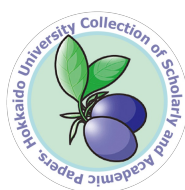




HOKKAIDO UNIVERSITY

Title	Multi-locus exon-primed intron crossing (EPIC) primer design for regional birds and algorithm design for a combination of introns
Author(s)	Kumakura, Keiko; Osanai, Yuta; Osa, Yuichi et al.
Citation	Japanese Journal of Veterinary Research, 66(4), 261-272
Issue Date	2018-11
DOI	https://doi.org/10.14943/jjvr.66.4.261
Doc URL	https://hdl.handle.net/2115/72019
Type	departmental bulletin paper
File Information	261-272 Daiji Endoh.pdf



Multi-locus exon-primed intron crossing (EPIC) primer design for regional birds and algorithm design for a combination of introns

Keiko Kumakura¹⁾, Yuta Osanai¹⁾, Yuichi Osa³⁾, Kei Fujii⁴⁾,
Mitsuhiko Asakawa²⁾, Kii Ushiyama²⁾, Manabu Onuma⁵⁾,
Hirotaka Igarashi⁶⁾ and Daiji Endoh^{1,*)}

¹⁾Laboratory of Radiation Biology, ²⁾Laboratory of Parasitology, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu, Hokkaido, 069–8501 Japan

³⁾Eastern Hokkaido Wildlife Station, Nature Conservation Division, Institute of Environmental Science, Hokkaido Research Organization, Nakahamacho, Kushiro, Hokkaido, 085–0027 Japan

⁴⁾OAT Agrio Co., Ltd., Hanamen Satoura Naruto, Tokushima 772–0021 Japan

⁵⁾Ecological Risk Assessment and Control Section, Center for Environmental Biology and Ecosystem, National Institute for Environmental Studies, 16–2, Onogawa, Tsukuba, Ibaraki, 305–8506, Japan

⁶⁾Laboratory of Small Animal Internal Medicine, School of Veterinary Medicine, Azabu University, Sagamihara, Kanagawa, 252–5201 Japan

Received for publication, February 9, 2018; accepted, April 11, 2018

Abstract

There are a variety of purposes for identifying the different species of wild animals and birds. It needs a method that can save cost and time for the investigation of many birds. One of which is DNA barcoding that has been used for species identification from sampled muscle, feather or feces. In this study, we developed a novel nuclear exon-primed intron-crossing (EPIC) markers to distinct local birds in Hokkaido area. In birds, the size of introns is smaller than that of mammals. Many introns are within the size range that is suitable for EPIC. To design the EPIC primers, we collated the genome sequences of chicken, turkey, zebra finch and flycatcher on which genes and exons were annotated. We tested the primers for 16 introns which could amplify the introns in 12 different species birds. On the analysis of amplified introns, the size of polymorphisms has been shown in some species of birds. While an intron in TBC1D22B gene showed an interspecies difference in the sizes of amplified introns, some of the size difference was too small for reproducible observation. Since a species could not be identified with one gene, attempts were made to identify species by combining multiple genes. Resultantly, the amplified products of 3 introns have distinguished 10 species of birds. It has been noted that the EPIC method is capable of distinguishing 10 species of local birds in Hokkaido and is therefore suggested to be utilized as a convenient means for multiple species identification.

Key Words: EPIC, polymorphic introns, species-identification, Hokkaido, Max-covering computer program

*Corresponding author: Daiji Endoh, Laboratory of Radiation Biology, School of Veterinary Medicine, Rakuno Gakuen University, Ebetsu, 069–8501 Japan
Phone: +81–11–388–4847. Fax: +81–11–387–5890. E-mail: dendoh@rakuno.ac.jp
doi: 10.14943/jjvr.66.4.261

Introduction

Surveys on the movement of wild birds have been documented for epidemiological research and environmental conservation^{2,14,25}. On the investigation of bird movement, one of the key steps for species-identification is from remnants of feces and feathers¹⁷. The nucleotide sequence of cytochrome oxidase I (COI) has been used as a marker gene in DNA barcoding. In the DNA barcoding, nucleotide sequences in COI were the indicator for species identification and the library of COI sequences has grown exponentially having reached around 8 million sequences in 2018 (2nd February 2018 at www.boldsystems.org). Of these sequences, more than 89,471 sequences belong to the mammalian and 6,379 to the bird species.

On the amplification step in DNA barcoding, the appearance of PCR-induced sequence artifacts is troublesome¹ while the nucleotide sequence that is derived from some species might have been lost if DNA from some other species contaminated. Barcoding based on COI that requires nucleotide sequence determination might not be appropriate for ecological survey because it needed high cost from the necessity of COI sequence on remnants of many birds²¹. As a countermeasure against such artifacts, another method was formulated for species identification. To analyze the alternative method, the genomic regions of the chromosomes were preferably examined¹⁶. A system that is dependent on polymorphic introns has provided broader information on nuclear polymorphisms involving many species^{16,21}. The PCR technique for detection of polymorphic introns that uses conserved exons surrounding polymorphic intron is named as exon-primed intron crossing (EPIC) method⁶. Primers for EPIC were designed on conserved exons surrounding polymorphic introns. In EPIC method, species may be presented from sizes of PCR products which can be determined even without the process of sequencing. The EPIC procedure has been suggested as a suitable method for surveillance of bird localization with many samples.

