Journal of Applied Ecology

DR JESSE FRANK ABRAMS (Orcid ID : 0000-0003-0411-8519)

DR RAHEL SOLLMANN (Orcid ID : 0000-0002-1607-2039)

DR DOUGLAS W. YU (Orcid ID : 0000-0001-8551-5609)

Shifting up a gear with iDNA: from mammal detection events to standardized surveys

Jesse F. Abrams[1]*, Lisa Hörig[1], Robert Brozovic[1], Jan Axtner[1], Alex Crampton-Platt[1], Azlan Mohamed[1], Seth T. Wong[1], Rahel Sollmann[2], Douglas W. Yu[3,4], and Andreas Wilting[1]

*corresponding author (abrams@izw-berlin.de)

[1] Leibniz Institute for Zoo and Wildlife Research, Alfred-Kowalke-Straße 17, 10315, Berlin, Germany

[2] Department of Wildlife, Fish, and Conservation Biology, University of California Davis, Davis, 95616 CA, USA

[3] State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 32 Jiaochang East Road, Kunming, Yunnan 650223, China

[4] School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR47TJ, UK

Running headline: Shifting gears in iDNA studies

Abstract

1. Invertebrate-derived DNA (iDNA), in combination with high throughput sequencing, has been proposed as a cost-efficient and powerful tool to survey vertebrate species.

   Previous studies, however, have only provided evidence that vertebrates can be detected using iDNA, but have not taken the next step of placing these detection events within a statistical framework that allows for robust biodiversity assessments.

2. Here, we compare concurrent iDNA and camera-trap surveys. Leeches were repeatedly collected in close vicinity to 64 camera-trap stations in Sabah, Malaysian Borneo. We analyse iDNA-derived mammalian detection events in a modern occupancy model that accounts for imperfect detection and compare the results with those from occupancy models parameterized with camera-trap-derived detection events. We also combine leech-iDNA and camera-trap data in a single occupancy model.

3. We found consistent estimates of occupancy probabilities produced by our camera-trap and leech datasets. This indicates that the metabarcoding of leech-iDNA method provides reasonable estimates of occupancy and may be a suitable method for studying and monitoring mammal species in tropical rainforests. However, we also show that a more extensive collection of leeches would be needed to assess mammal biodiversity with a robustness similar to that of camera traps. As certain taxa were only detected in leeches, we see great potential in complementing camera-trap studies with the iDNA approach, as long as the collection of leeches follows a robust and standardized sampling scheme.

4. *Synthesis and applications.* Here, we describe an approach to analyse detection records of mammals derived from leech samples using an occupancy framework that accounts for leech-specific factors influencing the detection probability. We further combined camera-trap and leech data, which lead to increased confidence in occupancy estimates. Our approach is not restricted to the processing of leech samples, but can be used for the analysis of other invertebrate DNA (iDNA) and environmental DNA (eDNA) data. Our study is the first step to shift the application of iDNA studies from opportunistic ad-hoc collections to the systematic surveys required for long-term management of wildlife populations.

Abstract (In Malay)

1. DNA yang diperolehi dari invertebrat (iDNA), dengan *high throughput sequencing*, telah dicadangkan sebagai sesuatu kaedah survei spesies vertebrata. Sebelum ini, kajian hanya menunjukkan bahawa spesies vertebrata boleh dikesan dengan kaedah ini, tetapi belum digunakkan dalam rangka kerja statistik untuk tujuan penilaian biodiversiti.

2. Untuk kajian ini, kami membanding data daripada survei iDNA dan perangkap kamera dalam jangka masa yang sama. Kami mengumpul pacat secara berulang dari 64 stesen perangkap kamera di Sabah, Malaysia. Kami menggunakan model *occupancy* untuk menganalisis data iDNA dan perangkap kamera secara berasingan. Kemudian, kami mengabung data dari iDNA dan perangkap kamera dalam satu model.

3. Kami mendapati bahawa anggaran *occupancy* dari data perangkap kamera dan iDNA adalah konsisten. Ini menunjukkan bahawa kaedah iDNA boleh digunakan untuk mengkaji dan memantau hidupan liar di hutan hujan tropika. Walau bagaimanapun, pengumpulan pacat perlu dibuat secara kerap lagi untuk memperolehi data yang seragam dengan perangkap kamera. Dengan adanya segelintir spesies yang hanya dikesan dengan iDNA, kaedah ini boleh digunakkan untuk melengkapi data dari perangkap kamera.

4. *Sinthesis dan aplikasi.* Kajian ini merupakan satu cara untuk menganalisis data yang diperolehi dari pacat dengan menggunakan rangka kerja *occupancy*. Kami juga mengabungkan data iDNA dan perangkap kamera, yang meningkatkan ketepatan anggaran *occupancy*. Kaedah ini tidak terhad kepada pacat saja, tetapi boleh digunakan untuk analisis iDNA yang lain dan juga DNA yang diperolehi dari alam sekitar (eDNA). Kajian kami merupakan langkah pertama untuk mengalihkan focus kaedah iDNA dari kaedah survei *ad-hoc* ke kaedah survey sistematik yang boleh digunakan untuk pemantauan hidupan liar dalam janka masa panjang.

1. Introduction

To halt further biodiversity loss, parties of the Convention of Biological Diversity (CBD) have agreed to track and report progress towards the Aichi Biodiversity targets. To do so, rigorous monitoring of wildlife, using fast and efficient tools to assess the status of biodiversity, is necessary (Bush *et al.*, 2017). The overarching framework for such assessments already exists: gather detection events, analyse these using modern statistical models, and track population status over time (Bush *et al.*, 2017). However, detecting species, particularly in tropical rainforests, remains a challenge as species are often secretive and occur in remote areas.

Today, numerous methods are used to gather detections of mammals, all of which are time- and labour-intensive. Camera trapping has proven to be the most labour-efficient method (Roberts, 2011) and now plays an important role in wildlife management, allowing researchers to record a wide range of species in remote terrain over long time periods (Abrams *et al.*, 2018; Burton *et al.*, 2015; Trolliet *et al.*, 2014). With the increased use of camera-trapping surveys, the methods for processing and statistically analysing the data have also advanced (Burton *et al.*, 2015; MacKenzie *et al.*, 2006). However, the use of camera traps remains limited by difficult setup and high capital and maintenance costs.

An alternative or complement is environmental DNA (eDNA), which refers to the DNA that can be collected from a variety of environmental samples such as soil, water or faeces (Bohmann *et al.*, 2014; Thomsen & Willerslev, 2015; Iishige *et al.*, 2017). Recent methodological advances, namely amplicon-based high-throughput sequencing or 'metabarcoding', now also allow the reliable reading of such DNA sources (Abrego *et al.*, 2018). Invertebrate-derived DNA (iDNA) is an offshoot of eDNA, where terrestrial vertebrates can be detected via their DNA that was ingested by invertebrates (Schnell *et al.*, 2012, 2015; Calvignac-Spencer *et al.*, 2013a, 2013b; Tessler *et al.*, 2018; Weiskopf *et al.*, 2018). Sanguivorous species such as leeches (Schnell *et al.*, 2012, 2018), mosquitos (Kent and Norris, 2005), or ticks (Gariepy *et al.*, 2012) are commonly used, and invertebrates that feed on vertebrate faecal matter or carcasses, such as dung beetles, blow flies, and carrion flies have also been employed (Calvignac-Spencer *et al.*, 2013a, 2013b; Lee, Sing, and Wilson, 2015; Rodgers *et al.*, 2017; Schubert *et al.*, 2015; Somervuo *et al.*, 2017). Although these initial studies provide proof of principle that vertebrates can be detected using iDNA, they have been restricted to opportunistic collections of invertebrates and only been used to compile species lists. The sampling and the analyses have not been carried out in ways that allow statistically robust assessments of species or community trends and species population status.

