

Estimation of Effective Population Size  
of Diploid Species Using the Molecular  
Mark and Recapture (MMR) MethodKaori Murase<sup>1\*</sup> and Joe Murase<sup>2</sup><sup>1</sup>Graduate School of Natural Sciences, Nagoya City University, Japan<sup>2</sup>Goshawk Protection Fund, 2-5-1 Hanawada, Utsunomiya, Tochigi 320-0027, Japan

## Article Information

Received date: Aug 24, 2017

Accepted date: Oct 03, 2017

Published date: Oct 06, 2017

## \*Corresponding author

Kaori Murase, Graduate School  
of Natural Sciences, Nagoya City  
University, 1 Yamanohata, Mizuho-cho,  
Mizuho-ku, Nagoya, Aichi 467-8501,  
Japan, Tel: +81-52-872-5802;  
Fax: +81-52-872-1531,  
Email: kmurase@nsc.nagoya-cu.ac.jp

Distributed under Creative Commons  
CC-BY 4.0

**Keywords** Biodiversity management;  
Conservation; Genetic marker; Genetic  
repertoire; Methodology; Local resource  
manager; Parentage relationship;  
Sampling strategy; Sex ratio

## Abstract

This study proposed a method for managing many diploid species together without constructing any statistical models for specific species. Our purpose was to estimate how many different genetic compositions exist in the adult generation involved in their reproduction (We referred this number as genetic repertoire) through our Molecular Mark and Recapture (MMR) method. First, we developed the MMR method for diploid species and proposed theoretical formulae to calculate the variance and confidence interval of the genetic repertoires. Second, we made three virtual diploid species (human or birds, harem-forming mammals, and plants), which included the first generation and the second generation, and then we conducted simulations to estimate the genetic repertoires of the first generation. Third, we showed a test study using microsatellite genotype data of wild boar. Our results showed that our methods would be useful, especially in tropical forests, because the method did not require highly sophisticated statistical models or much prior information for a species. Moreover, it was able to estimate the genetic repertoires with a one-time random sampling of the parent and offspring individuals. Furthermore, a decrease or increase in genetic repertoire would be detectable by increasing the number of random sample collections to twice or more. We consider it has great potential to enhance management methods of biodiversity by local people.

## Introduction

Although the importance of biodiversity conservation has been a focus for a long time (especially in tropical areas), local people and the staffs of national parks are limited by a lack of budget, analytical equipment, microscopes, specialists, and academic information (e.g., [1]). How do biostatisticians intend to contribute to such countries? There is not a biostatistician in all area such as islands or mountainous areas, even if a country has many biostatisticians.

Many existing studies have estimated the effective population size of a specific species (e.g., [2-4]) with the aim of improving the precision and accuracy of the estimation. These studies were performed with a sufficient budget, preliminary surveys, and a group of specialists of varied disciplines. Conversely, few studies (with the exception of [5]) have been concerned with discovering the tendency of effective population sizes in many species in areas where detailed examinations cannot be conducted (i.e., national parks in tropical areas that are known hot-spots of biodiversity (e.g., [6]) due to the following constraints: 1) Academic resource availability is limited; 2) There is not or a very small amount of prior information about a species of interest and 3) Too many species are candidates for future studies. We, therefore, think that local staff and people, who are a substantial manager of biodiversity in the hot-spots area, want to know how to manage the biodiversity, especially without any species specific statistical model.

What is needed in the scene of the biodiversity management is to grasp which species are going to increase or decrease in a limited ecosystem such as a national park as soon as possible. The number of adult individuals that can reproduce next generation may be a good indicator for the necessity, which can be estimated using traditional mark and recapture method [7]. However, by the number of the adult individuals only and without considering the genetic diversity of the adult population, the vulnerability of the species (i.e., the extinction risk) would not be assessed. Sea otters are one example: Once they decreased in the number. Although their population recovered by some conservation activities, a massive fatalities was occurred due to the insufficient genetic diversity which did not recovered well [8]. Then, the concept that excludes the adult individuals of the completely same or very similar genetic composition would be required. "The genetic repertoires" conforms to the concept. Generally, which species should be conserved in priority to other species is a very common and consistent issue [9]. If we can easily estimate the genetic repertoires of adult generation that participate in their reproduction, the genetic repertoires of many species can be monitored all together, which would provide a criterion for deciding which species should be prioritized for conservation. While many previous studies focused on one or some specific species with sufficient funds and specialists such as statisticians, there are few studies about the method for monitoring many species collectively.

Murase and Fukita [5] proposed a MMR (Molecular Mark and Recapture) method as a way to monitor many species at the same time. The method is summarized as follows: The individuals of parent and offspring generation are sampled randomly; Parent-offspring relationships are identified; the number of adult individuals of the parent generation that can reproduce next generation is calculated using the relationships. The number of adult individuals of the parent generation was written as the effective population size for the biodiversity management in Murase and Fukita [5]. The effective population size for the biodiversity management is different from the ones of evolutionary population genetics or coalescent theory. Evolutionary population genetics or coalescent theory focus on the process of mutation, the MMR method does not require the mutation at all. It is, of course, possible to consider for applied studies that take the mutation into account, the frequency of the mutation would be very low when we use parent and offspring generation only and study area is restricted such as national parks. Moreover, because we want to argue about usefulness of "MMR method" itself in this study, we want to exclude "the MMR method in consideration of mutation" which is derivative topics of our issue. For the staff of wildlife management, the number of adult individuals of the parent generation is important.

Although the previous MMR method has a lot of advantages, it is intended for clonal propagation species and is not so useful for field studies, for sexual reproduction is common in the field. We, then, considered for proposing a modified MMR method for the sexual reproduction species with the advantages kept.