On the design of EPIC primers, conserved exons were selected from the analysis of reference genomes^{6,16,21}. In birds genes including exons were suggested to be conserved through evolution^{4,7,10}. After the design of the primers, the primers can then be used for many species without their genomic information. Li *et al.* (2010) reported 12 EPIC markers in 13 teleosts of fishes wherein many of its genome had not yet been determined²¹. As shown in fishes, a restricted number of reference genomes in target class of organisms could be a preferable point for a design of EPIC primers. Birds have 4 reference genomes, and compared with that of the mammals, its genome is known to be conserved^{4,10}. Relatively, the short intron in the genome of birds is considered also to be appropriate for designing the EPIC primers^{4,30}. This method of detecting the short introns may be fitting as in the case of collecting materials for displays in museums²⁶.

As shown in the EPIC for fishes, a combination of EPIC primers for plural polymorphic introns could serve as a key step for usage of EPIC for surveillance purposes²¹. Ordinarily, polymorphic introns may vary from species to species and occasionally, may not also differ in some species. And hence, it is desirable to distinguish the maximum number of species with the lowest number of intron. This is because when the number of target species is increased, calculation by selection a computer program is necessary since reliance on gross observations may be unwieldy. Following the release of fragment genomic sequence together with the improvement in the next-generation sequencer, the number of species to be identified are likely to expand. From these backgrounds, an EPIC analysis of 12 birds in Hokkaido have been conducted. A computer program to select the minimum number of introns (locus in a genome) to distinguish the maximum number of species was also examined. Though the use of an algorithm, 3 introns for distinguishing 10 species of birds have been selected.

Materials and Methods

Genomic data and design of primers: Genomic data of chicken (*Gallus gallus domesticus*), turkey (*Meleagris gallopavo*), zebra finch (*Taeniopygia guttata*), and flycatcher (*Ficedula albicollis*), were obtained from NCBI Genome database (Feb 3, 2018). Original program (written with Ruby 2.5.0 D.E. sending on request) was made for cumulation of exons and introns with interspecies comparisons (Fig. S1). The program foremostly parses exons from each genome referring to an annotation of GenBank format file including positions and names of genes and exons. The program pick up nucleotide sequences of exons from each chromosomes file in fasta format. Exons were parsed from each species (158,376 exons from chicken, 150,092 from turkey, 144,565 from zebra finch and 153,091 from flycatcher). Homology of every exons were analyzed by USEARCH homology search program¹³⁾. The USEARCH program output homology data in BLAST8 format. Subsequently the program analyzed homology files to aligned each-species' exon to homologous Gallus exon. On this alignment, exons each of which was homologous to a Gallus exon was selected. The program assigns exon numbers and their position on chromosomes. The program also determined intron length from the positions of exons on the chromosomes. In the next step, the program selected exons which single homologs were found in all 4 referential genomes and subsequently select introns in 4 referential genome lesser than 1 kbp and length of introns of homologs were divergent in size (Table S1). After the determination of target introns, we designed primers for surrounding exons by Primer3-primer design program (Table S2)¹⁹⁾. One or two degenerate nucleotide were used on divergent positions between species.

Bird samples and extraction of DNA: Liver or muscle tissue and feather of black kite (*Milvus migrans*), jungle crow (*Corvus macrorhynchos*),

carion crow (*Corvus corone*), brown-eared bulbul (*Hypsipetes amaurotis*), turtle dove (*Streptopelia orientalis*), rock dove (*Columba livia*), sparrow (*Passer montanus*), swallow (*Hirundo rustica*), martin (*Delichon urbica*), starling (*Sturnus cineraceus*), black-backed wagtail (*Motacilla alba lugens*), and chicken (*Gallus gallus domesticus*) were obtained from dead birds lived in Hokkaido. Tissues and feathers were obtained from an individual each, respectively. DNA was extracted from liver or muscle tissues using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), according to manufacturer's instruction. Specifically, for the feathers, when the DNA was extracted from it, the feathers were cut into fine pieces using a pair of scissors or powdered by freeze-dry treatment and was subsequently ground with a mixer mill (MM300, Retsch, Haan, Germany), in the presence of 3 mm Zirconia-beads. Subsequently, the DNA of the feather was extracted using DNeasy Blood & Tissue Kit mentioned previously.

PCR, electrophoresis and fragment analysis: EPIC PCR was performed with GoTaq Green PCR Mix (Promega Corp., Madison, WI), with thermocycle: 1 cycle of 95°C for 2 min, followed by 33 cycles of denaturation (95°C for 30 s), annealing (59°C for 30 s), and extension (72°C for 30 s), and then a final long extension (72°C for 2 min).

PCR products were electrophoresed either in 1% agarose with 1 × TAE buffer or MultiNA-chip electrophoresis machine (SHIMAZU Corp., Kyoto, Japan), according to manufacturer's instructions²⁷⁾. Some PCR products were also subjected to fragment analysis with 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA), by GeneMapper software (Applied Biosystems)³¹⁾. On fragment analysis, the PCR products were diluted by a factor of 100.