Like all detection methods, e/iDNA is imperfect: the non-detection of a species by e/iDNA in a location does not prove the absence of that species in that location. Accounting for imperfect species detection requires a well-designed sampling scheme, combined with a statistical method known as occupancy modelling (MacKenzie *et al.*, 2002) to estimate the true spatial extent of species presence from detection events. Occupancy models estimate the probability of occupancy at any given site and have been widely used on camera-trap data (Burton *et al.*, 2015). Occupancy modelling (Fig. 1) uses detection/non-detection data collected from repeated sampling of the same locations. Under the assumption that the target vertebrate community does not change between sampling events, known as the 'closed-population' assumption, the repeated collection of species detection/non-detection data can be used to estimate the probability of species occurrence correcting for detection probability <1. Furthermore, both detection and occupancy probability can be modelled as functions of covariates.

Although occupancy modelling is widely used in wildlife studies, applications of occupancy models to eDNA studies have been limited. However, an increasing number of studies have recently applied occupancy modelling successfully to detection/non-detection data obtained through eDNA analyses (Pilliod *et al.*, 2013; Schmidt *et al.*, 2013; Ficetola *et al.*, 2014; Schmelzle & Kinziger, 2016; Hunter *et al.*, 2018; Wineland *et al.*, 2019). Based on these first eDNA occupancy studies Dorazio and Erickson (2018) recently developed an R package *ednaoccupancy*, which makes use of a multi-scale occupancy framework (Nichols *et al.* 2008), and accounts for the varying detection probabilities at different hierarchical levels (i.e. PCR replicates of one sample, sample replicates collected at the same station). In contrast to eDNA, occupancy modelling has only been proposed as a useful tool for iDNA data (Schnell *et al.*, 2015).

Here, we present an approach that makes this much-needed shift from proof-of-principle studies using iDNA to gather detection events to using iDNA as an input to statistically robust biodiversity assessment. We carried out standardized and repeated collections of terrestrial haematophagous leeches in Sabah, Malaysian Borneo, and we employed a molecular pipeline (Axtner *et al.*, 2019) that minimizes the risk of false-positive species detections. We compared leech-derived and camera-trap species richness estimates and investigated the detection bias of both methods towards smaller or larger species. We then analysed the leech-derived species detections within an occupancy-modelling framework that accounts for imperfect detection and compared the results to estimates of detection and occupancy probability from concurrently collected camera-trap data. Finally, we combined the leech iDNA and camera-trap data in a single occupancy model to evaluate the opportunities of combining e/iDNA data with conventionally collected biodiversity data.

2. Materials and Methods

2.1 Study area and data collection

We conducted camera-trap surveys and leech collection in the Deramakot Forest Reserve in Sabah, Malaysian Borneo (Fig. 2), which covers an area of approximately 55,500 hectares (ha) of mixed dipterocarp forest. A systematic camera-trapping survey was carried out from February - May 2015. Forty-six camera trap stations were deployed in a clustered design with each cluster consisting of a square of four camera-trap stations, spaced 500 m apart (Fig. 2). 16 clusters were established in a 4 x 4 formation with a distance of 1.5 km between cluster centres. Each station consisted of two Reconyx PC850 white-flash camera traps (a total of 128 camera traps) facing each other, operating 24 hours/day, and left in the forest for a minimum of 60 days (for details, see Abrams *et al.*, 2018). Two types of leeches, tiger and brown leeches, were collected concurrent to camera trapping. Although tiger leeches are described as *Haemadipsa picta* and brown leeches as *Haemadipsa zeylanica* (Fogden & Proctor, 1985), we refer to the types only because taxonomy within the genus *Haemadipsa* is currently not resolved (see Schnell *et al.*, 2015). Tiger leeches, the larger of the two types, live in small trees and bushes, while the smaller brown leeches occur mainly on the ground. This difference in behaviour may lead to different preferences in host species (Schnell *et al.*, 2015). Samples were taken within a 20 x 20 m sampling plot around the camera-trap stations. Sampling was repeated three times with approximately 30 days between sampling instances (at setup, check, and collection of camera traps). The leeches were immediately placed in RNAlater and stored at -20°C. All leeches of the same type (tiger or brown) from the same site and occasion were pooled and processed as one sample.

2.2 Laboratory procedures and taxonomic assignment

We implemented a novel e/iDNA workflow to extract raw species detections from leech iDNA (see Axtner *et al.* (2019) for a full description of our methods and Fig. S1 for an overview). In short, leech samples were first digested, and each sample was split into two extraction replicates, from which DNA extraction was carried out. Of each extraction replicate, we PCR-amplified three vertebrate mitochondrial markers, *12S*, *16S* and *cytochrome-b* twice. This resulted in 12 PCR replicates for each leech sample (2 technical replicates x 2 PCR replicates x 3 markers). We used a two-step, twin-tagging PCR strategy to produce double-labelled PCR libraries, which allowed us to implement a high throughput workflow and to minimize the risk of sample misidentification (Axtner *et al.*, 2019). PCR products were sequenced using Illumina MiSeq, and after sample demultiplexing and processing, we assigned each haplotype to a taxonomy using a curated reference database (Axtner *et al.*, 2019) and the PROTAX software (Somervuo *et al.*, 2016). We followed the *lax*

criteria by Axtner *et al.* (2019) for accepting an assignment, in which a species detection was accepted if it appeared in at least two PCR replicates.

2.3 Data analysis

To investigate if one method is more efficient (i.e. faster) at detecting species we computed mammal species accumulation curves in R using the function *specaccum* from the *vegan* package (Oksanen *et al.*, 2007) for the leech and camera-trap surveys. For instances where a species was not identifiable to the species level in one or both of our detection methods we included the species group (e.g., *Tragulus sp.*) in the accumulation curve. Earlier leech studies proposed that leeches might be better suited for the detection of smaller bodied species than are camera traps (Weiskopf *et al.*, 2018). We investigated whether detections in brown or tiger leeches were associated with host's body length by checking for correlation between the percentage of detections of a given species and the species body length in both tiger and brown datasets using Spearman's rank coefficient (rho). We also conducted the same analysis for camera-trap detections. Species head and body length data was taken from the PanTHERIA database (Jones *et al.*, 2009).

For the occupancy analysis, we used a subset of 11 species that were detected multiple times in both the leech and camera-trapping surveys. Detections for *Tragulus napu* and *Tragulus javanicus* were grouped together as *Tragulus sp.* and detections for *Hystrix brachyura* and *Hystrix crassispinis* were grouped together as *Hystrix sp.* for the occupancy analysis. We adopted the hierarchical formulation of occupancy models by Royle & Dorazio (2008) and used single-species, single-season models (MacKenzie *et al.*, 2002). We defined a total of six sampling occasions for the occupancy analysis of the leech data, based on the two types of leeches and the three sampling events.

