SCIENTIFIC REPORTS

OPEN

Demonstration of the potential of environmental DNA as a tool for the detection of avian species

Masayuki Ushio [1,2], Koichi Murata[3,4], Tetsuya Sado[5], Isao Nishiumi[6], Masamichi Takeshita[7], Wataru Iwasaki [7] & Masaki Miya [5]

Birds play unique functional roles in the maintenance of ecosystems, such as pollination and seed dispersal, and thus monitoring bird species diversity is a first step towards avoiding undesirable consequences of anthropogenic impacts on bird communities. In the present study, we hypothesized that birds, regardless of their main habitats, must have frequent contact with water and that tissues that contain their DNA that persists in the environment (environmental DNA; eDNA) could be used to detect the presence of avian species. To this end, we applied a set of universal PCR primers (MiBird, a modified version of fish/mammal universal primers) for metabarcoding avian eDNA. We confirmed the versatility of MiBird primers by performing *in silico* analyses and by amplifying DNAs extracted from bird tissues. Analyses of water samples from zoo cages of birds with known species composition suggested that the use of MiBird primers combined with Illumina MiSeq could successfully detect avian species from water samples. Additionally, analysis of water samples collected from a natural pond detected five avian species common to the sampling areas. The present findings suggest that avian eDNA metabarcoding would be a complementary detection/identification tool in cases where visual census of bird species is difficult.

Environmental DNA (eDNA) is genetic material that persists in an environment and is derived from organisms living there, and researchers have recently been using eDNA to detect the presence of macro-organisms, particularly those living in aquatic/semiaquatic ecosystems[1–5]. For example, several fish species inhabiting a river can be detected by amplifying and sequencing DNA fragments extracted from water samples[6] by using methodologies such as quantitative PCR and eDNA metabarcoding. Quantitative PCR requires the design of species-specific PCR primers and enables quantitative measurements of eDNA of target species[3,4,7,8], while eDNA metabarcoding, which has been becoming a common methodology in eDNA studies, uses a universal primer set and high-throughput sequencer (e.g., Illumina MiSeq) to enable qualitative detection of eDNA of multiple species belonging to a target taxon[1,2,9–11] (but see ref.[12]).

Although earlier studies mainly focused on detecting fish/amphibian species (i.e., organisms that have close associations with water), recent studies have shown that eDNA can be used to detect a diverse group of animals, including mammals[9,13,14], reptiles[15] and arthropods[10]. Detecting the presence of animals is possible even if their habitats are terrestrial[9,10,13,14] because animals must have, in general, frequent opportunities to contact water in order to live. The findings of these recent studies imply that any organism, regardless of its main habitat, can potentially be detected by using eDNA if we can design suitable primers that enable amplification and identification of DNA fragments of target organisms and if we can collect appropriate media that contain eDNA.

Wild birds represent an important part of the biodiversity in ecosystems, and they play a unique role in the maintenance of ecosystem functions. For example, in forest ecosystems, birds can contribute to maintenance of the tree community by seed dispersal and pollination, and to the reduction of herbivory by predation upon insect herbivores[16–18]. However, recent increases in anthropogenic impacts on ecosystems, e.g., urbanization and habitat

[1]PRESTO, Japan Science and Technology Agency, Kawaguchi, 332-0012, Japan. [2]Center for Ecological Research, Kyoto University, Otsu, 520-2113, Japan. [3]Yokohama Zoological Gardens ZOORASIA, Kanagawa, 241-0001, Japan. [4]College of Bioresource Sciences, Nihon University, Kanagawa, 252-0880, Japan. [5]Natural History Museum and Institute, Chiba, 260-8682, Japan. [6]Department of Zoology, National Museum of Nature and Science, Tsukuba, Ibaraki, 305-0005, Japan. [7]Department of Biological Sciences, The University of Tokyo, Tokyo, 113-0032, Japan. Correspondence and requests for materials should be addressed to M.U. (email: ong8181@gmail.com) or M.M. (email: miya@chiba-muse.or.jp)

Primer name	Information
Primers for the first PCR[a,b] (with MiSeq sequencing primer and six random bases)	
MiBird-U-F (forward)	*ACACTCTTTCCCTACACGACGCTCTTCCGATCT* **NNNNNN** GGGTTGGTAAATCTTGTGCCAGC
MiBird-U-R (reverse)	*GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT* **NNNNNN** CATAGTGGGGTATCTAATCCCAGTTTG
Primers for the second PCR[c,d]	
2nd PCR-F	<u>AATGATACGGCGACCACCGAGATCTACAC</u> **XXXXXXXX** *ACACTCTTTCCCTACACGACGCTCTTCCGATCT*
2nd PCR-R	<u>CAAGCAGAAGACGGCATACGAGAT</u> **XXXXXXXX** *GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*

Table 1. Detailed information for MiBird primers. [a]Italic characters indicate the MiSeq sequencing primers. [b]Bold Ns indicate random bases to improve the quality of MiSeq sequencing. [c]Bold Xs indicate index sequences to identify each sample. [d]Underlined characters indicate P5/P7 adapter sequences for MiSeq sequencing.

fragmentation, drive substantial declines in bird species diversity[19,20], which could have impacts on the ecological functions of birds. Monitoring bird species diversity is required for detecting such declines, and such detection is necessary for avoiding undesirable consequences in ecosystem functions due to the loss of avian biodiversity. To monitor bird species diversity, visual census is one of the most common methods[21], and considering the higher visibility of birds than that of fish and forest mammals, visual census is generally a successful method. However, if an alternative method can overcome limitations of visual census, such as low visibility at night or in a dense forest, and eliminate the requirement for taxonomic identification skill under field conditions, that method could be complementarily used for monitoring bird species diversity.

