# Large-scale Quantification of Vertebrate Biodiversity in Ailaoshan Nature Reserve from Leech iDNA

Yinqiu Ji[1,*], Christopher CM Baker[2,*,**], Yuanheng Li[1,2], Viorel D Popescu[3,4], Zhengyang Wang[2], Jiaxin Wang[1], Lin Wang[5,6], Chunying Wu[1], Chaolang Hua[7], Zhongxing Yang[7], Chunyan Yang[1], Charles CY Xu[8], Qingzhong Wen[7], Naomi E Pierce[2,**], and Douglas W Yu[1,9,10**]

[1]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, 32 Jiaochang Dong Lu, Kunming, Yunnan 650223 China
[2]Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge MA 02138 USA
[3]Department of Biological Sciences and Sustainability Studies Theme, Ohio University, Athens OH 45701 USA
[4]Center for Environmental Studies (CCMESI), University of Bucharest, Bucharest, Romania
[5]Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla 666303, China
[6]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Mengla 666303, China
[7]Yunnan Institute of Forest Inventory and Planning, 289 Renmin E Rd, Kunming Yunnan 650028 China
[8]Redpath Museum and Department of Biology, McGill University, 859 Sherbrooke Street West, Montreal, PQ H3A2K6 Canada
[9]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming Yunnan, 650223 China
[10]School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, Norfolk NR47TJ, UK
[*]These authors contributed equally to this work.
[**]Corresponding authors. CCMB: bakerccm@gmail.com; NEP: npierce@oeb.harvard.edu; DWY: dougwyu@mac.com

Keywords: environmental DNA, metabarcoding, occupancy, leeches, China, monitoring, conservation, biodiversity

## 1 Abstract

Environmental DNA (eDNA) has great potential to complement visual surveys, camera trapping, and bioacoustics in measuring biodiversity. We report here a large-scale attempt to use DNA from leech-ingested bloodmeals to estimate vertebrate occupancy at the scale of an entire protected area: the 677 km$^2$ Ailaoshan national-level nature reserve in Yunnan province, southwest China. We contracted 163 park rangers to collect leeches in 172 patrol areas, resulting in 30,468 total leeches, divided over 893 replicate samples. Replicate sampling from each patrol area was achieved by providing rangers with preservative-filled tubes in separate, self-sealing bags, inducing them to distribute collected leeches over multiple bags, which we could sequence separately. We show that combining leech-derived DNA data with Bayesian site-occupancy modeling can produce detailed and useful biodiversity inferences for a wide range of vertebrates in a reasonable timeframe. For example, in

Ailaoshan, sites at lower elevations and closer to the reserve edge have higher occupancy of domestic species (cows, sheep, goats) compared to sites at higher elevations and closer to the reserve interior. Muntjak deer show similar occupancy patterns to the domesticated species, but otherwise, most species of large mammal wildlife (e.g. sambar, black bear, serow, tufted deer) show the opposite pattern: greater occupancy at higher elevations and closer to the reserve interior. We conclude that leech-derived eDNA can be used to efficiently measure the effectiveness of protected areas in terms of vertebrate biodiversity outcomes and to help us optimize the deployment of management resources within reserves, by providing valuable information on the spatial distributions of vertebrate species and on the environmental and anthropogenic correlates of those distributions.

环境DNA (eDNA) 在生物多样性评估方面具有极大的潜力, 可以与现有的生物多样性监测方法, 如目视遇测法、红外相机监测法, 生物声学监测法等形成互补. 本研究首次利用蚂蝗eDNA(即从蚂蝗吸食的血液中提取的DNA) 对位于中国西南部云南省的哀牢山国家自然保护区进行了一个全局的脊椎动物多样性的评估. 在本研究中, 677平方公里的保护区被划分成172个巡逻区, 由163位护林员在巡视过程中采集了总共30468只蚂蝗, 这些蚂蝗根据采集样点, 采集时间及具体数量被合并成了893份样本. 我们将蚂蝗eDNA数据和贝叶斯位点占据模型相结合, 分析推断得到在一定时间范围内哀牢山脊椎动物各物种的具体分布图谱. 例如, 哀牢山海拔较低及靠近保护区边缘的地区, 相较于海拔较高及靠近保护区中心的地区, 具有更多的牛, 羊等家养动物; 在野生物种中, 赤麂呈现了与家养动物类似的分布趋势, 而其它大型哺乳动物(如水鹿、黑熊、苏门羚) 则相反, 在海拔较高及靠近保护区中心的区域分布更多. 本研究的结果显示基于蚂蝗的eDNA技术可以提供关于脊椎动物物种的空间分布以及环境和人类活动对这些物种分布的影响等有价值的信息, 让我们可以据此评估保护区对脊椎动物多样性的保护效率, 从而可以帮助优化保护区内管理资源的部署.

## 2 Introduction

*The difficulty of measuring the effectiveness of protected areas.* In 2010, the signatories of the Convention on Biological Diversity, including China, agreed to the 2011-2020 Aichi Biodiversity Targets [6]. Aichi Target 11 concerns the safeguarding of biodiversity and sets the goals of placing (A) 17% of terrestrial and inland water habitats in a system of nature reserves that are (B) ecologically representative, (C) well-connected, (D) equitable, and (E) effective. The world has nearly achieved goal A, with 15% of global land area now under national jurisdiction [13, 71]. China has to date also placed 15% (1.43 million $km^2$) of its land surface into nature reserves [78, 9]. Moreover, Wu *et al.* [76] have shown that, at least in western China, the reserve system covers most ecoregions, biodiversity priority areas, and natural vegetation types (goal B), and Ren *et al.* [52] have used time-series analyses of Landsat imagery to show that China's national-level nature reserves successfully prevent deforestation (goal E). China has therefore already demonstrated some considerable institutional capacity for achieving Aichi Target 11.

However, in southern and eastern China, the ecological representativeness of reserves is low (goal B) [77], many reserves are isolated (goal C) [76], there is little information on the impact of the reserves on local human populations (goal D) and, most importantly, we know little about whether the reserves are effective at protecting the species that live inside them (goal E). Our focus in this study is thus goal E, reserve effectiveness, because if reserves cannot protect their existing biodiversity endowments, the other four goals do

not matter [34, 22, 71, 5, 13]. We further focus on vertebrates (mammals, birds, amphibians, and reptiles), because one of the most important threats to vertebrate populations in China is overexploitation [79], which is undetectable using remote sensing methods, and thus especially difficult to measure.

Here, we ask if it is possible to design *efficient*, *repeated*, *direct*, *granular*, *understandable*, and *auditable* metrics of vertebrate-biodiversity statuses. While indirect measures, such as reports of staffing and budget levels [13] and expert-opinion surveys [34] can be used to estimate the aggregate status of reserves at broad spatial scales, we also need regularly updated (*repeated*) assessments of vertebrate populations themselves (*direct* and *granular*) in order to structure management incentives (e.g. job promotions) and thereby to increase and maintain management effectiveness. Moreover, it should be possible for these metrics to be checked in a rigorous way and thus validated by neutral third parties such as courts and the public (*auditable* and *understandable*), which is necessary for dispute resolution. Finally, because management resources are limited, such metrics should be cost-effective (*efficient*) [20, 80, 16].

*Emerging technologies for surveying vertebrate biodiversity at broad spatial scales.* Advances in and increased availability of technologies such as camera traps, bioacoustics, and environmental DNA (eDNA) allow for broad-scale biodiversity monitoring, via direct and auditable vertebrate-biodiversity metrics that complement current Earth Observation technologies (e.g. satellite remote sensing of land-cover change) [10]. In particular, camera traps (and increasingly, bioacoustics) have shown great promise in developing biodiversity indicators that meet the requirements of the Convention for Biological Diversity for broad-scale biodiversity monitoring [2, 5, 35, 49, 74, 24]. However, the costs of buying and deploying camera traps and bioacoustic devices places limitations on the area that they can be used to monitor. For example, Beaudrot *et al.* [5] recently reported that multi-year camera-trap surveys of 511 populations of terrestrial mammals and birds in fifteen tropical-forest protected areas did not detect "systematic declines in biodiversity (i.e. occupancy, richness, or evenness)." However, each of their camera-trap sets sampled between 140 to 320 $km^2$, which is only 1-2% of the area of the largest parks in their dataset, the obvious reason being the difficulty and expense of setting up a camera-trap network to cover large, difficult-to-access areas, exacerbated by theft and vandalism in some contexts [46, 24]. Furthermore, both camera traps and acoustic recorders may miss large portions of vertebrate species diversity; for example, amphibians, reptiles, and birds are not readily (if ever) captured on camera traps, and many mammals and reptiles may be easily missed via bioacoustic monitoring.

As such, eDNA has great potential to complement camera traps and acoustic recorders, while circumventing some of the logistical issues with deployment of field equipment, as well as the taxonomic bias in sampling. Here, we focus on iDNA, which is a subset of environmental DNA, or eDNA [7], as an emerging technology for broad spatial scale and broad taxonomic breadth biodiversity monitoring. iDNA is vertebrate DNA collected by invertebrate 'samplers,' including haematophagous parasites (leeches, mosquitoes, biting flies, ticks) and dung visitors (flies, dung beetles) [8, 11, 62]. iDNA methods are rapidly improving, with research focused on documenting the ranges of vertebrate species and their diseases that can be efficiently detected via iDNA [18, 25, 32, 61, 66], plus comparisons with camera trapping and other survey methods [1, 53, 72], and pipeline development [4, 23].

*Leech-derived iDNA.* We report here a large-scale attempt to use iDNA to estimate vertebrate occupancy at the scale of an entire protected area, the Ailaoshan national-level nature reserve in Yunnan province, southwest China. Ailaoshan covers 677 $km^2$, nearly the size of Singapore, and the Yunnan Forestry Service has previously attempted to monitor vertebrate diversity in the reserve via camera traps [28]. Our goal was to test whether it is realistic to scale up an iDNA survey within a realistic management setting, from sample collection and molecular labwork through bioinformatic processing and statistical analysis.

We had several reasons to test the use of leech-derived iDNA as a promising broad-scale monitoring technology. The two most important advantages concern efficiency. First, the personnel collecting leeches do not require specialized training. The Ailaoshan reserve is divided into 172 'patrol areas' that are each patrolled monthly by park rangers hired from neighboring villages, whom we contracted to collect terrestrial, haematophagous leeches during their rainy-season patrols. We were thus able to sample the entire reserve in a short period of time (2-3 months) at low cost. Second, leech sampling potentially provides an efficient way to correct for false negatives (i.e. the failure to detect species that are nonetheless truly present at that site). With leeches, false negatives can arise because (A) animals were not fed upon by leeches; (B) the leeches containing animal DNA were not captured; or (C) the species' DNA was not successfully amplified and/or associated with the correct taxon. Statistical models may be used to account for these sources of imperfect detection.

In this project, we use a hierarchical site-occupancy model [44], which distinguishes between the detection of a species' DNA and the true presence or absence of the species at a site. The goal of site-occupancy modeling is to infer whether non-detections represent a truly absent species or a false negative, by separately estimating the probability that a species is present at a site and the probability that a species is detected if it is present [44, 68]. Separating these probabilities relies on a replicated sampling design, with replicates taken in sufficiently close spatial and/or temporal proximity such that the underlying distribution of species presences or absences may be treated as fixed. We achieved *replicate samples per patrol area in just one patrol* by issuing each ranger with multiple, small storage tubes, inducing subsets of leeches to be stored in separate tubes [62], which we processed separately.

