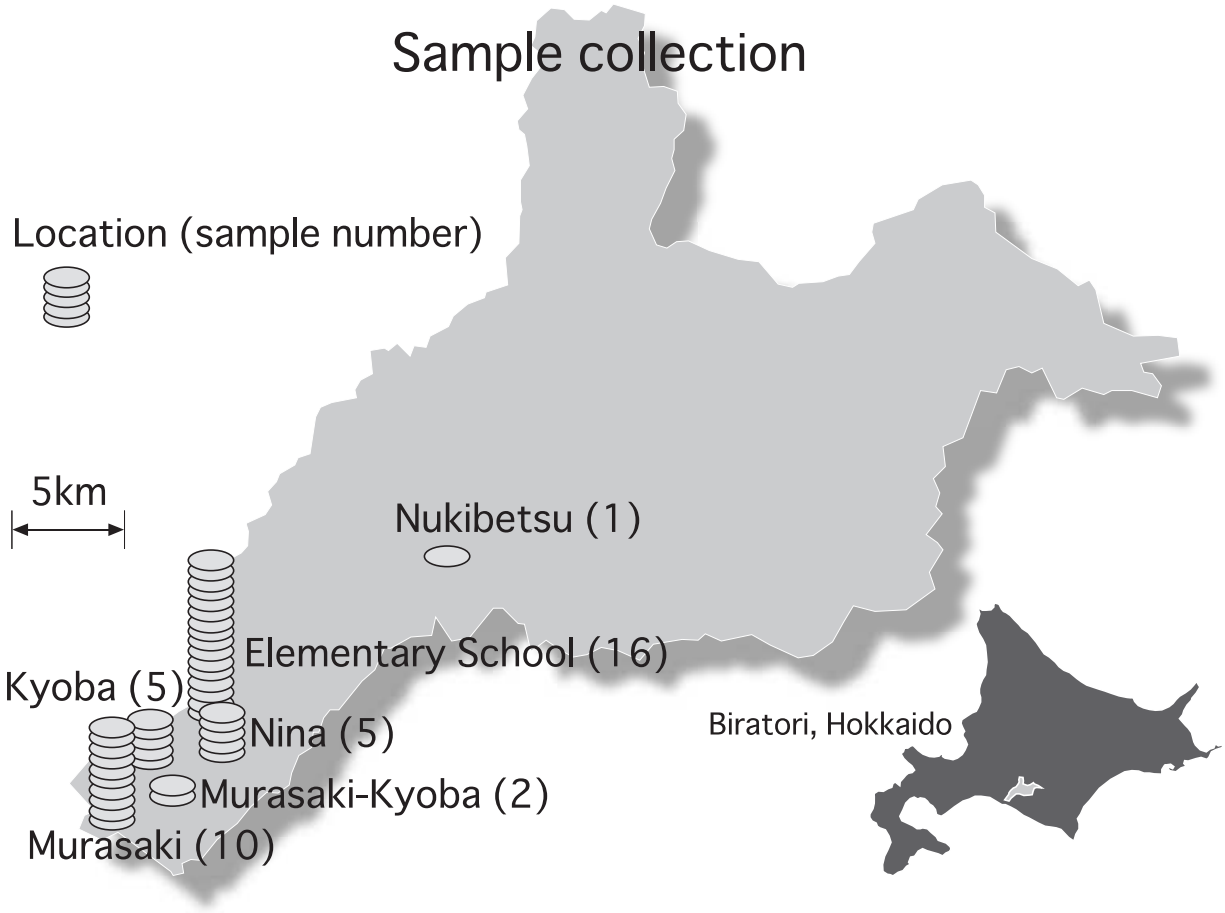


Figure 3 A map of Biratori indicating the location and sample sizes for each site.