In the present study, we tested the potential of eDNA as a tool for the detection of avian species. Previous eDNA surveys performed in marine ecosystems detected some avian species along with diverse fish/mammal species (2–4 avian species per study)[22–25], suggesting that more diverse avian species are potentially detectable using eDNA if suitable primers are designed. To this end, we modified a previously reported universal primer set for fish/mammals (MiFish/MiMammal[1,9]), such that the primer set accommodated bird-specific variations, and conducted avian eDNA metabarcoding. During the primer design, we did not try to eliminate the capability of detecting mammalian and other vertebrate species, because simultaneous detection of mammals and other vertebrates along with birds may be advantageous, especially for ecologists who are interested in co-occurrence patterns and potential interactions among various animal species. We performed a series of analyses to test the versatility of the designed primers: *In silico* examinations of the primers, amplification of extracted tissue DNAs of birds belonging to various taxa, and field tests by analyzing water samples from zoo cages containing birds of known species composition. Additionally, we briefly examined the usefulness of the new primer set using water samples from field samples with unknown bird species composition.

Methods

All of the critical information of our study is described below, but is also listed in Table S1 to facilitate comparisons with other studies, following the recommendations of Goldberg *et al*.[26]. All experiments were performed without direct captures of avian species, and carried out in accordance with the relevant guidelines and regulations. Also, all experimental protocols in the zoo were approved by Yokohama Zoological Gardens ZOORASIA.

Primer design. To facilitate design based on comparisons of diverse avian sequences, we first batch downloaded 410 avian sequences from RefSeq (https://www.ncbi.nlm.nih.gov/refseq/) on June 9, 2015. Then, a base composition for a selected position in the conservative region was shown in Mesquite[27]. The base compositions in selected characters were manually recorded in a spreadsheet for the primer design. In the primer design process, we considered a number of technical tips that enhance the primer annealing to the template without the use of degenerate bases[28]: primers include some G/C at the 3′-ends to strengthen primer-template annealing at this position, but a string of either Gs or Cs at the 3′-end should be avoided: considering the unconventional base pairing in the T/G bond, the designed primers use G rather than A when the template is variably C or T, and T rather than C when the template is A or G; G/C contents of the primers fall between 40 and 60%, with an almost identical melting temperature (T_m). T_m was calculated using a nearest-neighbour thermodynamic model implemented in OligoCalc[29].

We designed our primers by modifying previously developed MiFish/MiMammal primers[1,9], which corresponded to regions in the mitochondrial 12 S rRNA gene (insert length = *ca*. 171 bp), and we named our primers MiBird-U ("U" indicates "universal"). Primer sequences with MiSeq adaptors (for the first- and second-round PCR) are listed in Table 1.

In silico evaluation of interspecific variation of MiBird sequences.
The binding capacity of MiBird-U primers was computationally evaluated using the batch-downloaded 410 avian sequences. Using custom Ruby and Python scripts, the number of mismatches between MiBird-U primers and the 410 avian sequences as well as other non-target animal sequences (i.e., 741 mammalian, 197 amphibian, and 245 reptilian sequences) was calculated. Positions of base match/mismatch between MiBird-U primers and avian sequences were also examined using the downloaded avian sequences.

Common name	Scientific name	Order	Family	Accession No.
Spot-billed duck	*Anas zonorhyncha*	Anseriformes	Anatidae	LC104767
Grey nightjar	*Caprimulgus indicus*	Caprimulgiformes	Caprimulgidae	LC104768
Ancient murrelet	*Synthliboramphus antiquus*	Charadriiformes	Alcidae	LC104769
Lesser sand plover	*Charadrius mongolus*	Charadriiformes	Charadriidae	LC104770
Oriental turtle dove	*Streptopelia orientalis stimpsoni*	Columbiformes	Columbidae	LC104771
Lesser cuckoo	*Cuculus poliocephalus poliocephalus*	Cuculiformes	Cuculidae	LC104772
Northern Goshawk	*Accipiter gentilis fujiyamae*	Accipitriformes	Accipitridae	LC104773
Common kestrel	*Falco tinnunculus*	Falconiformes	Falconidae	LC104774
Chinese bamboo partridge	*Bambusicola thoracicus*	Galliformes	Phasianidae	LC104775
Red-throated loon	*Gavia stellata*	Gaviiformes	Gaviidae	LC104776
Red-crowned crane	*Grus japonensis*	Gruiformes	Gruidae	LC104777
Jungle crow	*Corvus macrorhynchos*	Passeriformes	Corvidae	LC104778
Eurasian sparrow	*Passer montanus*	Passeriformes	Passeridae	LC104779
Black-crowned night heron	*Nycticorax nycticorax*	Pelecaniformes	Ardeidae	LC104780
Great white pelican	*Pelecanus onocrotalus*	Pelecaniformes	Pelecanidae	LC104781
Great cormorant	*Phalacrocorax carbo hanedae*	Suliformes	Phalacrocoracidae	LC104782
Japanese pygmy woodpecker	*Dendrocopos kizuki*	Piciformes	Picidae	LC104783
Little grebe	*Tachybaptus ruficollis*	Podicipediformes	Podicipedidae	LC104784, LC104785
White-chinned petrel	*Procellaria aequinoctialis*	Procellariiformes	Procellariidae	LC104786
Short-tailed shearwater	*Puffinus tenuirostris*	Procellariiformes	Procellariidae	LC104787
King penguin	*Aptenodytes patagonicus*	Sphenisciformes	Spheniscidae	LC104788
Humboldt penguin	*Spheniscus humboldti*	Sphenisciformes	Spheniscidae	LC327059

Table 2. Extract DNAs used to test the performance of the MiBird primer set.

Interspecific differences within the amplified DNA sequences are required for assignment of taxonomic categories. Levels of interspecific variation in the target region (hereafter called 'MiBird sequence') across different taxonomic groups of birds were computationally evaluated using the 410 downloaded avian sequences. Among the sequences of the 410 avian, species with the deletion of primer regions (*Hemignathus munroi*, *Loxops coccineus* and *Arborophila rufipectus*) were excluded, and 407 MiBird sequences were extracted and subjected to calculation of pairwise edit distances using custom Python scripts. Pairwise inter-species edit distances were calculated for all species pairs, and pairwise inter-genus edit distances were calculated for pairs of species belonging to different genera. The edit distance quantifies dissimilarity of sequences in bioinformatics and is defined as the minimum number of single-nucleotide substitutions, insertions or deletions that are required to transform one sequence into the other.