Other potential advantages are that iDNA is likely to yield inferences about a broad range of taxa, as leeches feed on small and large mammals, birds, reptiles, and amphibians, including arboreal species; this provides a taxonomic breadth that is not typically captured via camera traps or bioacoustic surveys [35, 61, 66]. Also, DNA sequences can potentially distinguish some visually cryptic species [1] (although lack of species-level resolution also occurs with iDNA sequences). Finally, leeches can yield PCR-amplifiable DNA for at least four months after their last blood meal [63], which should improve the efficiency of leech iDNA by increasing the proportion of collected leeches that can yield information on their previous bloodmeal. On the other hand, leech iDNA persistence could also *decrease* the spatiotemporal resolution of vertebrate detections, since the potentially long period between leech capture and its previous feed affords more opportunity for the vertebrate hosts to have moved in and out of the sampling areas [62]).

In this study, we used metabarcoding [30] to detect vertebrate species sampled in the blood meals of wild leeches, and occupancy modeling to estimate the spatial distributions of those vertebrates throughout the 677 $km^2$ Ailaoshan reserve in Yunnan Province, China. We further identified environmental and human-related factors that correlated with these distri-

butions. We find that leech-derived iDNA data can capture plausible and useful occupancy patterns for an unusually wide range of vertebrates, including amphibians and birds that are unlikely to be sampled using alternative methods such as camera traps and bioacoustic surveys. We conclude that iDNA can contribute usefully characterizing the effectiveness of protected areas, by providing information on the spatial distributions and environmental and human-related correlates of vertebrate species, helping us to optimize management strategies within the reserve.

# 3 Methods

This section provides an overview of methods. Supplementary File S1 provides additional detailed descriptions of the leech collections, laboratory processing, bioinformatics pipeline, and site-occupancy modeling.

## 3.1 Field site

The long and narrow 677 km$^2$ Ailaoshan reserve runs northwest-to-southeast along a ridgeline for around 125 km (approx. 24.9°N 100.8°E to 24.0°N 101.5°E), averaging just 6 km wide along its length, with an elevation range of 422 to 3,157 m and an annual precipitation range of 1,000 to 1,860 mm, depending on altitude [82] (Figure 1a). Vegetation is subtropical, evergreen broadleaf forest, and the reserve is flanked by agricultural land on lower-elevation slopes in all directions. There are 261 villages within 5 km of the reserve border [83], with an estimated human population of over 20,000. After the reserve's establishment in 1981, a 1984-5 survey published a species list of 86 mammal, 323 bird, 39 reptile, and 26 amphibian species/subspecies [3]. Although investigators have since carried out one-off targeted surveys [75, 70, 39] and individual-species studies [43, 40, 38, 37, 33], there has never been a synoptic survey of vertebrate biodiversity. As a result, the current statuses and population trends of vertebrate species in the park are mostly unknown.

## 3.2 Leech collections

Samples were collected in the rainy season, from July to September 2016, by park rangers from the Ailaoshan Forestry Bureau. The nature reserve is divided into 172 non-overlapping patrol areas ranging in size from 0.5 to 12.5 km$^2$ (mean 3.9 $\pm$ sd 2.5 km$^2$). Each ranger was supplied with several small bags containing tubes filled with preservative. Rangers were asked to place any leeches they could collect opportunistically during their patrols (e.g. from the ground or clothing) into the tubes, in exchange for a one-off payment of RMB 300 ($\sim$ USD 43) for participation, plus RMB 100 if they caught one or more leeches. Multiple leeches could be placed into each tube, but the small tube sizes generally required the rangers to use multiple tubes for their collections.

A total of 30,468 leeches were collected by 163 rangers across all 172 patrol areas. When a bag of tubes contained $< 100$ total leeches, we reduced our DNA-extraction workload by pooling leeches from all tubes in the same plastic bag and treating them as one replicate. However, when a bag contained $\geq 100$ total leeches, we selectively pooled some of the

tubes in that bag to create five approximately equally sized replicates from the bag, to avoid any replicates containing an excessive number of leeches. Eighty-one per cent of bags contained $< 100$ leeches, and 78% of patrol areas consisted only of bags below the threshold. Each ranger's patrol thus typically returned multiple replicates, as rangers were issued multiple bags and these were always kept separate. After this pooling, the mean number of leeches per replicate was 34 (range 1 to 98), for a total of 893 replicates across the entire collection.

## 3.3 Environmental characteristics

We used ArcGIS Desktop 9.3 (Esri, Redlands, CA) and R v3.4.0 [50] to calculate characteristics of each patrol area from shapefiles. We created 30 m rasters for elevation, topographic position index (i.e. difference between each pixel and its surrounding pixels [27]), distance to nearest road, and distance to nearest stream. We then calculated the median of the raster values for each patrol area for use as predictors in our statistical modeling (Table 1 and S1). We also calculated distance to the Ailaoshan nature-reserve boundary as the distance of each patrol-area centroid to the nearest nature-reserve boundary.

**Table 1:** Environmental covariates

| Variable | Description | Mean ± SD | Min | Max |
|---|---|---|---|---|
| *elev* | median elevation (m) | 2,510 ± 210 | 1,690 | 2,900 |
| *TPI* | median topographic position index | 0.6 ± 3.5 | -12.0 | 20.0 |
| *road* | median distance to road (m) | 840 ± 640 | 60 | 2,870 |
| *stream* | median distance to stream (m) | 360 ± 180 | 90 | 1,010 |
| *reserve* | centroid distance to reserve boundary (m) | 1110 ± 670 | 150 | 3,900 |

## 3.4 Laboratory processing

We extracted DNA from each replicate, and then PCR-amplified two mitochondrial markers: one from the 16S rRNA (MT-RNR2) gene, and the other from the 12S rRNA (MT-RNR1) gene. We hereafter refer to these two markers as LSU (16S) and SSU (12S), respectively, referring to the ribosomal large subunit and small subunit that these genes code for. The LSU primers are designed to target mammals, and the SSU primers to amplify all vertebrates. A third primer pair targeting the standard cytochrome *c* oxidase I marker [36] was tested but not adopted in this study as it co-amplified leech DNA and consequently returned few vertebrate reads. Primers were ordered with sample-identifying tag sequences, and we used a fully-redundant twin-tagging strategy to identify and remove 'tag jumping' errors [60] using the DAMe protocol [81]. From our 893 replicate tubes, we successfully PCR-amplified in triplicate 661 samples using our LSU primers and 745 samples using our SSU primers. Successful amplifications were sent to Novogene (Beijing, China) for PCR-free library construction and 150 bp paired end sequencing on an Illumina HiSeq X Ten.

## 3.5 Bioinformatics pipeline

Three key features of our pipeline were the DAMe protocol [81], which uses independent PCR replicates to identify and remove tag-jumped and erroneous reads, the use of two independent markers (Figure S2), which provides an independent check on taxonomic assignments, and the PROTAX statistical 'wrapper' for taxonomic assignment [64, 65], which reduces overconfidence in taxonomic assignment when reference databases are incomplete, as they always are.

After DAMe filtering, we removed residual chimeras using VSEARCH v2.9.0 [54], clustered sequences into preliminary OTUs ('pre-OTUs') using SWARM v2.0 [45], and then used the R package `lulu` v0.1.0 [21] to merge pre-OTUs with high similarity and distribution across samples. We then used PROTAX to assign taxonomy to representative sequences from the merged pre-OTUs [64, 65], in which we benefited from recent additions to the mitochondrial reference database for Southeast Asian mammals [48]. We shared taxonomic information between the LSU and SSU datasets by making use of correlations between the datasets. To do this, we calculated pairwise correlations of SSU and LSU pre-OTUs across the 619 replicates for which both markers had been amplified and visualized the correlations as a network (Figure S2). If an SSU and an LSU pre-OTU occur in the same subset of replicates and are assigned the same higher-level taxonomies, the two pre-OTUs were deemed likely to have been amplified from the same set of leeches feeding on the same species. We manually inspected the network diagram and assigned such correlated pre-OTU pairs the same taxonomy.

We eliminated any pre-OTUs to which we were unable to assign a taxonomy; these pre-OTUs only accounted for 0.9% and 0.2% of reads in the LSU and SSU datasets respectively, and most likely represent sequencing errors rather than novel taxa. Within the LSU and SSU datasets, we merged pre-OTUs that had been assigned the same taxonomies, thus generating a final set of OTUs for each dataset. Finally, we removed the OTU identified as *Homo sapiens* from both datasets prior to analysis.

After excluding humans, the final LSU and SSU datasets comprised 18,502,593 and 84,951,011 reads respectively. These reads were assigned to a total of 72 OTUs across 740 replicates and 127 patrol areas in the SSU dataset, and 59 OTUs across 653 replicates and 126 patrol areas in the LSU dataset. We attached IUCN data for individual OTUs by using the R package `rredlist` v0.5.0 [12] to search for scientific names assigned by PROTAX (or synonyms where we were aware of nomenclature changes). For mammalian OTUs, we used the PanTHERIA database [31] to obtain data on adult body mass for each OTU; where species-level information was not available, we used the median adult body mass from the database for the lowest taxonomic group possible.

## 3.6 Site-occupancy modeling

We estimated separate multispecies site-occupancy models [17] using the LSU and SSU OTU tables. The hierarchical models that we used are an extension of the single-season occupancy model in [44]. For each species, the models explicitly capture (A) an 'ecological process' governing the (unobserved) presence or absence of the species in each patrol area; and (B) an 'observation process', governing whether we detect the species' DNA in each of

our replicate samples. The ecological and observation processes for individual species are linked in our model by imposing community-level priors over the parameters that describe the processes for each species.

For the ecological process, each species $i$ was assumed to be either present or absent in each patrol area $j$, and we used $z_{i,j}$ to denote this unobserved ecological state. We assumed the $z_{i,j}$ are constant across all replicates taken from patrol area $j$, consistent with the samples being taken at essentially the same point in time. $z_{i,j}$ was assumed to be a Bernoulli random variable governed by an occupancy parameter $\psi_{i,j}$, i.e. the probability that species $i$ was present in patrol area $j$:

$$z_{i,j} \sim Bernoulli(\psi_{i,j}). \tag{1}$$

After model selection (see Supplementary File S1 for details), we modelled occupancy $\psi_{i,j}$ as a function of elevation and distance from the reserve boundary:

$$logit(\psi_{i,j}) = \beta_{0i} + \beta_{1i} elev_j + \beta_{2i} reserve_j. \tag{2}$$

where $elev_j$ is the median elevation for the patrol area $j$, and $reserve_j$ is the distance from centroid of patrol area $j$ to the nature reserve boundary.

We modelled the observation as a Bernoulli process assuming imperfect detection but no false negatives:

$$y_{i,j,k} \sim Bernoulli(z_{i,j}.p_{i,j,k}), \tag{3}$$

where $y_{i,j,k}$ is the observed data, i.e. detection or non-detection of species $i$'s DNA in replicate $k$ from patrol area $j$.