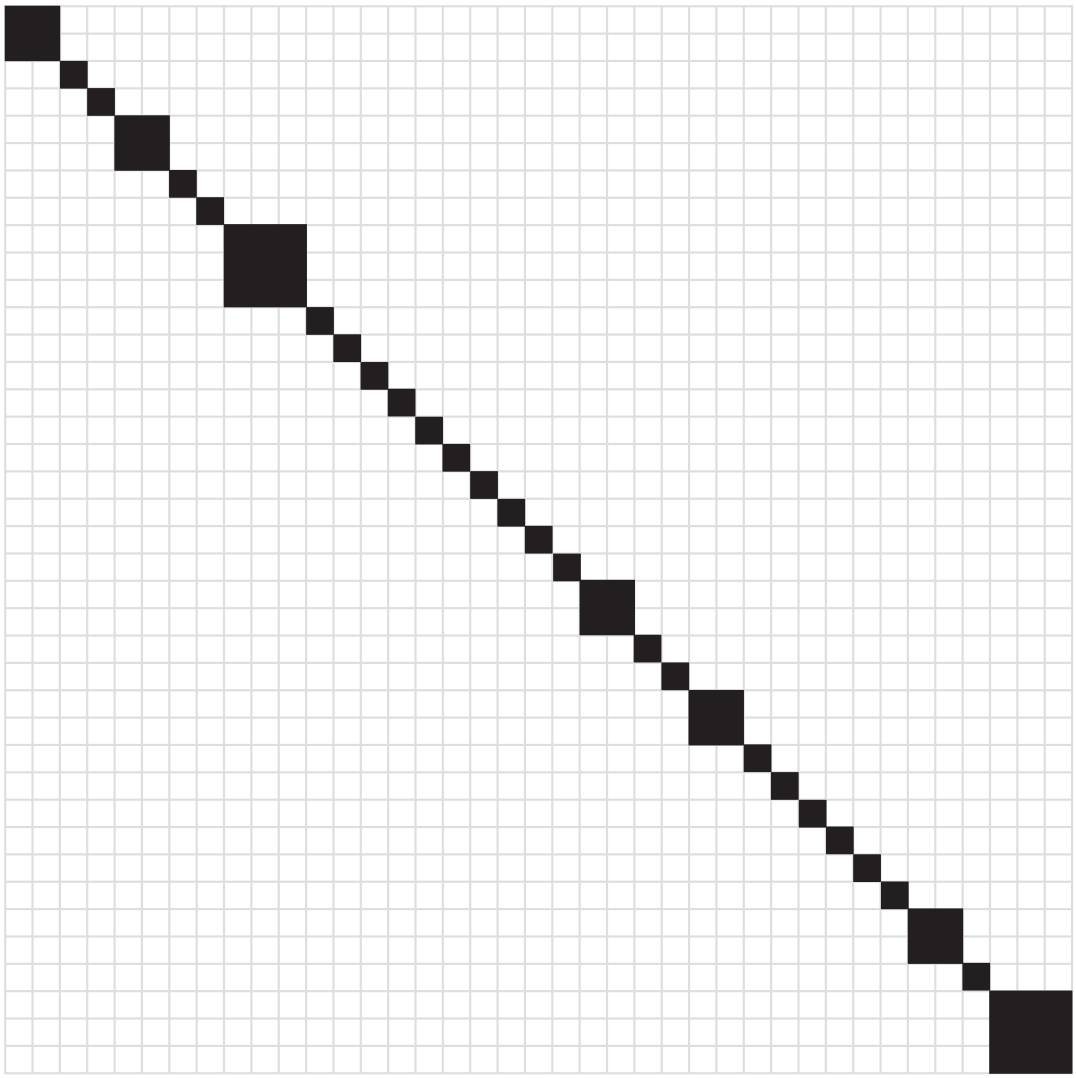

Figure 4 The resulted matrix for the number of families estimated by SLCA.