In addition, the binding capacity and the levels of interspecific variations of the target region were further evaluated using 'primerTree' package[30] of R version 3.3.1[31]. Briefly, primerTree performs the following analysis: (1) *In silico* PCR against sequences in the NCBI database; (2) retrieval of DNA sequences predicted to be amplified; (3) taxonomic identification of these sequences; (4) multiple DNA sequence alignment; (5) reconstruction of a phylogenetic tree and (6) visualization of the tree with taxonomic annotation. Thus, by using primerTree package, species whose sequences can be amplified, phylogenetic relationships among these amplified species, and interspecific variations in the amplified sequences are rapidly visualized. Further information and instructions for the primerTree package can be found in Cannon *et al.*[30].

Primer testing with extracted DNA. We tested the versatility of MiBird-U (no adapter sequences) using DNA extracted from 22 species representing major groups of birds (Table 2). Double-stranded DNA concentrations from those samples were measured with a NanoDrop Lite spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and the extracted DNA was diluted to $15\,\text{ng}\,\mu\text{l}^{-1}$ using Milli-Q water. PCR was carried out with 30 cycles of a $15\,\mu\text{l}$ reaction volume containing $4.5\,\mu\text{l}$ sterile distilled H_2O, $7.5\,\mu\text{l}$ $2 \times$ Gflex PCR Buffer (Mg^{2+}, dNTPs plus) (Takara, Otsu, Japan), $0.7\,\mu\text{l}$ of each primer ($5\,\mu\text{M}$), $0.3\,\mu\text{l}$ *Taq* polymerase (Tks Gflex DNA Polymerase; Takara) and $1.2\,\mu\text{l}$ template. The thermal cycle profile after an initial 1 min denaturation at $94\,°\text{C}$ was as follows: denaturation at $98\,°\text{C}$ for $10\,\text{s}$; annealing at $50\,°\text{C}$ for $10\,\text{s}$; and extension at $68\,°\text{C}$ for $10\,\text{s}$ with a final extension at the same temperature for 7 min.

Study site and water sampling for primer testing with eDNA from zoo samples. To test the versatility of the newly designed primers for metabarcoding avian eDNA, we sampled water from cages on 13 December 2016 in Yokohama Zoological Gardens ZOORASIA, Yokohama, Japan (35°29′42″ N, 139°31′35″ E), where we previously tested the usefulness of a universal primer set targeting mammals[9]. We chose the zoo as a sampling site because the information about avian species in a cage is precisely known, and because the zoo rears diverse taxonomic groups of animals (i.e., >100 animal species, including many mammals and birds). Thirteen cages, in which diverse taxonomic groups of birds were reared, were selected as sampling places (Table 3). Most of the

Common name	Species name	Order	Family
Steller's sea eagle	*Haliaeetus pelagicus*	Accipitriformes	Accipitridae
Black-tailed gull	*Larus crassirostris*	Charadriiformes	Laridae
Capercaillie	*Tetrao urogallus*	Galliformes	Tetraonidae
Lady Amherst's pheasant	*Chrysolophus amherstiae*	Galliformes	Phasianidae
Ruddy shelduck[a]	*Tadorna ferruginea*	Anseriformes	Anatidae
Temminck's tragopan	*Tragopan temminckii*	Galliformes	Phasianidae
Victoria crowned pigeon	*Goura victoria*	Columbiformes	Columbidae
Mandarin duck	*Aix galericulata*	Anseriformes	Anatidae
Humboldt penguin	*Spheniscus humboldti*	Sphenisciformes	Spheniscidae
Snowy owl	*Bubo scandiacus*	Strigiformes	Strigidae
Oriental white stork	*Ciconia boyciana*	Ciconiiformes	Ciconiidae
White-naped crane	*Grus vipio*	Gruiformes	Gruidae
Common crane	*Grus grus*	Gruiformes	Gruidae
Southern ground hornbill	*Bucorvus leadbeateri*	Coraciiformes	Bucerotidae
Harris's hawk	*Parabuteo unicinctus*	Accipitriformes	Accipitridae
Emu	*Dromaius novaehollandiae*	Struthioniformes	Casuariidae

Table 3. Classification of the target bird species in the Zoorasia experiment. [a]Species kept in a walk-through bird cage with other bird species.

target species were kept separately, but ruddy shelduck (*Tadorna ferruginea*) were kept in a walk through bird cage (hereafter, "the bird cage") with other bird species (i.e., Lady Amherst's pheasant [*Chrysolophus amherstiae*], Temminck's tragopan [*Tragopan temminckii*], Victoria crowned pigeon [*Goura victoria*] and mandarin duck [*Aix galericulata*]). Note that different individuals of Lady Amherst's pheasant, Temminck's tragopan, Victoria crowned pigeon and mandarin duck from those in the bird cage were separately kept (i.e., in different cages from the bird cage), and that each water sample of the bird species was collected from each cage.

Each 100–200 ml water sample was collected through a sterile φ0.45-μm Sterivex™ filter (Merck Millipore, Darmstadt, Germany) using a sterile 50-mL syringe (TERUMO, Tokyo, Japan). After the filtration, approximately 2 ml of RNAlater (ThermoFisher Scientific, Waltham, Massachusetts, USA) was injected into the Sterivex cartridge, and the filtered water samples were stored at 4 °C for up to one day until further processing. Three negative controls (distilled water) were taken to the zoo to monitor contaminations during water sampling, filtration and transport.

In addition to the survey in the zoo, we collected water samples from a pond adjacent to the Natural History Museum and Institute, Chiba (35°35′59″ N, 140°8′18″ E; Funada-ike Pond) to test the potential effectiveness of the MiBird primers under a field condition with unknown bird species composition. Water collections at the pond were performed in the same way as those performed in the zoo.