We allowed the conditional detection probability $p_{i,j,k}$ to vary across species and as a function of the number of leeches included in the replicate, $numleeches_{j,k}$:

$$logit(p_{i,j,k}) = \gamma_{0i} + \gamma_{1i} numleeches_{j,k}. \tag{4}$$

Finally, whereas equations (1) through (4) define a site-occupancy model for species $i$ alone, we united these species-specific model with community models for both ecological and detection processes:

$$\begin{aligned} \beta_{1i} &\sim N(\mu_{\beta_1}, \sigma_{\beta_1}) \\ \beta_{2i} &\sim N(\mu_{\beta_2}, \sigma_{\beta_2}) \\ \gamma_{1i} &\sim N(\mu_{\gamma_1}, \sigma_{\gamma_1}) \\ (\beta_{0i}, \gamma_{0i}) &\sim MVN([\mu_{\beta_0}, \mu_{\gamma_0}], [\sigma_{\beta_0}, \sigma_{\gamma_0}]) \end{aligned} \tag{5}$$

where $N(\ )$ and $MVN(\ )$ denote normal and multivariate normal distributions, with community-level hyperparameters $\mu_\bullet$ and $\sigma_\bullet$. We used a multivariate normal prior for $(\beta_{0i}, \gamma_{0i})$ to allow non-zero covariance between species' occupancy and detection probabilities, as we might expect if, for example, variation in abundance affects both probabilities [17]. These community models allow rare species effectively to borrow information from more common ones, producing a better overall ensemble of parameter estimates [59, 41, 17].

We estimated all model variants in a Bayesian framework with uninformative diffuse priors for all parameters and hyperparameters. We ran each model with three chains of 40,000 generations and a burn-in of 10,000, thinning results by a factor of 20. From the retained results, we calculated means for all model parameters of interest, as well as estimated species richness for each patrol area.

## 3.7 Statistical analyses

*OTUs.* To assess the comprehensiveness of our sampling, we used `vegan::specaccum` to generate rarefaction curves for each dataset, and for the two datasets combined, at the replicate level. We assessed total detectable diversity by plotting each of these curves and estimating their asymptotes visually.

*Species richness.* After examining occupancy and detection estimates for each species, we used histograms to visualize the distribution of estimated species richness per patrol area. We calculated median estimated species richness across the patrol areas for comparison with median observed species richness per patrol area and per replicate. We drew choropleths to visualize the spatial distribution of both observed and estimated species richness across the nature reserve.

We focused on community occupancy (i.e. the average occupancy probability across species) in order to examine the effect of elevation and distance to reserve boundary on species richness. To see the relationship between elevation and community occupancy, we took the community mean hyperparameter on the $\beta_{0i}$ (i.e. $\mu_{\beta 0}$) and adjusted it for elevation scaled by the community mean hyperparameter on the elevation coefficients $\beta_{1i}$ (i.e. $\mu_{\beta 1}$). Taking the inverse logit gave us community occupancy on the probability scale. This approach holds distance from reserve edge at zero, corresponding to the mean value in our data, since predictors were normalized prior to modeling. We varied elevation over the range of the data, and for each elevation value calculated a mean and 95% credible interval for the community occupancy by repeating the calculation over the posterior distribution for $\mu_{\beta 0}$ and $\mu_{\beta 1}$. We performed analogous calculations to examine the influence of distance from reserve edge on community occupancy, and the influence of leech quantity on community detection probability.

We compared three measures of species richness between the two datasets in order to assess the extent to which the two datasets agreed on variation in richness within Ailaoshan. First, the observed species richness in each replicate; second, the observed species richness in each patrol area; and third, the estimated species richness in each patrol area. For each of these measures, we computed the Pearson correlation between the datasets and tested the correlation coefficient against zero with a $t$-test. We also used Poisson GLMs to examine the relationship between each of these species richness measures and sampling effort: we regressed observed species richness per replicate against the log-transformed number of leeches per replicate, and we regressed both the observed and estimated species richnesses per patrol area against the log-transformed number of replicates per patrol area, testing the significance of the slope coefficient with a $t$-test.

*Community composition.* To assess variation in vertebrate community composition across Ailaoshan, we conducted for each dataset a principal components analysis (PCA) on the predicted communities in each patrol area, as captured by the posterior means of the eco-

logical states $z_{i,j}$. We assessed the dimensionality of our data by examining the fraction of total variance explained by each principal component. To assess the extent to which the two datasets identified common patterns of variation in community composition across the patrol areas, we performed a co-inertia analysis on the matrices of predicted species in each patrol area in each dataset using `ade4::coinertia` in R. We used the RV coefficient [19] to quantify co-inertia, testing its significance with the permutation test in `ade4::RV.rtest` with 999 permutations. Since the first principal component (PC1) explained a large fraction of the overall variance, we next focused on exploring variation along that axis. First, we performed a redundancy analysis (RDA) using the environmental parameters in Table 1, drawing biplots with patrol areas as points colored by location on PC1, and environmental covariates as arrows. Second, to visualize spatial variation in community composition, we drew maps of Ailaoshan with patrol areas colored by location on PC1. Third, we examined the principal component scores of individual species. For mammals over 10 kg adult body mass, we plotted additional biplots showing those species as arrows. For all species, we extracted scores along the first RDA axis (RDA1) and ranked them to visualize the species that tended to be found more commonly on the ends of the ecological gradient identified by the PCA and RDA analyses.

# 4 Results

## 4.1 OTUs

We identified 86 vertebrate OTUs across the LSU and SSU datasets, in addition to humans. Of these, the LSU dataset included 59 OTUs, and the SSU dataset contained 72 OTUs. Although the LSU primers target mammals, both the LSU and SSU primers amplified amphibians, birds, mammals, and reptiles, with the general-vertebrate SSU primers amplifying more bird OTUs (Figure 2a). Forty-five of the OTUs were common to both datasets, including those that were linked by their distribution across replicates (Figure S2)), leaving 14 OTUs unique to LSU and 27 OTUs unique to SSU. We identified 58 of our 86 OTUs to species level (45 LSU, 50 SSU). Table 2 lists the top taxa in each dataset by estimated occupancy.

Domesticated species featured heavily in our data, consistent with observed grazing of these species in the reserve. Domestic cattle (*Bos taurus*) were the most frequently detected taxon in both datasets, being identified from almost half of all patrol areas; domestic goats (*Capra hircus*) were also common, being identified from just under a third (Supplementary File S2). Domestic sheep (*Ovis aries*) were detected in 7.1% and 6.3% of patrol areas in the LSU and SSU datasets respectively.

Among the detected wild taxa, several are listed as threatened by the IUCN (Supplementary File S2). Among the mammals identified by our sequencing, three taxa have IUCN Vulnerable status: Asiatic black bears (*Ursus thibetanus*) were detected by both LSU and SSU datasets, while LSU also detected sambar (*Rusa unicolor*) and stump-tailed macaques (*Macaca arctoides*). Among the amphibians, the Yunnan Asian frog (*Nanorana unculuanus*), Yunnan spiny frog (*Nanorana yunnanensis*), piebald spiny frog (*Nanorana unculuanus*) and Chapa bug-eyed frog (*Theloderma bicolor*) are all listed as Endangered, while the Jingdong toothed toad (*Oreolalax jingdongensis*) holds Vulnerable status. At least some of

these taxa are widely present in Ailaoshan (Table 2), highlighting the value of this reserve for protecting these species.

In general, leech iDNA was more successful at detecting Ailaoshan's mammals and amphibians than its birds and reptiles. For this, we used unpublished, working species lists maintained by researchers at the Kunming Institute of Zoology (Supplementary File S6). Among mammals, 34 of the 127 species in Ailaoshan were detected, with nearly half the detections in the larger-bodied orders: Artiodactyla (8 of 11 species), Carnivora (7 of 18), and non-human primates (1 of 4). Of the smaller-bodied orders, we detected 14 of 41 Rodentia species (including both the porcupines, *Atherurus macrourus* and *Hystrix brachyura*), 2 of 24 Eulipotyphla species (shrews and allies), and no bats (0 of 25), rabbits (0 of 1), pangolins (0 of 1), or treeshrews (0 of 1). We also detected two unnamed OTUs assigned to Rodentia. Among amphibians, 12 of the 25 frog species (order Anura) known from Ailaoshan were detected, and so were both of the salamander species (family Salamandridae). We detected 13 more anuran OTUs that could not be assigned to species, including two assigned to genus *Kurixalus*, which has not been reported from Ailaoshan but which has a distribution that overlaps Yunnan (Supplementary File S6). Among reptiles, we detected only 3 unnamed OTUs, compared to 39 species known from Ailaoshan. One of our OTUs was assigned only to Squamata, and the others to families Scincidae and Viperidae respectively. Finally, among birds, 12 of the 462 bird species known from Ailaoshan were detected, plus 10 more OTUs that were assigned to genus or higher. Interestingly, of the 12 species-level OTUs, five were in the ground-feeding and terrestrial Phasianidae (pheasants and allies), out of 14 species known from Ailaoshan, and the other seven species-level OTUs are known to be part-time ground and understorey feeders.

The more common taxa had occupancy estimates of 0.38 to 0.76 (Table 2) and detection estimates in the range 0.1 to 0.6. Most taxa, however, were observed infrequently (median number of detections: 2 and 3 patrol areas in the LSU and SSU datasets, respectively). This was reflected in low occupancy and detection estimates for most taxa (Figure 2c) (median occupancy estimates: 0.22 and 0.21 in LSU and SSU, respectively; median detection estimates: 0.013 and 0.029 in LSU and SSU, respectively).

Supplementary File S2 lists all OTUs, including observed occupancy as well as their occupancy and detection estimates. Supplementary Files S3 and S4 provide the OTU representative sequences in FASTA format. Supplementary File S5 provides OTU tables with sample metadata. Supplementary File S6 provides the Ailaoshan species lists, with the matched and unmatched OTUs.

## 4.2 Species richness

The rarefaction curve for the combined LSU and SSU dataset suggested total *detectable* species richness on the order of 100 species (Figure 2b). Additional replicates might therefore be expected to capture around 25% more species, but it would likely require around double the number of replicates in the present study to capture them fully. The rarefaction curves for the individual datasets illustrate the value of using multiple primers: the combined data set produced observed species richness comparable to the SSU data with around 450 replicates, and comparable to the LSU data with around 250 replicates.

Per patrol area, estimated median species richness was 18 and 22 in the LSU and SSU

**Table 2:** Top OTUs by estimated occupancy in the **(a)** LSU and **(b)** SSU datasets. Each OTU is ranked according to the estimated value for its occupancy parameter $\beta_{0i}$, shown here as Occupancy after conversion to a probability with a logit transformation. 95% Bayesian confidence intervals (BCIs) for occupancy are provided in parentheses. Taxonomic information and IUCN Red List category are based on classification generated by PROTAX. IUCN categories: LC = Least Concern; NT = Near Threatened; VU = Vulnerable; EN = Endangered. Supplementary File S2 provides a complete list of OTUs.

