Development of microsatellite markers for a soricid water shrew, *Chimarrogale platycephalus*, and their successful use for individual identification

Haruka Yamazaki¹, Tomohiro Sekiya¹, Shun Nagayama², Kei Hirasawa², Keita Tokura², Akio Sasaki³, Hidetaka Ichyanagi⁴ and Koji Tojo¹,⁵,⁶*

¹Division of Science, Interdisciplinary Graduate School of Science and Technology, Shinshu University, Matsumoto, Nagano 390-8621, Japan
²Aquamarine Inawashiro Kingfishers Aquarium, Inawashiro, Fukushima 969-3283, Japan
³Amphi LLC, Kimino, Wakayama 640-1221, Japan
⁴Faculty of Advanced Science and Technology, Kumamoto University, Kumamoto, Kumamoto 860-8555, Japan
⁵Department of Biology, Faculty of Science, Shinshu University, Matsumoto, Nagano 390-8621, Japan
⁶Institute of Mountain Science, Shinshu University, Matsumoto, Nagano 390-8621, Japan

(Received 14 March 2020, accepted 6 June 2020; J-STAGE Advance published date: 3 October 2020)

The soricid water shrew *Chimarrogale platycephalus* is a mammalian species endemic to the Japanese Islands. The animals inhabit the islands of Honshu and Kyushu, and are considered to be extinct in Shikoku. Information on this water shrew from Honshu and Kyushu is scarce, and *C. platycephalus* is registered on many local governments’ red lists as an endangered species. There are very few studies on their ethology, ecology or phylogenetics, due to difficulties related to the shrews being both nocturnal and aquatic: to study *C. platycephalus*, field research must be conducted in mountain streams at night. To overcome these challenges, we previously established a genetic analysis method using the feces of *C. platycephalus*, as a result of which the amount of phylogenetic and phylogeographic data has increased and our understanding of the species has improved. In this study, microsatellite markers were developed, and analyses using markers for 21 loci were performed. Moreover, to confirm the ability of these 21 microsatellite markers to differentiate individuals, all markers were tested using fecal and tissue specimens from 12 individuals reared separately in an aquarium. Using as few as 12 of these loci, individual differentiation with 100% accuracy should be achievable. The development of microsatellite markers in this study and the establishment of individual identification methods should greatly contribute to future ecological, ethological, population genetics and biogeographical research on *C. platycephalus*.

Key words: feces, freshwater shrew, individual identification, next-generation sequencing (NGS), microsatellite marker

INTRODUCTION

The freshwater soricid shrew, *Chimarrogale platycephalus* (Eulipotyphla, Soricidae), is endemic to Japan but has been recorded on only two of the main islands, Honshu and Kyushu. There are no records of *C. platycephalus* in Hokkaido or Shikoku. The island of Hokkaido, which is located north of the Tsugaru Strait, functions as a distribution boundary for many mammals (e.g., Masuda, 1999; Ohdachi, 1999), and is thus considered to be situated outside of this species’ distribution area. On the other hand, the absence of distribution records in Shikoku suggests that *C. platycephalus* has become extinct there (Abe, 2003).
Chimarrogale platycephalus inhabits mountain streams, preying on benthic organisms such as aquatic insects and freshwater crabs, and also freshwater mountain fish (e.g., salmonid Salvelinus leucomaenis and cyprinid Rhynchocypris lagowskii) (Supplementary Fig. S1) (Abe, 2011). This is an ecologically important position, making C. platycephalus one of the top predators in mountain stream ecosystems. In Honshu and Kyushu, it is ranked within one of the categories in local government red lists, and is a target for conservation (Yokohata et al., 2008; Ministry of Environment, 2014).

As this shrew is extremely vulnerable to stress when captured by trapping (Ichikawa et al., 2004; Fujimoto et al., 2011), investigations of its behavior and ecology have been hindered. We therefore tried to establish methods and techniques for genetic structure analyses utilizing the shrew’s feces, and obtained concise results (Sekiya et al., 2017). That study established a method that allows us to investigate water shrews without harming them. Another previous study suggested that water shrews included four genetic lineages based on analysis of their mitochondrial DNA cyt b region (Iwasa and Abe, 2006). Unfortunately, however, no nuclear genes have been analyzed for this water shrew. Therefore, there was no knowledge regarding crossing or gene flow between the four putative genetic lineages. Genetic analyses of their nuclear DNA regions should provide more important insights.

In addition, there are many points of interest with respect to these animals. How large is their range and foraging activity at the individual level? Are there differences in behavior between males and females? What kind of family structure do they maintain? Is there territorial behavior? Answering these ecological and ethological questions is very important because this water shrew is a stream-inhabiting endangered mammal and is one of the highest-level species within the stream ecosystem hierarchy. Even though several research attempts have been made (Abe, 2003; Kitagaki, 2016), ecological and ethological knowledge remains scarce.

Therefore, we attempted to develop microsatellite markers (also known as simple sequence repeat (SSR) markers) that can enable identification of C. platycephalus individuals, and obtained 21 good microsatellite markers. In this study, we conducted experimental verification of 12 previously identified individuals, which were subsequently reared separately. We then extracted and purified their total genomic DNA from their feces in a blind test, and performed fragment analysis using the microsatellite markers we had developed. We confirmed that it is possible to identify reared individuals by their feces even under masked conditions.