We observed a difference in raw detection rates between tiger and brown leeches (see Fig. S2), so we included a categorical covariate on detection for the two leech types. The probability of detecting a mammal species in a leech DNA sample likely also depends on the number of leeches collected within this sample, as well as the number of other species detected in the sample, as species with low DNA amounts in the sample might not amplify in the PCR, if more abundant DNA of other species is present. Therefore, we included both the number of leeches per sample (we referred to this as *effort*[leech]), as well as the number of species detected per sample (referred as *detection*[leech]) as covariates on detection probability.

To compare the detection probabilities (p) and the occupancy probabilities (ψ) of the leech dataset (ψ_{leech}, p_{leech}) with the camera-trapping dataset (ψ_{ct}, p_{ct}), we prepared the camera-trap data for the same subset of species and stations for occupancy analysis using the R package *camtrapR* (Niedballa *et al.*, 2016) with an occasion length of seven days. We

also calculated the total number of independent detections. Detection events of the same species by the same camera trap were treated as independent detections if they occurred at least one hour apart from each other. Since not all occasions were sampled for the full seven days (e.g. due to camera-trap failure) we accounted for this ($effort^{CT}$) on the detection probability. Out of 128 camera traps, 8 cameras (6.3%) malfunctioned (failed to take photographs) for a certain time period and 2 cameras (1.6%) were disturbed by animals. We had no instances of complete station failure (the failure of both cameras placed at a station). Both effort and the number of species detected per sample are indexed by site (j) and occasion (k) (for a full model description see the Supporting Information).

We implemented the models in a Bayesian framework using JAGS (Plummer, 2003) accessed through R version 3.4.2 (R Core Team, 2017) using the package *jagsUI* (Kellner, 2015). We report results as posterior mean (in cases of skewed posterior distributions, the mode) and 95% Bayesian credible intervals (95BCI, the 2.5% and 97.5% percentiles of the posterior distribution). We evaluated the consistency of the leech-iDNA method when compared to the camera trap method, as well as the precision of the methods based on the 95BCI of detection and occupancy probabilities.

Last, we evaluated the value of combining camera-trapping and leech detections in a joint analysis. In the combined models, we included a categorical covariate for the detection method so that we could estimate the detection probabilities for the camera traps, brown leeches, and tiger leeches independently, but draw on all data sources to estimate occupancy.

Our survey design with the 500 - 1000 m spacing between camera-trap stations will most likely lead to spatial autocorrelation for some species, creating some bias in occupancy estimates (Legendre, 1993; Dormann, 2007). Further, both camera-trapping and leech collection constitute point based sampling in a continuous landscape. Application of occupancy models to such sampling schemes has been criticised for multiple reasons. Animal movement can make the target species temporarily unavailable for detection, so that occupancy in these situations is better interpreted as the probability that a site is used (MacKenzie and Royle, 2005). Regardless of interpretation, Efford and Dawson (2012) showed that occupancy estimates from continuous habitat can be confounded with animal density and home range size. Since we collected both leech and camera-trap data according to the same survey design, neither the spatial autocorrelation nor the difficulties introduced by sampling in continuous habitat should affect our within-species comparison of camera traps and iDNA.

3. Results

A total of 1,532 leeches (801 brown; 731 tiger) were collected during the survey, with the number of leeches sampled varying between stations and sampling occasions (Fig. 2). Leeches of both types were not detected at every site on every occasion. The number of leeches collected across all locations decreased from 762 to 576 to only 194 in the first, second, and third sampling occasions, respectively. Leeches of the same collecting occasion and type were pooled for a total number of 126 brown-leech and 116 tiger-leech samples (i.e. a collection tube). From these 242 samples, 196 mammal detection events were obtained after sequencing and bioinformatic processing. In 4,035 camera-trapping nights we obtained 2,733 independent mammal detection events. The camera-trap data had 31 mammals identifiable to the species level and 3 identifiable to genus, while the leech data had 22 mammal species, with some leech samples only identifiable to genus level for *Tragulus sp.*, *Hystrix sp.*, and *Macaca sp.* (Fig. 3a). All mammal species detected via leech iDNA are known to occur in the study area, and two of the species, binturong *Arctictis binturong* and domestic cat *Felis catus* were not detected by the camera traps. Additionally, using the leech iDNA, we were able to distinguish two species of mouse-deer, Java mouse-deer *Tragulus javanicus* and greater mouse-deer *Tragulus napu*, which was difficult and often impossible from the camera-trap photographs. In some cases we were not able to assign species-level taxonomies to sequences from the leech iDNA, due to an incomplete reference database or a low-confidence assignment from PROTAX (usually caused by low inter-specific sequence diversity). The most frequently identified species in the leech samples was sambar *Rusa unicolor* with 69 detections, followed by Bornean bearded pig *Sus barbatus* (22 detections, Table S1). The most identified species in the camera-trap samples were mouse-deer *Tragulus sp.* (669 detections), Bornean yellow muntjac *Muntiacus atherodes* (398), and Malay civet *Viverra tangalunga* (293). The species accumulation curve for the leech detection method showed a similar increasing trend to that of the camera-trap method (Fig. 3a), but did not reach its asymptote, contrary to the camera-based curves.

A comparison of the detections of mammal species in tiger and brown leeches revealed that the detection rate (detections per samples) in tiger leeches (1.15) was much higher than that of brown leeches (0.5). The smallest mammal we detected was the moonrat *Echinosorex gymnura* with a body mass of ~756 gr (Jones *et al.*, 2009), which was also detected in the camera-trap survey. We did, however, record a frog species (Black-spotted Sticky Frog *Kalophrynus pleurostigma*) in the leech dataset. The percentage of detections in the leech data was significantly positively correlated with the head and body length of the species (Fig. 3; Spearman's rho = 0.491; p-value = 0.0013). This pattern did not extend to the camera-trap data (rho = 0.115, p-value = 0.481, Fig. 3b).

Detection probability (Fig. 5a) varied between species and detection methods. Generally, the estimated detection probability for the camera-trap dataset had smaller CIs. In the leech dataset, estimates of detection probability of species with a low number of detections had high uncertainty (Fig. 5a). Detection probability in the tiger leeches was higher than in the brown leeches for all but two cases, Malay civet *Viverra tangalunga* and mouse-deer Bornean orangutan *Pongo pygmaeus* (Fig. 5a). The detection probability of brown leeches was lower than that of camera traps for all species. Tiger leeches had more success detecting certain species than did camera traps, with a higher estimated detection probability for 2 out of the 11 species. Although there are differences in the mean detection probabilities for brown and tiger leeches, the BCIs overlapped for all species, except for banteng *Bos javanicus*. Despite the higher number of detections in the camera-trap dataset, the occupancy models for the two survey methods generated mostly similar occupancy probabilities with overlapping confidence intervals (Fig. 5b) with the exception of mouse-deer *Tragulus sp.* and long-tailed porcupine *Trichys fasciculata* where the occupancy estimates are significantly lower for the leech dataset than for the camera trap dataset. The occupancy estimates from the camera-trap data had narrower BCI for all species except banteng *Bos javanicus*, long-tailed porcupine *Trichys fasciculata* and sambar *Rusa unicolor* (Figs. 5b and 6).

Occupancy estimates for the independently analysed leech and camera-trap datasets were similar to those from the combined analysis (Figs. 6, S3). When compared to the camera-trap only models, the combined dataset resulted in on average 12% narrower BCIs for occupancy probability estimates (Figs. 6, S3). The suite of 11 species had an average BCI of 0.508, 0.32, and 0.281 for the leech, camera-trap, and combined models, respectively.