Selection computer program for minimum set of introns with maximum number of distinguishable species: In order to exploit the differences in intron size different for each gene, combinations

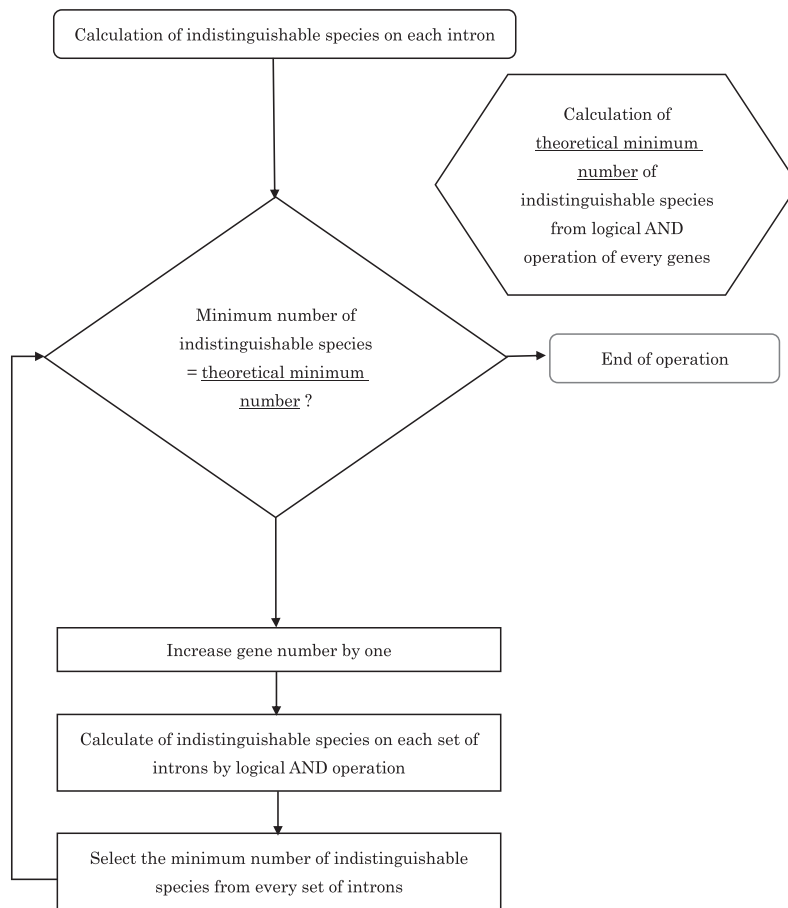


Fig. 1. Algorithm selecting the least combination of genes resulting maximum number of distinguishable species.

that obtain the most results with the smallest number of genes were examined (Fig. 1). As a precondition to study the method, the sizes of DNA fragments distinguishable on agarose gel were searched on the multiplex-PCR studies. Assuming that the size to be the center of the observation is about 100–300 bases, 25 bases are clearly distinguished also in the literatures^{11,24}, the least difference of sizes were set to 20 bases.

The computer program basically according to greedy algorithm to increase the number of distinguishable species¹². In the usual Greedy algorithm, the selection range is increased with the fact that a large number can be selected as an index, but in this research, reducing the number of indistinguishable species, which is a complementary set of distinguishable species, was used as an indicator of selection. For this

procedure, the computer program firstly created indistinguishable species-pair, comparing PCR products on every combinations species pairs in each introns. Secondly, the computer program determined the largest number of distinguishable species by combine results on every introns and set it to the attainment target. Before searching least set for maximum distinguish species, the computer program ascertain that single intron did not distinguish maximum number of species. On the check for distinguishable species, the computer program merged the lists of indistinguishable species by logical AND operation. For example, if intron A could not distinguish species X and Y, and intron B could not distinguish species Y and Z, combination of A and B could not distinguish Y ([X, Y] AND [Y, Z] = [Y]). Serially, number of introns were