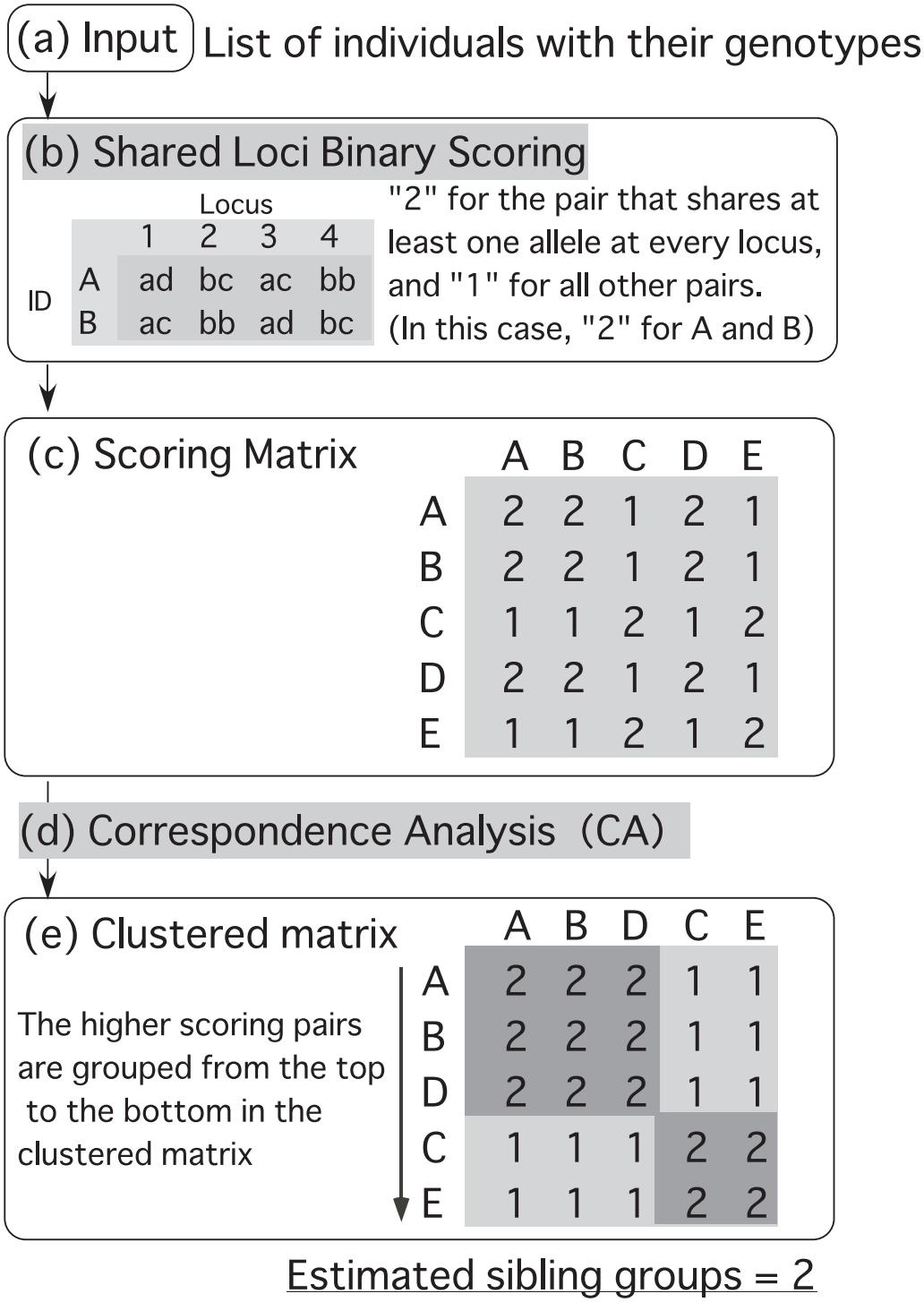


Figure 5 Flow of mSLCA algorithm.