4. Discussion

Overall, our results showed that 128 camera traps in 64 stations resulted in more detections and in higher species-richness estimates than did three leech collections at the same stations (Fig. 3a). Although requiring the same amount of field time (three visits), the sampling efforts were very different as the cameras were active for up to 64 nights. Because the initial rates of species accumulation for the camera-trap and leech data were quite similar (Fig. 3a, inset), we expect that increasing numbers of leech samples would lead to similar estimates of total species richness (i.e., similar asymptotes in species-accumulation curves) and decreased uncertainty in occupancy estimates.

Leech sampling success can vary due to season, weather, and microhabitat conditions. Thus, an understanding the factors that influence leech abundance is an important requirement to adequately design and execute a successful leech collection study

(Schnell *et al.*, 2015). Currently, little is known about the ecology of terrestrial leeches, but earlier studies indicated that they depend on humid conditions and survive the dry season by burying themselves in the soil (Nesemann and Sharma, 2001). In our study, the number of leeches collected varied largely between the stations and decreased with each sampling occasion (Fig. 2). The number of leeches collected was negatively correlated with the Keetch–Byram drought index (climate data from the Deramakot Forestry Office, see Fig. S4). This index increased during our study, which corresponds to a decrease in the amount of rainfall in the area between the sampling occasions, and suggests that leeches may be less available for sampling in dryer conditions, even within a moist tropical rainforest. It is also conceivable that each collection might have depleted a local leech population, and they may not be mobile enough to replenish the sampling quadrat before the next collection. The potential depletion of the local leech (or possibly other invertebrate) population is an aspect that must be carefully considered when employing an iDNA based survey method. Based on our data and the current knowledge about their ecology, leech collection may be most successful during the rainy season. Logistics, however, are likely more challenging in the rainy season. Further, camera trapping is often carried out in the dry season, which would prohibit concurrent camera trapping and leech sampling. The two methods, however, could also complement each other, allowing surveying the mammal community throughout the year. The resulting combined dataset could be analysed in a multi-season occupancy framework, investigating changes in mammal occurrence across seasons.

In contrast to Weiskopf *et al.* (2018), our data did not support the hypothesis that leeches are more suitable than camera traps to survey smaller mammal species. We detected no rats and mice, and although we detected one frog species in the leeches, our data showed a bias of leech detections towards larger bodied species, particularly ungulates. In fact, the proportion of ungulate detections in relation to detections of other mammals was higher in leeches than in camera traps (Fig. 4). Despite this apparent bias, we also detected several other mammal species using the leeches, such as the critically endangered Bornean orangutan, the critically endangered Sunda pangolin, and the vulnerable and primarily arboreal binturong (a species not recorded by the camera traps).

Our analysis did reveal a difference in mammal detection success between the two types of leeches. For 6 out of 11 species the posterior distributions of detection probability for brown and tiger leeches had an overlap of less than 50% (Table S2), where three of those species with an overlap of less than 10%. It is possible that the larger size of tiger leeches leads to a higher chance to amplify mammalian DNA (see also Weiskopf *et al.*, 2018). We also found a slightly stronger bias in tiger leeches towards larger bodied mammals (Fig. S2). This might be a result of their ecology, as tiger leeches live in small trees and bushes about 1 m off the ground (Lai, Nakano, & Chen, 2011), while brown

leeches mainly occur on the ground (Fogden and Proctor, 1985). Using an occupancy analysis where we estimated different detection probabilities for the two types of leeches allowed us to account for these differences. We are, however, aware that this might be challenging in other studies, due to the poorly known leech taxonomy (Schnell *et al.*, 2015) and the difficulties in distinguishing leech species in the field (Weiskopf *et al.*, 2018). Furthermore, as our results show the difference in detection probability for many target species was not significant and therefore separating leech species may not be a limiting factor. Overall, average detection probability of the two leech types was lower than that of the camera traps, but for two species, tiger leeches had higher detection probabilities than camera traps (Fig. 5). As very low detection probabilities can result in poor occupancy estimates (MacKenzie *et al.*, 2002), future leech studies might be restricted to species with higher detection probabilities, or will have to increase sampling effort.

Despite the above differences, we found consistent estimates of occupancy probability across our camera-trap and leech datasets with overlapping confidence intervals all but two species. For these two species, we had a significantly higher number of detections in camera traps than in leeches. The low detection rates in leeches possibly lead to problems in estimating occupancy probability. This indicates that the leech method provides reasonable estimates of occupancy and is thus a suitable method for studying and monitoring mammal species in tropical rainforests. The occupancy estimates from the camera-trap data, however, had narrower BCIs, which was likely a result of the overall larger dataset. The smallest BCIs and likely the most robust and accurate measure of occupancy were derived by combining the leech and camera-trap datasets, due to the increased total amount of data available. Similarly, other occupancy studies that have used multiple detection methods reported improved occupancy estimates (Iknayan *et al.*, 2014; Nichols *et al.*, 2008). The use of two detection methods allows researchers to collect more data, which will be especially beneficial in situations where detection probability is low and when rare species are the target, and in situations where both methods have complementary strengths. In camera-trapping studies, detection probability often depends on the way the camera traps are set up. Certain species such as larger felids are known to travel on roads or trails potentially increasing the probability of detection by camera-traps, whereas other species often avoid such features (Wearn *et al.*, 2013). While, according to our results, the leech iDNA method preferentially samples larger ungulates. A key point for practitioners trying to choose between these two methods is to consider which method is more efficient for sampling for their study area and species. Camera trapping might generate more detections, since cameras can be left in the field for months, but iDNA benefits from the ease of leech collection (i.e. with the help of local people). In particular, iDNA could be the only feasible method for gathering large numbers of vertebrate detections during one-off visits to remote

sites. One-off visits can yield data suitable for occupancy modelling if each sampling unit can be subdivided into several independent spatial replicates, which take the place of the typical temporal repeat visits (Guillera-Arroita, 2011). This, however, requires that if a species is present in one spatial replicate, it is present in all replicates - the spatial analogue to the closure assumption, which is likely to be violated when habitat is not homogeneous.

In this study, we were mainly interested in examining whether the analysis of leech-iDNA could be used as an efficient tool for biodiversity assessments. Our nested study design most likely resulted in spatial autocorrelation between observations of neighbouring stations for some species, particularly for wide ranging species such as the banteng or sun bears. This did not matter for our purpose of comparing the two types of data. Future studies that aim to apply the leech method for biodiversity assessment should take spatial autocorrelation into account.