Table 1. EPIC primers which produced PCR products on over 10 species

Primer name	Exon No.*	Primer sequence (5'-3')	Exon No.*	Primer sequence (5'-3')
TEAD3	10	CTCGAGTACTCYGCCTTCATGGAGG	11	CAAACTTGACGAGRAAGAAGGAGT
TBC1D22B	11	CAGTTTGCCTTTTCGYTGGATGAACA	12	GATCTCCTTCCGCCACTTGATCARG
TBC1D22B-i	11	CCTCTGGGACACCTACCAGGTA	12	AGAATCCTTCTGGCTCTGACTG
LOC395100	5	CTCCAAGGGCAGCTGYTGTGGAG	6	CYTCCCAGCTGATCCAAAGGTCCC
AHCYL1_1	9	TACGTACACVGAGATCGACCCCATCT	10	TTCRCTCAGCTTCACCACCCGAAAT
AHCYL1_2	4	AACATCTACTCCACCCAGAACGARG	5	ACCACCAGAARTCATCCTCTGACTC
PHTF1_1	12	TGTTTTTCTTYATGATGTGTGTGC	13	GACTGAARAGCTTGGCAAACAAAA
PHTF1_2	11	GAATGACTGYAAGAAGGTGGACATG	12	GCAACACACATCATRAAGAAAAACA
RSBN1_1	3	TCCCMACAGCTGATATGCCAAAATC	4	GGTCCGAGGTAGGTACTGSAGRT
AMPD1	5	AGAAATGAGCCAAAGGACCTGCCYT	6	TCTCCTCCATCTCATTKAGCATTTT
CSDE1	3	CTTCKGATCGCCGAAGTGGAAAACC	4	CTCTCTAAATTGTGAGGAACAGCAC
PDXDC1	20	TTAGAAAAGGGATACAGGAAGCACA	21	GAAKGGAGAAAACCAGTTCAGCACA
BFAR	6	YCCAGGAAAATCACTCTTTCTT	7	AAGGAGAATAATTCCAGAACAGAGAG
FAM20C_1	9	ACATGGATCGACATCACTAYGAAACC	10	AAGGCACCAATATGGAAAGTTCATC
FAM20C_2	7	AGGAAAACCTGGAGAAACCCYTGGA	8	GGGTGTTGTTTAACCTCTTCACAA
GET4_1	6	ACAGTRTTGTGTGAACAGTATCAACC	7	RAAGAAAAGCTGTCCTATTCTATCT
GET4_3	2	GCTGATCTGTCCATGYTGGTTTTGG	3	TTCCACAGGGTAATRGCTAGTAACTG

*: Exon number of chicken reference from Ensembl (<http://asia.ensembl.org/index.html>)

increased till the computer program found a set of introns distinguish maximum number of species. The computer program was written with Ruby 2.5.0 (Fig. S6) and performed in a Linux PC (Ubuntu 16.04, Core-i7, 32GB memory). The execution time was measured nine times.

Results

Introns on referential genomes

We picked up 144,565 to 158,376 exons from chicken, turkey, zebra finch and flycatcher genomes as reference genomes as candidates for EPIC primer design. Homology search indicated 59,274 exons were found in all four referential genomes (Table S1). In 53,381 exons, intron existed down of each in translation direction (Table S2). In these intron sets, 25,123 intron sets included only under 1kbp introns which were suitable for PCR amplification (Table S3). In these under 1kbp-intron sets, intron sizes were different each other in 19,821 sets of introns (Table S4). In these sets of introns, the difference of 14.8% intron set was less than 10 bases. Similarly, the

difference of 21.3% intron set was less than 20 bases which were suitable for MultiNA (data not shown).

For further experimentations, 16 pairs of exons were selected according to sizes of introns in four standard genomes and its diversities; and EPIC PCR was performed on the 12 species of birds. The designed primer sequences are shown in Table 1 (EPIC primers). The sixteen set of primers produced PCR product on 10 to 12 species. Sizes of PCR products analyzed by 1% agarose gel electrophoresis were indicated in Table 2. Primer set designed for TEAD3, TBC1D22B, AHCYL1_1, PHTF1_1, PHTF1_2, RSBN1_1, AMPD1, CSDE1, PDXDC1, BFAR, FAM20C_1, GET4_1 were amplified on 12 species. Primer sets designed for LOC395100, AHCYL1_2 introns amplify on 11 species and those designed for FAM20C_2, GET4_3 amplify on 10 species (Table 2). In some cases, multiple band were shown on the gel electrophoresis. From multiple bands, the heterozygosity of the intron sizes or artifacts during amplification were suggested. Moreover, it was decided that multiple bands could not be used for species identification.

Table 2. Sizes of EPIC-PCR products amplified with primer sets listed in Table 1

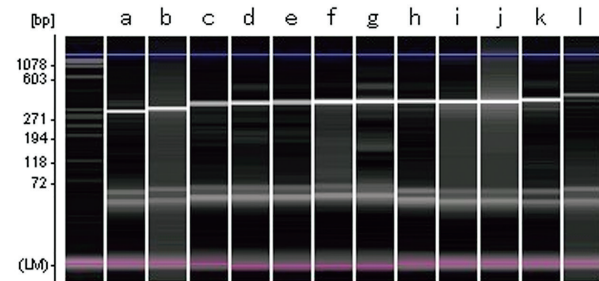
Primer name	CK	BK	JC	CC	BB	TD	RD	SP	SW	MT	SL	BW
TEAD3*	260	280	270	270	290	270	270	280	265	265	260	290
TBC1D22B*	365	400	380	380	350	320	320	370	360	355	360	380
LOC395100	M	M	M	M	M	M	M	M	800	800	M	-
AHCYL1_1*	450	400	400	400	400	400	400	400	400	400	400	400
AHCYL1_2	260	480	860	860	1000	460	460	1000	960	-	790	860
PHTF1_1*	450	220	450	450	500	210	220	450	500	500	450	500
PHTF1_2*	530	700	700	700	700	M	M	700	700	700	690	700
RSBN1_1*	680	990	990	M	1000	1000	M	M	1000	1000	M	M
AMPD1*	660	D	430	430	450	390	390	450	420	420	390	450
CSDE*1	M	330	D	D	350	320	320	350	370	370	370	350
PDXDC1*	190	200	200	200	200	200	200	200	200	200	200	200
BFAR*	860	860	860	860	600	840	840	800	M	600	730	800
FAM20C_1*	510	200	200	200	200	200	200	190	200	200	200	190
FAM20C_2	600	-	600	600	360	700	600	550	550	550	-	550
GET4_1*	380	235	M	M	M	235	240	M	230	230	M	M
GET4_3	-	1260	430	420	435	1260	-	200	435	435	M	430