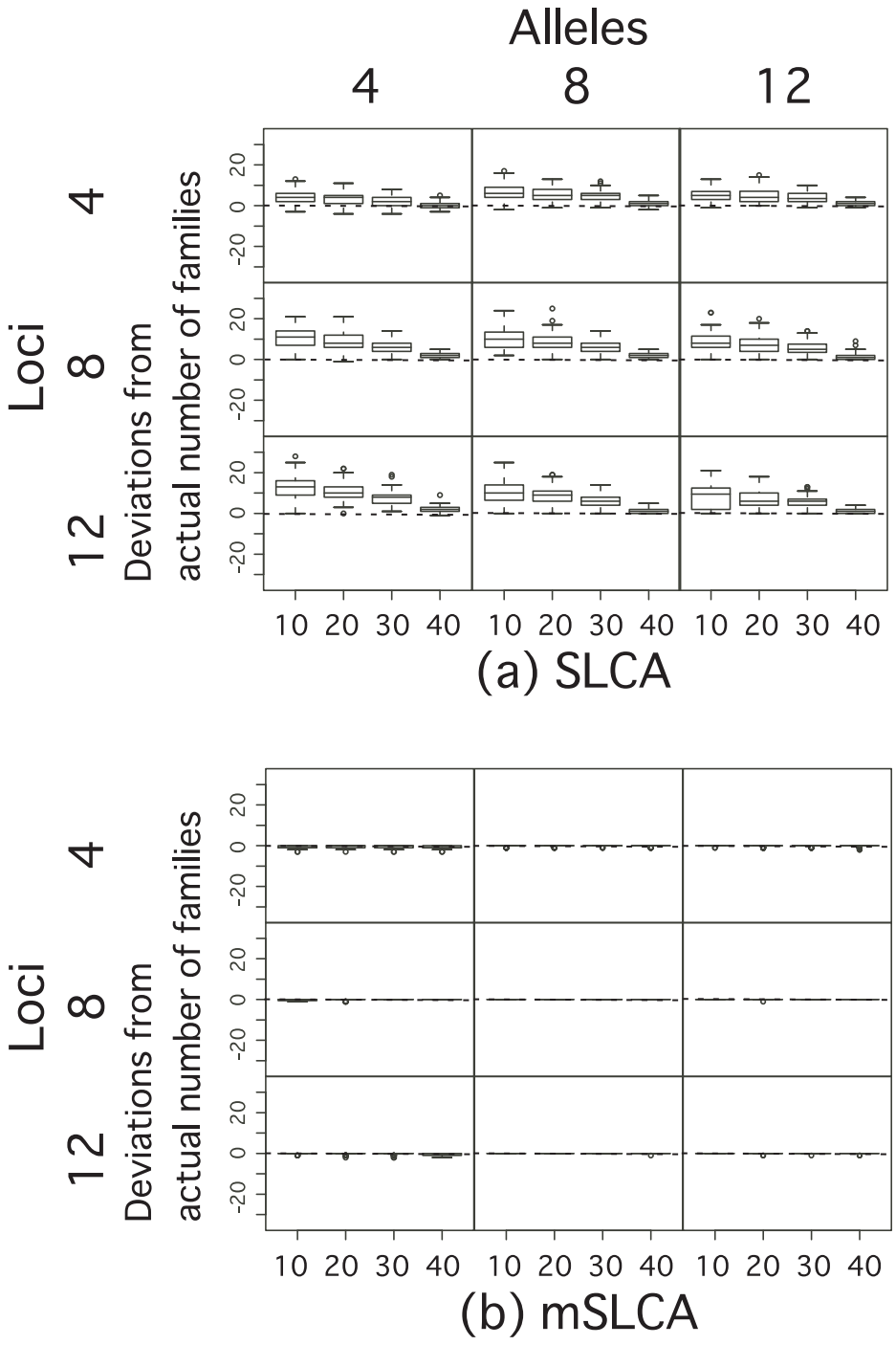


Figure 6 Boxplots of deviations from the actual number of families with 4 distributions of 50 individuals by SLCA and mSLCA.

