

1 Host selection of haematophagous leeches (*Haemadipsa japonica*): implications
2 for iDNA studies

3

4 Goro Hanya¹, Kaori Morishima², Tomoya Koide³, Yosuke Otani⁴, Shun Hongo⁵,
5 Takeaki Honda¹, Hiroki Okamura¹, Yuma Higo⁶, Masamichi Hattori⁷, Yuki
6 Kondo^{8,9}, Yosuke Kurihara^{1,10}, Sakura Jin¹¹, Aji Otake^{11,12}, Izumi Shiroishi¹,
7 Tomomi Takakuwa¹³, Hiroki Yamamoto¹⁴, Hanami Suzuki⁶, Hisashi Kajimura⁶,
8 Takashi Hayakawa^{1,15,16}, Nami Suzuki-Hashido¹⁷, Takafumi Nakano¹⁸

9

10 ¹ Primate Research Institute, Kyoto University, Inuyama, Japan

11 ² United Graduate School of Agricultural Science, Tokyo University of Agriculture
12 and Technology, Utsunomiya, Japan

13 ³ Fukuroi, Japan

14 ⁴ Center for the Study of Co* Design, Osaka University, Toyonaka, Japan

15 ⁵ The Center for African Area Studies, Kyoto University, Kyoto, Japan

16 ⁶ Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya,
17 Japan

18 ⁷ Graduate School of Natural Science and Technology, Gifu University, Gifu,
19 Japan

20 ⁸ Graduate School of Education, Gifu University, Gifu, Japan

21 ⁹ Graduate School of Science, Osaka City University, Osaka, Japan

22 ¹⁰ Center for Education and Research in Field Sciences, Faculty of Agriculture,
23 Shizuoka University, Hamamatsu, Japan

24 ¹¹ Faculty of Agriculture, Iwate University, Morioka, Japan

25 ¹² Graduate School of Human and Environmental Studies, Kyoto University,

26 Kyoto, Japan

27 ¹³ Faculty of Agriculture, Tokyo University of Agriculture and Technology, Fuchu,

28 Japan

29 ¹⁴ Graduate School of Letters, Kyoto University, Kyoto, Japan

30 ¹⁵ Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan

31 ¹⁶ Japan Monkey Centre, Inuyama, Japan

32 ¹⁷ Chubu University Academy of Emerging Sciences, Kasugai, Japan

33 ¹⁸ Graduate School of Science, Kyoto University, Kyoto, Japan

34

35 Correspondence to: G. Hanya: Primate Research Institute, Kyoto University,

36 Kanrin 41-2, Inuyama, Aichi, 484-8506 Japan. E-mail: hanya.goro.5z@

37 kyoto-u.ac.jp, Tel.: +81-568-63-0542, Fax: +81-568-63-0564

38

39 Running title: Host of haematophagous leeches

40

41

Abstract The development of an efficient and cost-effective method for monitoring animal populations or biodiversity is urgently needed, and invertebrate-derived DNA (iDNA) may offer a promising tool for assessing the diversity and other ecological information of vertebrates. We studied the host species of a haematophagous leech (*Haemadipsa japonica*) in Yakushima by genetic barcoding and compared the results with those for mammal composition revealed by camera trapping. We analyzed 119 samples using two sets of primers by Sanger sequencing and one set of primer by next generation sequencing. The proportion of the samples that were successfully sequenced and identified to at least one species was 11.8-24.3%, depending on the three different methods. In all of these three methods, most of the samples were identified as sika deer (18/20, 6/15, 16/29) or human (2/20, 7/15, 21/29). The non-human mammal host species composition was significantly different from that estimated by camera trapping. Sika deer was the main host, which may be related with their high abundance, large body size and terrestriality. Ten samples included DNA derived from multiple species of vertebrates. This may be due to the contamination of human DNA, but we also found DNA from deer, Japanese macaque and a frog in the same samples, suggesting the mixture of the two meals in the gut of the leech. Using *Haemadipsa japonica*-derived iDNA would not be suitable to make an inventory of species, but it may be useful to collect genetic information on the targeted species, due to their high host selectivity.

Key words: barcoding, haematophagy, Japanese macaque, monitoring, sika deer

67

68 **Introduction**

69 The development of an efficient and cost-effective method for monitoring animal
70 populations or biodiversity is crucial for rapid decision-making in conservation
71 and management of wildlife. Traditional detection-based monitoring is
72 time-consuming and labor-intensive (Whitesides et al. 1988; Hanya et al. 2003),
73 and the precision of identifying species or age-sex classes may depend on the
74 ability of the observers. Methods based on animal traces, such as feces, nests,
75 or footprints, enable researchers to collect more data per unit effort than
76 detection-based methods (Delibes et al. 2012; Hanya et al. 2017; Kanamori et al.
77 2017). However, one can only obtain limited information on the individual animal
78 that left the trace compared with direct observation (Hanya et al. 2017). Camera
79 trapping is a powerful method that can reveal not only the species or age-sex
80 classes but also the behavior of the filmed animals (Pebsworth and LaFleur
81 2014), and it requires minimum effort in the field. However, the ability to identify
82 age-sex classes also depends on the observers' abilities with detection-based
83 surveys. In addition, camera trapping can be time-consuming with respect to the
84 duration of setting cameras in the study site as well as the time needed for
85 analysis.

86 Environmental DNA, or eDNA, is an emerging and promising
87 monitoring tool for the study of biodiversity (Taberlet et al. 2012; Ficetola et al.
88 2016). Organisms leave their tissues in their environment (water, soil, etc.), and
89 researchers can collect the organisms' deposited DNA by collecting samples of
90 the environment. This can be used to ascertain the presence of a particular
91 species by PCR amplification using species-specific primers (Fukumoto et al.

2015) or to make an inventory of species by sequencing barcoding regions with universal primers of the taxon using a next-generation sequencer (Ishige et al. 2017; Sato et al. 2017). eDNA is mostly used for aquatic ecosystems, as the substrate (water) can stir the deposited DNA through the environment.

As a monitoring tool for vertebrates, much attention has recently been paid to invertebrate-derived DNA, or iDNA. The idea is to sample the invertebrates that collect vertebrates' DNA. Target invertebrates are those animals that eat the feces, blood or carrion of vertebrates, such as leeches, mosquitoes, carrion flies and blowflies (Calvignac-Spencer et al. 2013a). Calvignac-Spencer *et al.* (2013a) argued that iDNA can potentially be used for studies on species biodiversity, genetic information of a targeted species (e.g. genotyping and sexing), and disease transmission. Various species of invertebrates collect vertebrate DNA in a variety of ways, and each species selects vertebrate species in its own way. Therefore, we need to examine which invertebrates are appropriate for use in obtaining certain kinds of ecological information on vertebrates.

Aside from studies using haematophagous invertebrates (Kocher et al. 2017) and flies (Hoffmann et al., 2016; Hoffmann et al. 2017) to study diseases across landscapes and using flies to study diseases within primate social groups (Gogarten et al., 2019), iDNA has mostly been used to study the presence of a particular species or biodiversity (Schnell et al. 2012; Calvignac-Spencer et al. 2013b; Schubert et al. 2015; Lee et al. 2016; Perez-Flores et al. 2016; Rodgers et al. 2017; Schnell et al. 2018; Tessler et al. 2018; Axtner et al. 2019; Drinkwater et al. 2019). For example, Schnell *et al.* (2012) analyzed iDNA from 25 individuals of haematophagous leeches in Viet Nam and detected six species of

mammals, including two recently described species of rare muntjac and rabbit. Calvignac-Spencer *et al.* (2013b) collected 201 individuals of carrion flies in Madagascar and Côte d'Ivoire and detected 4 and 22 species of vertebrates (mostly mammals), respectively. Recently, Axtner *et al.* (2019), Abrams *et al.* (2019) and Schnell *et al.* (2018) examined large numbers of leeches (1,532 and 3,427) to confirm if iDNA from leeches can be used to monitor biodiversity in a species-rich tropical rain forest ecosystem. They suggested experimental and statistical workflows under the assumption of imperfect detection by iDNA.