This was done in collaboration with an aquarium that successfully raised individually isolated water shrews for this study. For each animal, individual identification was confirmed to be of sufficient precision by a combination of its microsatellite markers.

These reared shrews included individuals collected from closely located sites within the same mountain stream. Although they were genetically closely related to each other, it was still possible to identify them as specific individuals.

Thereby, we demonstrated that this technique makes it possible to identify individuals with high confidence. In this article, we detail the development of a robust process for selection of suitable microsatellite markers and introduce a practical example for the use of these markers. In the analysis of fecal samples taken from the field, interesting cases were detected in which multiple individuals shared the same “droppings site”; this finding is consistent with an observation in a previous study (Funakoshi et al., 2016). This study will contribute to future research on the water shrew, and perhaps also to its conservation.

MATERIALS AND METHODS

Microsatellite marker development using next-generation sequencing (NGS)

Materials A tail tissue specimen dissected from a C. platycephalus shrew captured in a series of biodiversity surveys in Shizuoka Prefecture was used to extract high purity total genomic DNA. This specimen was stored frozen in 99% ethanol.

Total genomic DNA extraction from C. platycephalus tissue A DNeasy Blood & Tissue Kit (QIAGEN) was used to extract total DNA from 5 mg of the C. platycephalus tissue. Extraction and purification were carried out largely according to the manufacturer’s protocol. The tissue was placed in 180 μl of ATL buffer, and 20 μl of proteinase K was added. A DNA lysate was obtained after the cells disaggregated after stirring for about 24 h at 56 °C. A total of 100 μl of AL buffer and 200 μl of 100% ethanol were added to the lysate, and by passing the suspension through a special filter only the DNA was collected on the filter. Thereafter, 500 μl of AW 1 washing solution was added and a centrifuge was used to wash impurities through the filter; this step was repeated using 500 μl of AW 2 washing solution. Finally, the DNA adsorbed on the filter was eluted with 200 μl of AE buffer to obtain the DNA solution.

Using a NanoVue Plus spectrophotometer (GE Healthcare), the concentration of the DNA was determined and adjusted using AE buffer to about 20 ng/μl, and thereafter the DNA was stored at 4 °C.

NGS analysis and the search for microsatellite sites The extracted and purified total genomic DNA of C. platycephalus was analyzed using NGS as described in
Microsatellite markers for a water shrew

detail below, and microsatellite-containing sequences were extracted from the obtained sequence data. To prepare a sequence library, the total genomic DNA was treated with the restriction enzyme EcoRI and ligated with P1 adapters. After that, shear and end repair were performed at random, and P2 adapters were ligated to the new fragments. By performing PCR using both P1 and P2 adapter sequences as a primer set, only DNA fragments with P1 and P2 adapters added to both ends were amplified. Among the fragments amplified by this PCR, those having a size of 300–500 bp were selected. By quantitative real-time PCR (qPCR), libraries of the genomic DNA sequences of C. platycephalus with an appropriate insert size and with effective concentrations of 2 nM or more were obtained by paired-end sequencing using the HiSeq X Ten system (Illumina). Sequence assembly was carried out using a Velvet Optimizer (Zerbino and Birney, 2008), after excluding low-quality data by filtering. Primer design to detect simple sequence repeats from the obtained assembled contigs of C. platycephalus genomic DNA and amplification of all detected repeat sequences was conducted using the software Primer 3 (Rozen and Skaletsky, 2000).

Selection of primers Based on the obtained sequence data, suitable primers for our microsatellite analyses were selected according to the following four conditions (cf. Suzuki et al., 2020).

Number of repeat units First, they were selected based on the number of microsatellite repetitions. The more repetitions, the more polymorphisms are likely to occur (Weber, 1990; Ellegren, 2000; Petit et al., 2005; Kelkar et al., 2008), but allelic drop-out increases in repeating sequences (Kirov et al., 2000; Buchan et al., 2005), which may cause problems such as an increase in stuttering (Hoffman and Amos, 2005) that results in multiple peaks around the correct allele. According to van Asch et al. (2010), it is recommended to select target instances of 11–16 repeats for trinucleotides. Therefore, we selected instances of eight or more repetitions in this study.

Repeat unit size Next, selection was based on the number of bases present in the repeating unit of microsatellites. In general, there are one to six bases within a microsatellite repeat, and the shorter the allele size range, the easier it is for multiplex PCR to analyze multiple loci at one time. On the other hand, stutter peaks tend to occur in short repeat units (Chambers and MacAvoy, 2000), which increases the possibility of erroneous identification of alleles (Levinson and Gutman, 1987; Meldgaard and Morling, 1997). Since it is difficult to perform multiplex PCR when long microsatellite repeat units of such as five or six bases are repeated, microsatellites of two to four bases were selected in this study.

T_m value of the primer The specificity of PCR decreases when the T_m value is lowered, and the possibility of amplifying the DNA of an organism other than C. platycephalus, and a region other than the target region, is increased. Therefore, in this study, the T_m value was set to around 60 °C. Furthermore, in multiplex PCR, which analyzes multiple loci at the same time, since multiple primer sets are used simultaneously, the T_m values were limited to between 60.0 and 62.6 °C in order to regulate the amplification efficiency of all the primer sets.

Specificity of the C. platycephalus DNA These sequences selected as primer candidates were then searched using the software Primer BLAST (Ye et al., 2012), and checked for homology with other organisms. In addition to fish and aquatic insects that C. platycephalus consumes as food, primers with high similarity to human sequences were excluded to avoid the risk of contamination. Also, any sequences homologous to bovine DNA were excluded, because bovine serum albumin (BSA) was added in the process of fecal DNA analysis.