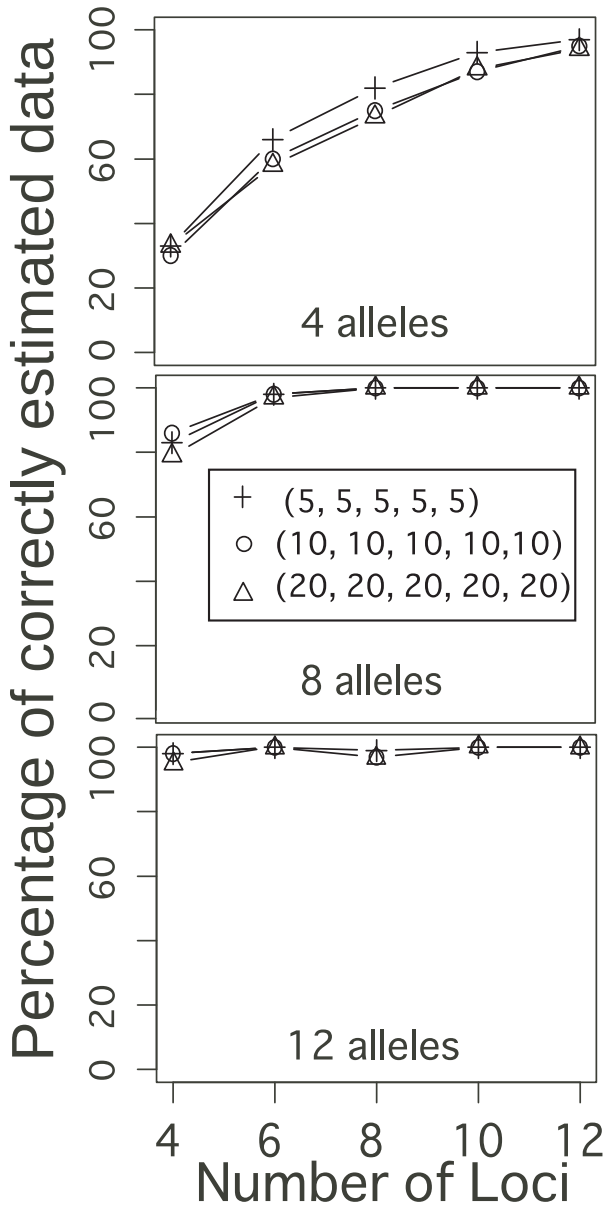


Figure 7 Effect of family size on accuracy.