To confirm the effectiveness of iDNA, one needs to study the vertebrate community both by iDNA and conventional methods, such as transect census or camera trapping, and compare the results. Such studies have been conducted for flies (Lee *et al.* 2016; Rodgers *et al.* 2017) and leeches (Weiskopf *et al.* 2018). These studies reveal that iDNA collection from flies or leeches for a few weeks could detect more species of mammals than could a line transect census or camera trapping conducted concurrently. However, iDNA missed some of the very abundant species in the habitat. These results indicate that iDNA is an imperfect yet cost-effective method for making an inventory of mammals.

Most of the previous studies on iDNA were conducted in species-rich ecosystems, which is appropriate for testing its validity as a tool of monitoring biodiversity. One of the unexplored factors that need to be examined to confirm the validity of this method is the feeding preferences of the invertebrates, which would bias the kinds of species detected by this method (Schnell *et al.* 2015; Schnell *et al.* 2018). To clarify this preference, one needs to compare the relative abundance of each species revealed by iDNA and conventional methods, such

as camera trapping. These kinds of comparisons would be easier for species-poor ecosystems, since there would be proportionally more data for each species, from either conventional methods or iDNA.

Another relatively unexplored topic is on how different analysis method affects the results. Next generation sequencing (NGS) technology has been applied in the iDNA studies (Calvignac-Spencer et al. 2013b; Abrams et al. 2019; Axtner et al. 2019). One obvious advantage of NGS is that it can detect the DNA of multiple species collected from one individual of an invertebrate. NGS may enable us to examine how the DNA of different species are mixed in one sample or a pool of invertebrates. However, as a monitoring tool, it may sometimes be better to apply a simpler technique, such as Sanger sequencing, which is still often used in iDNA studies (Perez-Flores et al. 2016; Tessler et al. 2018; Weiskopf et al. 2018). Using the NGS technique, it is possible to handle a large number of individual invertebrates by pooling samples of many individuals before extraction of DNA. However, by doing so, we need to discard information from individuals. Therefore, the optimal method may be different depending on the cost, sample size, and research questions. The selection of different sets of primers could also affect the results, but this effect has rarely been studied systematically. For example, blocking primer for the human sequence has been applied in iDNA studies (Calvignac-Spencer et al. 2013b; Schnell et al. 2018) to minimize the inevitable effect of contamination of human DNA. However, this may affect the sensitivity to detect some animals (Schnell et al. 2018), such as non-human primates that are phylogenetically close to humans.

Haematophagous leeches used for iDNA studies are classified in the family Haemadipsidae, within the suborder Hirudiniformes distributed throughout

Asia, Australasia, and Madagascar (Phillips and Siddall 2009; Nakano 2017). Compared with flying animals, they have advantages as a source of iDNA because they offer more DNA, and show high spatial resolution due to their lower movement ability (Schnell et al. 2015).

The purpose of this study is to reveal the host species of terrestrial haematophagous leeches (*Haemadipsa japonica*) by genetic barcoding in the coniferous forest of Yakushima, southern Japan, which has a simplified set of mammal fauna. Two previous studies on the host of this species have been conducted (Sasaki et al. 2005; Sasaki and Tani 2008) in northern and central Japan, but it remains unanswered whether the leech positively selects the host species or they opportunistically feed upon the mammals they encounter. If the proportion of mammal species revealed to be fed upon by the leeches was similar to that of the relative abundance of each mammal species, then, we could conclude that the iDNA from leeches provides a promising monitoring tool for knowing the relative abundance of each mammal species. On the other hand, if the leeches are highly selective, they could be used as a tool for collecting the genetic information of a particular species. In order to understand how the choice of different sets of primers affects the results, we analyzed the mixture of DNA from the leech and tissues of mammals of known concentration for two sets of primers. We also analyzed samples by both Sanger sequencing and NGS to examine the possibility of 'mixing' DNA from two or more different species, either naturally or through contamination.

Methods

Study site

This study was conducted in the coniferous forest of Yakushima, which is an island in the southwestern part of Japan (30°N, 131°E). We selected a study site in the western part of Yakushima with an altitude of 700-1300 m above sea level. The dominant species include warm-temperate evergreen broad-leaved trees, such as *Quercus acuta*, *Q. salicina*, *Distylium racemosum*, and conifers, such as *Cryptomeria japonica*, *Abies firma*, and *Tsuga sieboldii*. *Haemadipsa japonica* is the only known haemadipsid leech in Yakushima (Tani and Ishikawa 2005).

Collection of leeches

We collected leeches during a census of Japanese macaques (*Macaca fuscata*) (Hanya et al. 2003; Hanya et al. 2005) conducted in August 2016, 2017 and 2018, with the aid of volunteers. We set a 7.5 km² census area and divided it into 30 500 m * 500 m quadrats. We positioned one point observer at an observation point in each quadrat. The point observer stayed at the observation point from ca 6:00-7:00 to 16:00 and collected leeches while wearing a mask and gloves. The collected leeches were kept in a 1.5 mL tube filled with 1 mL of 99% ethanol at room temperature. We did not collect leeches that touched the skin of the observers to avoid contamination by human DNA. This was repeated 5-7 days at each point. We collected leeches at 25 out of the 30 observation points.

Camera trapping

We set 30 camera traps (Trophy Cam HD ® Bushnell, Model119436) in the study site. Cameras were set to record videos of 30 s and placed 50 m directly north or south from the observation points. We deployed cameras during the period of July 15-17, 2014, and retrieved them from June 28 to July 5, 2015. We changed

batteries and SD cards in August and December 2014 and in March 2015.

Details of the camera settings are described in Hanya *et al.* (2018). Even though the period when we set cameras does not overlap with the period when we collected leeches, the abundance of sika deer and Japanese macaques did not change considerably between the two periods, according to the census of these animals (Hanya et al., unpublished).

Sample processing and extraction of DNA

In September 2018, among the 126 samples collected over the three years, 92 leeches were dissected, and their gut (crop and crop ceca) was removed and ground with a bead crusher (Bead Smash 12 BS-12 ® WAKENYAKU) for 15-120 s at 4200 rpm after evaporating the remaining ethanol. The 27 samples became completely dry due to the evaporation of the ethanol, so the entire body was ground. The remaining seven samples were discarded by mistake or because they were rotten. To avoid cross-contamination among samples, the scalpel and tweezers were wiped with ethanol and then flame-cleaned each time different samples were handled. The desk was also cleaned with ethanol each time different samples were analyzed. We extracted DNA using QIAamp DNA Mini Kit ® Qiagen following the manufacturer's protocol. Dissection and DNA extraction were conducted at a different workspace from the subsequent analysis.