Multiple samplings are often impossible in practical field studies, which make it difficult to estimate the interval of the genetic repertoires (i.e., the minimum and maximum of how many genetically different individuals a species will contain). Thus, it would be significant to derive the theoretical variance and confidence interval of the genetic repertoires even if we were allowed to conduct sampling only once. Therefore, our first purpose is to develop the MMR method for diploid species and to propose theoretical formulae to measure the variance and confidence interval of the genetic repertoires.

Diploid species have different mating systems; however, Murase and Fukita [5] did not examine the impact of the MMR method precipitated by such differences. We should examine whether the MMR method for diploid species is applicable to the varied mating systems. Therefore, our second purpose is to examine the applicability of the MMR method for diploid species assuming three different mating systems. In this study, we assume three different populations that have three different mating systems: (1) Human; monogamous species such as humans or birds; (2) Animal; polygamous species such as lions or sea lions; and (3) Plant; monoecious species. The applicability of the MMR method was examined for each type of population. Additionally, how the variability of the three mating systems affects the estimates derived from the theoretical formulae (the variance and confidence interval of the genetic repertoires) is also discussed.

In wildlife studies, we sometimes need information for either gender (e.g., the effective population size of adult females or adult males) as well as the effective population size of both genders in total [10]. For example, if the sex ratio of the individuals with wings of a rare ant species inhabiting a national park is skewed toward males, the probability of the disappearance of the species from the site would

be considered to be high. When considering the conservation of the species, we may misjudge the species as a non-endangered species that has an adequate effective population size in this example if we concentrate our attention on the total effective population size. Thus, we should prepare a method that can separately estimate the genetic repertoires of adult females and males. Among the above three diploid species, *Plant* has no sex ratio because they are a monoecious species; however, in the *Human* and the *Animal* groups the sex ratio is often of ecological interest. The sex ratio of both adult and child individuals of the *Human* group is one to one. In the *Animal* group, the number of adult individuals that can participate in reproduction often differs between males and females because it is assumed that the small number of dominant males mate with many females. Although the sex ratio of young individuals of *Animal* species may be one to one, only strong male individuals survive; thus, the sex ratio again skews towards females. This type of lifecycle can be generally assumed. To conserve such wildlife, we have to estimate the genetic repertoires of both males and females. Thus, our third purpose of this study is to propose a method that separately estimates the genetic repertoires of adult males and adult females of *Animal* species. Additionally, the sex ratio of young available individuals often leans towards either sex (e.g., a species that has a life history of young males dispersing far away and young females staying at their birth place or humans collecting horned male individuals due to their economic value (e.g., ivories or antlers)). Thus, we add a discussion of whether the MMR method is able to estimate the genetic repertoires of adult males and adult females if the available young individuals are only male or female to the third purpose.

## Materials and Methods

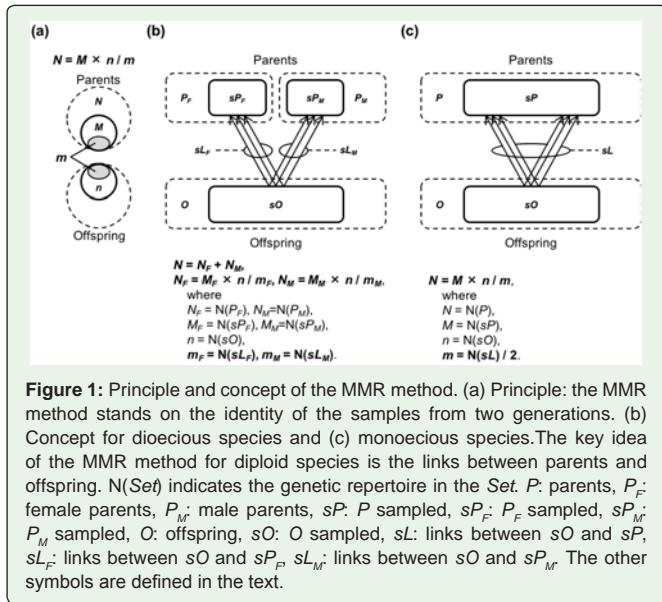
This paper proposes "the genetic repertoire" as an index for monitoring many species collectively. First, we describe a principle of the MMR method. MMR is the abbreviation of "Molecular Mark and Recapture". Second, we describe estimation methods of the MMR method using the three statistics that we had made. Third, we wrote our material which is simple virtual species in a computer. Fourth, we show the results of the three experiments (the three simulations) using the MMR method. Finally, we showed an example study using microsatellite genotype data of wild boar.

### The MMR method for diploid species

**Principle and concept:** As described in Murase and Fukita [5], the MMR method estimates the effective population size (that is the same as the number of genetic repertoires) of a parent generation. Murase and Fukita [5] examined haploid species with clonal propagation to establish the principle of the MMR method. In this study, we estimated the genetic repertoires of a parent generation of diploid species with panmictic reproduction. Equation 1 is used for the estimation,

$$\hat{N} = \frac{M \cdot n}{m}, \quad (1)$$

where  $\hat{N}$  is the estimated genetic repertoires of a parent generation,  $M$  and  $n$  are the genetic repertoires of sampled parents and offspring, respectively, and  $m$  is the genetic repertoire common to both the sampled parents and offspring. This principle is shown in Figure 1a. Although the concept of the MMR method for diploid species is a bit different between dioecious and monoecious species, in either case parentage links are used as the genetic identity between a sampled parent and a sampled offspring individual. The number