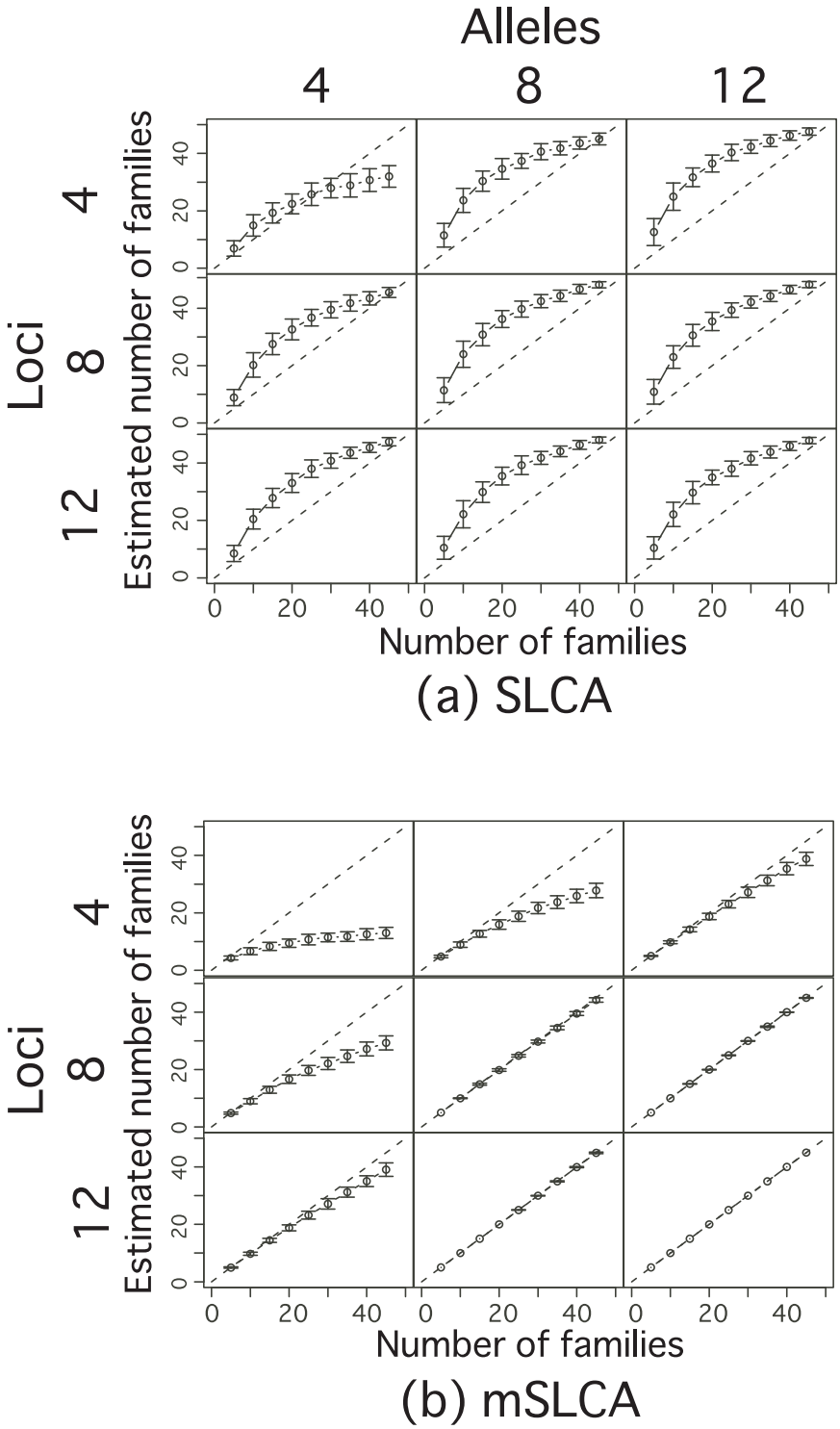


Figure 8 Mean and standard deviations of the estimated number of families with 50 individuals, using SLCA and mSLCA.