Sanger sequencing

First, we tried the original set of primers (16Smam1 and 16Smam2) used in Calvignac-Spencer et al. (2013b); however, we always obtained multiple bands and thus it was impossible to find an optimal PCR condition to amplify only a

single region. Therefore, we used two different sets of primers for Sanger sequencing, both of which amplify the 16S rRNA regions of the mitochondrial genome. For the primers SCPH02500 (5'-TTACCAAAACATCACCTCT-3') and SCPL02981 (5'-ATCCAACATCGAGGTCGTAA-3') (Matsui et al. 2007) (fragment size ca 503 bp), we conducted PCR under the following conditions: an initial denaturation at 94°C for 10 min, followed by 45 cycles of 10 s at 94°C, 30 s at the annealing temperature of 51°C, 1 min at 72°C, and a final extension at 72°C for 10 min. Polymerase chain reaction (PCR) were performed in 15 µl reaction volumes containing 10–15 ng genomic DNA, 1 × PCR buffer, 0.2 mM of each dNTPs, 1.5 mM MgCl₂, 0.2 µM of each primer and 0.5 U of GoTaq polymerase (Promega, Madison, WI, USA). PCR products were electrophoretically separated on a 2.0% agarose gel and visualized using ethidium bromide in 1× TAE; all products exhibiting a single DNA fragment were selected for sequencing. These selected products were then purified using ExoSAP-IT (Affymetrix, Cleveland, OH, USA). Direct sequencing of both sequence directions was conducted using an ABI PRISM BigDye Terminator version 3.1 Cycle Sequencing Kit (Applied Biosystems) on an ABI 3500 Genetic Analyzer. We also used a newly designed set of the following primers covering the region used in a fly-derived iDNA study (Calvignac-Spencer et al. 2013b): 16Smam1_out-F (5'-ATAAGACGAGAAGACCCTATGGAG-3'), and 16Smam2_out-R (5'-TGAAGTCAGATCACGTAGGACTTT-3'), and blocking primer of human DNA 16Smam1_out_blkhum (5'-CCTATGGAGCTTTAATTTATTAATGCAAACAGT-spacerC3-3') (fragment size: ca 366 bp). The blocking primer was designed so that the amplification of only human DNA, but not other mammal DNA, was inhibited (Vestheim and

Jarman 2008). PCR amplifications were prepared in a 25- μ l volume containing 1 μ l of template DNA, 2.5 μ l of 10xEx Taq buffer, 0.5 μ l of each 2.5 mM dNTP, 0.5 μ l of forward and reverse 10 μ M primer, 10 μ l of the blocking primer, 0.125 μ l of ExTaq polymerase (Takara Bio Inc.), 0.1 μ l of T4 gene 32 protein (Nippon Gene Co., Ltd.), and 8.275 μ l of dH₂O. Negative controls were included in the PCR and confirmed that it does not amplify. After confirming successful amplification by electrophoresis and purification by isopropanol precipitation, the PCR products were directly sequenced with a BigDye Terminator v3.1 Cycle Sequencing Kit and a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

We examined how the concentration of mammal-derived DNA affected the success of the PCR amplification by mixing DNA extracted from the tissues of leeches and mammals at various ratios. We measured the concentration of DNA in the solution extracted from the entire body of the *H. japonica* (in which it has been confirmed that mammal DNA was not amplified for any of the primer sets), blood of a wild Japanese macaque captured in Yakushima, commercially sold meat of a wild sika deer in Hokkaido (*Cervus nippon yesoensis*), and a finger of a wild small Japanese field mouse (*Apodemus argenteus*) captured in Inabu Field, Graduate School of Bioagricultural Sciences, Nagoya University, central Japan. Then, we diluted the solutions to 5.0 ng/ μ l and mixed mammal and leech DNA at ratios of 1:0 (positive control), 1:1, 1:10, 1:100, 1:1,000, 1:10,000, and 1:100,000. Next, we conducted PCR with the above two primer sets under the same conditions. We visualized the PCR products by electrophoresis. This was repeated five times for each mammal for each primer set.

292 Massively-parallel sequencing using MiSeq
293 For the primer set SCPH02500/SCPL02981, we also performed massively
294 -parallel sequencing of samples using one of the next-generation sequencing
295 (NGS) machines, MiSeq. PCR condition was slightly changed from Sanger
296 sequencing: the first denaturation at 94 °C for 10 min, 45 thermal cycles of
297 denaturation at 94 °C for 10 sec, annealing at 51 °C for 30 sec, and extension at
298 72 °C for 1 min, followed by the final extension at 72 °C for 10 min. PCR
299 amplifications were prepared in a 25- μ l volume containing 1 μ l of template DNA,
300 2.5 μ l of 10xEx Taq buffer, 0.5 μ l of each 2.5 mM dNTP, 0.5 μ l of each 10 μ M
301 primer, 0.2 μ l of ExTaq polymerase (Takara Bio Inc.) and 18.3 μ l of dH₂O.
302 Negative controls were included in the PCR, and we confirmed that it was not
303 amplified. We chose this condition and this primer set because we could amplify
304 more samples under this condition than that used for Sanger sequencing and
305 also more than the primer set
306 16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum. However, another
307 shorter region was also amplified, since there were two bands for some samples
308 (shorter band: ca 200 bp). For the NGS, we prioritized having more samples
309 rather than having only one band. For the 53 samples that were successfully
310 amplified, we purified the PCR products with Agencourt AMPure XP (Beckman
311 Coulter). Following procedures of library preparation, sequencing on MiSeq, and
312 bioinformatics were based on Hayakawa et al. (2018). We conducted the second
313 PCR to attach the specific dual indices and sequencing adapters with KAPA HiFi
314 HS ReadyMix (Nippon Genetics Co Ltd.) and Nextera XT Index Kit (Illumina,
315 Inc.) in a total volume of 25 μ l mixture containing 2.5 μ l each of forward and
316 reverse primers, and 2.5 μ l of purified first PCR solution, 12.5 μ l of KAPA, and 5

µl of water with 5 thermal cycles. The DNA concentration of the purified 2nd PCR products was measured by Qubit dsDNA HS Assay Kit, and the DNA solutions for all samples were pooled so that the estimated molarity from each sample was equal. We sequenced this library with other libraries unrelated to this study and 5% PhiX spike-in on an Illumina MiSeq using MiSeq Regent Kit v3 (600 cycles). The read lengths from the MiSeq run were 301 bp (forward sequences), 8 bp (forward indices), 8 bp (reverse indices), and 301 bp (reverse sequences). The MiSeq base calls were converted to FASTQ files using `configureBclToFastq.pl` implemented by the `bcl2fastq` conversion software v1.8.4 (Illumina) (options: `no-eamss`, `mismatches 0`, and `use-bases-mask Y300n,Y8,Y8,Y300n`). The FASTQ files were demultiplexed using `clsplitleq` in `Claident v0.2.2016.04.07` (Tanabe and Toju (2013); <https://www.claident.org>). Read pairs with low-quality index sequences were discarded, that is, where the index sequences included nucleotide(s) with a quality score < 30 (option: `minqualtag = 30`). Since quality scores of nucleotides in the reverse sequences were low in this run possibly due to low diversity of the library and batch effect, we used only the forward sequences.

Sequence data analysis

For the sequences obtained by Sanger sequencing, we assigned the identity of the obtained sequences through an online BLAST search (Basic Local Assignment Search Tool; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) on November 5, 2018. We used the 'Nucleotide collection (nr/nt)' option for the database. We chose the species that showed the highest identity but rejected them when the maximum identity was less than 98%. All of the samples showed similarity to the

sequence of the species distributed in Yakushima.

For the NGS data, to identify the species for each read, we first made a local database for the animal species known to be distributed in Yakushima (Table 1). They were 13 species of mammals native to Yakushima, three introduced mammal species, human, and one native frog species (*Rana tagoi yakushimensis*). The frog was included because it was detected by Sanger sequencing. For the two species of field mice (*A. speciosus* and *A. argenteus*), the sequence was not available in the database, so we newly sequenced the region and added to the database using the tissue samples collected in Inabu. Then, each read was queried to the database using the command *blastn* (NCBI). We discarded reads with <98% maximum identity or when the length of the matched sequence was less than 200 bp. Under this criterion, among the 12,038,444 reads, no similarity with the vertebrate database was found for 11,573,147 reads. When the number of reads for the sample identified as the vertebrate species was less than 100, we did not include the reads for further analysis because sample misidentification is inevitable on the Miseq platform (Kircher et al. 2012) (Supplemental Material 1). None of the remaining 465,297 reads showed maximum identity >98% for two or more species.