**Figure 1:** Principle and concept of the MMR method. (a) Principle: the MMR method stands on the identity of the samples from two generations. (b) Concept for dioecious species and (c) monoecious species. The key idea of the MMR method for diploid species is the links between parents and offspring. N(Set) indicates the genetic repertoire in the Set. P: parents, P<sub>F</sub>: female parents, P<sub>M</sub>: male parents, sP: P sampled, sP<sub>F</sub>: P<sub>F</sub> sampled, sP<sub>M</sub>: P<sub>M</sub> sampled, O: offspring, sO: O sampled, sL: links between sO and sP, sL<sub>F</sub>: links between sO and sP<sub>F</sub>, sL<sub>M</sub>: links between sO and sP<sub>M</sub>. The other symbols are defined in the text.

of links is related to  $m$  as an essential random variable. This concept is shown in Figure 1b and 1c. In the case of the dioecious species, female parent-offspring links ( $m_F$ ) and male parent-offspring links ( $m_M$ ) are separately determined, and the genetic repertoires of the female parents ( $\hat{N}_F$ ) and the male parents ( $\hat{N}_M$ ) are also separately estimated. In the monoecious species, a maximum of two parentage links can exist between one offspring individual and the parents sampled; thus, the number of links divided by two is employed as  $m$ . Example calculations are shown in Figure 2a and 2b. Importantly, individuals who cannot be discriminated genetically are treated as one individual because the MMR method estimates the genetic repertoire.

**Expectation and variance of parent-offspring links:** In dioecious species, the probability of a link between an offspring individual and female parent individual sampled is  $p_F$ , when the sampling rate of the female parent individual is  $p_F$  ( $0 \leq p_F \leq 1$ ). The female parent individuals are sampled with replacements because the female parent individual linked with an offspring individual is not removed from the sample set. A female can be the parent of many offspring individuals. Thus, the number of links between female parents and offspring individuals follows a binomial distribution of

$$E(m_F) = np_F \quad V(m_F) = np_F(1 - p_F),$$

where  $m_F$  is a random variable representing the number of links and  $n$  is the number of offspring sampled. The idea is the same when considering the links between sampled male parents and offspring individuals:

$$E(m_M) = np_M \quad V(m_M) = np_M(1 - p_M),$$

where  $m_M$  is a random variable representing the number of links between male parents and offspring individuals and  $p_M$  ( $0 \leq p_M \leq 1$ ) is the sampling rate of the male parents. In the case of monoecious species,

$$E(m) = E\left(\frac{X}{2}\right) = \frac{1}{2}E(X) = \frac{1}{2} \cdot 2np = np,$$

$$V(m) = V\left(\frac{X}{2}\right) = \frac{1}{2^2}V(X) = \frac{1}{2^2} \cdot 2np(1 - p) = \frac{1}{2} \cdot np(1 - p),$$

where  $X$  is the number of parent-offspring links.

Variance and confidence interval of the number of parents: In the case of dioecious species, we estimate  $\hat{N}_F$  and  $\hat{N}_M$  separately and then calculate  $\hat{N}$  and its variance as

$$\hat{N} = \hat{N}_F + \hat{N}_M,$$

$$V(\hat{N}) = V(\hat{N}_F) + V(\hat{N}_M).$$

$\hat{N}_F$  is a function of  $m_F$  and is written as

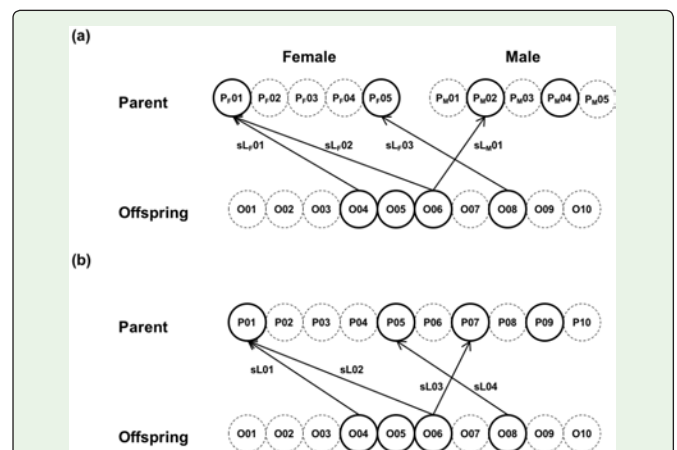
$$\hat{N}_F = \frac{M_F \cdot n}{m_F} = f(m_F),$$

where  $\hat{M}_F$  is the number of genetic repertoires of female parents sampled.  $V(\hat{N}_F)$  is derived as follows based on Seber (1982):

$$V(\hat{N}_F) = V\{f(m_F)\} \approx V(m_F) \cdot \left(\frac{df}{dm_F}\right)^2 = np_F(1 - p_F) \cdot \frac{M_F^2 \cdot n^2}{m_F^4}.$$

We substitute  $\hat{p}_F = \frac{m_F}{n}$  for  $p_F$  and then obtain

$$V(\hat{N}_F) \approx \left\{n \cdot \frac{m_F}{n} \cdot \left(1 - \frac{m_F}{n}\right)\right\} \cdot \frac{M_F^2 \cdot n^2}{m_F^4} = \left\{\frac{m_F(n - m_F)}{n}\right\} \cdot \frac{M_F^2 \cdot n^2}{m_F^4} = \frac{M_F^2 \cdot n \cdot (n - m_F)}{m_F^3}.$$



**Figure 2:** Example calculations of the MMR method. (a) An example of dioecious species. Three links between female parents and offspring (sL<sub>F</sub>01 - sL<sub>F</sub>03) and one link between male parents and offspring (sL<sub>M</sub>01) are identified by a parentage analysis using genotype data. The genetic repertoires of parents ( $N$ ) is calculated as  $N_F = M_F \times n / m_F = 2 \times 4 / 3 = 2.67$ ,  $N_M = M_M \times n / m_M = 2 \times 4 / 1 = 8$ ,  $N = N_F + N_M = 10.67$ . (b) An example of monoecious species. Four links are identified (sL01 - sL04). The genetic repertoires of the parents is calculated as  $N = M \times n / m = 4 \times 4 / 2 = 8$  because  $m$  is the number of links divided by two. In both cases, the true value of  $N$  is 10.