DNA extraction. The Sterivex filter cartridges were taken back to the laboratory, and DNA was extracted from the filters using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following a protocol described and illustrated in Miya *et al.*[32]. Briefly, the RNAlater-supplemented solution was removed under a vacuum using the QIAvac system (Qiagen, Hilden, Germany). Proteinase-K solution (20 μl), phosphate buffered saline (PBS) (220 μl) and buffer AL (200 μl) were mixed, and 440 μl of the mixture was added to each filter cartridge. The materials on the filter cartridges were subjected to cell-lysis conditions by incubating the filters on a rotary shaker (at a speed of 20 rpm) at 50 °C for 20 min. The incubated mixture was transferred into a new 2-ml tube, and the collected DNA was purified using a DNeasy Blood and Tissue Kit following the manufacturer's protocol. After the purification, DNA was eluted using 100 μl of the elution buffer provided with the kit.

Paired-end library preparation. Prior to the library preparation, work-spaces and equipment were sterilized. Filtered pipet tips were used, and separation of pre- and post-PCR samples was carried out to safeguard against cross-contamination. We also employed two negative controls (i.e., PCR negative controls) to monitor contamination during the experiments.

The first-round PCR (first PCR) was carried out with a 12-μl reaction volume containing 6.0 μl of 2 × KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA), 0.7 μl of MiBird primer (5 μM primer F/R, w/ adaptor and six random bases; Tables 1), 2.6 μl of sterilized distilled H$_2$O and 2.0 μl of template. The thermal cycle profile after an initial 3 min denaturation at 95 °C was as follows (35 cycles): denaturation at 98 °C for 20 s; annealing at 65 °C for 15 s; and extension at 72 °C for 15 s, with a final extension at the same temperature for 5 min. We performed triplicate first-PCR, and these replicate products were pooled in order to mitigate the PCR dropouts. The pooled first PCR products were purified using AMPure XP (PCR product: AMPure XP beads = 1:0.8; Beckman Coulter, Brea, California, USA). The pooled, purified, and 10-fold diluted first PCR products were used as templates for the second-round PCR.

The second-round PCR (second PCR) was carried out with a 24-μl reaction volume containing 12 μl of 2 × KAPA HiFi HotStart ReadyMix, 1.4 μl of each primer (5 μM primer F/R; Table 1), 7.2 μl of sterilized distilled H$_2$O and 2.0 μl of template. Different combinations of forward and reverse indices were used for different templates (samples) for massively parallel sequencing with MiSeq. The thermal cycle profile after an initial 3 min
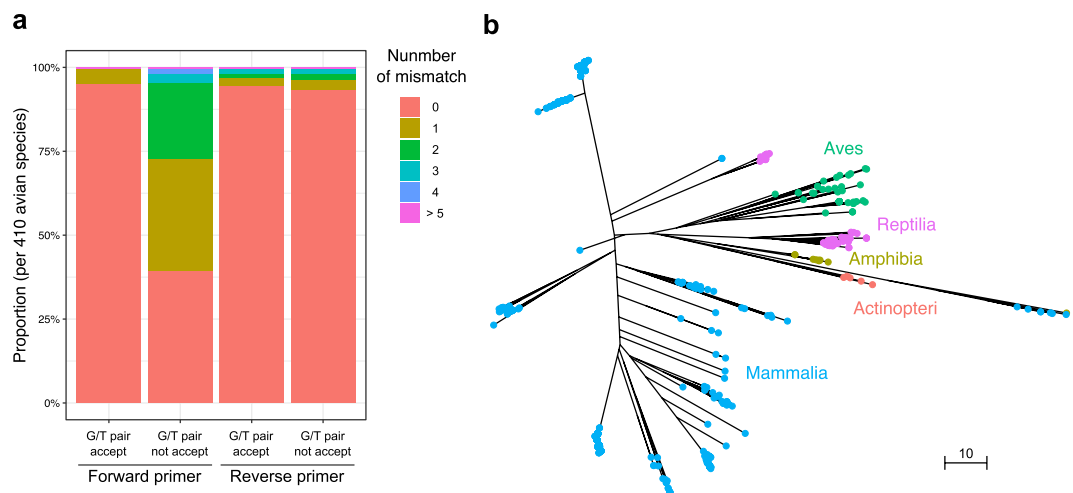
a

b

Figure 1. *In silico* evaluations of MiBird-U primers. Binding capacity of MiBird-U primers (**a**). y-axis represents the proportion of avian species that showed 0, 1, 2, 3, 4, or >5 mismatches (indicated by different colours) with MiBird-U F/R primers. The total number of avian species evaluated was 410. The phylogenetic tree was constructed for species that can be amplified using MiBird-U primers (**b**). A total of 2,000 sequences were retrieved from the database to construct the phylogenetic tree. Different classes are represented by filled circles with different colours. Lengths of branches correspond to the differences in sequences. Bar indicates edit distance.

denaturation at 95 °C was as follows (12 cycles): denaturation at 98 °C for 20 s; combined annealing and extension at 72 °C (shuttle PCR) for 15 s, with a final extension at 72 °C for 5 min.

The indexed second PCR products were mixed at equimolar concentrations to produce equivalent sequencing depth from all samples and the pooled library was purified using AMPure XP. Target-sized DNA of the purified library (*ca*. 370 bp) was excised using E-Gel SizeSelect (ThermoFisher Scientific, Waltham, MA, USA). The double-stranded DNA concentration of the library was quantified using a Qubit dsDNA HS assay kit and a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The double-stranded DNA concentration of the library was then adjusted to 4 nM using Milli-Q water and the DNA was applied to the MiSeq platform (Illumina, San Diego, CA, USA). The sequencing was performed using a MiSeq Reagent Kit Nano v2 for 2×150 bp PE (Illumina, San Diego, CA, USA).