Statistical analysis

Even though we found that mammals were not the only host for leeches, we had no data on the relative abundance of humans and frogs in the study site. Therefore, we compared the relative abundance of only non-human mammals detected by camera traps and genetic analysis of the leeches. We tested whether the proportion in the number of the host mammal species estimated by

genetic analysis differed significantly from the expected number of detections calculated based on the relative abundance of each mammal species, which was estimated by camera trapping. For the results of host identification, we examined the following three cases: primer set 16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum (16Smam, hereafter) (Sanger sequence), SCPH02500/SCPL02981 (SCPH, hereafter) (Sanger sequence), and SCPH02500/SCPL02981 (NGS). Raw numbers of filmed animals may bias the estimates of abundance for each species. First, movies filmed at short intervals might capture the same individuals repeatedly and thus over-estimate its abundance. To avoid such bias, we can discard additional movies taken within a short interval (e.g. 30 min). However, for group-living animals, such as Japanese macaques, repeated filming may reflect a larger group size, and thus the crude number of filming events may be a better index. Second, more than one individual can be filmed in a single movie. Third, even though the frequency of filming events may reflect the encounter rate for the leech with the animal species, it does not account for the biomass of the species. Considering all of these possibilities, we examined the following four parameters as indices of animal abundance: (1) crude number of filming events, (2) number of filming events after discarding other movies taken within 30 minutes, (3) average number of individuals filmed in one movie x (2), (4) body mass of the species x (3). Data on body size were taken from Agetsuma *et al.* (2011) for sika deer (*Cervus nippon yakushimae*), from Mori (1979) for Japanese macaques and Ohdachi et al. (2009) for other species. See Supplemental Material 3 for detail. The expected number of samples was less than five for most of the mammals. To make expected the numbers more than five, we combined all of

the non-deer mammals and compared this amalgam with deer, the most common species. We compared the observed (by iDNA) and expected (from camera trapping) number of samples between the sika deer and the other non-human mammals using the χ^2 test by R 3.2.2 (R_Core_Team 2015).

GH, KM, TK, NSH, and TH conducted genetic analysis, GH, YO SH, TH, HO, and YH set and maintained the camera traps, GH, YO, TH, MH, YK, YK, SJ, AO, IS, TT, and HY collected the leeches, HS and HK collected the mice specimen, GH and YK collected the macaque specimen, and GH, KM, and TN dissected the leeches.

Results

The experiment on mixing mammal and leech DNA indicated that the threshold DNA concentration below which the mammal DNA could not be detected was different between the primer sets and also by the mammal species (Fig. 1, Table 2). It was possible to detect mammals with a lower concentration of mammal DNA using the primer set 16Smam than using the primer set SCPH. Sika deer was detected at a lower concentration than Japanese macaques and mice in the primer set 16Smam, but opposite was true for the primer set SCPH.

For the primer set 16Smam, 20 out of 119 samples (16.8%) were successfully sequenced and identified to one species. For the primer set SCPH, it was 15 samples (12.6%). For the NGS using primer set SCPH, 29 samples (24.3%) included at least 100 reads for the sample that was successfully identified as vertebrate species. Forty-four samples showed positive results for at least one of the three methods. Among them, seven samples showed positive

results for all of the three methods, seven for two methods, and the remaining 30 samples showed positive results by only one method (Supplemental Materials 2). There was no case in which the results of different methods were contradictory to each other.

Most of the sequences were identified as sika deer or human. For the primer set 16Smam, 18 samples were identified as sika deer and two as human. For the primer set SCPH, seven were identified as human, six as sika deer, one as frog (*R. tagoi*), and one as raccoon dog (*Nyctereutes procyonoides*). For the NGS using primer set SCPH, 16 samples included reads identified as sika deer, 21 samples for human, two samples as Japanese macaque, and two samples for the frog. For ten samples, more than one species were identified per sample, even when infrequent reads were excluded. Deer and human were detected for eight samples, frog, human and deer were detected for one sample, and deer, macaque and frog were detected for one sample.

Camera traps functioned for 8658 camera-days in total. We obtained 2542 videos of sika deer, 636 of Japanese macaques, 653 of field mice, 22 of Japanese weasels, and 14 of feral dogs.

The proportion of the non-human mammal host species of the leeches was significantly different from the relative abundance calculated from the camera trap data (Table 2). The proportion of the host species (deer vs non-deer) was significantly different from the abundance of the mammals when we calculated the expected values based on the crude number of filming events, the number of filming events after discarding repeatedly filmed animals, or the number of filming events multiplied by the mean number of filmed individuals. This tendency was statistically significant for 16Smam (Sanger sequence) and

NGS but not for SCPH (Sanger sequence), due to the small number of samples identified as non-human mammals. When we took body mass into consideration, however, we could not reject the null hypothesis that leeches feed upon mammals based on their relative biomass.

Discussion

Host selection by *Haemadipsa japonica* in Yakushima

We found that the composition of mammal species revealed by camera trapping and the host species of the leech *Haemadipsa japonica* differed. Sika deer was more frequently identified as hosts of the leech than expected by their detection rates by the camera traps. Field mice were not detected and Japanese macaques were only infrequently detected from the leeches, even though their detection rates by camera trapping were not so low. However, when the large body size of a sika deer was considered, we could not reject the null hypothesis that the leeches feed on deer more often than on other species due to their highest relative biomass among the mammals here.

The fact that the sika deer was the major host of leeches in Yakushima may be related to the large abundance, body mass and terrestriality of the deer, and the effects of these factors were considered in the statistical analysis. The probability of encountering the leeches is expected to be the highest for the most abundant species. Respiratory CO₂ is one of the suggested cues that leeches use to detect mammals (Yoshiba 1996), and large-bodied animals are likely to emit more such cues. Terrestrial animals would offer more opportunity for leeches to attack than partially arboreal field mice or Japanese macaques (Hanya et al. 2007; Sakamoto et al. 2012). The (partial) arboreality of mice and

macaques was already reflected in their detection rates by the cameras, since the cameras were set near the ground. High grooming capability of Japanese macaques may also related with the low detection rate of this species in leeches. The main target of grooming by this species is the eggs of louse, each of which is only 0.6 mm in length (Tanaka and Takefushi 1993). Considering the high dexterity of their fingers, it would be easy for them to remove leeches from their skin. Actually, researchers in Yakushima observed two cases of Japanese macaques removing a leech from their own or their infant's body in the lowland of Yakushima (Yosuke Kurihara and Mari Nishikawa, unpublished data). The grooming done by deer functions to remove ticks (Heine et al. 2017), but because they use their mouth for grooming, their manipulation ability would be lower than that of macaques. The difference in hair length and the major posture taken on the ground may also affect the accessibility for leeches. Further research based on the behavioral observation of both leeches and host animals would be interesting to explore the reason on the host preference by the leeches.

Comparisons between the results of NGS and Sanger sequencing

Even though the main conclusion, i.e., that sika deer is the major host for this leech, was the same among the three different methods, the results were not always the same when examining this for each sample. Concerning the mismatches of the two Sanger sequencing methods, the main reason would be the blocking primer of human sequence and the lower sensitivity of SCPH than 16Smam, as shown by the mixing experiment. In addition, the primer set 16Smam may not be able to detect the frog sequence (#293).

There were 15 cases in which we detected the host by NGS but not by Sanger sequencing. When human DNA is dominant in the sample, amplification is blocked by the primer set 16Smam but not by the primer set SPCH in NGS (possibly in the cases of #210, 229, 230, 233, 235, 236, 237, 246, 250, 252, 268). Alternatively, when the DNA of multiple species of non-human vertebrates are mixed, Sanger sequencing is likely to be inhibited (#226, 271).