*: Primers which could amplify each gene in every 12 species. CK: chicken, BK: black kite, JC: jungle crow, CC: carrion crow, BB: brown-eared bulbul, TD: turtle dove, RD: rock dove, SP: sparrow, SW: swallow, MT: martin, SL: starling, BW: black-backed wagtail, M: Multiple bands, D: Double bands.

Identification of species from single intron TBC1D22B

Size differences on PCR product size suggested the maximum distinguishable avian species were 10 in 12 species. Two sets of avian species, jungle crow v.s carrion crow and swallow v.s martin showed small differences (≤ 5 base) on the electrophoresis (Table 2). Because these species are closely related to phylogenetic trees, it may be difficult to find its genetic polymorphism (Fig. S3). The search proceeded with the 10 species as the largest distinguishable species. Size differences were reproductively detected on MultiNA chip electrophoresis (Fig. 2). MultiNA machine predicts molecular size from the moving speed of each band and thus, in some case, the machine detects size differences even if one cannot detect differences on a graphical image which was produced by MultiNA. Table 3 shows the predicted molecular sizes of PCR products. While, the MultiNA-predicted sizes were not the same as those determined from sequences of the PCR products by Sanger method (Table 3), both MultiNA and Sanger sequence results were reproducible in other individuals of the same species (data not shown). Sizes of the TBC1D22B

A



B

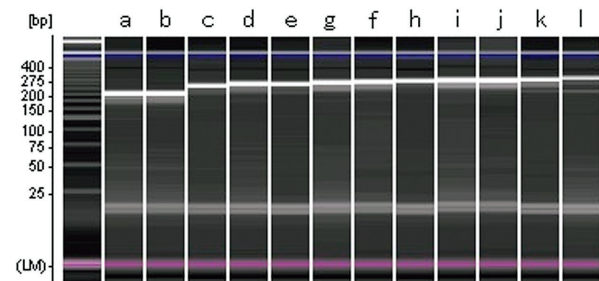


Fig. 2. MultiNA chip electrophoresis of TBC1D22B-primer (A) or TBC1D22B-i-primer amplified PCR. Species are rock dove (a), turtle dove (b), brown-eared bulbul (c), martin (d), swallow (e), starling (f), chicken (g), sparrow (h), carrion crow (i), jungle crow (j), black-backed wagtail (k) and black kite (l). Molecular sized were predicted from internal markers listed in left-side of the graphical view. *: The order of the fragment sizes in chicken and starling was different from that of TBC1D22B original primers (Fig. 2).

Table 3. Molecular sizes of TBC1D22B PCR products predicted by MultiNA and determined sequences by Sanger-method

		RD	TD	BB	MT	SW	SL	CK	SP	CC	JC	BW	BK
MultiNA	Tissue DNA	322	324	351	355	358	361	363	370	378	380	379	403
	Feather DNA	–	–	346	366	368	371	375	383*	380	382	375*	398
GeneMapper		320.8	320.7	355.0	361.6	362.7	375.9	372.9	377.2	383.1	381.1	386.7	400.4
Sanger Sequence		321	321	354	–	366	374	375	378	384	384	388	–

RD: rock dove, TD: turtle dove, BB: brown-eared bulbul, MT: martin, SW: swallow, SL: starling, CK: chicken, SP: sparrow, CC: carrion crow, JC: jungle crow, BW: black-backed wagtail, BK: black kite, –: not determined.

Table 4. Molecular sizes of TBC1D22B_i PCR products predicted by MultiNA and determined sequences by Sanger-method

		RD	TD	BB	MT	SW	SL	CK	SP	CC	JC	BW	BK
MultiNA	Tissue DNA	200	200	237	248	247	255	255	266	269	268	273	285
	Feather DNA	203	204	237	248	248	260	259	269	268	267	271	287
Sanger Sequence		200	200	235	–	244	256*	254*	259	263	263	269	–

RD: rock dove, TD: turtle dove, BB: brown-eared bulbul, MT: martin, SW: swallow, SL: starling, CK: chicken, SP: sparrow, CC: carrion crow, JC: jungle crow, BW: black-backed wagtail, BK: black kite, –: not determined.

PCR products could not separate turtle dove from rock dove, carrion crow from jungle crow, swallow from martin and chicken from starling. In these sets of species, turtle dove and rock dove, carrion crow and jungle crow, swallow and martin are closely related species. We also analyzed the PCR products by capillary electrophoresis and GeneMapper fragment-analysis software (Table 3). Predicted sizes from GeneMapper were different from both MultiNA and Sanger sequences. On some species, size order was different from both Sanger sequence and MultiNA-determined molecular sizes. Thus the 10 avian species could not be identified according to TBC1D22B intron. Resultantly, the 10 avian species could not be identified using any of single intron.