$V(\hat{N}_M)$  is also derived in the same way with the substitution of  $\hat{p}_M = \frac{m_M}{n}$  for  $p_M$ . Thus,  $V(\hat{N})$  is written as

$$V(\hat{N}) \approx \frac{M_F^2 \cdot n \cdot (n - m_F)}{m_F^3} + \frac{M_M^2 \cdot n \cdot (n - m_M)}{m_M^3}. \quad (2)$$

In the case of monoecious species,  $V(\hat{N})$  is derived as follows:

$$V(\hat{N}) = V\{f(m)\} \approx V(m) \cdot \left(\frac{df}{dm}\right)^2 = \frac{1}{2}np(1-p) \cdot \frac{M^2 \cdot n^2}{m^4} = \frac{M^2 \cdot n \cdot (n-m)}{2m^3}, \quad (3)$$

where  $M$  is the number of genetic repertoires of parents sampled. The distribution of  $\hat{N}$  asymptotically approaches a normal distribution with increasing  $\hat{N}$ ; therefore, the 95% confidence interval is calculated as

$$\hat{N} \pm 1.96\sqrt{V(\hat{N})} \quad (4)$$

### Simulation data sets of diploid species

We generated three data sets corresponding to the three mating systems using R 3.0.2 [11]. Each data set contained individuals of the parent and offspring generations.

**Human:** A dioecious species with a 1 to 1 the ratio of males to females. The fitness of all individuals was the same. One male and one female of the parent generation made a pair, and then the female had one male and one female individual of the offspring generation. Humans or some types of birds were assumed in this system. A total of 100 males and 100 females were generated for the parent generation. The sex ratio and population size of the offspring generation were the same as the parent generation.

**Animal:** A dioecious species with a 1 to 3 ratio of males to females contributing to reproduction, whereas the sex ratio of the offspring born is 1 to 1. The fitness of the males and females was distributed differently. One male of the parent generation had 1-22 individuals of the offspring generation; this number was obtained from a random variable following a geometric distribution, but 1 was used instead when 0 was drawn because the geometric distribution was a discrete distribution starting from 0, whereas any male in the parent generation was assumed to contribute to the offspring generation. One female of the parent generation had 1-7 individuals of the offspring generation; the number was obtained from a random variable following a Poisson distribution, but 1 was used instead when 0 was drawn for a similar reason. Wild boar or deer were assumed in this system. A total of 50 males and 150 females were generated in the parent generation. A total of 150 males and 150 females were generated in the offspring generation.

**Plant:** A monoecious species. The fitness of each individual was distributed unequally. One individual of the parent generation had 1 or more individuals of the offspring generation; the number was obtained from a random variable following a geometric distribution, but 1 was used instead when 0 was drawn for the same reason

described in *Animal*. A total of 200 individuals were generated in the parent generation. The offspring generation included a population size of 1000, which was five-fold higher than the parent generation.

### Simulation experiments and statistical evaluation

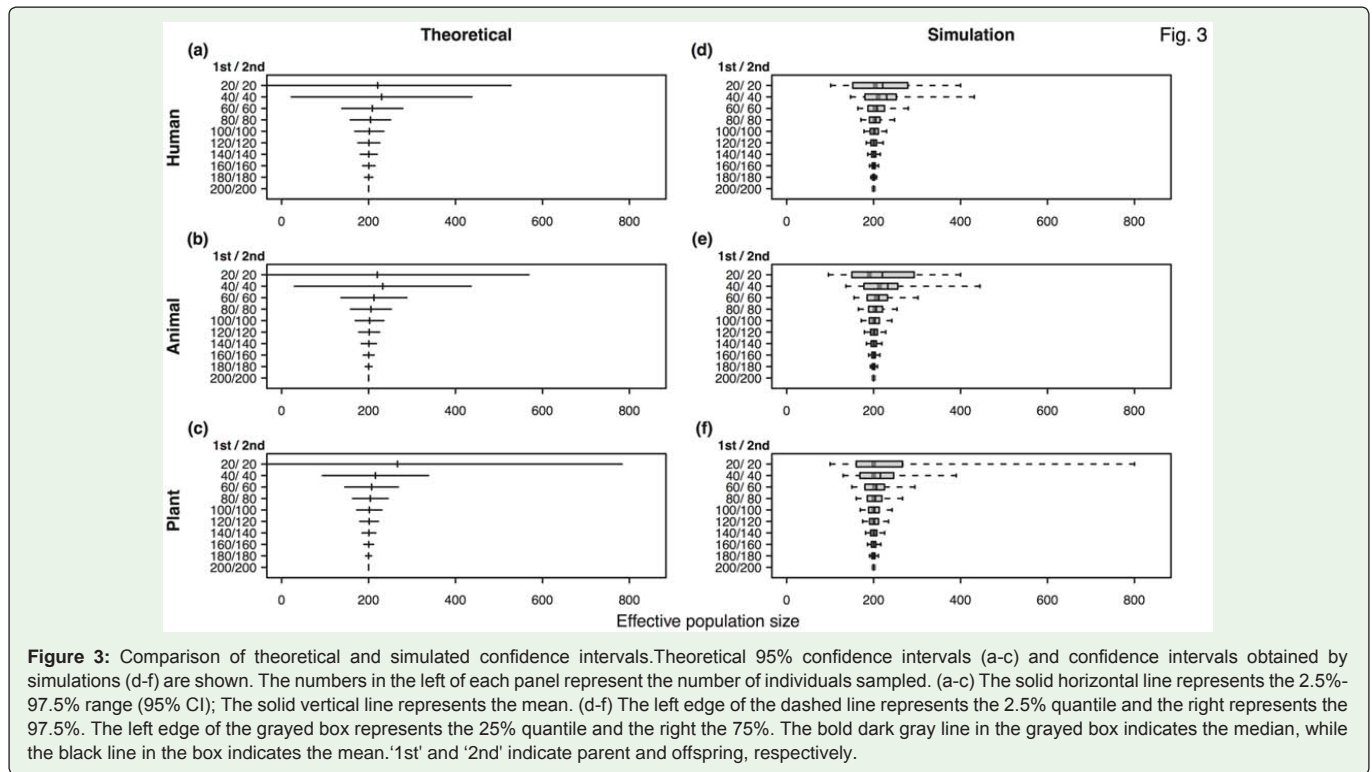
**Procedure of each simulation experiment:** In all simulation experiments, the first samples were drawn from the parent generation randomly without replacement, and the number of samples ( $M_F$  and  $M_M$  for dioecious species or  $M$  for monoecious species) was counted. Second, samples were drawn from the offspring generation randomly without replacement, and the number of samples ( $n$ ) was counted. When duplicated samples of the offspring generation were found (i.e., when they originated from the same parents), only one of the samples was used. Third, the number of links between the parent and the offspring generations ( $m_F$  and  $m_M$  for dioecious species or  $m$  for monoecious species) was counted. Last, the number of individuals of the parent generation ( $\hat{N}_F$  and  $\hat{N}_M$  for dioecious species or  $\hat{N}$  for monoecious species) was estimated as described in Figures 1 and 2. When no links were found (i.e.,  $m_F$ ,  $m_M$ , or  $m$  was zero), we denoted the situation as 'Inf' because  $\hat{N}_F$ ,  $\hat{N}_M$ , or  $\hat{N}$  became infinite. We repeated the steps 1000 times under the same conditions.