There were also instance of the opposite case (N=15) in which hosts were detected by Sanger sequencing but not by NGS. The reason for this remains unknown, but it might be related to the reads that could not be assigned to any vertebrate sequences. These would probably result from the shorter fragments obtained in the 1st PCR of NGS (see methods). Shorter fragments are easier to bind flow cells than longer ones, which might have prevented the detection of vertebrate-derived DNA in NGS. Unfortunately, in our system, we fail to find a primer set and optimal PCR conditions that can amplify only one region of a short fragment length. We can expect shorter fragments to lower the rate of sequencing error and improve amplification success (Axtner et al. 2019). In the future, thorough examinations of various primer sets will be necessary (Axtner et al. 2019).

Detecting multiple host species from one individual of leech

There were ten cases in which multiple vertebrate species were detected from a single individual leech. Since we have excluded results that can arise by misidentification on the Miseq, this could either be because of contamination or mixture of the blood of the last and the one previous meal. For the majority of cases (eight), human DNA was detected, so the possibility of contamination

during sampling, processing, extraction and amplification cannot be ruled out. However, since there were two cases in which DNA of deer, macaque and frog were simultaneously detected, we assume it is possible that the DNA of the last two or three meals were detected. These two samples (#226 and 271) were dissected by different persons at different desks, with the DNA extracted and amplified at different times, so it is unlikely that DNA was cross-contaminated from one sample to another. The frog (*R. t. yakushimensis*) is endemic to Yakushima, and they have never been handled in our laboratory, so it is also unlikely that the sample was contaminated by the frog DNA in the laboratory. Due to the small number of samples, it is difficult to conclude that DNA from both meals could have been detected. Further examination with many more samples, or a feeding experiment on multiple species by captive leeches, is required.

Potential bias

We need to examine whether any biases lowered the probability of detection of a particular species, both for camera trapping and genetic analysis.

Sollmann *et al.* (2013) summarized the problems of regarding detection rates of different species in camera trapping as an index of the relative abundance of each species. For example, partially arboreal species and species with smaller home ranges are less represented than expected from their relative abundance. Large species are more likely to be filmed more often because the sensor of the camera can detect larger animals more easily. In Yakushima, all of these factors would bias results toward higher detection of the largest and terrestrial sika deer. Therefore, the proportion of sika deer in our camera trapping may have been an overestimation of the relative abundance of deer.

However, we found that, in spite of this potential bias, the species composition of the hosts of leeches was even more biased toward deer. Therefore, our conclusion of the selective feeding upon deer by leeches remains unchanged, even considering the bias regarding camera trapping.

The sequence of a particular species may not be amplified with the sets of primers used in this study. We need to examine this possibility in particular for macaques, because we used blocking primer for human sequence was included in the primer set 16Smam. We designed the blocking primer so that the sequences were different between the macaque and human for the corresponding sequence with the blocking primer (Table 1). According to the mixing experiment of leech and mammal DNA, we confirmed that macaque DNA can be amplified even with blocking primer of human sequence. The experiment also showed that the DNA of Japanese macaques may likely be amplified less than that of sika deer using the primer set 16Smam but more than that of deer using the primer set SCPH (Fig. 1). However, underrepresentation of Japanese macaques was consistent between the two primers, so we believe the result was robust.

Hosts of *Haemadipsa* leeches in other regions and implications for iDNA studies

Combined with the two previous studies conducted in other parts of Japan, the main host of *Haemadipsa japonica* is Artiodactyla, but studies in the tropics indicate that the hosts of other *Haemadipsa* leeches are more diversified. In Akita, northern Japan, the hosts of *H. japonica* were identified as Japanese serow (*Capricornis crispus*) for 50 samples and large terrestrial birds (*Phasianus*

versicolor or *Syrnaticus soemmerringii*) for 5 samples (Sasaki et al. 2005). In Kanagawa, central Japan, hosts were identified as sika deer for 17 samples and wild boar (*Sus scrofa*) for 14 samples (Sasaki and Tani 2008). Artiodactyla is abundant in many forested landscapes in Japan (Saito and Koike 2013; Ikeda et al. 2016). They are all terrestrial and large-bodied and not as dexterous in the removal of external parasite as Japanese macaques (Ohdachi et al. 2009). However, in Vietnam, the host species found for the congeneric unidentified leech was more diversified, including Artiodactyla (pig, serow, cow and muntjac), Lagomorpha (rabbit) and Carnivora (badger) (Schnell et al. 2012). In a study in Bangladesh, 12 species of non-human mammals were detected by *Haemadipsa* leech-derived iDNA, including Artiodactyla (cow, pig, and muntjac), Rodentia (mouse), Carnivora (dog and civet), Primates (macaque), and Scandentia (tree shrew). To reveal the reasons for the differences among these studies, it would be necessary to conduct studies similar to the present one that examine mammal communities both by leech-derived iDNA and conventional monitoring (e.g. camera trapping) in more species-rich ecosystems.

Our finding that primates were not fed upon by *Haemadipsa* leeches may not be a general tendency. In Bangladesh, the proportion of rhesus macaques (*Macaca mulatta*) included in the hosts of leeches was more than that found by camera trapping (Weiskopf et al. 2018). It is difficult to examine the selectivity of the leeches in that study because the mammal diversity is much higher than in Yakushima. The feeding strategy of leeches can vary with the species even within the same genus (Gasiorek and Rozycka 2017), as well as with the behavior of the host animals, such as strata use (Reed and Bidner 2004). Much more study would be needed to generalize the findings on the

preferred and avoided host taxa of leeches.

The results found in Japan, including those of this study, suggest that the iDNA from *Haemadipsa japonica* is not appropriate for make an inventory of mammals, unless the sample size is very large. This species may be more suitable as a material to provide genetic information on a particular species of ungulates, such as sika deer or Japanese serow. In the future, we need to examine whether the iDNA could be used for genetic population monitoring, such as sex determination and individual identification. Data on the life history and behavior of leeches would also be needed to interpret the results, such as the time interval between feeding and sampling (Schnell et al. 2015). However, the studies in Vietnam, Bangladesh and Borneo indicated that the hosts were taxonomically variable (Schnell et al. 2012; Weiskopf et al. 2018; Axtner et al. 2019), so other *Haemadipsa* leeches might be suitable for making a species inventory.

In conclusion, *Haemadipsa japonica* showed strong bias for sika deer as a host over other mammals, compared with the relative abundance estimated by camera trapping. This leech species would be more suitable to collect genetic information for the deer, rather than making a species inventory in its habitat.

Acknowledgements

We would like to thank the members of 'Yakuzaru-Chosa-Tai' (Yakushima Macaque Research Group) in 2012-2018 who voluntarily joined in the census and participated in the sample collection and processing, maintenance of the camera traps and checking the videos. We would also like to thank our friends

and colleagues in Yakushima for their hospitality and help and Miho Hakukawa for her assistance in genetic analysis. The Sarugoya Committee and the Wildlife Research Center of Kyoto University provided us excellent facilities. Dr. Calvignac-Spencer and other anonymous reviewers' constructive comments were really useful in revising the earlier version of the manuscript. Ken Kondo and many other people contributed crowdfunding for our census team. This study was permitted by the Yakushima Forest Ecosystem Conservation Center and Yakushima National Park, Kagoshima Prefecture (capture of Japanese macaques) and Aichi Prefecture (live-trapping of field mice). The study was financed in part by MEXT Grant-in-Aid for Challenging Exploratory Research (25650145 and 15K14604), Scientific Research B (25291100), and Promotion of Joint International Research (Fostering Joint International Research) (15KK0256) to GH.