Amplification of DNA fragments by PCR PCR was finally carried out using the primers and PCR conditions selected as described above. Successful DNA amplification was confirmed by electrophoresis of the PCR product. The genomic DNA from a tissue specimen of C. platycephalus was used as the template DNA for PCR, and KOD FX Neo (TOYOBO) as the DNA polymerase. At that point, four fluorescent markers to detect the microsatellite sequences were added, i.e., the FAM, PET, NED and HEX sets (Applied Biosystems). For PCR, the thermal cycler PC 350 (Astec) was used. PCR conditions were as follows: pre-heating for 2 min at 94 °C, followed by 28 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s and extension at 68 °C for 30 s, with a final extension at 68 °C for 5 min. Electrophoresis was then carried out to confirm the presence or absence of amplification of the target DNA fragments.

Sequencing All of the primer sets in which DNA amplification was confirmed by electrophoresis were used and sequencing was carried out. Hi-Di Formamide (Applied Biosystems) and a 500 LIZ size standard (Applied Biosystems) were added to the PCR product and sequencing was conducted using an automated DNA sequencer (ABI 3130xl Genetic Analyzer; Perkin Elmer/Applied Biosystems). Using the software Genemapper ver. 4.1 (Applied Biosystems), the size of the DNA sequences was confirmed, alleles were determined, and genotyping was performed.

Establishment of an individual identification method using microsatellite markers

Materials The feces of 12 C. platycephalus individu-
als separately reared in an aquarium (Aquamarine Inawashiro Kingfishers Aquarium, Inawashiro, Fukushima Prefecture, Japan) were used as samples for gene analysis (Fig. 1, Supplementary Table S1). Feces were collected during the periods from December 7–20, 2017 and January 10 to February 18, 2019. Two to four feces samples per individual were collected on different days, and a total of 29 samples were obtained (Supplementary Fig. S2). Approximately half the volume of each feces sample (100–200 mg) was immersed in 3 ml of InhibitEX Buffer (QIAGEN) and stored at room temperature, and then mixed with a vortex mixer just before extracting its total genomic DNA. In addition, tissue samples from six of these 12 individuals that died under rearing conditions between 2017 and 2019 were collected and added to the analysis.

Total genomic DNA extraction, purification and genetic analyses from C. platycephalus feces A QIAamp Fast DNA Stool Mini Kit (QIAGEN) was used for total genomic DNA extraction and purification from fecal samples of C. platycephalus. We extracted and purified the DNA according to methods developed in our previous research (Sekiya et al., 2017), which improved on the protocol recommended by the product manufacturer.

Impurities were removed by centrifuging with InhibitEX Buffer and the feces-turbid suspension. Proteinase K (25 μl) and AL buffer (600 μl) were added to 600 μl of the resultant supernatant after centrifuging, and the mixture was incubated at 70 °C for 10 min to allow the cells to disaggregate and obtain a lysate containing DNA. Next, 600 μl of 100% ethanol was added to this lysate, and DNA was collected on a special filter by passing the lysate through the filter. Washing solutions were added and centrifuged successively to remove impurities. Finally, the DNA was purified by eluting it from the filter using 100 μl of ATE buffer. The concentration of total genomic DNA extracted in this way was measured using a NanoVue Plus spectrophotometer (GE Healthcare) and adjusted to about 20 ng/μl with ATE buffer, and the DNA was stored at 4 °C.

Gene analyses were performed on the feces-extracted samples in the same manner as performed after DNA extraction and purification from the tissue samples. However, since fecal samples contain substances that inhibit DNA analysis, PCR was carried out using MightyAmp DNA Polymerase ver. 3 (Takara), which is capable of resistance to nonspecific amplification and smear suppression against such PCR-inhibiting substances. In addition, BSA, which is a PCR stabilizer, was added. PCR conditions were as follows: pre-heating for 2 min at 98 °C, followed by 35 to 38 cycles of denaturation at 98 °C for 10 s, annealing at 60 °C for 30 s and extension at 68 °C for 1 min, with a final extension at 68 °C for 7 min. Each experiment was repeated at least twice to ensure the reliability of the results. If two samples failed to confirm PCR copy amplification of DNA fragments, the results were excluded from the analyses. If the two results differed or the peak was small, verification was carried out by further repeating the analysis, and if the results were not stable after multiple repetitions they were excluded. Using the software GeneMapper ver. 4.1 (Applied Biosystems), the sizes of the DNA sequences were confirmed, alleles were determined, and genotyping was performed. This analysis was conducted individually for each locus; multiplex PCR to analyze the plurality of loci at the same time was not conducted, because DNA yield extracted from feces is not high and the DNA is likely to be fragmented.

Analysis by GENECAP To statistically determine whether multiple fecal samples from the same individual could be confirmed to be from the same individual by this microsatellite analysis, the software GENECAP version 1.4 (Wilberg and Dreher, 2004) was used to compare the
detected mismatch alleles per sample and the probability of identity (PID) was calculated. The PID is the probability that two genotypes that are not identical happen to coincide, the sib-PID (sibling-PID) is the probability that two individuals have a blood relationship, and the HW-PID (Hardy–Weinberg-PID) is the probability that there is no blood relationship to each other. The actual PID is HW-PID > PID > sib-PID. Particularly in mammals, a sib-PID < 0.01 is considered to be desirable (Waits et al., 2001). In this study, GENECAP analysis was done using the default setting. “Match probability” was set to “0.05”, and the “Method to use” setting was selected by analyzing “sib”. By analyzing multiple fecal specimens and tissue specimens, we identified loci considered to be “allelic drop-out” and “allelic drop-in”. In addition, we also calculated the PID excluding both of these (i.e., “Case 2” in Table 4).