**Simulation subjects and sampling types:** First, to examine whether the MMR method could be applied to diploid species regardless of the differences in the three mating systems (*Human*, *Animal*, and *Plant*), we conducted simulations and evaluated the number of individuals in the parent generation. The numbers of samples in the parent and offspring generations were set as identical (from 20 to 200 by 20). We compared the theoretical 95% confidence intervals (CI) obtained by Equation 4 and the 95% intervals defined as 2.5%-97.5% of the quantile value of  $\hat{N}$  obtained by the simulation experiments to confirm the validity of the theory of the MMR method. The theoretical 95% CI was calculated using the average of  $\hat{N}$  and  $V(\hat{N})$  obtained by the simulation experiments provided that they were finite values. Next, we conducted additional simulation experiments to explore how we should sample the two generations. We fixed the number of samples of the parent or the offspring generation. When the number of samples of the parent generation was fixed to 20, the offspring generation varied from 20 to 200 by 20. Conversely, when the offspring generation was fixed to 20, the parent generation varied from 20 to 200 by 20. These settings were applied to the three mating systems.

**Separate estimation of males and females:** To examine whether the MMR method for diploid species was able to separately estimate the genetic repertoires of either gender, we conducted detailed simulation experiments. The *Animal* dataset was used because its sex ratio was skewed towards females. The sex of the offspring individuals sampled was identified, and then the following estimations were conducted: the genetic repertoires of adult females using offspring of both sexes, adult females using female offspring only, adult females using male offspring only, adult males using offspring of both sexes, adult males using female offspring only, and adult males using male offspring only.

**Statistical evaluation:** We calculated four statistics to evaluate the result of the simulation experiments based on the MMR method. The first statistic was the mean value of  $\hat{N}$ , which represented the genetic repertoires of the parent generation. The second statistic was the





median of  $\hat{N}$ . The third was the 95% confidence interval (CI) of the quantile of  $\hat{N}$ . This interval consisted of the 2.5% and 97.5% values of the 1,000  $\hat{N}$  values in ascending order. The fourth was the coefficient of variation (CV) of  $\hat{N}$ . The number of 'Inf' was also counted. All of the simulation experiments were conducted using R 3.0.2 [11].

**An example study using microsatellite genotype data of wild boar**

We used microsatellite genotype data of adult Japanese wild boar [12] together with that of young individuals (unpub.) for our example study. The largest population consisted of 29 adult (15 male and 14 female parent individuals) and 19 young individuals (See Web Appendix A for details). Seven father-offspring links and five mother-offspring links were identified using Cervus 3.0.7 [13], which determines parent-offspring relationships based on likelihood ratios with the level of confidence calculated by simulation. Then, we estimated the genetic repertoires of male and female parent individuals respectively using our MMR method. We also calculated their variance and confidence interval of the genetic repertoires of male and female parents, respectively.