References

- Abrams, J. F., Hörig, L. A., Brozovic, R., Axtner, J., Crampton - Platt, A., Azlan, M., Wong, S. T., Sollmann, R., Yu, D. W. & Wilting, A. (2019). Shifting up a gear with iDNA: From mammal detection events to standardised surveys. *Journal of Applied Ecology*, 56, 1637–1648.
- Agetsuma, N., Agetsuma-Yanagihara, Y. & Takafumi, H. (2011). Food habits of Japanese deer in an evergreen forest: Litter-feeding deer. *Mammalian Biology*, 76, 201-207.
- Axtner, J., Crampton-Platt, A., Hörig, L. A., Azlan, M., Xu, C. C. Y., Yu, D. W. & Wilting, A. (2019). An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies. *GigaScience*, 8, 1-17.

- 642 Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P. & Schubert, G. (2013a).
643 An invertebrate stomach's view on vertebrate ecology: Certain invertebrates
644 could be used as "vertebrate samplers" and deliver DNA-based information on
645 many aspects of vertebrate ecology. *Bioessays*, 35, 1004-1013.
- 646 Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kuhl, H., Boesch, C., Kappeler, P.
647 M., Metzger, S., Schubert, G. & Leendertz, F. H. (2013b). Carrion fly-derived
648 DNA as a tool for comprehensive and cost-effective assessment of mammalian
649 biodiversity. *Molecular Ecology*, 22, 915-924.
- 650 Delibes, M., Calzada, J., Clavero, M., Fernandez, N., Gutierrez-Exposito, C.,
651 Revilla, E. & Roman, J. (2012). The Near Threatened Eurasian otter *Lutra lutra*
652 in Morocco: no sign of recovery. *Oryx*, 46, 249-252.
- 653 Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E.,
654 Gilbert, M. T. P. & Rossiter, S. J. (2019). Using metabarcoding to compare the
655 suitability of two blood-feeding leech species for sampling mammalian diversity
656 in North Borneo. *Molecular Ecology Resources*, 19, 105-117.
- 657 Ficetola, G. F., Taberlet, P. & Coissac, E. (2016). How to limit false positives in
658 environmental DNA and metabarcoding? *Molecular Ecology Resources*, 16,
659 604-607.
- 660 Fukumoto, S., Ushimaru, A. & Minamoto, T. (2015). A basin-scale application of
661 environmental DNA assessment for rare endemic species and closely related
662 exotic species in rivers: a case study of giant salamanders in Japan. *Journal of*
663 *Applied Ecology*, 52, 358-365.
- 664 Gasiorek, P. & Rozycka, H. (2017). Feeding strategies and competition between
665 terrestrial *Haemadipsa* leeches (Euhirudinea: Arhynchobdellida) in Danum
666 Valley rainforest (Borneo, Sabah). *Folia Parasitologica*, 64, 7.

667 Gogarten, J.F., D  x, A., Mubemba, B., Pl  h, K., Hoffmann, C., Mielke, A.,
668 M  ller-Tiburtius, J., Sachse, A., Wittig, R. M., Calvignac-Spencer, S. & Leendertz,
669 F. H. (2019) Tropical rainforest flies carrying pathogens form stable associations
670 with social nonhuman primates. *Molecular Ecology*, in press.

671 Hanya, G., Yoshihiro, S., Zamma, K., Kubo, R. & Takahata, Y. (2003). New
672 method to census primate groups: estimating group density of Japanese
673 macaques by point census. *American Journal of Primatology*, 60, 43-56.

674 Hanya, G., Zamma, K., Hayaishi, S., Yoshihiro, S., Tsuruya, Y., Sugaya, S.,
675 Kanaoka, M. M., Hayakawa, S. & Takahata, Y. (2005). Comparisons of food
676 availability and group density of Japanese macaques in primary, naturally
677 regenerated, and plantation forests. *American Journal of Primatology*, 66,
678 245-262.

679 Hanya, G., Kiyono, M. & Hayaishi, S. (2007). Behavioral thermoregulation of wild
680 Japanese macaques: Comparisons between two subpopulations. *American*
681 *Journal of Primatology*, 69, 802-815.

682 Hanya, G., Naito, S., Namioka, E., Ueda, Y., Sato, Y., Pastrana, J., He, T., Yan,
683 X., Saito, M., Costa, R., Allanic, M., Honda, T., Kurihara, Y., Yumoto, T. &
684 Hayakawa, T. (2017). Morphometric and genetic determination of age class and
685 sex for fecal pellets of sika deer (*Cervus nippon*). *Mammal Study*, 42, 239-246.

686 Hanya, G., Otani, Y., Hongo, S., Honda, T., Okamura, H. & Higo, Y. (2018).
687 Activity of wild Japanese macaques in Yakushima revealed by camera trapping:
688 Patterns with respect to season, daily period and rainfall. *Plos One*, 13,
689 e0190631.

690 Hayakawa, T., Sawada, A., Tanabe, A. S., Fukuda, S., Kishida, T., Kurihara, Y.,
691 Matsushima, K., Liu, J., Akomo-Okoue, E. F., Gravena, W., Kashima, M., Suzuki,

692 M., Kadowaki, K., Suzumura, T., Inoue, E., Sugiura, H., Hanya, G. & Agata, K.
693 (2018). Improving the standards for gut microbiome analysis of fecal samples:
694 insights from the field biology of Japanese macaques on Yakushima Island.
695 *Primates*, 59, 423-436.

696 Heine, K. B., DeVries, P. J. & Penz, C. M. (2017). Parasitism and grooming
697 behavior of a natural white-tailed deer population in Alabama. *Ethology Ecology*
698 & *Evolution*, 29, 292-303.

699 Hoffmann, C., Stockhausen, M., Merkel, K., Calvignac-Spencer, S., & Leendertz,
700 F.H. (2016). Assessing the feasibility of fly based surveillance of wildlife
701 infectious diseases. *Scientific Reports*, 6, 37952.

702 Hoffmann, C., Zimmermann, F. et al. (2017). Persistent anthrax as a major driver
703 of wildlife mortality in a tropical rainforest. *Nature* 548, 82.

704 Ikeda, T., Uchida, K., Matsuura, Y., Takahashi, H., Yoshida, T., Kaji, K. & Koizumi,
705 I. (2016). Seasonal and Diel Activity Patterns of Eight Sympatric Mammals in
706 Northern Japan Revealed by an Intensive Camera-Trap Survey. *Plos One*, 11,
707 16.

708 Ishige, T., Miya, M., Ushio, M., Sado, T., Ushioda, M., Maebashi, K., Yonechi, R.,
709 Lagan, P. & Matsubayashi, H. (2017). Tropical-forest mammals as detected by
710 environmental DNA at natural saltlicks in Borneo. *Biological Conservation*, 210,
711 281-285.

712 Kanamori, T., Kuze, N., Bernard, H., Malim, T. P. & Kohshima, S. (2017).
713 Fluctuations of population density in Bornean orangutans (*Pongo pygmaeus*
714 *morio*) related to fruit availability in the Danum Valley, Sabah, Malaysia: a
715 10-year record including two mast fruitings and three other peak fruitings.
716 *Primates*, 58, 225-235.

717 Kircher, M., Sawyer, S. & Meyer, M. (2012). Double indexing overcomes
718 inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids*
719 *Research*, 40, 8.

720 Kocher, A., de Thoisy, B., Catzeflis, F., Valiere, S., Banuls, A. L. & Muriénne, J.
721 (2017). iDNA screening: Disease vectors as vertebrate samplers. *Molecular*
722 *Ecology*, 26, 6478-6486.

723 Lee, P. S., Gan, H. M., Clements, G. R. & Wilson, J. J. (2016). Field calibration of
724 blowfly-derived DNA against traditional methods for assessing mammal diversity
725 in tropical forests. *Genome*, 59, 1008-1022.