Data processing and taxonomic assignment. The overall quality of the MiSeq reads was evaluated using the programs Fastqc (available from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and SUGAR[33]. After confirming the lack of technical errors in the MiSeq sequencing, low-quality tails were trimmed from each read using DynamicTrim.pl from the SolexaQa software package[34] with a cut-off threshold set at a Phred score of 10 ($=10^{-1}$ error rate). The tail-trimmed pair-end reads were assembled using the software FLASH with a minimum overlap of 10 bp. The assembled reads were further filtered by custom Perl scripts in order to remove reads with either ambiguous sites or those showing unusual lengths compared to the expected size of the PCR amplicons. Finally, the software TagCleaner[35] was used to remove primer sequences with a maximum of three-base mismatches and to transform the FASTQ format into FASTA (see Table S2 for the numbers of reads remained after these pre-processing).

The pre-processed reads from the above custom pipeline were dereplicated using UCLUST[36], with the number of identical reads added to the header line of the FASTA formatted data file. Those sequences represented by at least 10 identical reads were subjected to the downstream analyses, and the remaining under-represented sequences (with less than 10 identical reads) were subjected to pairwise alignment using UCLUST. If the latter sequences observed for less than 10 reads showed at least 99% identity with one of the former reads (one or two nucleotide differences), they were operationally considered as identical (owing to sequencing or PCR errors and/or actual nucleotide variations in the populations).

The processed reads were subjected to local BLASTN searches[37] against a custom-made database. The custom database was generated by downloading all whole mitogenome sequences from Sarcopterygii deposited in NCBI Organelle Genome Resources (http://www.ncbi.nlm.nih.gov/genomes/OrganelleResource.cgi?taxid=8287). As of 15 March 2016, this database covered 1,881 species across a wide range of families and genera (including birds, mammals, reptiles and amphibians). In addition, the custom database was supplemented by all whole and partial fish mitogenome sequences deposited in MitoFish[38] in order to cover fish detection (note that MiBird primers amplify fish sequences as well; see Fig. 1).

The top BLAST hit with a sequence identity of at least 97% and *E*-value threshold of 10^{-5} was applied to species assignments of each representative sequence. Reliability of the species assignments was evaluated based on the ratio of total alignment length and number of mismatch bases between the query and reference sequences. For example, if a query sequence was aligned to the top BLAST hit sequence with an alignment length of 150 bp with one mismatch present, the ratio was calculated as $150/(1+1)$. The value one was added to the denominator to

MiBird-U-F	*G*	*G*	*G*	*T*	*T*	*G*	*G*	*T*	*A*	*A*	*A*	*T*	*C*	*T*	*T*	*G*	*T*	*G*	*C*	*C*	*A*	*G*	*C*				
A	1	4	1	1	0	1	0	0	**397**	**407**	**407**	0	0	0	0	0	0	0	0	0	**407**	0	0				
C	0	0	0	61	129	0	0	0	2	0	0	7	**407**	155	0	0	0	0	**407**	**407**	0	0	**407**				
G	**406**	**403**	**406**	0	0	**406**	**407**	0	1	0	0	0	0	0	0	**407**	0	**407**	0	0	0	**407**	0				
T	0	0	0	**345**	**278**	0	0	**407**	7	0	0	**400**	0	**252**	**407**	0	**407**	0	0	0	0	0	0				
MiBird-U-R	*C*	*A*	*T*	*A*	*G*	*T*	*G*	*G*	*G*	*G*	*T*	*A*	*T*	*C*	*T*	*A*	*A*	*T*	*C*	*C*	*C*	*A*	*G*	*T*	*T*	*T*	*G*
A	0	**407**	0	**406**	0	0	0	2	0	0	0	**407**	0	1	0	**407**	**407**	0	0	0	0	**406**	0	1	0	0	0
C	**407**	0	0	0	0	1	0	0	0	0	0	0	0	**406**	0	0	0	0	**407**	**406**	**404**	0	0	1	0	5	1
G	0	0	0	0	**407**	0	**407**	**405**	**407**	**407**	0	0	1	0	0	0	0	0	0	0	0	1	**405**	0	0	0	**406**
T	0	0	**407**	1	0	**406**	0	0	0	0	**407**	0	**406**	0	**407**	0	0	**407**	0	1	3	0	2	**405**	**407**	**402**	0

Table 4. Nucleotide sequences of the universal primers (MiBird-U) and base compositions of the selected 407 avian species. Among the downloaded sequences, 3 species with the deletion of primer regions were excluded, resulting in the sequences of 407 avian species. Italic bases indicate primer sequences. Numbers indicate the number of avian species of which base matches with A, C, G, or T listed in the left column. Bold numbers indicate the number of avian species of which base matches with that of the primer.

Edit distance	0	1	2	3	4	≥5	Total
Frequency distributions of the inter-specific/genus edit distances of the insert sequence							
Species	27	50	67	113	187	82,177	82,621 combinations
Genus	10	23	40	80	157		

Table 5. Frequency distributions of the interspecific edit distances of the primer set against bird sequences. Pairwise inter-species edit distances were calculated for all species pairs, and pairwise inter-genus edit distances were calculated for pairs of species belonging to different genera.

avoid zero-divisors. This value (e.g., 150/(1 + 1)) was calculated for the top and second-highest BLAST hit species, and the ratio score between these values was used as a comparable indicator of the species assignment. Results from the BLAST searches were automatically tabulated, with scientific names, common names, total number of reads and representative sequences noted in an HTML format. The above bioinformatics pipeline from data pre-processing through taxonomic assignment is available in supplements in a previous study[1]. Also, the above bioinformatic pipeline can be performed on a website. For more detailed information, please see http://mitofish. aori.u-tokyo.ac.jp/mifish. Please note that the pipeline implemented in the website currently uses the custom fish database and does not aim to detect avian species (confirmed on 20 September 2017).

Data availability. DDBJ Accession numbers of the DNA sequences analyzed in the present study are DRA006196 (Submission ID), PRJDB4990 (BioProject ID) and SAMD00096837–SAMD00096858 (BioSample ID).