Microsatellite analysis of fecal specimens collected from droppings sites of wild C. platycephalus: We examined whether it is possible to analyze total genomic DNA not only from shrew feces collected in breeding environments, but also that from feces collected from wild populations. Between August 2017 and March 2018, at three mountain streams inhabited by water shrews, two or three pieces of feces were collected from each droppings site, and microsatellite analysis was performed using a total of eight fecal specimens. The three habitats targeted in these analyses were the Kamimura-gawa River in Iida City, Nagano Prefecture (3 pieces of feces), the Na-gawa River, Matsumoto City, Nagano Prefecture (3 pieces), and the Kouzura-gawa River, Mimata Town, Miyazaki Prefecture (2 pieces) (Supplementary Table S2).

Structure analysis of C. platycephalus using microsatellite markers: All 12 reared C. platycephalus individuals used in this study were collected from small areas within Fukushima Prefecture, Japan. Therefore, they were expected to be genetically closely related to each other. To verify this, we performed structure analysis using the 21 microsatellite markers newly developed in this study. The number of genetic clusters (K) of individuals was estimated by a model-based Bayesian assignment method using Structure 2.3.4 (Pritchard et al., 2000). In the structure analysis using the newly developed markers, fecal samples collected from geographically distant areas were also added for comparison (Nagano and Miyazaki prefectures; Supplementary Table S2). Using this method, independent assessments are made for each cluster without prior information about the origin of the population. The log-likelihood values [Mean LnP(K)] (Pritchard et al., 2000) were estimated for the given K between 1 and 10 with 10 independent runs set by 500,000 Markov chain Monte Carlo iterations, following a burn-in period of 100,000 iterations. The highest hierarchical cluster level was determined by comparing the log-likelihood estimates for different values of K using Structure Harvester v0.6.94 (Earl and vonHoldt, 2012).

RESULTS AND DISCUSSION

Development of microsatellite markers: As a result of paired-end sequencing by HiSeq X Ten, a total of 14,866,986,000 bp of contig (raw sequence) data were obtained, within which 14,726,967,300 bp (99.1%) were evaluated as being “clean data (correct rate: > 99%)”. Furthermore, following subsequent assembly, the sequences of 339,864,595 bp of the genomic DNA of C. platycephalus were obtained. Among these, microsatellite sequences were detected at 11,727 sites. Of these, we were able to design appropriate primers for 11,176 of the microsatellite sites (Supplementary Table S3).

Selection of microsatellite sites for actual analyses: From the sequences containing the 11,176 microsatellites for which appropriate primers were designed, selection of target sites for our microsatellite analyses was performed based on the following four criteria: 1) number of repeat units, 2) size of repeat units, 3) \( T_m \) value, and 4) specificity of the C. platycephalus sequence. As a result, 50 microsatellite markers were selected as candidates. PCR and subsequent electrophoresis were performed using these 50 primer sets to determine whether DNA amplification could be performed without problems. For sequences in which DNA amplification could be confirmed, DNA sequencing was performed. PCR was performed using the 50 sets of primers that were selected followed by sequencing and fragment analyses, and suitable DNA amplification was obtained at 21 loci (by 21 of the primer sets). Thus, these 21 DNA loci were considered to be useful as microsatellite markers (Table 1). In addition, as a result of the sequencing of these 21 microsatellite sites of C. platycephalus individuals, suitable genetic polymorphisms among the individuals were identified.

Individual identification using the 21 microsatellite markers: In this study, fecal specimens from individuals of C. platycephalus reared separately were used. Twenty-nine fecal specimens from 12 C. platycephalus individuals had the 21 microsatellite sites analyzed, and this was successful for 26 of the 29 specimens (Table 2). For the three unsuccessful samples, it was concluded that there were problems with the condition of the feces DNA sample. In general, DNA analysis from fecal specimens is difficult because the quantity of DNA contained in the stool is very small. Also, analysis of nuclear DNA is particularly difficult because the number of copies of nuclear DNA is smaller than that of mitochondrial DNA. Furthermore, in the feces of C. platycephalus excreted near a
H. YAMAZAKI et al.

Water area, hydrolysis of DNA is likely to occur, which is considered to make analysis more difficult. As a result, the successful analysis of 26 of the 29 specimens is of great significance. Even among the 26 fecal specimens that could be analyzed, some microsatellite sites could not be used. Nevertheless, although the number of loci analyzed from fecal samples varied from 15 to 21 loci, we obtained sufficiently reliable results for each of the sequenced gene loci analyzed. As described below, our analyses yielded data that allowed individual identification even from fecal specimens of *C. platycephalus*. Six tissue samples were added to the original 26, and analysis was performed with 32 samples in total.