726 Matsui, A., Rakotondraparany, F., Hasegawa, M. & Horai, S. (2007).
727 Determination of a complete lemur mitochondrial genome from feces. *Mammal*
728 *Study*, 32, 7-16.

729 Mori, A. (1979). Analysis of population changes by measurement of body weight
730 in the Koshima troop of Japanese monkeys. *Primates*, 20, 371-398.

731 Nakano, T. 2017. Diversity of Leeches from Japan: Recent Progress in
732 Macrophagous and Blood-Feeding Taxa. . In: (Motokawa, M. and Kajihara, H.
733 (eds)) *Species Diversity of Animals in Japan Diversity and Commonality in*
734 *Animals*, pp. Springer, Tokyo

735 Ohdachi, S., Ishibashi, Y., Iwasa, M. & Saitoh, T. 2009. *The Wild Mammals of*
736 *Japan*. Shoukadoh, Kyoto

737 Pebsworth, P. A. & LaFleur, M. (2014). Advancing Primate Research and
738 Conservation Through the Use of Camera Traps: Introduction to the Special
739 Issue. *International Journal of Primatology*, 35, 825-840.

740 Perez-Flores, J., Rueda-Calderon, H., Kvist, S., Siddall, M. E. &
741 Ocegüera-Figueroa, A. (2016). From the Worm in a Bottle of Mezcal: iDNA

- 742 Confirmation of a Leech Parasitizing the Antillean Manatee. Journal of
743 Parasitology, 102, 553-555.
- 744 Phillips, A. J. & Siddall, M. E. (2009). Poly-paraphyly of Hirudinidae: many
745 lineages of medicinal leeches. BMC Evolutionary Biology, 9, 11.
- 746 R_Core_Team (2015) R: A language and environment for statistical computing.
747 R Foundation for Statistical Computing Vienna, Austria, pp.
- 748 Reed, K. E. & Bidner, L. R. (2004). Primate communities: past, present, and
749 possible future. Yearbook of Physical Anthropology, 47, 2-39.
- 750 Rodgers, T. W., Xu, C. C. Y., Giacalone, J., Kapheim, K. M., Saltonstall, K.,
751 Vargas, M., Yu, D. W., Somervuo, P., McMillan, W. O. & Jansen, P. A. (2017).
752 Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys:
753 Evidence from a known tropical mammal community. Molecular Ecology
754 Resources, 17, e133-e145.
- 755 Saito, M. & Koike, F. (2013). Distribution of Wild Mammal Assemblages along an
756 Urban-Rural-Forest Landscape Gradient in Warm-Temperate East Asia. Plos
757 One, 8, 1-11.
- 758 Sakamoto, S. H., Suzuki, S. N., Degawa, Y., Koshimoto, C. & Suzuki, R. O.
759 (2012). Seasonal habitat partitioning between sympatric terrestrial and
760 semi-arboreal Japanese wood mice, *Apodemus speciosus* and *A. argenteus* in
761 spatially heterogeneous environment. Mammal Study, 37, 261-272.
- 762 Sasaki, O., Saito, H. & Harada, M. (2005). A survey of host animals of land leech
763 *Haemadipsa zeylanica* var. *japonica* (Whitman) in Akita Prefecture revealed by
764 PCR-SSCP analysis of 28S rRNA genes. Medical Entomology and Zoology, 56,
765 79-84.
- 766 Sasaki, O. & Tani, S. (2008). Sika deer and wild boar are possible host animals

767 of the land leech *Haemadipsa zeylanica* var. *japonica* (Whitman) in Kanagawa
768 Prefecture based on PCR-SSCP analysis of 28S rRNA. Medical Entomology
769 and Zoology, 59, 25-28.

770 Sato, H., Sogo, Y., Doi, H. & Yamanaka, H. (2017). Usefulness and limitations of
771 sample pooling for environmental DNA metabarcoding of freshwater fish
772 communities. Scientific Reports, 7, 12.

773 Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R. D.,
774 Willerslev, E., Bertelsen, M. F. & Gilbert, M. T. P. (2012). Screening mammal
775 biodiversity using DNA from leeches. Current Biology, 22, R262-R263.

776 Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Yu, D. W.,
777 Wilting, A. & Gilbert, M. T. P. (2015). iDNA from terrestrial haematophagous
778 leeches as a wildlife surveying and monitoring tool - prospects, pitfalls and
779 avenues to be developed. Frontiers in Zoology, 12, 14.

780 Schnell, I. B., Bohmann, K., Schultze, S. E., Richter, S. R., Murray, D. C.,
781 Sinding, M. H. S., Bass, D., Cadle, J. E., Campbell, M. J., Dolch, R., Edwards, D.
782 P., Gray, T. N. E., Hansen, T., Hoa, A. N. Q., Noer, C. L., Heise-Pavlov, S.,
783 Pedersen, A. F. S., Ramamonjisoa, J. C., Siddall, M. E., Tilker, A., Traeholt, C.,
784 Wilkinson, N., Woodcock, P., Yu, D. W., Bertelsen, M. F., Bunce, M. & Gilbert, M.
785 T. P. (2018). Debugging diversity - a pan-continental exploration of the potential
786 of terrestrial blood-feeding leeches as a vertebrate monitoring tool. Molecular
787 Ecology Resources, 18, 1282-1298.

788 Schubert, G., Stockhausen, M., Hoffmann, C., Merkel, K., Vigilant, L., Leendertz,
789 F. H. & Calvignac-Spencer, S. (2015). Targeted detection of mammalian species
790 using carrion fly-derived DNA. Molecular Ecology Resources, 15, 285-294.

791 Sollmann, R., Mohamed, A., Samejima, H. & Wilting, A. (2013). Risky business

- 792 or simple solution - Relative abundance indices from camera-trapping. Biological
793 Conservation, 159, 405-412.
- 794 Taberlet, P., Coissac, E., Hajibabaei, M. & Rieseberg, L. H. (2012).
795 Environmental DNA. Molecular Ecology, 21, 1789-1793.
- 796 Tanabe, A. S. & Toju, H. (2013). Two New Computational Methods for Universal
797 DNA Barcoding: A Benchmark Using Barcode Sequences of Bacteria, Archaea,
798 Animals, Fungi, and Land Plants. Plos One, 8, 11.
- 799 Tanaka, I. & Takefushi, H. (1993). Elimination of external parasites (lice) is the
800 primary function of grooming in free-ranging Japanese macaques.
801 Anthropological Science, 101, 187-193.
- 802 Tani, S. & Ishikawa, E. (2005). Biology and control of land leech, *Haemadipsa*
803 *zeylanica japonica* (in Japanese) Forest Pest, 54, 87-95.
- 804 Tessler, M., Weiskopf, S. R., Berniker, L., Hersch, R., McCarthy, K. P., Yu, D. W.
805 & Siddall, M. E. (2018). Bloodlines: mammals, leeches, and conservation in
806 southern Asia. Systematics and Biodiversity, 16, 488-496.
- 807 Vestheim, H. & Jarman, S. N. (2008). Blocking primers to enhance PCR
808 amplification of rare sequences in mixed samples - a case study on prey DNA in
809 Antarctic krill stomachs. Frontiers in Zoology, 5, 11.
- 810 Weiskopf, S. R., McCarthy, K. P., Tessler, M., Rahman, H. A., McCarthy, J. L.,
811 Hersch, R., Faisal, M. M. & Siddall, M. E. (2018). Using terrestrial
812 haematophagous leeches to enhance tropical biodiversity monitoring
813 programmes in Bangladesh. Journal of Applied Ecology, 55, 2071-2081.
- 814 Whitesides, G. H., Oates, J. F., Green, S. M. & Kluberanz, R. P. (1988).
815 Estimating primate densities from transects in a West African rain forest: a
816 comparison of techniques. Journal Of Animal Ecology, 57, 345-367.