Results and Discussion
Tests of versatility of designed primers in silico and using extracted DNA.
First, the performance of MiBird-U primers was tested *in silico* (Fig. 1 and Tables 4 and 5). When G/T pairs were accepted, MiBird-U-F and -R perfectly matched 390 (95.1%) and 388 (94.6%) species among 410 species tested, respectively, and 99.5% and 96.8% of the 410 species showed at most 1 mismatch (Fig. 1a). Among the avian sequences tested, all species showed no mismatch at the 3′-end of MiBird-U-F, and most species (>98.7%) showed no mismatch at the 3′-end of MiBird-U-R (Table 4). In addition, inter-specific differences in the edit distance were calculated and 82,177 out of 82,621 combinations (99.5%) showed edit distance larger than 5 (Table 5). These analyses suggested that the target region of most avian species can be amplified using MiBird-U primers, and that the amplified sequences contain sufficient information required for assignment of taxonomic categories.

To examine the range of species that can be amplified using MiBird-U primers, we performed an analysis with the primerTree package[30]. The results confirmed that the primers can amplify avian species (Fig. 1b). MiBird-U primers can also amplify a diverse group of mammalian species in addition to amphibian, reptilian and fish species (Fig. 1b), which is not surprising because MiBird-U primers were produced by modifying fish/mammal-targeting universal primers. The potential of MiBird-U primers to amplify mammalian, amphibian, and reptilian species was also confirmed by *in silico* test of the binding capacity of MiBird-U primers (Table S3). The capacity of MiBird-U primers to detect mammalian and other species might be useful when simultaneous detection of these animals is desired (e.g., when one tries to study co-occurrence patterns and potential interactions among animals).

Second, the performance of MiBird-U primers was evaluated using 22 extracted avian DNA samples. All of the extracted DNA samples were successfully amplified, and the resultant sequences were deposited in the DDBJ/ EMBLE/GenBank databases (Table 2). Together, the results of *in silico* tests and the amplification of extracted DNAs suggested that MiBird-U primers are capable of amplifying/identifying DNA fragments derived from diverse avian species.

Primer testing with eDNA from field water samples.   MiSeq sequencing and data pre-processing generated 656,472 sequences from 21 samples (including 3 field negative controls and 2 PCR negative controls) (Table 5). In general, the quality of sequences produced by our experiment was high (i.e., most raw reads passed the filtering process; Table S2).

Among the 16 water samples from zoo cages examined here, all avian species were successfully detected (Table 6). Briefly, eDNA samples of the Steller's sea eagle (*Haliaeetus pelagicus*), capercaillie (*Tetrao urogallus*), white-naped crane (*Grus vipio*), common crane (*Grus grus*) and southern ground hornbill (*Bucorvus leadbeateri*) generated high numbers of sequence reads, and 64.9–94.9% of total sequence reads were assigned to the target avian species. Samples from cages of the black-tailed gull (*Larus crassirostris*), Humboldt penguin (*Spheniscus humboldti*), snowy owl (*Bubo scandiacus*), Oriental white stork (*Ciconia boyciana*), Harris's hawk (*Parabuteo unicinctus*) and emu (*Dromaius novaehollandiae*) generated fewer sequence reads, and 1.4–28.8% of total sequence reads were assigned to the target avian species. The reason for these variations in the proportions of sequence reads from target avian species is not known, but as discussed in the previous study[9], the observed levels of variations were not surprising because detection of animals' sequences relies on contacts of animals with water and because opportunities for animals to contact water would depend on animals' behaviour. These considerations imply that the proportion of sequence reads from a particular avian species would be inherently spatially and temporally stochastic to some extent (see also results of mammalian eDNA metabarcoding in Ushio *et al.*[9]). It is not surprising that sequences of the Lady Amherst's pheasant, ruddy shelduck, Temminck's tragopan, Victoria crowned pigeon and mandarin duck were detected in the ruddy shelduck sample (Table 6) because all of these five species were kept in the bird cage where the ruddy shelduck sample was collected.

In addition to the target avian species, we frequently detected many non-target species (Table 6 and S4). For example, sequences of the Steller's sea eagle were frequently detected in other samples, e.g., the Victoria crowned pigeon, Oriental white stork, Humboldt penguin and so on (Table 6). As our field negative controls generated no target bird sequences (Table 6), it does not seem likely that the detection of the sea eagle in other samples was due to cross-contamination during sampling or experiments. One possible reason for the detection of non-target avian species include the spatial closeness of the eagle's cage and the other cages. For instance, the cages of the Victoria crowned pigeon (i.e., the bird cage) and Humboldt penguin were located close to the eagle's cage, and thus it is possible that the eagle's feathers and other tissues could be transported (e.g., via wind) to other cages. Also, zoo staff frequently moved among cages, and they were possible transporters (e.g., through their shoe sole) of materials containing DNA of non-target species.

Other frequently detected non-target species were falcated teal (*Mareca falcata*), common shelduck (*Tadorna tadorna*), common moorhen (*Gallinula chloropus*), fishes and humans (Table S4). The falcated teal, shelduck and moorhen were not kept in cages, but wild common moorhens and close relatives of the duck and shelducks (i.e., Eurasian wigeon [*Anas penelope*] and common pochard [*Aythya ferina*], respectively) are commonly observed in the regulating pond on-site of sampling region, and thus their DNA might have contaminated zoo cages (possibly via feathers or other tissues) and thus have been detected by the metabarcoding.

The frequently detected fish species here are also species that are commonly observed in Japan, and the zoo uses waters from a natural lake and rivers. Therefore, the fish sequences might have been derived from water under natural conditions. Detection of many human sequences was not surprising considering that visitors to the zoo and staff members, who are potential sources of human sequences, are almost always near the cages. It is also be possible that contaminations of human and fish DNA happened under the laboratory conditions (Table S4), because in our lab fish DNAs were routinely processed and humans were often working (i.e., carry-over contaminations). Specifically, ocean fish sequences were detected from zoo samples despite the efforts for decontamination, and these contaminants are likely due to previous work in the same lab. The sequences of these obvious non-target taxa (i.e., humans, fish, and potential non-target carry-over contaminations) may be excluded from further statistical analyses[25] if one may be interested in ecological interpretations of the results.