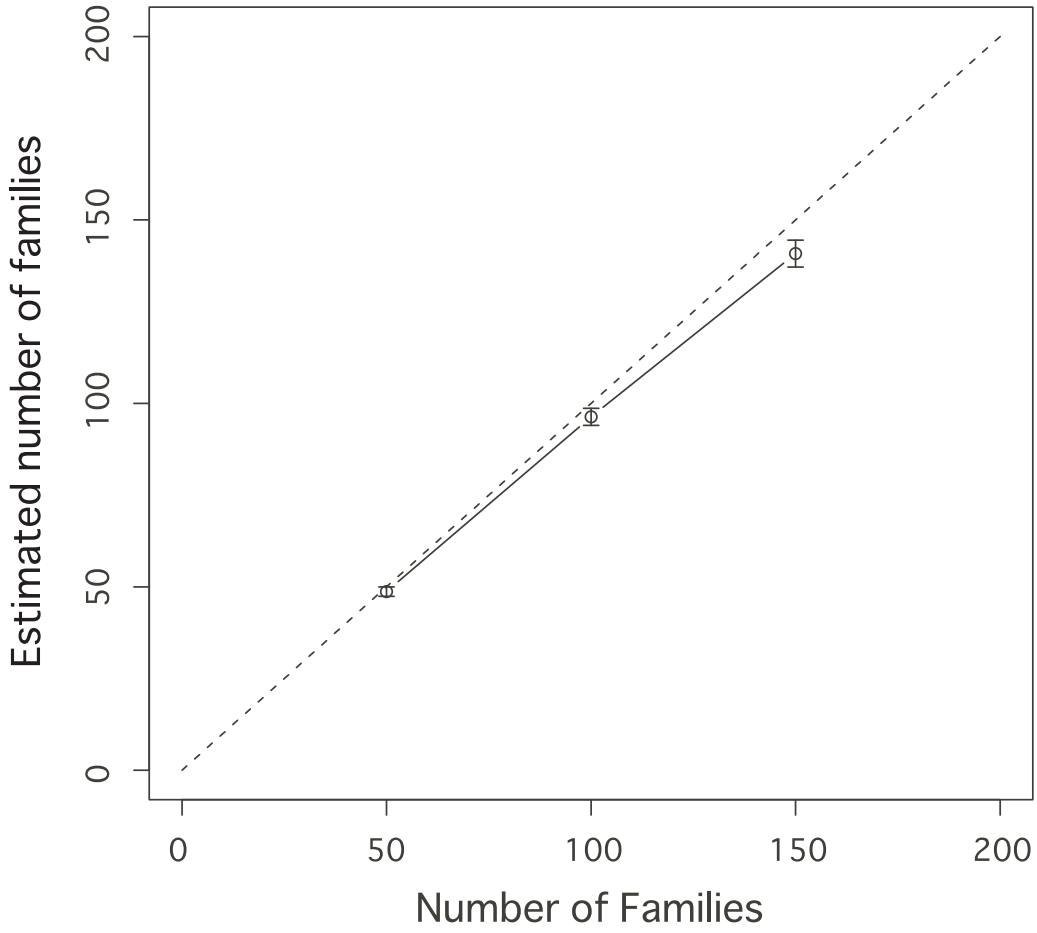


Figure 9 Mean and standard deviations of the estimated number of families with 200 individuals.