We also acknowledge that leeches themselves move, and thus, our sampling locations might not coincide exactly with vertebrate presence locations. Although precise ecological information about the movement of terrestrial leeches is unavailable, leeches are believed to be mostly quiescent between feeding events (Schnell *et al.*, 2018). The consistency in occupancy modelling results from camera traps and leeches suggests that the potential movement of leeches does not cause significant bias in our study, but we note that we did not include any habitat covariates in our occupancy analysis. Leech movement could cause potential problems when exploring species-habitat relationships, particularly at a fine scale in heterogeneous habitats. This problem increases for flying invertebrates, such as mosquitos, tsetse flies or carrion flies (Calvignac-Spencer *et al.*, 2013a), that likely move over larger distances of up to a few kilometres (Verdonschot & Besse-Lototskaya, 2014), or for eDNA samples which are regularly transported away from their original deposition sites to their collection sites by currents or wind. A vertebrate obviously must be in front of the camera at the moment its photo is taken. With e/iDNA, the vertebrate does not have to be at the collection location, but could have been at any distance that the sample has moved since deposition, potentially leading to wrong inferences about species habitat preferences. Therefore, leeches, as well as ticks, present an advantage over other, more mobile invertebrates or samples taken from streams and rivers. Further real-world complications with e/iDNA that must be considered in future studies are, for example, different habitat and/or feeding preferences of invertebrates that affect species detection probabilities. Proper sampling design and statistical modelling must therefore be used to correct for the extra uncertainty introduced by the use of e/iDNA samples so that the efficiency benefit of e/iDNA can be properly exploited.

5. Conclusions

Our results are a promising indication that use of iDNA can help to overcome difficulties in surveying and monitoring terrestrial mammals in tropical rainforests. The species accumulation curves and occupancy estimates indicate that the leech iDNA method performed similarly to the well-established camera-trap approach. The iDNA approach, however, was limited by sample sizes, and it may be challenging to collect sufficient samples to achieve accuracy in estimates comparable to that from camera-trapping. This suggests that the collection of iDNA may be best used to supplement camera-trap surveys. Leeches helped to detect a few species that were not detected during the camera-trap survey; allowed us to distinguish between similar species that could not be differentiated in photographs; and combining leech and camera-trap data in a single model improved estimates of occupancy estimates. The main challenge for upcoming studies relying solely on iDNA appears to be the collection of a sufficient number of samples, which may be helped by the use of invertebrate traps. In conclusion, iDNA presents a promising approach for systematically surveying wildlife populations, but future studies need to consider (a) potential sample size limitations and (b) idiosyncrasies of the detection data, such as the potential mismatch of detection and presence location, or factors influencing detection probability. In combining systematic leech surveys with occupancy modelling while accounting for differences in detection due to leech type, numbers, and detections of other species, we hope to highlight this approach to wildlife ecologists as a new sampling tool, and to molecular ecologists as a robust analytical framework for e/iDNA.

Data accessibility

Data and scripts supporting this manuscript are available from Dryad Digital Repository https://doi.org/10.5061/dryad.4558p38 (Abrams *et al.*, 2019). We have anonymized both species and station names as it contains sensitive information about endangered and

threatened species. For requests for the non-anonymized data for academic purposes please contact abrams@izw-berlin.de or wilting@izw-berlin.de.

Conflict of interest declarations

DY is a co-founder of NatureMetrics, which is a UK-based private company that provides metabarcoding services. ACP is employed by NatureMetrics. The other authors declare no conflicts of interests.

Authors' contributions

AW and RS designed the study; AM and STW performed the sampling; JFA, LH, JA, ACP, RB, and DWY analysed data; JFA and AW lead writing the manuscript. All authors contributed to drafts and gave final approval for publication.

References

Abrams, J.F., Axtner, J., Bhagwat, T., Mohamed, A., Nguyen, A., Niedballa. J., Sollmann, R., Tilker, A.R., & Wilting, A. (2018). Studying terrestrial mammals in tropical rainforests. A user guide for camera-trapping and environmental DNA. Leibniz-IZW, Berlin, Germany.

Abrams, J.F., Hörig, L., Brozovic, R., Axtner, J., Crampton-Platt, A., Mohamed, A., Wong, S.T., Sollmann, R., Yu, D.W., & Wilting, A. (2019). Data from: Shifting up a gear with iDNA: from mammal detection events to standardized surveys. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.4558p38

Abrego, N., Norros, V., Halme, P., Somervuo, P., Ali-Kovero, H., & Ovaskainen, O. (2018). Give me a sample of air and I will tell which species are found from your region: Molecular identification of fungi from airborne spore samples. *Molecular Ecology Resources, 18*(3), 511-524.

Axtner, J., Crampton-Platt, A., Hörig, L. A., Xu, C. C., Douglas, W. Y., & Wilting, A. (2019). An efficient and improved laboratory workflow and tetrapod database for larger scale eDNA studies. *GigaScience*. doi: 10.1093/gigascience/giz029

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., and de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358-367.

Burton, A.C., Neilson, E., Moreira, D., Ladle, A., Steenweg, R., Fisher, J.T., Bayne, E., and Boutin, S. (2015). Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes. *Journal of Applied Ecology*, *52*(3), 675-685.

Bush, A., Sollmann, R., Wilting, A., Bohmann, K., Cole, B., Balzter, H., Martius, C., Zlinszky, A., Calvignac-Spencer, S., Cobbold, C.A. & Dawson, T.P. (2017). Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution*, *1*(7), 0176.

Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappeler, P.M., Metzger, S., Schubert, G. & Leendertz, F.H. (2013a). Carrion fly derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology, 22*(4), 915-924.

Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. P., & Schubert, G. (2013b). An invertebrate stomach's view on vertebrate ecology: Certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. *Bioessays, 35*(11), 1004-1013.

Dorazio, R. M., & Erickson, R. A. (2018). ednaoccupancy: An r package for multiscale occupancy modelling of environmental DNA data. Molecular ecology resources, 18(2), 368-380.

Dormann, C. F. (2007). Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography, 16*(2), 129-138.

Efford, M. G., & Dawson, D. K. (2012). Occupancy in continuous habitat. *Ecosphere, 3*(4), 1-15.

Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguet Covex, C., De Barba, M., Gielly, L., Lopes, C. M., Boyer, F., Pompanon, F., Rayé, G., & Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources, 15*(3), 543-556.

Fogden, S.C.L. Proctor, J. (1985). Notes on the Feeding of Land Leeches (Haemadipsa Zeylanica Moore and H. picta Moore) in Gunung Mulu National Park, Sarawak. *Biotropica 17*, 172-174.

Gariepy, T.D., Lindsay, R., Ogden, N., & Gregory, T.R. (2012). Identifying the last supper: utility of the DNA barcode library for bloodmeal identification in ticks. *Molecular Ecology Resources*, *12*(4), 646-652.

Guillera-Arroita, G. (2011). Impact of sampling with replacement in occupancy studies with spatial replication. *Methods in Ecology and Evolution, 2*(4), 401-406.

Hunter, M. E., Meigs-Friend, G., Ferrante, J. A., Kamla, A. T., Dorazio, R. M., Diagne, L. K., Luna, F., Lanyon, J. M., & Reid, J. P. (2018). Surveys of environmental DNA (eDNA): a new approach to estimate occurrence in Vulnerable manatee populations. *Endangered Species Research, 35*, 101-111.

Iknayan, K.J., Tingley, M.W., Furnas, B.J. & Beissinger, S.R. (2014) Detecting diversity: emerging methods to estimate species diversity. *Trends in Ecology & Evolution, 29*, 97–106.

IIshige, Taichiro, Masaki Miya, Masayuki Ushio, Tetsuya Sado, Masaharu Ushioda, Kaori Maebashi, Risako Yonechi, Peter Lagan, & Matsubayashi, H. (2017). Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biological Conservation, 210*, 281-285.

Jones, K.E., Bielby, J., Cardillo, M., Fritz, S.A., O'Dell, J., Orme, C.D.L., Safi, K., Sechrest, W., Boakes, E.H., Carbone, C. & Connolly, C. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology, 90*(9), 2648-2648.