**(a) LSU**

| Rank | Scientific name | Common name | IUCN category | Occupancy (95% BCI) |
|---|---|---|---|---|
| 1 | *Bombina maxima* | Yunnan firebelly toad (大蹼铃蟾) | – | 0.658 (0.486 - 0.821) |
| 2 | *Bufo pageoti* | Tonkin toad (缅甸溪蟾) | NT | 0.655 (0.489 - 0.808) |
| 3 | *Capra hircus* | domestic goat (山羊) | – | 0.623 (0.450 - 0.800) |
| 4 | *Rhacophorus* sp1 | – | – | 0.585 (0.352 - 0.836) |
| 5 | *Bos taurus* | domestic cattle (黄牛) | – | 0.584 (0.462 - 0.704) |
| 6 | *Nanorana unculuanus* | Yunnan Asian frog (棘肛蛙) | EN | 0.475 (0.354 - 0.599) |
| 7 | *Tylototriton verrucosus* | Himalayan salamander (棕黑疣螈) | LC | 0.448 (0.185 - 0.810) |
| 8 | *Muntiacus muntjak* | southern red muntjac (赤麂) | LC | 0.434 (0.212 - 0.747) |
| 9 | *Nanorana yunnanensis* | Yunnan spiny frog (云南棘蛙) | EN | 0.405 (0.159 - 0.789) |
| 10 | *Prionailurus bengalensis* | leopard cat (豹猫) | LC | 0.377 (0.103 - 0.856) |

**(b) SSU**

| Rank | Scientific name | Common name | IUCN category | Occupancy (95% BCI) |
|---|---|---|---|---|
| 1 | *Tylototriton verrucosus* | Himalayan salamander (棕黑疣螈) | LC | 0.761 (0.460 - 0.961) |
| 2 | Megophryidae sp6 | – | – | 0.721 (0.337 - 0.960) |
| 3 | *Bufo pageoti* | Tonkin toad (缅甸溪蟾) | NT | 0.705 (0.522 - 0.864) |
| 4 | Megophryidae sp3 | – | – | 0.682 (0.461 - 0.870) |
| 5 | Leiothrichidae sp1 | – | – | 0.658 (0.363 - 0.929) |
| 6 | Megophryidae sp5 | – | – | 0.645 (0.485 - 0.808) |
| 7 | *Bos taurus* | domestic cattle (黄牛) | – | 0.624 (0.486 - 0.755) |
| 8 | *Capra hircus* | domestic goat (山羊) | – | 0.603 (0.441 - 0.779) |
| 9 | *Bombina maxima* | Yunnan firebelly toad (大蹼铃蟾) | – | 0.597 (0.445 - 0.757) |
| 10 | *Leptobrachium ailaonicum* | Ailao moustache toad (哀牢髭蟾) | NT | 0.597 (0.261 - 0.926) |

datasets, respectively, compared to observed median species richnesses of 3 and 4 species per patrol area (Figures 3a and 3b). Per replicate, observed median species were 1 and 2 in the LSU and SSU datasets, respectively, and the median numbers of replicates per patrol area were 3 and 4, respectively.

Almost half of all patrol areas had no observed species, either because they were not sampled, or because of inadequate labelling of samples (Figures 3c and 3d). Our occupancy model, however, provided species richness estimates for all patrol areas, both with and without observed values (Figures 3e and 3f). Both datasets indicated that species richness was highest in the southern third of the Ailaoshan Nature Reserve.

Both LSU and SSU datasets showed a positive relationship between community mean occupancy and elevation (Figures S3a and S3b). Distance to reserve boundary did not show a strong relationship with community mean occupancy, despite being retained in our final model selection (Figures S3c and S3d). Topographic position index, distance to road, and distance to stream were excluded at the model selection stage. Number of leeches was positively related to community mean detection in both datasets (Figures S3e and S3f).

There was good agreement on species richness between the LSU and SSU datasets. Observed species richness in the two datasets was positively correlated at the grain of both individual replicates ($r = 0.64$, $t_{616} = 21.2$, $p < 0.001$; Figure S4a) and patrol areas ($r = 0.89$, $t_{120} = 20.8$, $p < 0.001$; Figure S4c). The observed data showed the effect of sampling effort: replicates with more leeches tended to contain more species (Figure S4b), as did patrol areas with more replicates (Figure S4d). Estimated species richness was also highly correlated between the two datasets (Figure S4e), but the effect of sampling effort was no longer apparent, since our model compensates for variation in both leech quantity and number of replicates (Figure S4f).

## 4.3 Community composition

Almost half of the variation among sites in predicted occupancy (i.e. the $z_{i,j}$s) was captured by the first principal component axis (LSU: 49.8% variation explained; SSU: 45.6% variation explained). Comparing the two datasets revealed significant co-inertia (RV coefficient [19] 0.888, $p \leq 0.001$), indicating that the two datasets yield similar pictures of variation in composition among sites. Redundancy analysis showed that most of this variation is driven by a combination of distance to reserve boundary, elevation, and distance to road, with a positive correlation between the latter two variables (Figures 4a and 4b). Distance to the nearest stream and topographic position index explained little of the variation. Practically, this meant that both our LSU and SSU models detected differences in community composition between the northern two-thirds and the southern one-third of the nature reserve (Figures 4c and 4d), with the southern one-third containing a larger amount of less accessible forest at higher elevation and/or farther from roads.

Among mammals over 10 kg, species with negative RDA1 scores included domestic cow (*B. taurus*), domestic sheep (*O. aries*), domestic goat (*C. hircus*), and muntjak (*Muntiacus muntjak*) (Figures S5 and S6); these species were therefore more likely to occur in forest closer to the reserve edge and/or at lower elevation and nearer to roads. In contrast, species such as tufted deer (*Elaphodus cephalophus*), sambar (*R. unicolor*), serow (*Capricornis milneedwardsii*), Asiatic black bear (*U. thibetanus*), and wild boar (*Sus scrofa*) had positive

RDA1 species scores and were thus more likely to occur in more central, higher-elevation forest (Figures 4e and 4f).

Among mammals below 10 kg, and birds, most species were also estimated to have greater occupancy in more central, higher-elevation forest, including the Asian red-cheeked squirrel (*Dremomys rufigenis*) and the leopard cat (*Prionailurus bengalensis*) (Figures S5 and S6). However, some small-mammal species including the Himalayan field rat (*Rattus nitidus*) fared better in reserve-edge, lower-elevation forest. Amphibians showed a mix of responses, with some species such as the near-threatened Tonkin toad (*Bufo pageoti*) more common in less accessible areas at higher elevations, but others such as the fire-bellied toad (*Bombina maxima*) more common in reserve-edge, lower-elevation forest.

# 5 Discussion

Here we have demonstrated that metabarcoding of iDNA from bulk-collected leeches is an effective way to survey vertebrate biodiversity, requiring untrained forest rangers only 2-3 months to capture distribution information on mammals and amphibians, and to a much lesser extent, birds and reptiles, across a topographically challenging, 677 $km^2$ nature reserve, with a mean sampling unit of 3.9 $km^2$ (Figure 1). Our study is the most granular and broadest-scale biodiversity survey using iDNA to date, and the results show that the reserve does provide protected space for vertebrate species of high conservation value, mostly in its core area. However, the results also highlight the vulnerability of the rest of the reserve to degradation arising from human activity (i.e. farming, livestock, and possibly poaching) (Figures 3, 4). This study thus provides a vertebrate biodiversity baseline for the Ailaoshan Nature Reserve, and future surveys can test for change in occupancy as a proxy for effectiveness, as argued by Beaudrot *et al.* [5]. In contrast, the most recent camera-trap study in Ailaoshan [28], run by researchers, surveyed two patrol areas and detected 10 mammal species and 10 bird species and thus could not measure reserve effectiveness. Our study also functions as a progress report on the use of iDNA in a real-world management setting and highlights areas for improvement in iDNA monitoring going forward.

## 5.1 Vertebrate biodiversity in Ailaoshan

Our iDNA survey recovered 86 species of mammals, amphibians, birds, and reptiles, plus humans. Many replicates contained evidence of common wildlife species, or domesticated taxa, including cattle. The dataset also included many less common taxa that would have not been detected without targeted traditional surveys, including 15 species recognized by the IUCN as near-threatened or threatened (e.g. Asiatic black bears, *U. thibetanus*; sambar, *R. unicolor*; stump-tailed macaques, *M. arctoides*).

Occupancy modeling indicated that vertebrate species richness was greatest in the higher-elevation portions of Ailaoshan (Figure S3). This contrasts with most studies of altitudinal species-richness gradients, which find a hump-shaped or decreasing relationship between elevation and richness [51]. Our result likely reflects lower levels of anthropogenic disturbance in the higher, less-accessible parts of the park, and may involve species being driven from

their preferred lower-elevation areas into less suitable habitat to escape human encroachment [67].

Both elevation and distance to reserve edge were important predictors of vertebrate community composition (plus distance to roads, but this is likely an effect of its correlation with elevation) (Figure 4ab). Examining the distribution of individual taxa revealed that many species, especially birds and small mammals, had higher occupancy at higher elevation and in the reserve interior (Figures 4ef, S5, S6). These species include several that are IUCN near-threatened or threatened species: stump-tailed macaque (*Macaca arctoides*), tufted deer (*E. cephalophus*), sambar (*R. unicolor*), serow (*C. milneedwardsii*), and Asiatic black bear (*U. thibetanus*). Some or all of these species are likely sensitive to habitat alteration along the reserve edge, to poaching, to competition with domestic animals (e.g. most ungulates), and/or may be prone to human-wildlife conflict (e.g. Asiatic black bear) in degraded areas where livestock use mixes with conservation areas. In contrast, a few wild species, like the southern red muntjak (*M. muntjak*), appear to do better in reserve-edge areas.

## 5.2 Using iDNA for biodiversity monitoring

Two key benefits of leech-iDNA surveys are (A) the ability to survey across a wider range of vertebrate taxa and body sizes than is possible for any other method (here, mammals, amphibians, and phasianid birds) and (B) the feasibility of contracting large numbers of minimally trained collectors. Both benefits result in time and cost savings, and the latter benefit, in our estimation, finally makes it operationally feasible to survey the entire Ailaoshan reserve on a regular basis. However, these benefits are partially offset by a greater laboratory workload (which could be mitigated in part by automation), challenges over the design of sampling incentives, and iDNA-specific sampling error and biases.

*Design of sampling incentives.* Sampling with the assistance of forest rangers proved to be a feasible and cost-effective way to collect leeches from across the entire reserve with good levels of replication. This is despite the fact that the rangers were hired locally from neighbouring villages surrounding the park and did not report to a central location. Instead, forestry officials brought boxes of hip packs to groups of rangers around the park in June-July 2016, issued instructions verbally, and retrieved the packs after September. Provisioning the packs with tubes distributed over multiple self-sealing bags naturally enforced replicate sampling without the need for much training [62]. This approach also made it feasible for replicates from each patrol area to be collected at a single time point, removing the possibility that occupancy might change between temporal replicates [1] (although, for logistical reasons, collections from different patrol areas took place over a period of three months).

Collection of metadata, however, was less successful, as many samples had information on the collecting ranger but not the patrol area. In future sampling, metadata submission could be made a condition of payment, and a subset of senior rangers should be trained on metadata collection. A longer-range possibility is to outfit rangers with a GPS app on their cell phones. That said, our occupancy modeling framework deals well with missing data, and we must be wary of creating incentives to fabricate information. For instance, we decided against paying on a per-leech or per-tube basis, because this could incentivize rangers to collect outside the reserve. We found that a fixed payment, plus paying a small

bonus for at least one leech collected, worked well, and we have since used this structure in other rounds of leech sampling. We do expect to need to increase future payments.