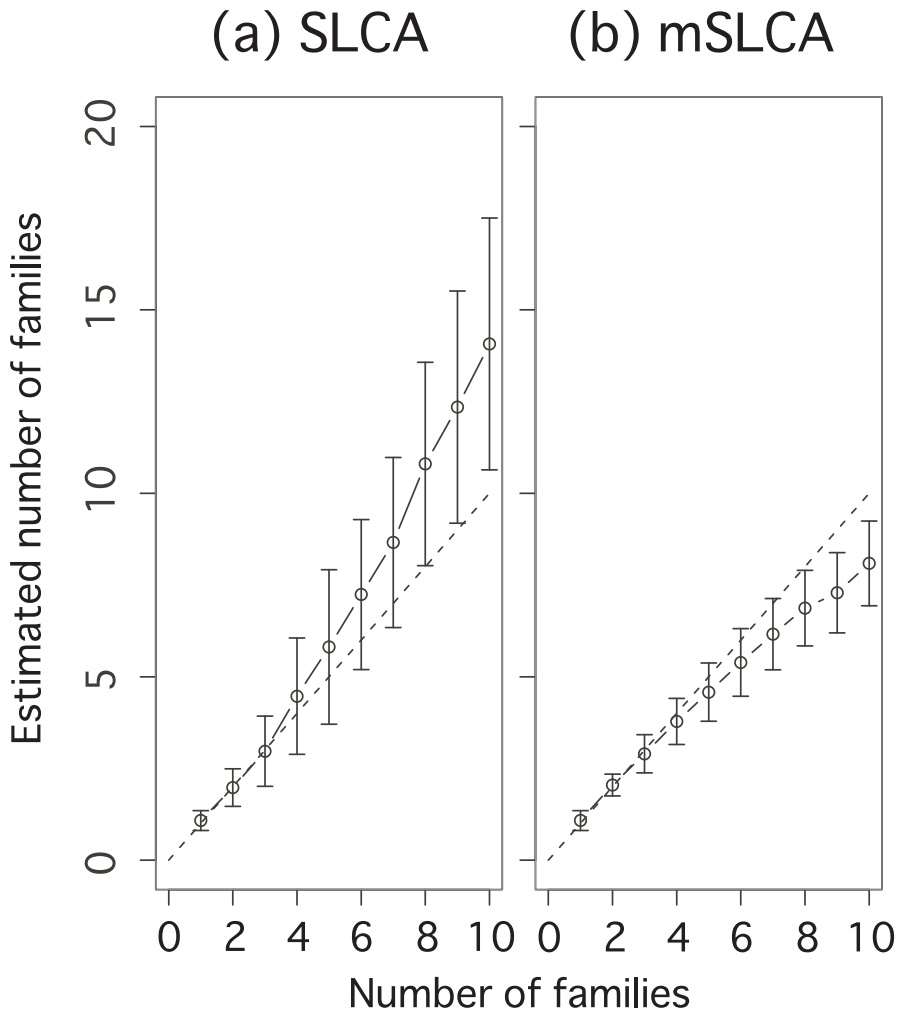


Figure 10 Mean and standard deviations of the estimated number of families calculated with *B. terrestris* data, using SLCA and mSLCA.

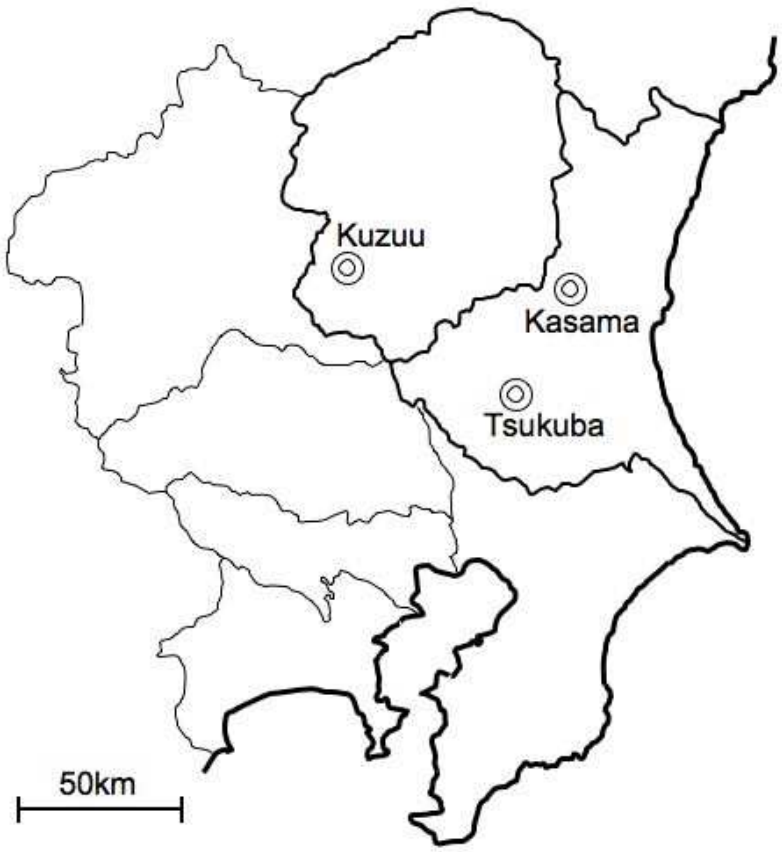


Figure 11 Locations of the study area of Tsukuba, Kasama, and Kuzuu