Establishment of an individual identification method using GENECAP Microsatellite analysis from fecal specimens of different individuals identified different alleles depending on the locus (Table 3). However, no polymorphisms were found among the three markers SSR_Cp10, Cp17 and Cp18. When all the analyzed loci were compared and examined, the probability of correct individual identification became sib-PID < 0.01 when at least 12 loci were used. It is considered that the positive identification probability for mammals should be sib-PID < 0.01 (Waits et al., 2001). The results of these microsatellite-based analyses were 100% consistent with the actual excretion recorded for the examined individuals. Even when only eight microsatellite sites were used, a sib-PID value of 0.0124 was obtained (i.e., the MT individual of “Case 2” in Table 4), giving further confidence in positive identification of individuals being possible. Of the 21 microsatellite sites selected in this study, if 9–10 sites are analyzed, individual identification should be possible. Although microsatellite markers are extremely sensitive, when we use fecal specimens collected from the field, especially for specimens with a small number of analyzed loci or only homozygous alleles, we should consider allelic drop-out. The *C. platycephalus* individuals targeted for analysis in this study for individual identification were all collected from close regional populations, including some from the same mountain stream in Fukushima Prefecture (Fig. 2, Supplementary Table S1). As a result of the structure analysis of these fecal specimens combined with fecal specimens collected from other regions, it was clearly shown that the specimens targeted for our individual identification test were all from genetically closely related populations (Fig. 3). Even under these challenging conditions (i.e., even for individual identification experiments between genetically closely related individuals), being able to identify the individual with an accuracy of 100% demonstrates

Table 1. Characteristics of 21 selected microsatellite markers and their associated primers for identification of individual *C. platycephalus* water shrews

| Locus | Repeat unit | Primer sequence (5′→3′) | Tm (°C) | Primer sequence (3′→5′) | Tm (°C) | Tag |
|-------|------------|-------------------------|--------|-------------------------|--------|-----|
| SSR_Cp01 | (AAC)10 | AGGATGAAGCATGCTGTTGCTGCC | 60.1 | TCCTAATCCACAAATGGAATC | 60.1 | FAM |
| SSR_Cp02 | (ATA)10 | GCCTTGCTTTTTAGGATTTTG | 60.6 | AAACACATCTGTAATTTTCATCAA | 60.2 | FAM |
| SSR_Cp03 | (AAT)10 | ACTGAGCCAGCGGACTGCTTAAT | 60.2 | AGATAAGTTAATGCGGACTGCT | 60.4 | NED |
| SSR_Cp04 | (AAT)11 | AACAAGCGAAAGGATGACGTTT | 60.3 | ATGCAAGCCCTTGTAGCTCCTAG | 61.0 | PET |
| SSR_Cp05 | (AC)15 | CACCCCATCCCCCATATAATA | 61.5 | CTCAACCAACCACATCCTAGTTT | 60.6 | FAM |
| SSR_Cp06 | (GT)15 | TGGTTCTGGTTTGGGAATAATGAC | 60.5 | TGCGTTATAGTTTGGAGGCTGCT | 60.6 | NED |
| SSR_Cp07 | (AC)15 | GAAGTCTGCGGTGGCAACTC | 62.2 | GGTGCAATCTGCTGTATAACGTG | 62.6 | PET |
| SSR_Cp08 | (AAT)9 | ACCAGATGTAACCTAAAAGCAA | 60.0 | AGAAGTCGATCTCCTGACATT | 60.1 | HEX |
| SSR_Cp09 | (TTA)9 | GCTAAGCCTATCTTTCCCAGT | 60.5 | CACACTGCGATCTGAGCTAGCAGC | 60.4 | HEX |
| SSR_Cp10 | (ATG)9 | GAATATATCTGAGATGAGCTG | 60.1 | TAGCCTTTATGGCCAGTATAAA | 60.0 | NED |
| SSR_Cp11 | (AC)14 | CTCTGAGGACTGCGGCTGTTTG | 60.8 | TAATCTCTAGTGACAGTCTTGG | 60.6 | PET |
| SSR_Cp12 | (CAAA)8 | ATATTGCCATCTCTGAGCACC | 60.1 | TCCCCATAGCGGTATTATCTCATATT | 60.1 | HEX |
| SSR_Cp13 | (AGAC)8 | GGTGAACTCCAATAATGCGAATA | 60.1 | CCCCATACCTCTGAGCTCCTCT | 60.7 | NED |
| SSR_Cp14 | (AAT)8 | AGTTATTTCTCAGTGGCTGTTTGG | 60.9 | ATTTGCTCTGCTGTTCTATTCCAGA | 60.1 | HEX |
| SSR_Cp15 | (TATT)8 | TGCTAATCTCCTGATGTTTCAAAG | 60.5 | CAGAAACAGCAATAATCCCTGAG | 60.1 | FAM |
| SSR_Cp16 | (TATC)8 | CCAATAGAACTCATTTGACAGC | 60.0 | CTGAGTACGGATACACCCCTGAG | 60.0 | FAM |
| SSR_Cp17 | (CAA)8 | ATACATCAGATGCAATCCAAAG | 60.3 | AGACTGAGCTGCTGAGTACAAAG | 60.1 | PET |
| SSR_Cp18 | (AGC)8 | GGTCTATGATATACGTGG | 61.1 | CAGCTCTGCTGCTTACAGGCC | 61.6 | HEX |
| SSR_Cp19 | (GAA)8 | TGTTATAGAAGTGTGATGCGC | 60.4 | ACCTAAGGCACATGCTTCTCATCA | 60.1 | HEX |
| SSR_Cp20 | (GAA)8 | TGTTATAGAAGTGTGATGCGC | 60.3 | AGCAGGAACTTGGCTCCTTTCCC | 60.0 | PET |
| SSR_Cp21 | (TATT)8 | CTCAGTCCTGAGTGGCGACATCAA | 60.5 | TTGAGCTAGCTGTCCTCATCCAAG | 60.2 | NED |
the highly accurate results obtained using these microsatellite markers. Moreover, we were able to perform microsatellite analysis on geographically distant populations (i.e., Nagano Prefecture, Honshu and Miyazaki Prefecture, Kyushu; Supplementary Table S2). The results of structure analysis clearly showed genetic dif-