Kellner, K. (2015). jagsUI: A wrapper around 'rjags' to streamline 'JAGS'analyses. R package 1.3. 7.

Kent, R. J., & Norris, D. E. (2005). Identification of mammalian blood meals in mosquitoes by a multiplexed polymerase chain reaction targeting cytochrome B. *The American Journal of Tropical Medicine and Hygiene, 73*(2), 336-342.

Lai, Y.-T., Nakano, T., & Chen, J.-H. (2011). Three species of land leeches from Taiwan, Haemadipsa rjukjuana comb. n., a new record for Haemadipsa picta Moore, and an updated description of Tritetrabdella taiwana (Oka). *ZooKeys*, (139), 1.

Lee, P. S., Sing, K. W., & Wilson, J. J. (2015). Reading mammal diversity from flies: the persistence period of amplifiable mammal mtDNA in blowfly guts (Chrysomya megacephala) and a new DNA mini-barcode target. *PloS one*, *10*(4), e0123871.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm?. *Ecology, 74*(6), 1659-1673.

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Andrew Royle, J., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248-2255.

MacKenzie, D. I., & Royle, J. A. (2005). Designing occupancy studies: General advice and allocating survey effort. *Journal of Applied Ecology, 42*(6), 1105–1114.

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. A., & Hines, J. E. (2006). Occupancy modeling and estimation. Academic Press, San Diego, CA, 324.

Nesemann, H., & Sharma, S. (2001). Leeches of the suborder Hirudiniformes (Hirudinea: Haemopidae, Hirudinidae, Haemadipsidae) from the Ganga watershed (Nepal, India: Bihar). *Annalen des Naturhistorischen Museums in Wien. Serie B für Botanik und Zoologie*, 77-88.

Nichols, J.D., Bailey, L.L., O'Connell Jr, A.F., Talancy, N.W., Grant, E.H.C., Gilbert, A.T., Annand, E.M., Husband, T.P. & Hines, J. E. (2008). Multi-scale occupancy estimation and modelling using multiple detection methods. *Journal of Applied Ecology, 45*(5), 1321-1329.

Niedballa, J., Sollmann, R., Courtiol, A., & Wilting, A. (2016). camtrapR: an R package for efficient camera trap data management. *Methods in Ecology and Evolution*, *7*(12), 1457-1462.

Oksanen, J., F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs & Helene Wagner (2017). vegan: Community Ecology Package. R package version 2.4-5.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences, 70*(8), 1123-1130.

Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing* (Vol. 124, No. 125.10).

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roberts, N. J. (2011). Investigation into survey techniques of large mammals: surveyor competence and camera-trapping vs. transect-sampling. *Bioscience Horizons, 4*(1), 40-49.

Rodgers, T.W., Xu, C., Giacalone, J., Kapheim, K., Saltonstall, K., Vargas, M., Yu, D., Somervuo, P., McMillan, W.O., & Jansen, P. (2017). Carrion Fly-Derived DNA Metabarcoding Is An Effective Tool For Mammal Surveys: Evidence From A Known Tropical Mammal Community. *Molecular Ecology Resources, 17*(6), e133–e145.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*. Elsevier.

Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources, 16*(4), 895-908.

Schmidt, B. R., Kery, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution, 4*(7), 646-653.

Schnell, I.B., Thomsen, P.F., Wilkinson, N., Rasmussen, M., Jensen, L.R.D., Willerslev, E., Bertelsen, M.F., & Gilbert, M.T.P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, *22*(8), R262-R263.

Schnell, I.B., Sollmann, R., Calvignac-Spencer, S., Siddall, M.E., Yu, D.W., Wilting, A., & Gilbert, M.T.P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool – prospects, pitfalls and avenues to be developed. *Frontiers in Zoology* , *12*(1), 24.

Schnell, I.B., Bohmann, K., Schultze, S.E., Richter, S.R., Murray, D.C., Sinding, M.H.S., Bass, D., Cadle, J.E., Campbell, M.J., Dolch, R. & Edwards, D.P. (2018). Debugging diversity – a pan-continental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool. *Molecular Ecology Resources,* 1–17.

Schubert, G., Stockhausen, M., Hoffmann, C., Merkel, K., Vigilant, L., Leendertz, F. H., & Calvignac-Spencer, S. (2015). Targeted detection of mammalian species using carrion fly-derived DNA. *Molecular Ecology Resources*, *15*(2), 285-294.

Somervuo, P., Koskela, S., Pennanen, J., Nilsson, H. & Ovaskainen, O. (2016) Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics, 32*, 2920–292 7.

Somervuo, P., Yu, D.W., Xu, C.C.Y., Ji, Y., Hultman, J., Wirta, H., & Ovaskainen, O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution*, 8, 398–407.

Tessler, M., Weiskopf, S. R., Berniker, L., Hersch, R., Mccarthy, K. P., Yu, D. W., & Siddall, M. E. (2018). Bloodlines: mammals, leeches, and conservation in southern Asia. *Systematics and Biodiversity*, 1-9.

Thomsen, P.F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18.

Trolliet, F., Vermeulen, C., Huynen, M. C., & Hambuckers, A. (2014). Use of camera traps for wildlife studies: a review. *Biotechnologie, Agronomie, Société et Environnement, 18*(3), 446-454.

Verdonschot, P. F., & Besse-Lototskaya, A. A. (2014). Flight distance of mosquitoes (Culicidae): a metadata analysis to support the management of barrier zones around rewetted and newly constructed wetlands. *Limnologica-Ecology and Management of Inland Waters, 45,* 69-79.

Wearn, O. R., Rowcliffe, J. M., Carbone, C., Bernard, H., & Ewers, R. M. (2013). Assessing the status of wild felids in a highly-disturbed commercial forest reserve in Borneo and the implications for camera trap survey design. *PLoS One, 8*(11), e77598.

Weiskopf, S.R., McCarthy, K.P., Tessler, M., Rahman, H.A., McCarthy, J.L., Hersch, R., Faisal, M.M., & Siddall, M. E. (2018). Using terrestrial haematophagous leeches to enhance tropical biodiversity monitoring programmes in Bangladesh. *Journal of Applied Ecology, 55*(4), 2071-2081.

Wineland, S. M., Welch, S. M., Pauley, T. K., Apodaca, J. J., Olszack, M., Mosher, J. J., Holmes, J. N., & Waldron, J. L. (2019). Using environmental DNA and occupancy modelling to identify drivers of eastern hellbender (Cryptobranchus alleganiensis alleganiensis) extirpation. *Freshwater Biology, 64*(1), 208-221.

Figure captions

Figure 1: Structure of the occupancy model, which consists of an ecological process and an observation process, through which the ecological process is filtered. Factors that affect the ecological state (presence / absence) are related to habitat, while factors that affect the detection state (detection / non-detection) can be related to both habitat and survey conditions (survey method, effort, etc.).