*Error and bias in iDNA sampling.* There are several potential sources of error in our study. One is the lag time between a leech's last feed and our sampling, which can be up to a few months [63]). While the retention of blood meal DNA facilitates detection of animals, it also means that detected DNA does not necessarily reflect current occupancy. Animal hosts may leave the patrol area between the feeding event and our sampling, and even leeches may disperse widely if carried on hosts such as birds that can travel long distances [15], potentially blurring the spatial resolution of our results. Our data show that the leeches we collected mostly feed on hosts that probably remain within one patrol area or, at most, move between adjacent areas (e.g. frogs), so our broad conclusions about the overall distributions of wild and domesticated species in Ailaoshan (Figures 3, 4) are unlikely to be seriously affected. Further, the collection of all replicate samples from a location at essentially the same time limits the potential for leech or host movements to violate the site-occupancy model assumption that species occupancy remains constant across replicates (i.e., the 'population closure' assumption [55, 62]). Nonetheless, the lag time restricts the suitability of leech iDNA for detecting very rapid change, occurring on the order of a few months, though longer term trends should still be detectable [62].

A second source of error is the possibility of systematic differences across patrol areas in leech communities, coupled with differing diet preferences among leech species, which could produce spurious spatial patterns of occupancy. For instance, if leech species differ with elevation (which we did not include as a detection covariate), and high elevation leech species tend to feed more on frogs and less on cattle, this would give the appearance of change in these species' occupancy with elevation. The large number of leeches in our sample made it infeasible to identify them individually, although the geographic location of our field site and the uniform morphology of the leeches is consistent with all the leeches being in the genus *Haemadipsa* [66], the taxonomy of which is poorly resolved. *Haemadipsa* are known to feed widely [66, 61], probably because they are opportunistic, sit-and-wait parasites, and published evidence for dietary differences across species is at most only suggestive. Tessler *et al.*'s [66] diet study of 750 leeches across 15 DNA-barcode clades of *Haemadipsa* reported that "no pattern was evident between leeches of a given clade and their prey," given that multiple clades were each found to have fed on birds and on multiple mammalian orders. Even for the two most different *Haemadipsa* species, brown and tiger leeches, only mild differences in detection probabilities have been reported [1, 18]. Given this evidence, we conclude tentatively that differences in leech diets are unlikely to account for any of the major results in this study. Although it would be possible to metabarcode a leech marker from the same DNA extracts for use as a detection covariate in our occupancy model, it is likely that leech presence/absence information would be of limited value for samples consisting of many pooled leeches, and unlikely to justify the additional sequencing costs.

A third potential source of error is the choice of PCR primers and genetic markers, which may prevent some taxa from being detected even when their DNA is present, e.g. due to non-amplification at the PCR stage. We addressed this problem in part by using data from two marker genes. More than half of the OTUs were detected by both markers, and high correlation in species richness and co-inertia of community composition between the datasets suggested that broad ecological inferences would not have been strongly affected

had either marker been chosen by itself (Figures 3, 4). On the other hand, the primers clearly differed in their ability to amplify DNA from certain species. For example, we detected the stump-tailed macaque (*M. arctoides*) in the LSU dataset in three different patrol areas, with 2,700, 170,066, and 245,477 reads. But there was no obvious SSU equivalent, with no OTUs (other than humans) assigned to the order Primates in the SSU dataset. Of course, we do not know what additional taxa would have been detected by yet other primers, and ultimately we must be careful to restrict inferences from our model to taxa that we know can be detected. In the future, a metagenomic approach that uses baits to enrich target regions should be tested as an alternative to PCR [42], one benefit being the opportunity to use the cytochrome *c* oxidase I barcode, for which databases are well populated [29].

Finally, the use of leech iDNA will naturally exclude taxa that are not well represented in leech blood meals. Studies have reported lower iDNA detection rates for many species compared to camera trapping, though iDNA appears to be better at detecting smaller-bodied species of mammal [63, 53, 72, 26, 67], and, in our study, amphibians. With sufficiently large samples, taxa that are present infrequently may still be detected, and their low detection rates accounted for using site-occupancy modeling. However any taxa that are not detected cannot be modeled. When leech sampling is the rate-limiting step, Abrams *et al.* [1] recommend using leech-iDNA to supplement camera-trap data and increase confidence in occupancy estimates. For instance, Tilker *et al.* [67] recently ran a camera-trap survey at 139 stations (17,393 trap-nights) over five protected areas in Vietnam and Laos, spanning 900 $km^2$, and supplemented the camera data with iDNA from 2,043 leeches from 93 of the stations. The camera-trap data were limited to 23 terrestrial mammal species, with squirrels and large rodents being the smallest organisms detected, and generally produced more species detections. However, leech iDNA provided the sole detections of marbled cat (*Pardofelis marmorata*) and doubled the detections of Owston's civet (*Chrotogale owstoni*) and Asian black bear (*U. thibetanus*). Similar to our results, Tilker *et al.* [67] reported that wild mammal species occupancy increased with remoteness and elevation. However, as Gogarten *et al.* [26] have found, camera-trap and fly-iDNA data classify habitats similarly, even when the two monitoring methods detect largely different communities (only 6% to 43% of species were found by both methods in any given location). This suggests that different components of the mammal community contain similar ecological information, a result that has also been found when comparing metabarcoded insects to visual bird and mammal surveys [30]. In our case, the large sample size and the considerably wider taxonomic range than possibly achievable using camera traps alone, allowed us to parameterise an occupancy model using only leech-iDNA.

*Site-occupancy modeling.* Our statistical modeling approach worked well to identify correlates of detection and occupancy at the level of the community as well as individual species. Most taxa were detected infrequently, and individually, they provided little insight into detection and occupancy rates, as it is difficult to distinguish low detection rates (i.e. crypsis) from low occupancy (i.e. rarity). However, by integrating these infrequent detections into community models of occupancy and detection, and sharing information across species and patrol areas, the entire dataset was able to produce a broad picture of vertebrate diversity across Ailaoshan. This modeling approach dealt well with missing data, demonstrating the usefulness of occupancy models in a Bayesian framework for dealing with the imperfect datasets that are to be expected with surveys across broad areas and relying on limited resources.

While in this study we focused our modeling attention on correcting for false negatives, false positives are also possible, e.g. due to lab contamination or taxonomic misassignment, and may cause serious bias in the estimation of biodiversity [56]. Hierarchical models may, in principle, also be used to correct for false positives, though they have in practice proven challenging to estimate without additional information about the false positive detection process [47]. Although false negatives are likely to be a more serious problem in our dataset than false positives, it may be valuable for future work to explore corrections for false positives through modeling – especially those arising from incorrect identifications, in which details of the taxonomic assignment methods might help inform statistical correction.

### 5.3 iDNA: a promising biodiversity monitoring tool

Many protected areas are under-resourced and under-staffed [13], and costly monitoring activities are rarely prioritized, making it difficult to assess the effectiveness of reserves for protecting biodiversity. iDNA metabarcoding could help relieve some of these constraints, by making it possible to achieve both broad-scale coverage and fine spatio-temporal resolution. To assess the effectiveness of Ailaoshan nature reserve at reaching its policy and management targets, and to identify changes in species richness and patterns of occurrence of species, future sampling can now rely on the baseline established by this study. Our work can also guide future monitoring to identify underlying sources of environmental change, anthropogenic influences, and overall wildlife community dynamics. We also recommend using these results to guide the design of targeted camera-trap and bioacoustic monitoring campaigns inside Ailaoshan, both to independently test our results with species that are amenable to being recorded with these methods (e.g. mammals, ground-dwelling birds), and to improve the accuracy of occupancy and detection estimates [1]. These monitoring methods could also be used for estimating population sizes and population trends for some species using an occupancy modeling framework [58, 57, 73].

As we prepare to replace the Aichi Biodiversity Targets with a new post-2020 framework, there has been a call to focus on measured conservation outcomes such as occupancy, abundance, and population trends, in addition to current targets on area and representativeness of protected areas [69]. Implementing outcome-focused biodiversity targets capable of detecting trends will require technological innovation so that biodiversity can be monitored repeatedly and at a fine scale over large spatial extents [16]. Our study provides an example of how the information found in environmental DNA sources can be feasibly scaled up and interpreted in a useful way, thus complementing biodiversity information that is being revealed by technological innovation more broadly [10].

## 6 Funding and Acknowledgments

We thank Jiang Xuelong, Yang Xiaojun, Che Jing, Li Xueyou, Chen Hongman, and Wu Fei for Ailaoshan species lists, and Michael Tessler and Mark Siddall for information on leech distributions. CCMB, YHL, ZYW, DWY, and NEP were supported by the Harvard Global Institute. CLH and QZW were supported by Research and Application Demonstration on Key Technology of Primary Forest Resources Investigation and Monitoring in Yunnan Province (2013CA004). YQJ, JXW, LW, CYW, CYY, CCYX, and DWY were supported

by the National Natural Science Foundation of China (41661144002, 31670536, 31400470, 31500305, 31872963); the Key Research Program of Frontier Sciences, Chinese Academy of Sciences (QYZDY-SSW-SMC024); the Bureau of International Cooperation (GJHZ1754); the Strategic Priority Research Program, Chinese Academy of Sciences (XDA20050202, XDB31000000); the Ministry of Science and Technology of China (2012FY110800); and the Biodiversity Investigation, Observation and Assessment Program (2019-2023), Ministry of Ecology and Environment of China (8-2-3-4-11). DWY was also supported by a Leverhulme Trust Research Fellowship. VDP was supported by the Ohio University Department of Biological Sciences and the Sustainability Studies Theme. The computations in this paper were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University.

# References

[1] J. F. Abrams, L. Hörig, R. Brozovic, J. Axtner, A. Crampton-Platt, *et al.* "Shifting up a gear with iDNA: from mammal detection events to standardized surveys". *Journal of Applied Ecology* 18.3 (2019), pp. 511–512.

[2] T. M. Aide, C. Corrada-Bravo, M. Campos-Cerqueira, C. Milan, G. Vega, *et al.* "Real-time bioacoustics monitoring and automated species identification". *PeerJ* 1 (2013), e103.

[3] I. group of Ailaoshan Nature Reserve. *Comprehensive Survey of Ailaoshan Nature Reserve*. Kunming, Yunnan: Yunnan Ethnic Press, 1988.

[4] J. Axtner, A. Crampton-Platt, L. A. Hörig, A. Mohamed, C. C. Y. Xu, *et al.* "An efficient and robust laboratory workflow and tetrapod database for larger scale environmental DNA studies". *GigaScience* 8.4 (Apr. 2019), pp. 646–17.

[5] L. Beaudrot, J. A. Ahumada, T. O'Brien, P. Alvarez-Loayza, K. Boekee, *et al.* "Standardized Assessment of Biodiversity Trends in Tropical Forest Protected Areas: The End is not in Sight". *PLOS Biology* 14.1 (Jan. 2016), e1002357.

[6] C. on Biological Diversity. *Aichi Biodiversity Targets*. 2010. URL: %5Curl%7Bhttps://www.cbd.int/sp/targets%7D (visited on ).