Species identification by multi-locus EPIC

An attempt to maximally identify the species by combining multiple introns was done in here since it could not be distinguished with the single intron. Since there are cases where many species are targeted by EPIC method, algorithms were prepared that can select a combination of a large

number of introns as well as a minimum number of introns for many target species. A number of indistinguishable species were used as indices for distinguishing species. In order to determine the maximum number of species from intron sizes, combinations of introns were searched which resulted in the smallest number of indistinguishable pair of species. The algorithm firstly created indistinguishable species-pair and search the least set for minimizing indistinguishable species with logical AND operations.

When the computer program used for 2 introns, the smallest number of indistinguishable species pairs were shown on two introns-combinations: AMPD 1 - TBC 1 D 22 B and AHCYL 1_2 - BFAR 4 (Table 5A). There were no particular tendency on the combination of indistinguishable species. When the combination was made with 3 introns, the smallest number of indistinguishable species pairs became 2 (Table 5B). The smallest number of indistinguishable species pairs with 3 introns were the same as that with all introns. Thus, the combination of three introns is

Table 5. A combination of species that could not be distinguished even by combination of multiple introns

intron_set	Table 5A: Combination of 2 introns															
	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK	CK
intron_set	JC	CC	MT	BK	MT	BK	TD	RD	BK	BK	CC	SW	BK	BK	CC	SW
	JC	CC	MT	BK	MT	BK	TD	RD	BK	BK	CC	SW	BK	BK	CC	SW
AMPD1 TBC1D22B																
AHCYL1_2 BFAR																
AHCYL1_2 PHTF1_1																
AHCYL1_2 FAM20C_2																
AMPD1 PHTF1_1																
AHCYL1_2 AMPD1																
BFAR PHTF1_1																
AHCYL1_2 TBC1D22B																
AHCYL1_2 CSDE1																
AMPD1 BFAR																

Table 5B: Combination of 3 introns

intron_set	JC									
	CC	CC	MT	BK	MT	BK	TD	RD	BK	CC
AHCYL1_2 BFAR FAM20C_2										
AHCYL1_2 PHTF1_1 TEAD3										
AHCYL1_2 BFAR TEAD3										
AHCYL1_2 AMPD1 TBC1D22B										
AMPD1 PHTF1_1 TBC1D22B										
AMPD1 FAM20C_2 TBC1D22B										
AMPD1 GET4_3 TBC1D22B										
AHCYL1_2 AMPD1 PHTF1_1										
AHCYL1_2 CSDE1 PHTF1_1										
AHCYL1_2 FAM20C_2 PHTF1_1										

*The first row and the second row indicate a combination of species that could not be distinguished with each combination of introns. Combinations of introns were shown separated by vertical bars. CK: chicken, BK: black kite, JC: jungle crow, CC: carrion crow, BB: brown-eared bulbul, TD: turtle dove, RD: rock dove, SP: sparrow, SW: swallow, MT: martin, SL: starling, BW: black-backed wagtail, ×: species cannot be distinguished using the intron set indicated leftmost column of the row.

suggested being the best for 12 species of birds. The calculation for minimum combination took 0.667 ± 0.013 seconds.

Discussions

As shown in Table 3, the differences in predictions for TBC1D22B the PCR product sizes by Sanger sequences, GeneMapper or MartiNA were presented. In Sanger sequence determination, a capillary sequencer corrected original electrophoresis results according to manufacturer's database for DNA-sequence^{15,31}. Hence, the nucleotide number was found to be most accurate in three predictions. For size prediction with GeneMapper, the Fluorescent-labelling treatments resulted in notable differences from that of the Sanger sequence²². MartiNA predicted reproducible sizes that differ from those by Sanger sequence²⁷. Previous authors have reported that small error between runs and differences of the predicted size from the actual size can occur⁸. Error in base length prediction by electrophoresis are indicated by the manufacturer (<https://www.an.shimadzu.co.jp/bio/mce/multina/spec.htm>, in Japanese) to be at 5% of the nucleotide sequence for the range of 100 bases to 500 bases. The results of these reports and the present study predicted sizes by MultiNA may differ from the actual size, but the predicted sizes were reproducible between trials. Consequently, predicted sizes of TBC1D22B PCR products by MultiNA were found to be imprecise, while differences between avian species were reproducible.

On species-identification of wild birds, the EPIC method needs as an alternative method for DNA barcoding. NCBI Taxonomy (<https://www.ncbi.nlm.nih.gov/Taxonomy>) listed 15,417 species for aves with over ten thousand bird species that are also listed in Avibase database (<https://avibase.bsc-eoc.org/avibase.jsp>). On the species that cannot be determined by DNA barcoding, it was described by Barreira *et al.* (2016) that DNA

barcoding could not determine about 4% of target species⁵. From all birds species in taxa and DNA barcoding ability, DNA barcoding cannot determine at least 650 bird species. Alternative methods against DNA barcoding for species-identification has been discussed in fishes and EPIC has been recognized as a suitable for complement for DNA barcoding²¹. Barcoding with COI requires the amplification process of the COI gene for sequencing, but in PCR it is possible that the less frequent base sequence disappears through amplification steps. Thus, it is desirable to use other methods in combination for improvement reproducibility and reliability¹. In addition, the high cost of the Barcoding suggested to be the problem of species-identification using COI.