Figure 2: (A) The study site, Deramakot Forest Reserve, in Malaysian Borneo, with the locations of the 64 camera-trap and leech sampling stations (white circles). (B) The number of leeches sampled (indicated by the size of the circles) in Deramakot on the three sampling occasions. (C) The number of detections for camera-trap and leech surveys (represented by the size of the circles).
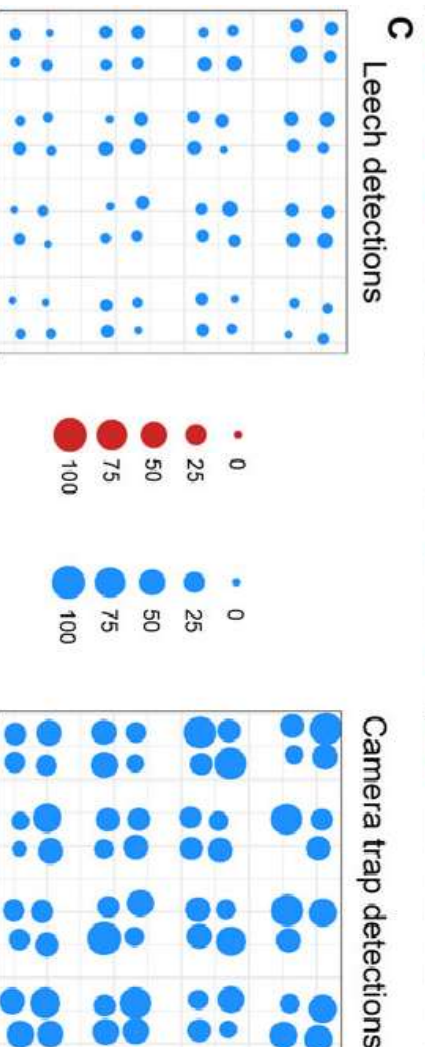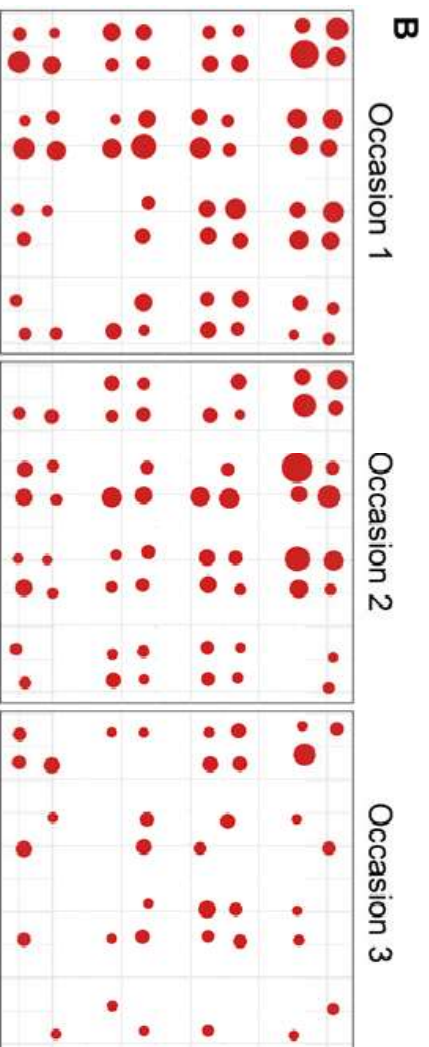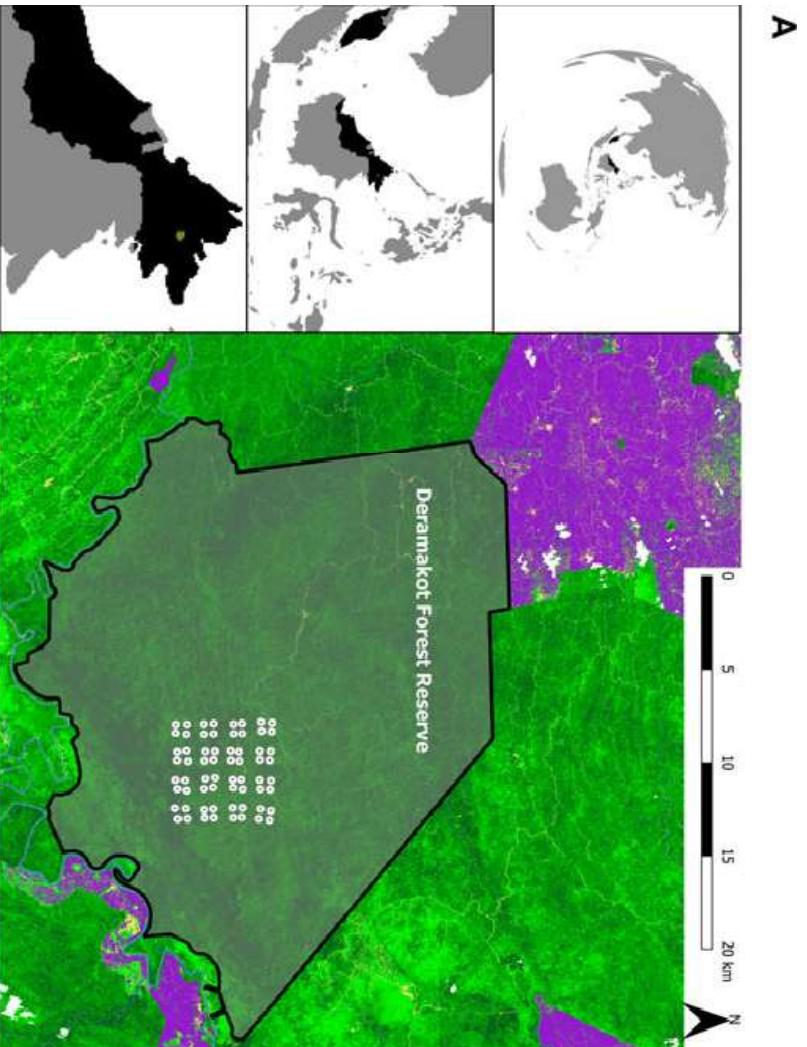
Figure 3: (A) Species-accumulation curves constructed for camera-trap detections (red line) and leech detections (light blue). The main plot shows the mammal species-accumulation up to the total of 2,733 camera-trap detections while the inset shows the accumulation up to the first 250 detections. (B) Correlation between detections and the average head and body length of species. The solid lines represent the best-fit line; the grey shaded areas represent the 95% confidence intervals.
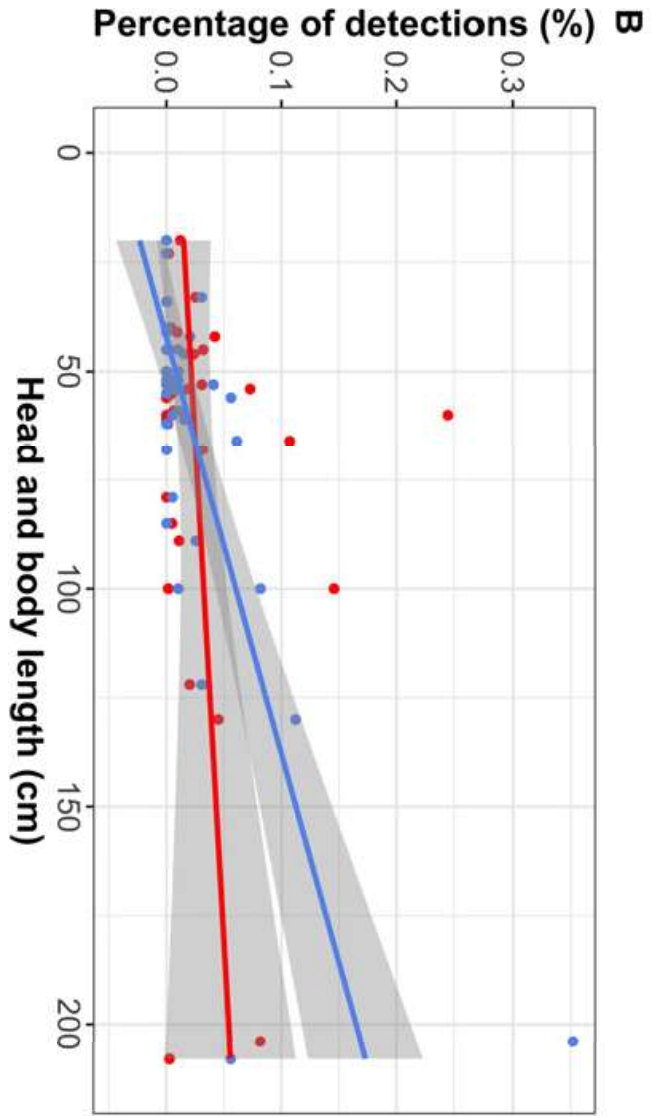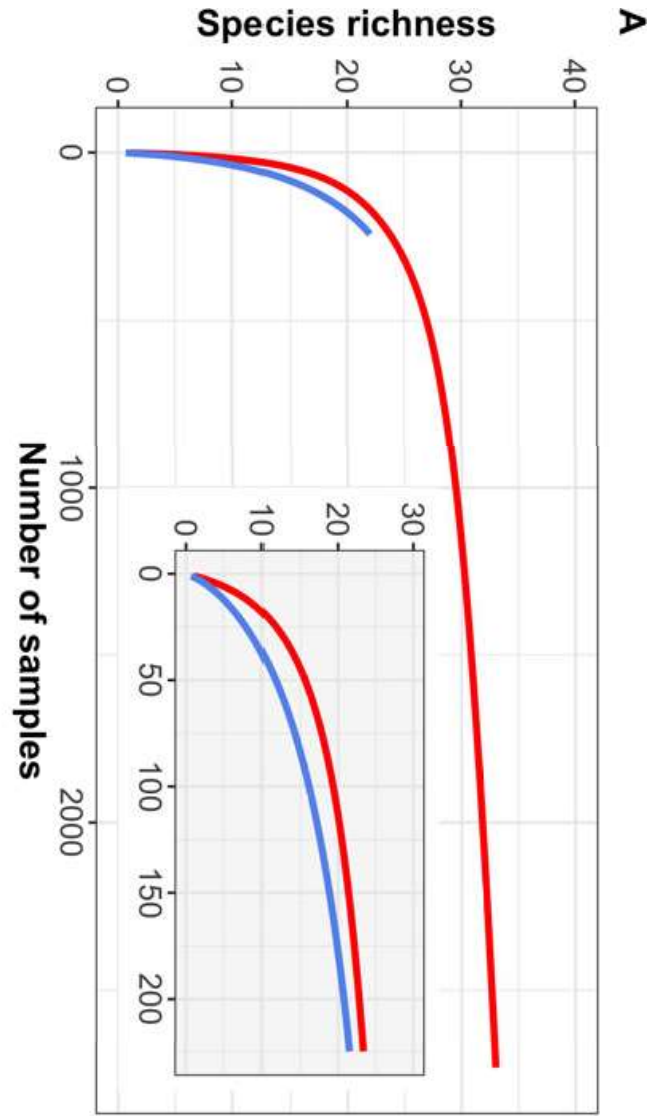
Figure 4: (A) Mammal species compositions of photographs from 1,334 camera-trap nights which generated 2,733 detections (left, red bars) and from 242 pooled leech samples, which generated 196 detections (right, light blue bars). (B) Percentage of successful detections of species in 116 pooled tiger leech samples, which generated 133 detections (right, orange bars) and 126 pooled brown leech samples, which generated 63 detections (left, brown bars).