**Results**

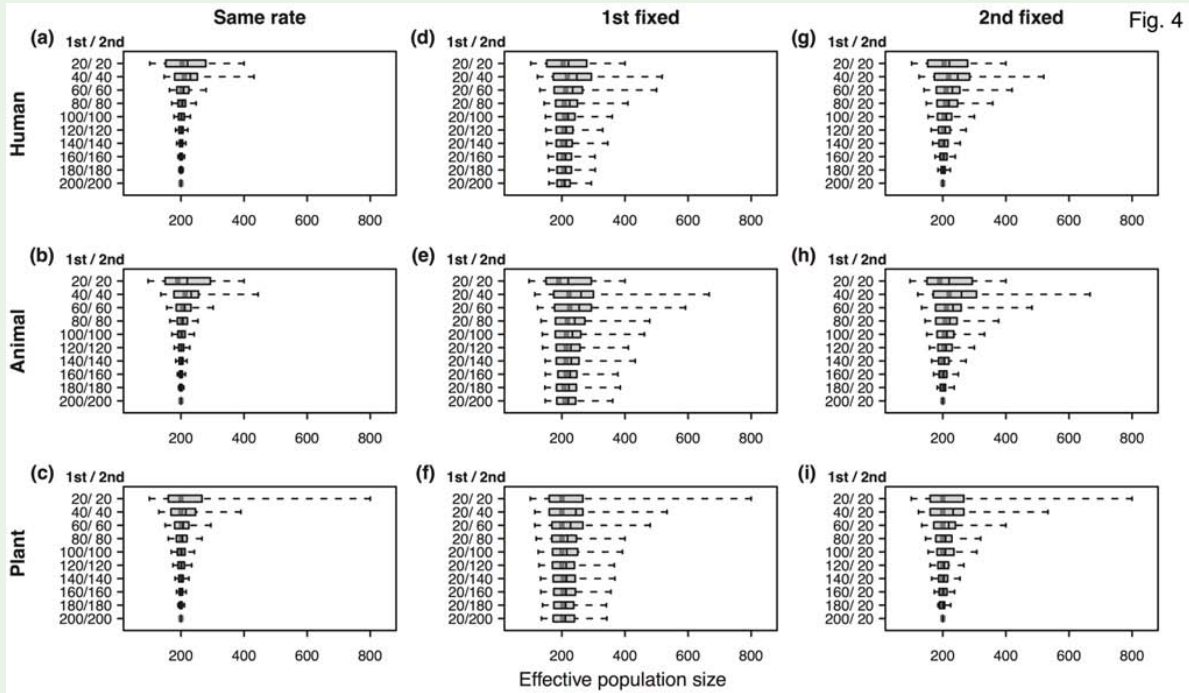
The results of the simulation experiments based on the three mating systems are shown in Figure 3 and Web Table 1. All of the results showed that the means of  $\hat{N}$  were slightly biased in the cases with small sample sizes, but the biases rapidly decreased with the increase in the sample number. In most cases, the medians were not biased. In the cases of *Human* and *Animal*, the theoretical 95% Confidence Intervals (CIs) and the 95% range of  $\hat{N}$  were very close

to one another, with the exception of the small sample size cases. In the case of *Plant*, the theoretical 95% CIs and the simulated 95% range were close even in cases with a small sample size. In the smallest sample size cases (denoted '20/ 20' in Figure 3) of *Human* and *Animal*, their 95% ranges were narrower than those of the '40/ 40' cases. Their means and medians were also smaller values.

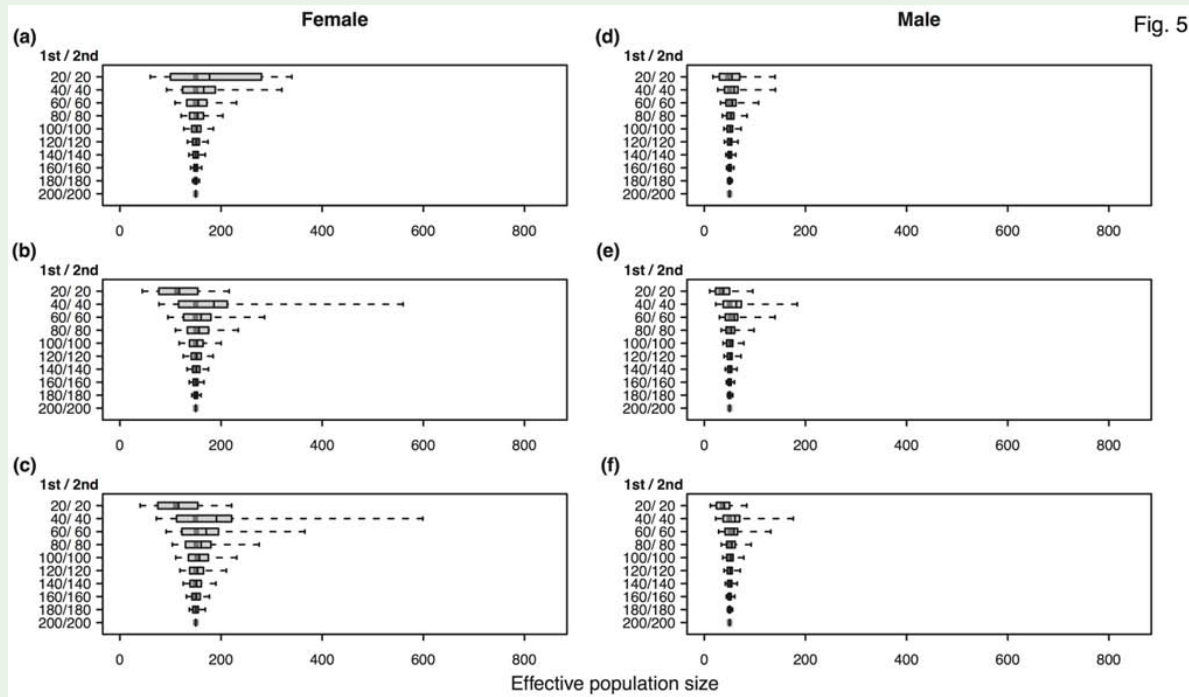
The results of the simulation experiments based on the three sampling strategies are shown in Figure 4 and Web Table 2. When the parents' sample was limited to a small number, the precision of the estimated genetic repertoires was not improved, even in the cases with the larger sample size (Figure 4d-f and CVs of Web Table 2d-f). Conversely, when the offspring was limited to small number, the precision was improved by the increase in the parents' sample numbers (Figure 4g-i and CVs of Web Table 2g-i).