### Table 2. Number of microsatellite markers analyzed for each fecal specimen of *C. platycephalus*

| No. of individual (symbol) | No. of each sample | Specimen type used for genetic analysis | No. of marker loci analyzed |
|----------------------------|--------------------|----------------------------------------|-----------------------------|
| 01 (A)                    |                    | A1 F                                    | 17                          |
|                            |                    | A2 F                                    | *                          |
|                            |                    | A3 T                                    | 21                          |
| 02 (K)                    |                    | K1 F                                    | 19                          |
|                            |                    | K2 F                                    | 19                          |
|                            |                    | K3 T                                    | 21                          |
| 03 (M)                    |                    | M1 F                                    | 20                          |
|                            |                    | M2 F                                    | 20                          |
|                            |                    | M3 F                                    | *                          |
|                            |                    | M4 T                                    | 21                          |
| 04 (S)                    |                    | S1 F                                    | 16                          |
|                            |                    | S2 F                                    | 16                          |
|                            |                    | S3 F                                    | 20                          |
| 05 (U)                    |                    | U1 F                                    | 20                          |
|                            |                    | U2 F                                    | 21                          |
|                            |                    | U3 F                                    | *                          |
|                            |                    | U4 T                                    | 21                          |
| 06 (Y)                    |                    | Y1 F                                    | 20                          |
|                            |                    | Y2 F                                    | 20                          |
|                            |                    | Y3 T                                    | 21                          |
| 07 (Z)                    |                    | Z1 F                                    | 21                          |
|                            |                    | Z2 F                                    | 21                          |
|                            |                    | Z3 F                                    | 20                          |
|                            |                    | Z4 T                                    | 21                          |
| 08 (KR)                   |                    | KR1 F                                   | 21                          |
|                            |                    | KR2 F                                   | 21                          |
| 09 (GM)                   |                    | GM1 F                                   | 16                          |
|                            |                    | GM2 F                                   | 21                          |
| 10 (MT)                   |                    | MT1 F                                   | 15                          |
|                            |                    | MT2 F                                   | 18                          |
|                            |                    | MT3 F                                   | 21                          |
| 11 (MS)                   |                    | MS1 F                                   | 18                          |
|                            |                    | MS2 F                                   | 21                          |
| 12 (WR)                   |                    | WR1 F                                   | 20                          |
|                            |                    | WR2 F                                   | 20                          |

F, feces; T, tissue from the same individual after the animal’s death; *, sample could not be analyzed.

### Table 3. Analysis results for each microsatellite marker

| Locus     | No. of samples analyzed | No. of alleles detected | Marker size range (bp) |
|-----------|-------------------------|-------------------------|------------------------|
| SSR_Cp1   | 31                      | 2                       | 167–173                |
| SSR_Cp2   | 28                      | 5                       | 169–193                |
| SSR_Cp3   | 28                      | 5                       | 177–195                |
| SSR_Cp4   | 32                      | 3                       | 165–171                |
| SSR_Cp5   | 32                      | 7                       | 103–122                |
| SSR_Cp6   | 25                      | 9                       | 151–183                |
| SSR_Cp7   | 32                      | 6                       | 123–161                |
| SSR_Cp8   | 32                      | 9                       | 117–141                |
| SSR_Cp9   | 31                      | 5                       | 148–164                |
| SSR_Cp10  | 29                      | 1                       | 166                    |
| SSR_Cp11  | 23                      | 9                       | 166–198                |
| SSR_Cp12  | 32                      | 4                       | 150–166                |
| SSR_Cp13  | 31                      | 5                       | 98–114                 |
| SSR_Cp14  | 30                      | 7                       | 168–192                |
| SSR_Cp15  | 26                      | 8                       | 158–198                |
| SSR_Cp16  | 31                      | 7                       | 155–179                |
| SSR_Cp17  | 32                      | 1                       | 142                    |
| SSR_Cp18  | 32                      | 1                       | 143                    |
| SSR_Cp19  | 29                      | 2                       | 178–181                |
| SSR_Cp20  | 31                      | 2                       | 136–139                |
| SSR_Cp21  | 31                      | 8                       | 170–194                |

### Table 4. Validation of microsatellite marker-based individual identification

| Sample matches | Sib-PID | HW-PID |
|----------------|---------|--------|
| Case 1         |         |        |
| A1, A3         | 1.28E-05| 7.04E-16|
| K1, K2, K3     | 5.97E-06| 1.17E-16|
| M2, M4         | 2.36E-06| 4.18E-17|
| U1, U2, U4     | 3.85E-06| 2.54E-16|
| Y1, Y3         | 8.85E-07| 8.39E-19|
| Z1, Z2, Z3, Z4 | 1.86E-06| 6.49E-19|
| MS1, MS2       | 2.49E-05| 5.24E-14|
| WR1, WR2       | 5.18E-07| 1.41E-20|
| Case 2         |         |        |
| M1, M2, M4     | 2.07E-06| 2.78E-18|
| S1, S2, S3     | 1.45E-03| 3.53E-08|
| Y1, Y2, Y3     | 4.62E-06| 1.35E-16|
| GM1, GM2       | 1.64E-04| 1.85E-11|
| KR1, KR2       | 1.10E-06| 6.71E-18|
| MT1, MT2, MT3  | 1.24E-02| 1.94E-05|