The nucleotide sequences of EPIC primers based on the exons were suggested to be evolutionary conserved^{10,30}. Assuming from conservative status of the exons, EPIC primers could be designed using the known exon sequences of standard species. The success rate of primers designed in this study based on four standard bird genome information was only about 37% in the trials (Table S4). This low success rate might be due to the inappropriateness of the primer-design algorithm. Most primers were designed on the end of the exon and therefore, other method/s for selecting consensus sequence might be necessary. If the information on exon sequence increased and selection algorithm for consensus sequence improved, the improved PCR success rate can be obtained. In the current situation, reference can be obtained both from the published genomic sequences and the improved primer-design algorithms. NCBI Genome site disclosed genomic sequences on 96 species of birds (<https://www.ncbi.nlm.nih.gov/genome>). A transcripts and exon prediction software Augustus predicts introns on those sequences^{28,29}. Further, new generation of sequencer such as Nanopore sequencer might speed up genome analysis of wild animals^{18,20}. Thus, EPIC method suggested to be the prime alternative method for DNA

barcoding.

To examine the use of EPIC method in birds, one should consider a number of subject species. Although NCBI Taxonomy published 15,714 avian species, an actual necessity of species refer to ecological studies which identified species on a certain region. DNA barcoding studies described 600 species for North America, 500 species for Argentina and 561 species for the low-latitude tropical regions⁵⁾. For the narrow region, 141 species for Netherlands, 154 species for Korea investigated by DNA barcoding²⁾. From these regional reports, identification of about 150 species might be suitable for a local region such as Hokkaido. As for the number of available introns for EPIC, from the presence in reference genome, sizes of the introns and polymorphism of the sizes, a number of subject introns calculated to be 19,000 to 25,000 (Table S1). Assuming 20,000 introns subjected to determine 150 species, the calculation time required to get the optimal combination of introns predicted to be 30,000 times longer than this research. From the result that 0.7 second to get the least combination for maximum range of species-identification, we will get the optimum combination in 21,000 seconds (= 5.8 hours). The calculation length is acceptable for a series of an experiment. From the discussion above, the EPIC method is the suitable method for DNA-mediated determination of avian species in each area.

In the observation in four standard genomes, many differences in intron size were less than 10 bases. Thus, a method detect 10-base difference will improve the efficiency of the EPIC method. In this study, MultiNA chip gel electrophoresis detected for a difference of 5 bases but capillary sequencer showed different results (Table 3). In whole trials of EPIC (Table S3), about 30% of intron-size difference between species could be detected only with MultiNA chip gel electrophoresis. If one uses the capillary sequencer with Gene Mapper software, EPIC method can return more reproducible results. While, capillary sequencer method needs high cost and may

reduce applicability of EPIC. In recent years, Nguyen *et al.* (2016) described high accuracy methods for predicting the size of PCR products by image-analysis of electrophoresis with Monte Carlo simulation²³⁾. There have been reported on the improvement of DNA band-analysis by digital image-analysis^{3,9)}. With this, it is expected that EPIC can become appreciable for many species with low cost and short working time. It is suggested that EPIC method with the selection algorithm for intron-combination be utilized as a method for species identification.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number JP24580430 (D.E.), the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT)-Supported Program for the Strategic Research Foundation at Private Universities, 2013–2017 (D.E.) and also supported in part by the priority research (2502-11E004) of Hokkaido Research Organization (Y.O.).

Supplemental data

Supplemental data associated with this article can be found, in the online version, at <http://dx.doi.org/10.14943/jjvr.66.4.261>

References

- 1) Acinas SG, Sarma-Rupavtarm R, Kiepac-Ceraiv, Polz MF. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Applied and environmental microbiology* 71, 8966–8969, 2005
- 2) Aliabadian M, Beentjes KK, Roselaar CS, van Brandwijk H, Nijman V, Vonk R. DNA barcoding of Dutch birds. *ZooKeys*, 25–48, 2013
- 3) Antiabong JF, Ngoepe MG, Abechi AS. Semi-quantitative digital analysis of polymerase