More in depth results of the simulation experiments of *Animal* were shown in Figure 5 and Web Table 3. The MMR method was successfully able to separately estimate the genetic repertoires of female and male parents (Figure 5a, d and Web Table 3a, d). Moreover, the genetic repertoires of the parents of each sex was also estimated using female or male offspring individuals alone (Figure 5b, c, e, f and Web Table 3b, c, e, f), although naturally its precision was lower than when both sexes of the offspring individuals were used together. When the offspring individuals sampled contained no females, we were unable to estimate the genetic repertoires of the parents using female offspring individuals; then, this case was denoted as 'Inf.' No male offspring cases were treated in this manner.

In our example study,  $\hat{N}_F$  and  $\hat{N}_M$  were calculated as 53.2 and 40.7, respectively. Their 95% confidence intervals were calculated as 13.2-93.2 and 16.7-64.7, respectively.



**Figure 4:** Distribution of the genetic repertoires estimated for three mating systems. The genetic repertoires of the parent generation are estimated according to three mating systems and three sampling strategies. The numbers in the left of each panel represent the number of individuals sampled. The left edge of the dashed line represents the 2.5% quantile and the right represents the 97.5% quantile. The left edge of the grayed box represents the 25% quantile and the right edge represents the 75% quantile. The bold dark gray line in the grayed box indicates the median, while the black line in the box indicates the mean. '1st' and '2nd' indicate parent and offspring, respectively.



**Figure 5:** Distribution of the genetic repertoires estimated for each sex. The *Animal* data set is used. The numbers of females and males in the parent generation are separately estimated. (a) Females using offspring of both sexes. (b) Females using female offspring only. (c) Females using male offspring only. (d) Males using offspring of both sexes. (e) Males using female offspring only. (f) Males using male offspring only. The meanings of the elements in the figure are described in the captions of Figures 3 and 4. '1st' and '2nd' indicate parent and offspring, respectively.

## Discussion

First, we discuss our first purpose. We achieved our first purpose, which was to develop the MMR method that had remained a conceptual proposal into a practical method for diploid species and to propose the theoretical formulae to obtain the variance and confidence interval of the genetic repertoires. We expect that the method described in this study will be used in varied fields, such as biodiversity management, pest management, and human disease management, because this method retains the advantageous points of the previous MMR method (e.g., cost effectiveness and ease of use). The differences between this study and the clonal organisms of Murase and Fukita [5] was that we found a way to calculate mothers and fathers separately and unified the duplicated haplotypes (i.e., we dealt with the number of haplotype repertoires). Moreover, by using the theoretical formulae proposed in this paper, we were able to estimate the range of the genetic repertoires using only a one-time survey. This step is very meaningful for field research. Repeatedly sampling rare species may decrease their reproduction rate due to frequent human visits to their reproduction sites, even if non-invasive surveys (e.g., dung collection) are conducted. For example, parents of raptors such as goshawks often leave their nests on alert against approaching humans, and their eggs or nestlings are predated by crows [14]. Thus, estimating the range of the genetic repertoires without a requirement for repeated sampling would be an important perspective for wildlife conservation field research.

Next, we discuss our second purpose, which was to examine the applicability of the MMR method for diploid species assuming three different mating systems. In each mating system, the means and medians were very close to the true values of the genetic repertoires (Figure 3 and Web Table 1); these values could be estimated using the MMR method even if there was a difference in the mating systems. Comparing the number of individuals sampled in any mating system showed that the estimation was more accurate when we increased the number of parents (but not the number of offspring) when the number of total individual samples was kept constant. This is important information when we consider research plans. For example, when both parent and offspring samples are available and we are allowed to slightly increase the number of samples, we expect that the precision of the estimates will be improved if the number of parent individuals is increased without waiting for the result of the genetic analysis. Moreover, based on the results of the simulation experiments in this study, the precision of the estimates was decreased in the order of *Human*, *Animal*, and *Plant*. This result indicates that the mating system affects the precision of the genetic repertoires estimation, although it can be estimable regardless of the mating system. This decrease is likely due to the ease of identification of parents from offspring individuals. The precision of the genetic repertoires is dependent only on  $n$ ,  $M$ , and  $m$  in the derived formulae. That is, in every mating system the precision of the estimate takes the same value in the stage that the genetic repertoires of the samples are identified. What is important is the difference between the number of offspring individuals sampled and their genetic repertoires; these differences create a situation in which the precision of the effective population size estimate differs among the different mating systems when it is estimated with the same number of individuals sampled.

Finally, we discuss our third purpose. We achieved our third purpose, which was to propose a method that separately estimated the genetic repertoires of adult males and adult females of *Animal* species. We were able to estimate the genetic repertoires of males or females using only the method described in this study. Moreover, our purpose was also to discuss whether the MMR method was able to estimate the genetic repertoires of adult males and adult females if the available young individuals were only males or females, and the results showed that this purpose was generally estimable. We discuss the result of the estimation of the genetic repertoires of the adult males and females using the *Animal* data generated in this study. In both the adult females and adult males, the estimation precision was increased when all offspring individuals sampled were used compared to when the offspring individuals of either sex were used alone. This result did not contradict the law of large numbers (i.e., the larger number of samples yielded the better estimation precision). Moreover, this tendency increased in the case of the adult females compared to the adult males. This finding would represent a concern for populations with sex ratios skewed toward females in the parent generation. Although some biases were observed in the cases with small sample numbers, the MMR method for diploid species did not misestimate the order of magnitude and generally provided a good estimation of the genetic repertoires in practice.

In summary, the bias against the true value for our results was a maximum of 33.2 % (Web Table 2c, f,i; the mean value was 266.32) and the maximum range of the 95% CI was 700 (Web Table 2c, f, i), which was 3.50-fold higher than the true value and contained the true value of 200. Additionally, the bias and the 95% CI range were decreased when the sampling rate increased. Consequently, the mean and median approached the true value of 200 (Figures 3 and 4). Thus, although some bias was introduced by the low sampling rate, we were able to state that the MMR method proposed in this study was developed for diploid species and generally provided good estimates.