[7] K. Bohmann, A. Evans, M. T. P. Gilbert, G. R. Carvalho, S. Creer, *et al.* "Environmental DNA for wildlife biology and biodiversity monitoring". *Trends in Ecology and Evolution* 29.6 (2014), pp. 358–367. DOI: 10.1016/j.tree.2014.04.003.

[8] K. Bohmann, I. B. Schnell, and M. T. P. Gilbert. "When bugs reveal biodiversity". *Molecular Ecology* 22.4 (Feb. 2013), pp. 909–911.

[9] B. A. Bryan, L. Gao, Y. Q. Ye, X. F. Sun, J. D. Connor, *et al.* "China's response to a national land-system sustainability emergency". *Nature* 559.7713 (2018), pp. 193–204. DOI: 10.1038/s41586-018-0280-2.

[10] A. Bush, R. Sollmann, A. Wilting, K. Bohmann, B. Cole, *et al.* "Connecting Earth observation to high-throughput biodiversity data". *Nature Ecology and Evolution* 1.7 (2017). DOI: 10.1038/s41559-017-0176.

[11] S. Calvignac-Spencer, F. H. Leendertz, M. T. P. Gilbert, and G. Schubert. "An invertebrate stomach's view on vertebrate ecology". *BioEssays* 35.11 (Nov. 2013), pp. 1004–1013.

[12] S. Chamberlain. *rredlist: 'IUCN' Red List Client*. R package version 0.5.0. 2018. URL: https://CRAN.R-project.org/package=rredlist.

[13] L. Coad, J. E. Watson, J. Geldmann, N. D. Burgess, F. Leverington, *et al.* "Widespread shortfalls in protected area resourcing undermine efforts to conserve biodiversity". *Frontiers in Ecology and the Environment* 17.5 (May 2019), pp. 259–264.

[14] G. Csárdi and T. Nepusz. "The igraph software package for complex network research". *InterJournal* Complex Systems (2006), p. 1695. URL: http://igraph.org.

[15] R. W. Davies, L. R. Linton, and F. J. Wrona. "Passive Dispersal of Four Species of Freshwater Leeches (Hirudinoidea) by Ducks". *Freshwater Invertebrate Biology* 1.4 (1982), pp. 40–44. DOI: 10.2307/1467140.

[16] T. Dietz, E. Ostrom, and P. C. Stern. "The Struggle to Govern the Commons". *Science* 302.5652 (Dec. 2003), pp. 1907–1912.

[17] R. M. Dorazio, J. A. Royle, B. Soderstrom, and A. Glimskar. "Estimating species richness and accumulation by modeling species occurrence and detectability". *Ecology* 87.4 (2006), pp. 842–854. DOI: 10.1890/0012-9658(2006)87[842:Esraab]2.0.Co;2.

[18] R. Drinkwater, I. B. Schnell, K. Bohmann, H. Bernard, G. Veron, *et al.* "Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo". *Molecular Ecology Resources* 19.1 (2019), pp. 105–117. DOI: 10.1111/1755-0998.12943.

[19] Y. Escoufier. "Le Traitement des Variables Vectorielles". *Biometrics* 29.4 (1973), pp. 751–760. DOI: 10.2307/2529140.

[20] P. J. Ferraro, T. Uchida, and J. M. Conrad. "Price Premiums for Eco-friendly Commodities: Are 'Green' Markets the Best Way to Protect Endangered Ecosystems?" *Environmental and Resource Economics* 32.3 (Nov. 2005), pp. 419–438.

[21] T. G. Frøslev, R. Kjøller, H. H. Bruun, R. Ejrnæs, A. K. Brunbjerg, *et al.* "Algorithm for post-clustering curation of DNA amplicon data yields reliable biodiversity estimates". *Nature Communications* 8 (2017). DOI: 10.1038/s41467-017-01312-x.

[22] J. Geldmann, M. Barnes, L. Coad, I. D. Craigie, M. Hockings, *et al.* "Effectiveness of terrestrial protected areas in reducing habitat loss and population declines". *Biological Conservation* 161 (May 2013), pp. 230–238.

[23] C. P. D. T. Gillett, A. J. Johnson, I. Barr, and J. Hulcr. "Metagenomic sequencing of dung beetle intestinal contents directly detects and identifies mammalian fauna". *bioRxiv* (Sept. 2016), p. 074849.

[24] P. Glover-Kapfer, C. A. Soto-Navarro, and O. R. Wearn. "Camera-trapping version 3.0: current constraints and future priorities for development". *Remote Sensing in Ecology and Conservation* 5.3 (Dec. 2018), pp. 209–223.

[25] J. F. Gogarten, A. Düx, B. Mubemba, K. Pléh, C. Hoffmann, *et al.* "Tropical rainforest flies carrying pathogens form stable associations with social nonhuman primates". *Molecular Ecology* 215.3 (July 2019), pp. 403–17.

[26] J. F. Gogarten, C. Hoffmann, M. Arandjelovic, A. Sachse, K. Merkel, *et al.* "Fly-derived DNA and camera traps are complementary tools for assessing mammalian biodiversity". *Environmental DNA* 9.1 (Oct. 2019), pp. 79–14.

[27] A. Guisan, S. B. Weiss, and A. D. Weiss. "GLM versus CCA spatial modeling of plant species distribution". *Plant Ecology* 143.1 (1999), pp. 107–122. DOI: 10.1023/a:1009841519580.

[28] X. He, K. Luo, Z. Y. Lu, and L. X. Lin. "Preliminary camera-trapping survey on wild mammals and birds in Ailaoshan National Nature Reserve, Yunnan Province, China". *Acta Theriologica Sinica* 38.3 (2018), pp. 318–322.

[29] P. D. N. Hebert, P. M. Hollingsworth, and M. Hajibabaei. "From writing to reading the encyclopedia of life". *Philosophical Transactions of the Royal Society B: Biological Sciences* 371.1702 (2016), p. 20150321. DOI: 10.1098/rstb.2015.0321.

[30] Y. Q. Ji, L. Ashton, S. M. Pedley, D. P. Edwards, Y. Tang, *et al.* "Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding". *Ecology Letters* 16.10 (2013), pp. 1245–1257. DOI: 10.1111/ele.12162.

[31] K. E. Jones, J. Bielby, M. Cardillo, S. A. Fritz, J. O'Dell, *et al.* "PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals". *Ecology* 90.9 (2009), pp. 2648–2648. DOI: 10.1890/08-1494.1.

[32] A. Kocher, B. de Thoisy, F. Catzeflis, S. Valière, A.-L. Bañuls, *et al.* "iDNA screening: Disease vectors as vertebrate samplers". *Molecular Ecology* 26.22 (2017), pp. 6478–6486. DOI: 10.1111/mec.14362.

[33] D. Kong, F. Wu, P. Shan, J. Gao, D. Yan, *et al.* "Status and distribution changes of the endangered Green Peafowl (Pavo muticus) in China over the past three decades (1990s–2017)". *Avian Research* 9.1 (May 2018), p. 427.

[34] W. F. Laurance, D. C. Useche, J. Rendeiro, M. Kalka, C. J. A. Bradshaw, *et al.* "Averting biodiversity collapse in tropical forest protected areas". *Nature* 489.7415 (Sept. 2012), pp. 290–294.

[35] L. Lellouch, S. Pavoine, F. Jiguet, H. Glotin, and J. Sueur. "Monitoring temporal change of bird communities with dissimilarity acoustic indices". *Methods in Ecology and Evolution* 5.6 (June 2014), pp. 495–505. DOI: 10.1111/2041-210X.12178.

[36] M. Leray, J. Y. Yang, C. P. Meyer, S. C. Mills, N. Agudelo, *et al.* "A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents". *Frontiers in Zoology* 10 (June 2013), p. 34.

[37] G. Li, X. Yang, H. Zhang, and W. Li. "Population and distribution of western black crested gibbon (*Nomascus concolor*) at Ailao Mountain, Xinping, Yunnan". *Zoological Research* 32.6 (2011), pp. 675–683. DOI: 10.3724/SP.J.1141.2011.06675.

[38] H. Li. "The Distribution and Perniciousness of *Rhabdophis subminiatus* at Ailaoshan in Xinping County". *Hubei Agricultural Sciences* 50.4 (2011), pp. 800–801. DOI: CNKI:SUN:HBNY.0.2011-04-051.

[39] H. Li, X. Zhang, D. Rao, and H. Zhang. "Research on the Reptiles Diversity in the East of Xinping Ailaoshan Nature Reserve". *Hubei Agricultural Sciences* 51.16 (2012), pp. 3557–3559. DOI: 10.3969/j.issn.0439-8114.2012.16.039.

[40] H. Li, H. Zhu, L. Wang, and J. Liu. "Biological characteristics and protection of *Tylototriton shanjing* at Mount Ailao in Xinping". *Journal of Chongqing College of Education* 23.6 (2010), pp. 16–18. DOI: 10.3969/j.issn.1008-6390.2010.06.004.

[41] W. A. Link and J. R. Sauer. "Extremes in Ecology: Avoiding the Misleading Effects of Sampling Variation in Summary Analyses". *Ecology* 77.5 (1996), pp. 1633–1640. DOI: 10.2307/2265557.

[42] S. Liu, X. Wang, L. Xie, M. Tan, Z. Li, *et al.* "Mitochondrial capture enriches mito-DNA 100 fold, enabling PCR-free mitogenomics biodiversity analysis". *Molecular Ecology Resources* 16.2 (Mar. 2016), pp. 470–479.

[43] W. S. Luo, S. Y. Zhao, Z. Q. Luo, and Q. Wang. "Population and distribution of *Nomascus concolor* in Jingdong jurisdiction of Ailaoshan National Nature Reserve". *Sichuan Journal of Zoology* 26.3 (2007), pp. 600–603.

[44] D. I. MacKenzie, J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, *et al.* "Estimating site occupancy rates when detection probabilities are less than one". *Ecology* 83.8 (2002), pp. 2248–2255.

[45] F. Mahe, T. Rognes, C. Quince, C. de Vargas, and M. Dunthorn. "Swarm v2: highly-scalable and high-resolution amplicon clustering". *PeerJ* 3 (2015). DOI: 10.7717/peerj.1420.

[46] P. D. Meek, G. A. Ballard, J. Sparkes, M. Robinson, B. Nesbitt, *et al.* "Camera trap theft and vandalism: occurrence, cost, prevention and implications for wildlife research and management". *Remote Sensing in Ecology and Conservation* 20 (Oct. 2018), pp. 222–9.

[47] D. A. Miller, J. D. Nichols, B. T. McClintock, E. H. C. Grant, L. L. Bailey, *et al.* "Improving occupancy estimation when two types of observational error occur: non-detection and species misidentification". *Ecology* 92.7 (2011), pp. 1422–1428. DOI: 10.1890/10-1396.1.

[48] F. Mohd Salleh, J. Ramos-Madrigal, F. Peñaloza, S. Liu, S. S. Mikkel-Holger, *et al.* "An expanded mammal mitogenome dataset from Southeast Asia". *GigaScience* 6.8 (Aug. 2017), pp. 1–8.