Lastly, in order to test the usefulness of MiBird primers under a natural field condition, we performed a metabarcoding study using a water sample from a pond adjacent to the Natural History Museum and Institute, Chiba (Funada-ike Pond). As a result of MiSeq sequencing, 14,873 reads of avian species were generated from three water samples, and five avian species (common shoveler [*Anas clypeata*], 883 reads; falcated teal, 3,246 reads; common moorhen, 9,260 reads; light-vented bulbul [*Pycnonotus sinensis*], 745 reads; and common shelduck, 739 reads) were detected. As a systematic monitoring of the bird community (e.g., frequent visual observation) has not been performed in the study site, rigorous validation of the metabarcoding study was not possible. Some avian species detected, i.e., light-vented bulbuls, common shelducks and falcated teals, are rare, or not reported, in this region, suggesting that these species were misidentified. These possible misidentifications are likely to be attributable to a lack of reference sequences and/or insufficient inter-species differences in the amplified DNA region (i.e., partial 12 S mitochondrial region) (see also ref.[14]). Light-vented bulbuls, common shelducks and falcated teals are relatives of brown-eared bulbuls (*Hypsipetes amaurotis*), common pochards (*Aythya ferina*) and Eurasian wigeons (*Anas penelope*), respectively, and these relatives are indeed common inhabitants in the sampling region. Together, these results suggest that MiBird primers were capable of detecting bird species under a field condition, but at the same time, improvements of reference sequence databases, further validations of MiBird primers, and careful interpretations are necessary.

Conclusion

A proof-of-concept that eDNA metabarcoding can potentially detect avian species has been already demonstrated in previous studies[22–25], and in the present study we explicitly demonstrated the potential and usefulness of avian eDNA metabarcoding using our new primer set and MiSeq platform. Describing and monitoring the diversity of bird species, as well as other animals, is one of the critical steps in ecosystem conservation and management, but it can be laborious, costly and incomplete if one relies on a few traditional survey methods. The eDNA metabarcoding approach

Common name of bird living in cage	Scientific name	Haliaeetus	Larus	Tetrao	Chrysolophus	Tadorna	Tragopan	Goura	Aix	Spheniscus	Bubo	Ciconia	G. vipio	G. grus	Bucorvus	Parabuteo	Dromaius	Non-target sequences[b,c]	Total sequences	% target living in cage
Steller's sea eagle	Haliaeetus pelagicus	**28,448**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3,238	31,686	89.8
Black-tailed gull	Larus crassirostris	0	**4,437**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17,922	22,359	19.8
Capercaillie	Tetrao urogallus	0	0	**36,095**	0	0	0	0	0	0	0	0	13	0	0	0	0	19,477	55,585	64.9
Lady Amherst's pheasant	Chrysolophus amherstiae	0	0	0	**39,151**	0	0	0	25	0	0	0	0	15	0	0	0	8,130	47,321	82.7
Ruddy shelduck[a]	Tadorna ferruginea	0	0	0	138	**2,848**	209	2,750	7,939	0	0	0	0	0	0	0	0	12,498	26,382	10.8
Temminck's tragopan	Tragopan temminckii	0	0	376	0	0	**57,072**	0	0	0	0	0	0	0	0	0	0	1,944	59,392	96.1
Victoria crowned pigeon	Goura victoria	11,023	0	0	0	0	0	**2,186**	24	0	0	0	0	0	0	0	0	13,108	26,341	8.3
Mandarin duck	Aix galericulata	254	0	0	0	65	51	34	**13,465**	0	0	0	0	0	0	0	0	12,272	26,141	51.5
Humboldt penguin	Spheniscus humboldti	1,905	0	0	0	289	0	85	428	**1,834**	0	0	0	0	0	0	0	28,091	32,632	5.6
Snowy owl	Bubo scandiacus	901	0	0	0	65	0	64	191	0	**425**	0	0	0	0	0	0	28,566	30,212	1.4
Oriental white stork	Ciconia boyciana	7,258	63	0	0	0	0	103	258	0	0	**3,072**	0	0	0	0	0	22,815	33,569	9.2
White-naped crane	Grus vipio	186	0	527	0	22	0	36	33	0	0	0	**59,678**	0	31	0	0	2,390	62,903	94.9
Common crane	Grus grus	0	0	0	548	0	0	0	0	0	0	0	0	**52,717**	0	0	0	2,586	55,851	94.4
Southern ground hornbill	Bucorvus leadbeateri	148	0	0	0	0	15	0	0	0	0	0	0	0	**36,955**	0	0	4,900	42,018	88.0
Harris's hawk	Parabuteo unicinctus	1,227	0	0	0	77	0	0	332	0	0	0	0	0	0	**306**	0	20,338	22,280	1.4
Emu	Dromaius novaehollandiae	396	0	0	0	67	0	25	146	0	0	0	22	0	0	0	**1,647**	3,412	5,715	28.8
Field NC		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27,218	27,218	
Field NC		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7,977	7,977	
Field NC		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40,890	40,890	
PCR NC		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
PCR NC		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Total sequence		51,746	4,500	36,998	39,837	3,433	57,347	5,283	22,841	1,834	425	3,072	59,713	52,732	36,986	306	1,647	277,772	656,472	

Table 6. Sequence reads of detected species from water samples collected in the zoo. Bold numbers indicate sequence reads of a target species. [a]Species kept in a walk-through bird cage. [b]See Table S4 for the contents of non-target sequences. [c]See Table S3 for the contents of non-target sequences.

presented here is non-invasive and efficient. Moreover, as information of non-target organisms (e.g., invertebrates and microbes in our case) is also encoded in eDNA, analyzing eDNA of organisms from multiple taxa might be useful for studying co-occurrence patterns and even potential interactions among organisms (e.g., bird-insect interactions). In conclusion, we propose that the eDNA metabarcoding approach can serve as an efficient alternative for taking a snapshot of bird diversity and could potentially contribute to effective ecosystem conservation and management.