Case 1: not excluding allelic drop-out/drop-in.
Case 2: excluding allelic drop-out/drop-in.
differentiation at the regional level (Fig. 3). Since these distant local populations are known to constitute different genetic lineages according to mitochondrial DNA cyt b region analysis (Iwasa and Abe, 2006), it was confirmed that the microsatellite markers developed in this study are also effective in genetically differentiating individuals within local populations. Thus, these markers are highly versatile and are not limited to intraspecific lineages or area groups. This study is the first to discuss the genetic structure of nuclear DNA in water shrews. In the future, it will be possible to estimate their evolutionary history in more detail by increasing the number of targeted populations.

In general, microsatellite analysis errors are likely to arise in cases where allelic drop-out has occurred (Kirov et al., 2000; Buchan et al., 2005). This is a phenomenon in which an apparent homozygous state occurs because the amplification of one gene fails despite the fact that the allele is in a heterozygous state. In this study, for the targeted fecal specimens of *C. platycephalus*, allelic drop-out detection was particularly pronounced. This tendency is considered to be caused by the extremely small amount of DNA contained in fecal specimens of *C. platycephalus*. However, in many specimens, by carrying out multiple repeated analyses, an allele peak was identified correctly when evaluating from homozygous
to heterozygous at the same site. From this, especially for microsatellite analysis of fecal specimens, it is clearly important to carry out repeated analyses to improve reliability by eliminating misidentified allele peaks. The samples for which we were only able to analyze a small number of loci tended to have allelic drop-out status, and in such cases, allelic drop-out with multiple loci was often observed. Regarding samples observed with only homozygosity, allelic drop-out may have been detected. On the other hand, the allelic drop-in phenomenon was observed only at one locus of one sample. In general, allelic drop-in has low reproducibility and often has irregular waveforms in sequencing, so it can be judged relatively easily. In microsatellite analysis, at least in fecal samples, multiplex PCR should be avoided.

When performing microsatellite analysis using fecal specimens, better results can be achieved by paying attention to the following points: 1) DNA yield is low and the quality is not good; 2) phenomena such as allelic drop-out and allelic drop-in are likely to occur; and 3) multiplex PCR should be avoided, based on the above two factors.

Specific examples of microsatellite analysis of C. platycephalus fecal samples from wild populations As a result of collecting and analyzing multiple C. platycephalus fecal samples from the same droppings sites in each of three populations, interesting new findings were obtained. In each sample, the number of microsatellite loci successfully analyzed was 5–20 (Supplementary Table S2). In the field, water shrews excrete feces on humid rocks, and if the feces collected are not be fresh the analysis efficiency might be reduced. However, at all three study sites where this microsatellite analysis was performed, multiple samples and loci were successfully analyzed. Surprisingly, it became clear that the feces of multiple individuals were excreted at the same droppings site. This characteristic phenomenon was confirmed in all three populations analyzed. Although social structure in eupotyphlan mammals is not yet well understood, it is known that in the few species studied, the animals are highly isolated in terms of behavior and have exclusive territories except during courtship and mating (Abe and Yokohata, 1998; Miura, 1998). The sharing of feces dropping sites by multiple individuals revealed in this study is a unique case among eupotyphlan mammals, and is an interesting phenomenon that should be examined more closely in the future.

Finally, the key finding of this study is that identification of individuals from C. platycephalus feces is possible without using invasive means such as capture. Therefore, we consider that the microsatellite markers developed in this study can contribute to the accumulation of fundamental ecological and ethological knowledge of C. platycephalus, and also to the conservation of this and other endangered species.

**AUTHOR CONTRIBUTIONS**

K. Tojo, H. Y. and T. S. conceived and designed the research. All the authors conducted sampling. S. N., K. H. and K. Tokura reared individuals and collected feces. H. Y. and T. S. conducted genetic analyses, and subsequent development of the microsatellite markers using NGS. H. Y. performed genotyping using microsatellite markers. All the authors discussed the composition of the paper, which was mainly written by K. Tojo and Y. H.

We thank Mr. T. Motoki (EAC Corp., Matsumoto), Dr. A. Ichikawa (BOGA Ltd., Matsumoto), Dr. T. Suzuki, Dr. M. Takenaka and Mr. K. Yano (Shinshu University) for their cooperation with the field research and collection of specimens. This study was supported by the Japan Society for the Promotion of Science KAKENHI (JP23657064 and JP16K14807 to K. Tojo), grants from the River Environment Fund (2018-5211-015 to K. Tojo) of River and Watershed Environment Management, and a research grant from the Institute of Mountain Science, Shinshu University (K. Tojo).

**REFERENCES**

Abe, H. (2003) Trapping, habitat, and activity of the Japanese water shrew, Chimarrogale platycephala. Mammalian Science 43, 51–65 (in Japanese).

Abe, H. (2011) Stomach contents of the Japanese water shrew, Chimarrogale platycephala. Mammalian Science 51, 311–313 (in Japanese).