817 Yoshiba, S. (1996). Medical-zoological characteristics of the land leech,
818 *Haemadipsa zeylanica japonica* Whitman, 1886, which explosively propagated
819 in the southern part of Boso Peninsula: Chiefly from the periodical fixed point
820 observations on its population for 10 years (in Japanese). Annual Report of the
821 Marine Ecosystems Research Center of Chiba University, 16, 34-53.

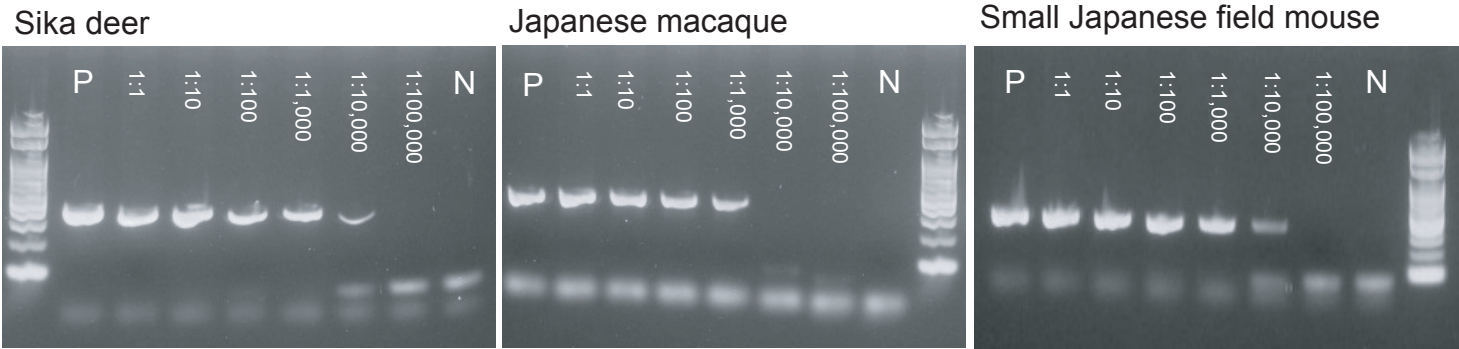
822

823

824 Legend of figures

825 Fig. 1. A sample of gel image of the electrophoresis of the PCR amplicons of the
826 mixture experiment of leech and mammal-derived DNA of various
827 concentrations. P is positive control (only mammal DNA added), N is negative
828 control (no DNA added), and the ratio indicate the ratio of the amount of DNA
829 of mammal and leech.

(a) 16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum



(b) SCPH02500/SCPL02981

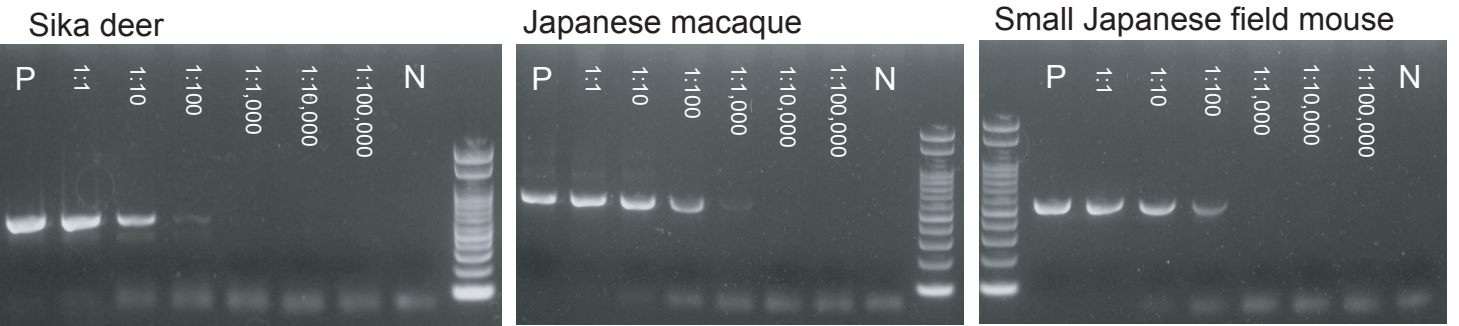


Table 1. Comparisons of the primer sequences and the corresponding regions of the native and introduced mammals and a frog in Yakushima

[illegible]

(Table1 continues)

[illegible]

(Table 1 continues)

[illegible]

Table 2. Results of the mixing experiment of mammal and leech DNA

(a) 16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum

English name	Scientific name	Ratio of the mammal (top) and leech (bottom) DNA						
		1	1	1	1	1	1	1
		0 (positive control)	1	10	100	1,000	10,000	100,000
Sika deer	<i>Cervus nippon</i>	5	5	5	5	5	5	4
Japanese macaque	<i>Macaca fuscata</i>	5	5	5	5	5	4	2
Small Japanese field mouse	<i>Apodemus argenteus</i>	5	5	5	5	5	5	0

(b) SCPH02500/SCPL02981

English name	Scientific name	Ratio of the mammal (top) and leech (bottom) DNA						
		1	1	1	1	1	1	1
		0 (positive control)	1	10	100	1,000	10,000	100,000
Sika deer	<i>Cervus nippon</i>	5	5	5	5	3	1	0
Japanese macaque	<i>Macaca fuscata</i>	5	5	5	5	5	3	0
Small Japanese field mouse	<i>Apodemus argenteus</i>	5	5	5	5	4	2	0

Each cell indicates the number of trials showing positive amplification out of five trials for each mammal species for each primer set.

Table 3. Results of the χ^2 test on the host selection by the leeches

(a) Number of leech samples detecting each non-human mammal species

	Sika deer	Japanese macaque	Mice	Weasel	Dog	Raccoon dog
	<i>Cervus nippon</i>	<i>Macaca fuscata</i>	<i>Apodemus</i> spp.	<i>Mustela itasi</i>	<i>Canis lupus</i>	<i>Nyctereutes procyonoides</i>
16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum, Sanger sequence	18	0	0	0	0	0
SCPH02500/SCPL02981, Sanger sequence	6	0	0	0	0	1
SCPH02500/SCPL02981, next generation	16	2	0	0	0	0

(b) Expected proportion of samples estimated from the data of camera trapping

	Sika deer	Japanese macaque	Mice	Weasel	Dog	Raccoon dog
	<i>Cervus nippon</i>	<i>Macaca fuscata</i>	<i>Apodemus</i> spp.	<i>Mustela itasi</i>	<i>Canis lupus</i>	<i>Nyctereutes procyonoides</i>
(1) Crude number of filmed events	0.66	0.16	0.17	0.01	0.00	0.00
(2) Repeatedly filmed movie excluded	0.57	0.19	0.22	0.01	0.01	0.00
(3) Number of individuals in a movie	0.55	0.23	0.21	0.01	0.01	0.00
(4) Body mass considered	0.89	0.11	0.00	0.00	0.01	0.00

(c) Results of the χ^2 test on the host selection by the leeches between deer and non-deer mammals

	16Smam1_out-F/16Smam2_out-R/16Smam1_out_blkhum, Sanger sequence		SCPH02500/SCPL02981, Sanger sequence		SCPH02500/SCPL02981, next generation sequence	
	χ^2	P	χ^2	P	χ^2	P
(1) Crude nuber of filmed events	9.46	0.002	1.24	0.265	4.34	0.037
(2) Repeatedly filmed movie excluded	13.46	0.000	2.32	0.128	7.37	0.007
(3) Number of individuals in a movie	14.73	0.000	2.62	0.105	8.35	0.004
(4) Body mass considered	2.12	0.145	0.73	0.394	0.00	1.000

Degree of freedom was 1 for all the analyses.