References

1. Miya, M. *et al*. MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: detection of more than 230 subtropical marine species. *R. Soc. open Sci.* **2**, 150088 (2015).
2. Bista, I. *et al*. Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nat. Commun.* **8**, 14087 (2017).
3. Fukumoto, S., Ushimaru, A. & Minamoto, T. A basin-scale application of environmental DNA assessment for rare endemic species and closely related exotic species in rivers: a case study of giant salamanders in Japan. *J. Appl. Ecol.* **52**, 358–365 (2015).
4. Ficetola, G. F., Miaud, C., Pompanon, F. & Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **4**, 423–5 (2008).
5. Kelly, R. P. *et al*. Harnessing DNA to improve environmental management. *Science (80).* **344**, 1455–6 (2014).
6. Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N. & Kawabata, Z. Surveillance of fish species composition using environmentalDNA. *Limnology* **13**, 193–197 (2011).
7. Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. Estimation of fish biomass using environmental DNA. *PLoS One* **7**, e35868 (2012).
8. Yamamoto, S. *et al*. Environmental DNA as a 'Snapshot' of Fish Distribution: A Case Study of Japanese Jack Mackerel in Maizuru Bay, Sea of Japan. *PLoS One* **11**, e0149786 (2016).
9. Ushio, M. *et al*. Environmental DNA enables detection of terrestrial mammals from forest pond water. *Mol. Ecol. Resour* https://doi.org/10.1111/1755-0998.12690 (2017).
10. Deiner, K., Fronhofer, E. A., Mächler, E., Walser, J.-C. & Altermatt, F. Environmental DNA reveals that rivers are conveyer belts of biodiversity information. *Nat. Commun.* **7**, 12544 (2016).
11. Evans, N. T. *et al*. Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Mol. Ecol. Resour.* **16**, 29–41 (2016).
12. Ushio, M. *et al*. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *bioRxiv* 113472 https://doi.org/10.1101/113472 (2017).
13. Rodgers, T. W. & Mock, K. E. Drinking water as a source of environmental DNA for the detection of terrestrial wildlife species. *Conserv. Genet. Resour.* **7**, 693–696 (2015).
14. Ishige, T. *et al*. Tropical-forest mammals as detected by environmental DNA at natural saltlicks in Borneo. *Biol. Conserv.* **210**, 281–285 (2017).
15. Hunter, M. E. *et al*. Environmental DNA (eDNA) sampling improves occurrence and detection estimates of invasive burmese pythons. *PLoS One* **10**, e0121655 (2015).
16. Anderson, S. H., Kelly, D., Ladley, J. J., Molloy, S. & Terry, J. Cascading Effects of Bird Functional Extinction Reduce Pollination and Plant Density. *Science (80).* **331**, 1068–1071 (2011).
17. Sethi, P. & Howe, H. F. Recruitment of Hornbill-Dispersed Trees in Hunted and Logged Forests of the Indian Eastern Himalaya. *Conserv. Biol.* **23**, 710–718 (2009).
18. Van Bael, S. A., Brawn, J. D. & Robinson, S. K. Birds defend trees from herbivores in a Neotropical forest canopy. *Proc. Natl. Acad. Sci.* **100**, 8304–8307 (2003).
19. Bregman, T. P., Sekercioglu, C. H. & Tobias, J. A. Global patterns and predictors of bird species responses to forest fragmentation: implications for ecosystem function and conservation. *Biol. Conserv.* **169**, 372–383 (2014).
20. Aronson, M. F. J. *et al*. A global analysis of the impacts of urbanization on bird and plant diversity reveals key anthropogenic drivers. *Proc. R. Soc. B Biol. Sci.* **281**, 20133330–20133330 (2014).
21. Bibby, C., Burgess, N., David Hill & Simon Mustoe. *Bird census techniques*. (Academic Press, 2000).
22. Thomsen, P. F. *et al*. Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS One* **7**, e41732 (2012).
23. Thomsen, P. F. *et al*. Monitoring endangered freshwater biodiversity using environmental DNA. *Mol. Ecol.* **21**, 2565–2573 (2012).
24. Thomsen, P. F. *et al*. Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *PLoS One* **11**, e0165252 (2016).
25. Port, J. A. *et al*. Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Mol. Ecol.* **25**, 527–541 (2016).
26. Goldberg, C. S. *et al*. Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods Ecol. Evol.* **7**, 1299–1307 (2016).
27. Maddison, W. P. & Maddison, D. R. Mesquite: a modular system for evolutionary analysis (2011).
28. Palumbi, S. R. in *Molecular Systematics* (eds. Hills, D. M., Moritz, C. & Mable, B. K.) 205–247 (Sinauer, 1996).
29. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res.* **35**, W43–W46 (2007).
30. Cannon, M. V. *et al*. In silico assessment of primers for eDNA studies using PrimerTree and application to characterize the biodiversity surrounding the Cuyahoga River. *Sci. Rep.* **6**, 22908 (2016).
31. R Core Team. R: A Language and Environment for Statistical Computing. (2016).
32. Miya, M. *et al*. Use of a Filter Cartridge for Filtration of Water Samples and Extraction of EnvironmentalDNA. *J. Vis. Exp*. e54741–e54741, https://doi.org/10.3791/54741 (2016).
33. Sato, Y. *et al*. SUGAR: graphical user interface-based data refiner for high-throughput DNA sequencing. *BMC Genomics* **15**, 664 (2014).
34. Cox, M. P. *et al*. SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics* **11**, 485 (2010).
35. Schmieder, R., Lim, Y. W., Rohwer, F. & Edwards, R. TagCleaner: Identification and removal of tag sequences from genomic and metagenomic datasets. *BMC Bioinformatics* **11**, 341 (2010).
36. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
37. Camacho, C. *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
38. Iwasaki, W. *et al*. MitoFish and MitoAnnotator: a mitochondrial genome database of fish with an accurate and automatic annotation pipeline. *Mol. Biol. Evol.* **30**, 2531–40 (2013).

Acknowledgements

Author Contributions

M.U. and M.M. conceived and designed research; M.U., I.N. and K.M. performed sampling; M.U., I.N., T.S. and M.M. performed experiments; M.U., M.T. and W.I. performed data analysis; M.U. and M.M. wrote the early draft and completed it with significant inputs from all authors.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-22817-5.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.