Abe, H., and Yokohata, Y. (1998) The Natural History of Insectivora (Mammalia) in Japan. Hiba Society of Natural History, Hiroshima, Japan (in Japanese).

Buchan, J. C., Archie, E. A., van Horn, R. C., Moss, C. J., and Alberts, S. C. (2005) Locus effects and sources of error in noninvasive genotyping. Mol. Ecol. Notes 5, 680–683.

Chambers, G. K., and MacAvoy, E. S. (2000) Microsatellites: consensus and controversy. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 126, 455–476.

Earl, D. A., and vonHoldt, B. M. (2012) STRUCTURE HAP- VESTER: a website and program for visualizing STRUC- TURE output and implementing the Evanno method. Conserv. Genet. Resour. 4, 359–361.

Ellegren, H. (2000) Microsatellite mutations in the germine: implications for evolutionary inference. Trends Genet. 16, 551–558.

Fujimoto, R., Ando, M., and Ogawa, H. (2011) Efficient capture methods for the Japanese water shrew Chimarrogale platy- cephal. J. Agric. Sci., Tokyo Univ. Agric. 55, 290–296 (in Japanese).

Funakoshi, K., Inadome, T., and Okada, S. (2016) Distribution of the Japanese water shrew Chimarrogale platycephalus, in Kagoshima Prefecture, Japan. Nature of Kagoshima 42, 1–4 (in Japanese).

Hoffman, J. I., and Amos, W. (2005) Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Mol. Ecol. 14, 599–612.

Ichikawa, A., Nakamura, H., and Yoshida, T. (2004) Use of plastic containers as a new strategy to survey the distribution of Japanese water shrew Chimarrogale platycephala (Temminck). Jpn. J. Environ. Entomol. Zool. 15, 169–177 (in Japanese with English abstract).

Iwasa, M. A., and Abe, H. (2006) Colonization history of the Japanese water shrew Chimarrogale platycephala, in the Japanese Islands. Acta Theriol. 51, 29–38.

Kelkar, Y. D., Tyekucheva, S., Chiaromonte, F., and Makova, K.
D. (2008) The genome-wide determinants of human and chimpanzee microsatellite evolution. Genome Res. 18, 30–38.

Kirov, G., Williams, N., Sham, P., Craddock, N., and Owen, M. J. (2000) Pooled genotyping of microsatellite markers in parent-offspring trios. Genome Res. 10, 105–115.

Kitagaki, K. (2016) Development of a device for outdoor recording of hunting behavior of Japanese water shrew, Chimarrogale platycephala. Natural Environmental Science Research 29, 25–30 (in Japanese with English abstract).

Levinson, G., and Gutman, G. A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4, 203–221.

Masuda, R. (1999) Blakiston’s line and genetic characteristics of Carnivora in Japan. Mammalian Science 39, 351–358 (in Japanese).

Meldgaard, M., and Morling, N. (1997) Detection and quantitative characterization of artificial extra peaks following polymerase chain reaction amplification of 14 short tandem repeat systems used in forensic investigations. Electrophoresis 18, 1928–1935.

Ministry of Environment (2014) Red Data Book 2014—Threatened Wildlife of Japan—Vol 1, Mammalia. Gyouei Corporation, Tokyo, Japan (in Japanese).

Miura, S. (1998) Ethology: Behavior and Social Organization of the Mammals 4. University of Tokyo Press, Tokyo, Japan (in Japanese).

Ohdachi, S. (1999) Some problems about Blakiston’s line in insectivores : special reference to shrews. Mammalian Science 39, 329–336 (in Japanese).

Petit, R. J., Deguilloux, M.-P., Chat, J., Grivet, D., Garnier-Géré, P., and Vendramin, G. G. (2005) Standardizing for microsatellite length in comparisons of genetic diversity. Mol. Ecol. 14, 885–890.

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

Rozen, S., and Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. 132, 365–386.

Sekiya, T., Ichiyanagi, H., and Tojo, K. (2017) Establishing of genetic analyses methods of feces from the water shrew, Chimarrogale platycephalus (Erinaceidae, Eulipotyphla). JSM Biol. 2, 1010.

Suzuki, T., Hirao, A., Takenaka, M., Yano, K., and Tojo, K. (2020) Development of microsatellite markers for a giant water bug, Appasus japonicus, distributed in East Asia. Genes Genet. Syst., in press.

van Asch, B., Pinheiro, R., Pereira, R., Alves, C., Pereira, V., Pereira, F., Gusmão, L., and Amorim, A. (2010) A framework for the development of STR genotyping in domestic animal species: characterization and population study of 12 canine X-chromosome loci. Electrophoresis 31, 303–308.

Waits, L. P., Luikart, G., and Taberlet, P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. Mol. Ecol. 10, 249–256.

Weber, J. L. (1990) Informativeness of human (dC-dA)\(\cdot\)(dG-dT)\(_n\) polymorphisms. Genomics 7, 524–530.

Wilberg, M. J., and Dreher, B. P. (2004) GENECAP: a program for analysis of multilocus genotype data for non-invasive sampling and capture-recapture population estimation. Mol. Ecol. Notes 4, 783–785.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics 13, 134.

Yokohata, Y., Kawada, S., and Ichiyanagi, H. (2008) A report on the workshop “Recent advances in the ecology and conservation of the Japanese water shrew, Chimarrogale platycephala”. Mammalian Science 48, 175–176 (in Japanese).