[49] T. G. O'Brien, J. E. M. Baillie, L. Krueger, and M. Cuke. "The Wildlife Picture Index: monitoring top trophic levels". *Animal Conservation* 13.4 (Aug. 2010), pp. 335–343. DOI: 10.1111/j.1469-1795.2010.00357.x.

[50] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2019. URL: https://www.R-project.org/.

[51] C. Rahbek. "The role of spatial scale and the perception of large-scale species-richness patterns". *Ecology Letters* 8.2 (2005), pp. 224–239. DOI: 10.1111/j.1461-0248.2004.00701.x.

[52] G. Ren, S. S. Young, L. Wang, W. Wang, Y. Long, *et al.* "Effectiveness of China's National Forest Protection Program and nature reserves". *Conservation Biology* 29.5 (2015), pp. 1368–1377. DOI: 10.1111/cobi.12561.

[53] T. W. Rodgers, C. C. Y. Xu, J. Giacalone, K. M. Kapheim, K. Saltonstall, *et al.* "Carrion fly-derived DNA metabarcoding is an effective tool for mammal surveys: Evidence from a known tropical mammal community". *Molecular Ecology Resources* 17.6 (Nov. 2017), e133–e145.

[54] T. Rognes, T. Flouri, B. Nichols, C. Quince, and F. Mahé. "VSEARCH: a versatile open source tool for metagenomics". *PeerJ* 4 (2016), e2584.

[55] C. T. Rota, R. J. Fletcher Jr, R. M. Dorazio, and M. G. Betts. "Occupancy estimation and the closure assumption". *Journal of Applied Ecology* 46.6 (2009), pp. 1173–1181. DOI: 10.1111/j.1365-2664.2009.01734.x.

[56] J. A. Royle and W. A. Link. "Generalized site occupancy models allowing for false positive and false negative errors". *Ecology* 87.4 (2006), pp. 835–841. DOI: 10.1890/0012-9658(2006)87[835:Gsomaf]2.0.Co;2.

[57] J. A. Royle. "N-Mixture Models for Estimating Population Size from Spatially Replicated Counts". *Biometrics* 60.1 (2004), pp. 108–115. DOI: 10.1111/j.0006-341X.2004.00142.x.

[58] J. A. Royle and J. D. Nichols. "Estimating abundance from repeated presence–absence data or point counts". *Ecology* 84.3 (2003), pp. 777–790. DOI: 10.1890/0012-9658(2003)084[0777:EAFRPA]2.0.CO;2.

[59] D. B. Rubin. "Bayesianly Justifiable and Relevant Frequency Calculations for the Applied Statistician". *The Annals of Statistics* 12.4 (1984), pp. 1151–1172.

[60] I. B. Schnell, K. Bohmann, and M. T. P. Gilbert. "Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies". *Molecular Ecology Resources* 15.6 (2015), pp. 1289–1303. DOI: 10.1111/1755-0998.12402.

[61] I. B. Schnell, K. Bohmann, S. E. Schultze, S. R. Richter, D. C. Murray, *et al.* "Debugging diversity - a pan-continental exploration of the potential of terrestrial blood-feeding leeches as a vertebrate monitoring tool". *Molecular Ecology Resources* 18.6 (2018), pp. 1282–1298. DOI: 10.1111/1755-0998.12912.

[62] I. B. Schnell, R. Sollmann, S. Calvignac-Spencer, M. E. Siddall, D. W. Yu, *et al.* "iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool – prospects, pitfalls and avenues to be developed". *Frontiers in Zoology* 12.1 (Oct. 2015), p. 302.

[63] I. B. Schnell, P. F. Thomsen, N. Wilkinson, M. Rasmussen, L. R. Jensen, *et al.* "Screening mammal biodiversity using DNA from leeches". *Current Biology* 22.8 (2012), R262–R263. DOI: 10.1016/j.cub.2012.02.058.

[64] P. Somervuo, S. Koskela, J. Pennanen, R. H. Nilsson, and O. Ovaskainen. "Unbiased probabilistic taxonomic classification for DNA barcoding". *Bioinformatics* 32.19 (2016), pp. 2920–2927. DOI: 10.1093/bioinformatics/btw346.

[65] P. Somervuo, D. W. Yu, C. C. Y. Xu, Y. Q. Ji, J. Hultman, *et al.* "Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding". *Methods in Ecology and Evolution* 8.4 (2017), pp. 398–407. DOI: 10.1111/2041-210x.12721.

[66] M. Tessler, S. R. Weiskopf, L. B. S. and, and 2018. "Bloodlines: Mammals, leeches, and conservation in southern Asia". *Systematics and Biodiversity* 16.5 (2018), pp. 488–496.

[67] A. Tilker, J. F. Abrams, A. Nguyen, L. Hörig, J. Axtner, *et al.* "Identifying conservation priorities in a defaunated tropical biodiversity hotspot". *Diversity And Distributions* 10.1 (Jan. 2020), pp. 100331–100315.

[68] A. J. Tyre, B. Tenhumberg, S. A. Field, D. Niejalke, K. Parris, *et al.* "Improving precision and reducing bias in biological surveys: estimating false-negative error rates". *Ecological Applications* 13.6 (Dec. 2003), pp. 1790–1801.

[69] P. Visconti, S. H. M. Butchart, T. M. Brooks, P. F. Langhammer, D. Marnewick, *et al.* "Protected area targets post-2020". *Science* (2019).

[70] Z. J. Wang, C. Carpenter, and S. S. Young. "Bird distribution and conservation in the Ailao Mountains, Yunnan, China". *Biological Conservation* 92.1 (2000), pp. 45–57. DOI: 10.1016/S0006-3207(99)00058-0.

[71] J. E. M. Watson, E. S. Darling, O. Venter, M. Maron, J. Walston, *et al.* "Bolder science needed now for protected areas". *Conservation Biology* 30.2 (2016), pp. 243–248. DOI: 10.1111/cobi.12645.

[72] S. R. Weiskopf, K. P. McCarthy, M. Tessler, H. A. Rahman, J. L. McCarthy, *et al.* "Using terrestrial haematophagous leeches to enhance tropical biodiversity monitoring programmes in Bangladesh". *Journal of Applied Ecology* 55.4 (2018), pp. 2071–2081. DOI: 10.1111/1365-2664.13111.

[73] C. M. Wood, V. D. Popescu, H. Klinck, J. J. Keane, R. Gutiérrez, *et al.* "Detecting small changes in populations at landscape scales: a bioacoustic site-occupancy framework". *Ecological Indicators* 98 (2019), pp. 492–507. DOI: 10.1016/j.ecolind.2018.11.018.

[74] P. H. Wrege, E. D. Rowland, S. Keen, and Y. Shiu. "Acoustic monitoring for conservation in tropical forests: examples from forest elephants". *Methods in Ecology and Evolution* 8.10 (Feb. 2017), pp. 1292–1301.

[75] D. L. Wu and C. C. Luo. "Effect of human activity on community structure of small mammals in Ailao Mountain". *Zoological Research* 14.1 (1993), pp. 35–41.

[76] R. Wu, H. P. Possingham, G. Yu, T. Jin, J. Wang, *et al.* "Strengthening China's national biodiversity strategy to attain an ecological civilization". *Conservation Letters* 68.2 (June 2019), e12660.

[77] R. Wu, S. Zhang, D. W. Yu, P. Zhao, X. Li, *et al.* "Effectiveness of China's nature reserves in representing ecological diversity". *Frontiers in Ecology and Evolution* 9 (2011), pp. 383–389.

[78] W. H. Xu, Y. Xiao, J. J. Zhang, W. Yang, L. Zhang, *et al.* "Strengthening protected areas for biodiversity and ecosystem services in China". *PNAS* 114.7 (2017), pp. 1601–1606. DOI: 10.1073/pnas.1620503114.

[79] L. Yiming and D. S. Wilcove. "Threats to Vertebrate Species in China and the United States". *BioScience* 55.2 (2005), pp. 147–153. DOI: 10.1641/0006-3568(2005)055[0147:TTVSIC]2.0.CO;2.

[80] A. Zabel and B. Roe. "Optimal design of pro-conservation incentives". *Ecological Economics* 69 (2009), pp. 126–134.

[81] M. L. Zepeda-Mendoza, K. Bohmann, A. Carmona Baez, and M. T. Gilbert. "DAMe: a toolkit for the initial processing of datasets with PCR replicates of double-tagged amplicons for DNA metabarcoding analyses". *BMC Research Notes* 9 (2016), p. 255. DOI: 10.1186/s13104-016-2064-9. Downloaded 9 August 2019 from forked version at https://github.com/shyamsg/DAMe.

[82] K. Y. Zhang, Y. P. Zhang, Y. H. Liu, and Y. R. Li. "Vertical distribution characteristics of rainfall in the Ailao mountain". *Scientia Geographica Sinica* 14.2 (1994), pp. 144–150.

[83] Z. Q. Zhang. "Status quo of the biodiversity of Ailaoshan Nature Reserve and countermearsures for protection and management". *Forest Inventory and Planning* 32.3 (2007), pp. 68–70.

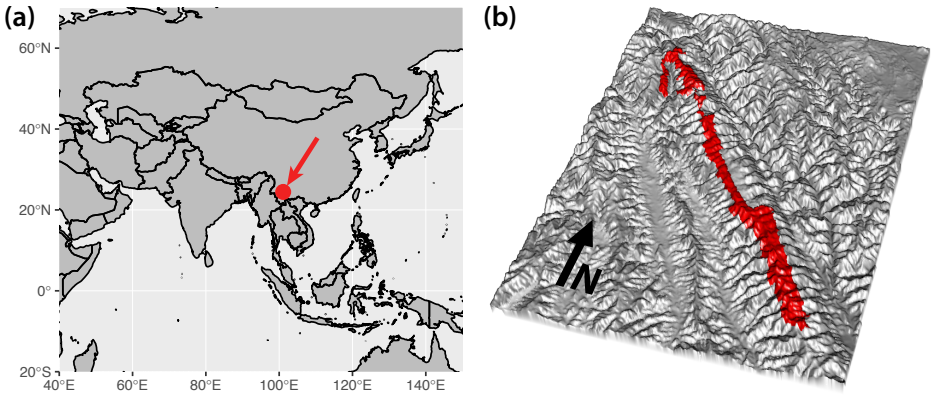


**Figure 1:** **(a)** Ailaoshan Nature Reserve is located in Yunnan Province, China. **(b)** Ailaoshan Nature Reserve runs northwest-to-southeast along a ridgeline for around 125 km, but averages just 6 km across along its entire length.