Next, we discuss the ecological factors that influence the estimation accuracy from the statistical point of view. By comparing the results of the three different mating systems, we found that the following two factors affected the estimation accuracy: (1) Fitness distribution of the parents and (2) Balance of the number of individuals sampled between the parents and offspring.

The fitness distributions of the parents used in the simulation of this study are shown in Web Figure 1. The deviance from the theoretical variance seemed to be larger when the fitness distribution was wider. This is caused by the higher possibility of obtaining offspring individuals that have the same genetic composition when the difference in fitness among the parent individuals is larger (i.e., the most dominant father and mother would have many offspring individuals with the same genetic contents). Our method counts the genetic repertoires of the offspring (i.e., we add one count to the genetic repertoire when many offspring individuals have the same genetic composition). Consequently, even if we sample the same number of offspring individuals, we will have different genetic repertoires when different mating systems are considered. The theoretical consideration and formulation are necessary in future work. However, the deviance seemed to be small (Figure 3, Web Table 1) and would not be a practical problem.

With regards to the balance of the parents and offspring, when we fix the total number of individuals sampled to be 120 (e.g., 60 parents/60 offspring), we achieve a slightly better estimate compared with 20/100 or 100/20 (Web Table 1). Ratios of 80/80 vs. 20/140 or 140/20 also exhibited the same tendency. The *Animal* and the *Plant* groups gave better estimates in the 40/40 setting, while the *Human* estimate was better at the 60/20 setting. Generally, the precision of the estimates tended to be high when the correct equilibrium of the number of parent and offspring individuals sampled was reached.

Therefore, we may reasonably conclude from the above results that the MMR method provides a good estimate of the genetic repertoires of typical diploid species. The estimates will be accurate in that at least the order of magnitude and the 95% confidential interval will contain the true value in the center, although its precision will vary depending on the mating systems or the sex ratios of the sampled individuals. Improved precision can be expected by sampling more from the parent generation in any mating system if the situation permits. The MMR method for diploid species would be useful in situations in which many species need to be managed together in national parks of tropical areas because the MMR method does not require highly sophisticated statistical models or much prior information for a species. Moreover, it can estimate the genetic repertoires with a one-time random sampling of the parent and offspring individuals. Furthermore, a decrease or increase in the genetic repertoires would be possible by increasing the number of random sample collections to two. Even if the national park lacks PCR equipment, the task can be delegated to an external institution such as a private company. Indeed, the cost of DNA extraction or sequencing is becoming more inexpensive every day. We believe that the management conducted by local people is quite essential for the conservation of the invaluable biodiversity of the global ecosystem, which contains a variety of local ecosystems.

## Supplementary Materials

Web Appendix 1, Web Figure 1, Web Tables 1-3 are available with this paper at the SM Journal of Biometrics & Biostatistics website.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 16K00575 and Grant-in-aid for research of Nagoya City University. We are grateful to Yasuo Ukai and the member of the Laboratory of Biometrics and Bioinformatics of University of Tokyo, for helpful comments.

## References

1. Barber PH, Ablan-Lagman MCA, Ambariyanto, Berlinck RG, Cahyani D, Crandall ED, Ravago-Gotanco R, et al. Advancing biodiversity research in developing countries: the need for changing paradigms. *Bulletin of marine science*. 2014; 90: 187-210.
2. Gerber BD, Karpanty SM, Kelly MJ. Evaluating the potential biases in carnivore capture–recapture studies associated with the use of lure and varying density estimation techniques using photographic-sampling data of the Malagasy civet. *Population Ecology*. 2012; 54: 43-54.
3. Tufto J, Lande R, Ringsby TH, Engen S, Saether BE, Walla TR, et al. Estimating Brownian motion dispersal rate, longevity and population density from spatially explicit mark-recapture data on tropical butterflies. *J Anim Ecol*. 2012; 81: 756-769.
4. Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, et al. Estimating animal population density using passive acoustics. *Biological reviews*. 2013; 88: 287-309.
5. Murase K, Fukita M. Biodiversity management under limiting conditions: estimating effective population size using the Molecular Mark and Recapture (MMR) method. *Sociobiology*. 2012; 59: 165-182.
6. Sodhi NS, Koh LP, Brook BW, Ng PK. Southeast Asian biodiversity: an impending disaster. *Trends in ecology and evolution*. 2004; 19: 654-660.
7. Seber GAF. *The estimation of animal abundance and related parameters*, 2nd Edition. London: Griffin. 1982.
8. Larson S, Jameson R, Etnier M, Fleming M, Bentzen P. Loss of genetic diversity in sea otters (*Enhydra lutris*) associated with the fur trade of the 18th and 19th centuries. *Molecular Ecology*. 2002; 11: 1899-1903.
9. Myers N, Mittermeier RA, Mittermeier CG, Da Fonseca GA, Kent J. Biodiversity hotspots for conservation priorities. *Nature*. 2000; 403: 853-858.
10. Frankham R, Briscoe DA, Ballou JD. *Introduction to conservation genetics*. Cambridge: Cambridge University Press. 2002.
11. R Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2013.
12. Murase K, Horie R, Saito M, Koganezawa M, Sato T, Kaji K. Integrating analyses of population genetics and space-time information for wildlife management: an empirical study on Japanese wild boar populations. *Mammal Study*. 2015; 40: 61-78.
13. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology*. 2007; 16: 1099-1106.
14. Squires JR, Kennedy PL. Northern Goshawk ecology: an assessment of current knowledge and information needs for conservation and management. *Studies in Avian Biology*. 2006; 31: 8-62.