Figure 5: Occupancy and detection probabilities estimated by null single-species occupancy models for independent camera-trap and leech survey data, which include 64 stations. (A) Camera-trap (circles), tiger (triangles), brown (squares) detection probabilities are plotted with their 95% Bayesian CIs. (B) Estimated occupancy probabilities for camera-trap models (x-axis) are plotted against the estimated occupancy probabilities for leech models (y-axis). The vertical and horizontal bars indicate 95% Bayesian CIs for the leech survey and camera-trap survey, respectively. The black line is the best-fit line; the grey shaded area represents the 95% confidence interval.

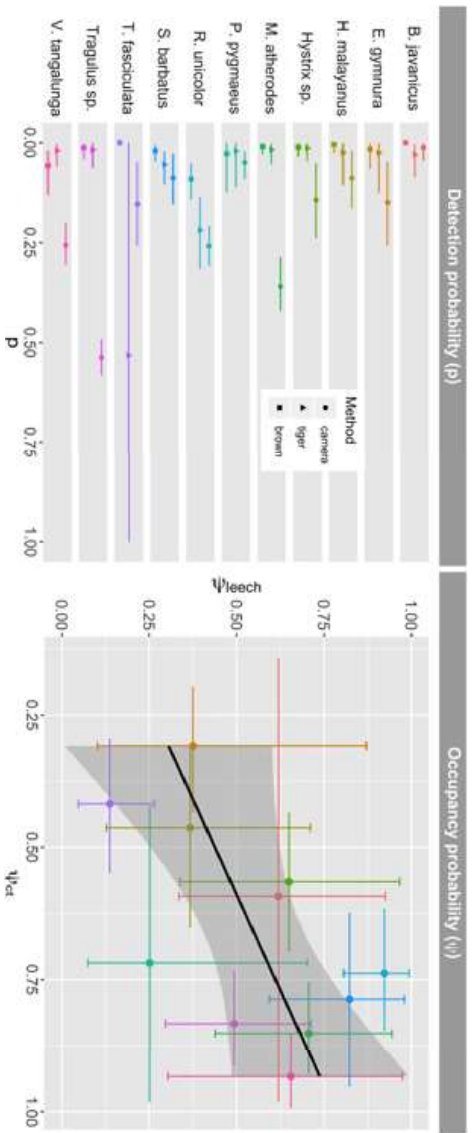**Figure 6**: The size of the 95% Bayesian CIs for modelled occupancy probability for 11 species (represented by the coloured points) for the leech iDNA, camera-trap, and combined datasets. The black dots for each dataset represent the average size of the occupancy probability CIs. The grey shaded areas represent the distribution and probability density of the size of the confidence intervals estimated from the single species models.

Ecological process

Imperfect detection process

Observation / Input

Environmental factors affecting presence / absence

Not present (1 − ψ)

Present (ψ)

True occupancy state

Not detected

detection method

number of leeches

number of nights

number of species detected

Factors affecting detectability

Not detected (1 − p)

Detected (p)

Site

Occasion

A

B

Occasion 1   Occasion 2   Occasion 3

C

Leech detections

Camera trap detections

Uncertainty in occupancy estimate

0.00 · 0.25 · 0.50 · 0.75 · 1.00 ·

leech

camera

Dataset

combined

Species

- B. javanicus
- E. gymnura
- H. malayanus
- Hystr x sp.
- M. atherodes
- P. pygmaeus
- R. unicolor
- S. barbatus
- T. fasciculata
- Tragulus sp.
- V. targalunga