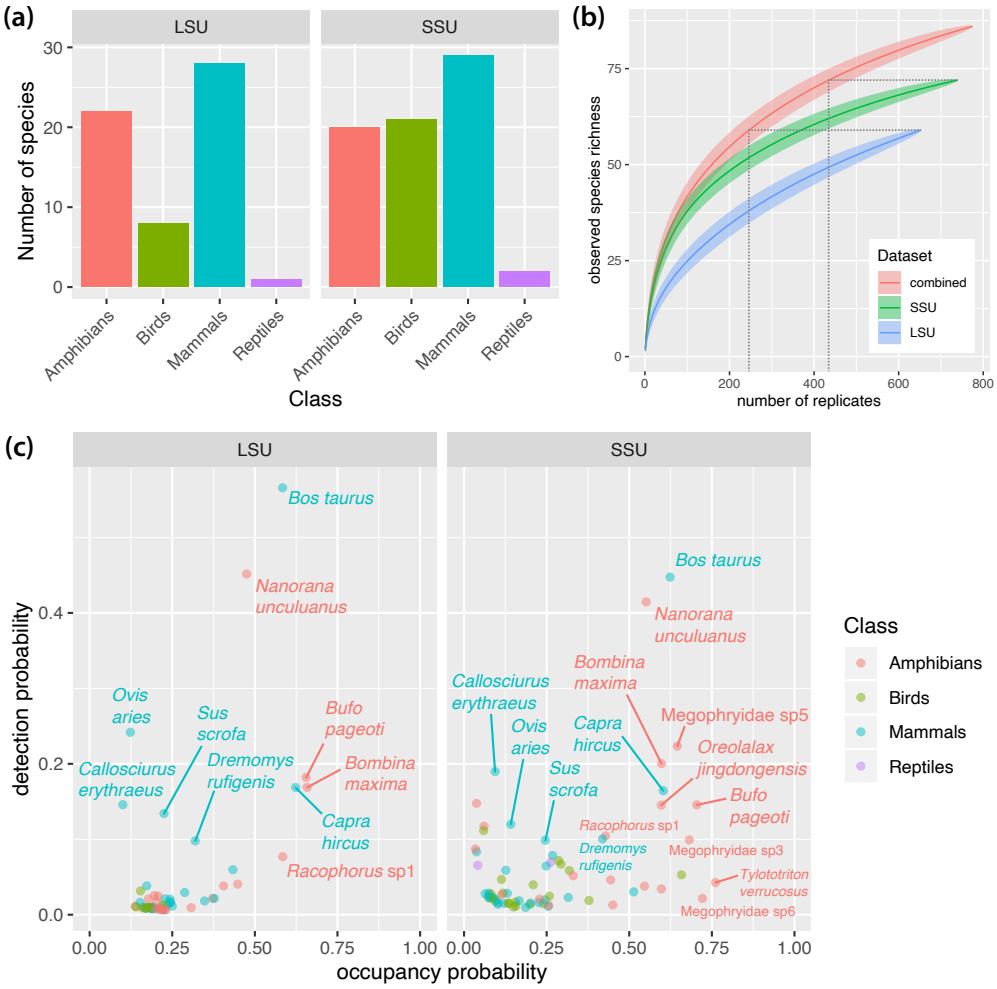


**Figure 2:** **(a)** Distribution of detected OTUs in each dataset by taxonomic class. **(b)** Rarefaction curves derived from resampling of replicates. Error bands show standard deviations. Dotted lines indicate the sampling depths required, using both LSU and SSU primers, to obtain the same total OTU richness observed with either LSU primers or SSU primers alone. **(c)** Estimated occupancy and detection estimates for each OTU. Taxa with low occupancy and detection probabilities are unlabelled for clarity; see Supplementary File S1 for full listing of results.

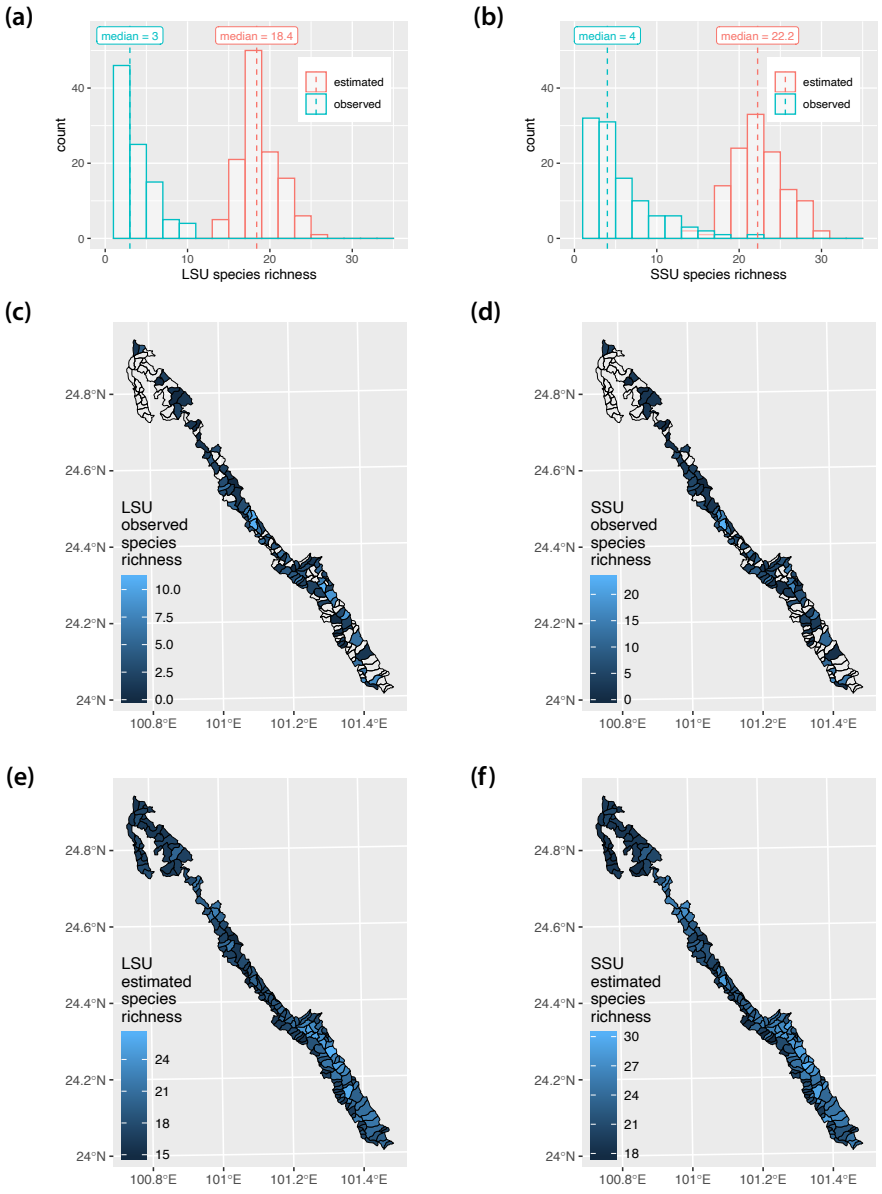


**Figure 3:** **(a,b)** Histograms of observed and estimated species richness per patrol area in the SSU and LSU datasets respectively. Dashed lines show median values. **(c,d)** Observed species richness in each patrol area in the SSU and LSU datasets respectively. Note missing data from approximately half of the patrol areas. Data with missing patrol area IDs are not represented in this figure, though they are incorporated in our occupancy model. **(e,f)** Estimated species richness for each patrol area in the SSU and LSU datasets respectively. Note that our occupancy model provides estimates for patrol areas with missing data, in addition to augmenting observed values to account for false negatives.

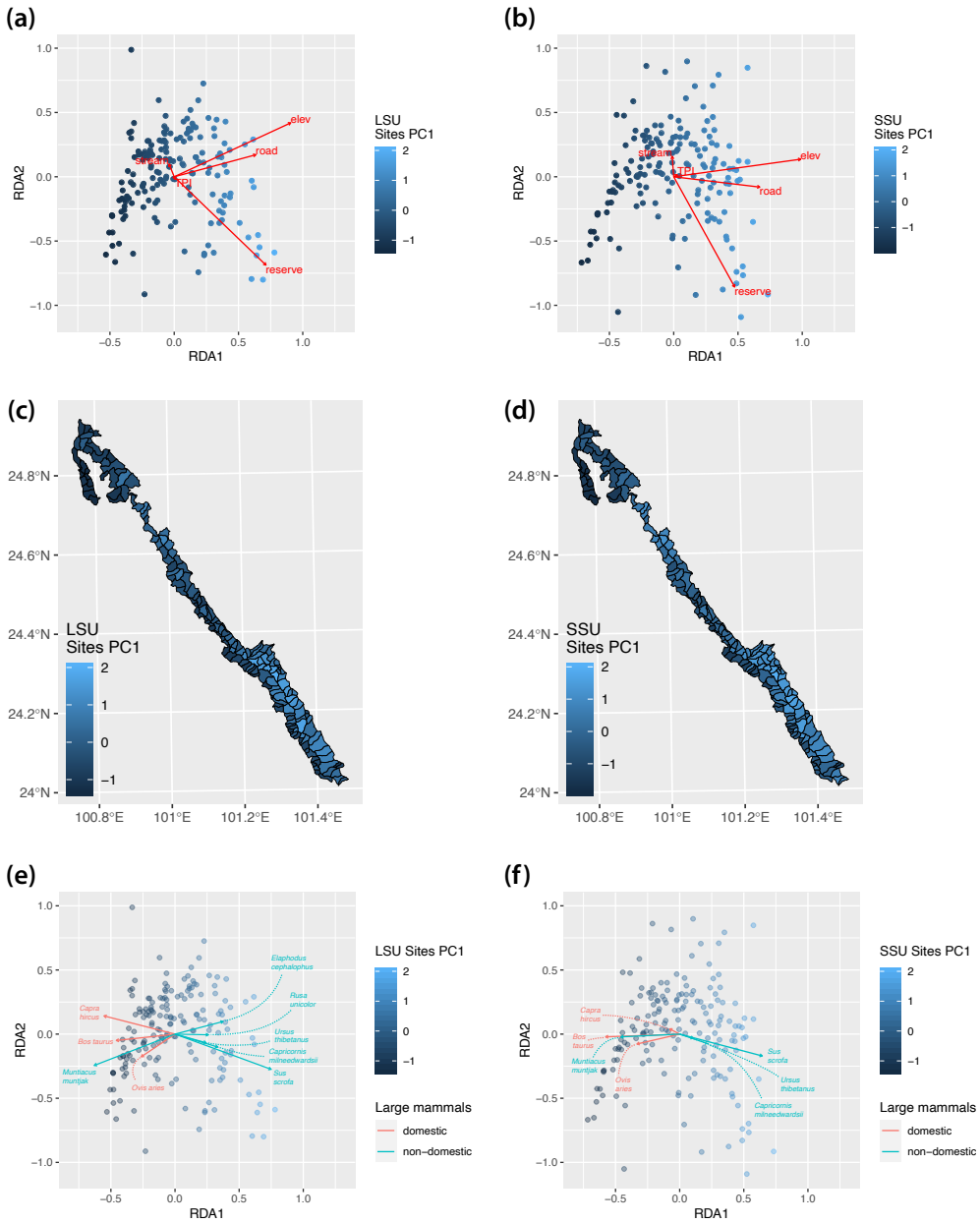


**Figure 4:** **(a,b)** Redundancy analysis biplots showing environmental covariates. Each point represents a single patrol area, colored according to position of the patrol area on the first principal component axis (PC1). **(c,d)** Site maps showing mapping onto PC1. Community composition differs most strongly between the northern two-thirds of the reserve and the southern one-third. **(e,f)** Redundancy analysis biplots showing mammal species >10 kg adult body mass, both domestic and non-domestic. Each point represents a single patrol area, colored according to position of the patrol area on the first principal component axis (PC1). Dotted lines connect some of the species names to their corresponding arrows.