- chain reaction-electrophoresis gel: Potential applications in low-income veterinary laboratories. *Vet World.*, 9, 935–939, 2016
- 4) Backstrom N, Karaïskou N, Leder EH, Gustafsson L, Primmer CR, Qvarnstrom A, Ellegren H. A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. *Genetics*, 179, 1479–1495, 2008
 - 5) Barreira AS, Lijtmaer DA, Tubaro PL. The multiple applications of DNA barcodes in avian evolutionary studies. *Genome*, 59, 899–911, 2016
 - 6) Bierne N, Lehnert SA, Bédier E, Bonhomme F, Moore SS. Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. *Mol Ecol.*, 9, 233–235, 2000
 - 7) Burt DW, Bruley C, Dunn IC, Jones CT, Ramage A, Law AS, Morrice DR, Paton IR, Smith J, Windsor D, Sazanov A, Fries R, Waddington D. The dynamics of chromosome evolution in birds and mammals. *Nature* 402, 411–413, 1999
 - 8) Burrell A, Foy C, Burns M. Applicability of three alternative instruments for food authenticity analysis: GMO identification. *Biotechnol Res Int.* 2011: 838232, doi: 10.4061/2011/838232, 2011
 - 9) Caridade CR, Margal AS, Mendonga T, Pessoa AM, Pereira S. An automatic Method to identify and extract information of DNA bands in Gel Electrophoresis images. *Conf Proc IEEE Eng Med Biol Soc* 1024–1027, 2009
 - 10) Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le Ann, Bouffard P, Burt DW, Crasta O, Crooijmans RP, Cooper K, Coulombe RA, De S, Delany ME, Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MA, Harkins TT, Herrero J, Hoffmann S, Megens HJ, Jiang A, de Jong P, Kaiser P, Kim H, Kim KW, Kim S, Langenberger D, Lee MK, Lee T, Mane S, Marcais G, Marz M, McElroy AP, Modise T, Nefedov M, Notredame C, Paton IR, Payne WS, Perteau G, Prickett D, Puiu D, Qiao D, Raineri E, Ruffier M, Salzberg SL, Schatz MC, Scheuring C, Schmidt CJ, Schroeder S, Searle SM, Smith EJ, Smith J, Sonstegard TS, Stadler PF, Tafer H, Tu ZJ, Van Tassell CP, Vilella AJ, Williams KP, Yorke JA, Zhang L, Zhang HB, Zhang X, Zhang Y, Reed KM. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* 8 pii: e1000475, 2010
 - 11) Datukishvili N, Kutateladze T, Gabriadze I, Bitskinashvili K, Vishnepolsky B. New multiplex PCR methods for rapid screening of genetically modified organisms in foods. *Front Microbiol* 6, 757, 2015
 - 12) Doi K, Imai H. A Greedy Algorithm for Minimizing the Number of Primers in Multiple PCR Experiments. *Genome Inform Ser Workshop Genome Inform* 10, 73–82, 1999
 - 13) Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461, 2010
 - 14) Fouchier RAM, Munster VJ. Epidemiology of low pathogenic avian influenza viruses in wild birds. *Revue scientifique et technique* 28, 49–50, 2009
 - 15) Giddings MC, Brumley RL Jr, Haker M, Smith LM. An adaptive, object oriented strategy for base calling in DNA sequence analysis. *Nucleic Acids Res* 21, 4530–4540, 1993
 - 16) Gostel MR, Weeks A. Development of novel, exon-primed intron-crossing (EPIC) markers from EST databases and evaluation of their phylogenetic utility in Commiphora (*Burseraceae*). *Appl Plant Sci* 2 pii: apps.1300098, 2014
 - 17) Hebert PD, Stoeckle MY, Zemlak TS, Francis CM. Identification of birds through DNA barcodes. *PLoS biology* 2, e312, 2004
 - 18) Jenjaroenpun P, Wongsurawat T, Pereira R, Patumcharoenpol P, Ussery DW, Nielsen J, Nookaew I. Complete genomic and transcriptional landscape analysis using third-generation sequencing: a case study of *Saccharomyces cerevisiae* CEN.PK113-7D. *Nucleic Acids Res* 46, e38, 2018
 - 19) Koressaar T, Remm M. Enhancements and modifications of primer design program Primer3. *Bioinformatics* 23, 1289–1291, 2007
 - 20) de Lannoy C, de Ridder D, Risse J. The long reads ahead: de novo genome assembly using the MinION. Version 2. *F1000Res* 6, 1083, 2017
 - 21) Li C, Riethoven JJ, Ma L. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. *BMC Evol Biol* 10, 90, 2010
 - 22) McEvoy CR, Seshadri R, Firgaira FA. Large DNA fragment sizing using native acrylamide gels on an automated DNA sequencer and GENESCAN software. *Biotechniques* 25, 464–470, 1998
 - 23) Nguyen PV, Ghezel A, Hsueh YC, Boudier T, Gan SK, Lee HK. Optimal processing for gel

- electrophoresis images: Applying Monte Carlo Tree Search in GelApp. *Electrophoresis*, 37, 2208–2216, 2016
- 24) Nunes AR, Alves BEB, Pereira HW, Nascimento YM, Morais IC, Fernandes JV, Araujo JM, Lanza DC. Improved reverse transcription-polymerase chain reaction assay for the detection of flaviviruses with semi-nested primers for discrimination between dengue virus serotypes and Zika virus. *Mem Inst Oswaldo Cruz*. 113, e170393, 2018
 - 25) Schäffer S, Zachos FE, Koblmüller S. Opening the treasure chest: A DNA-barcoding primer set for most higher taxa of Central European birds and mammals from museum collections. *PLoS One* 12, e0174449, 2017
 - 26) Paul SH. *Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics*. Oxford University Press, Oxford. 2011.
 - 27) Shimadzu. Microchip Electrophoresis system for DNA/RNA, MCE-202 MultiNA Instruction Manual Operating Procedure, 292-28464D, 2008
 - 28) Stanke M, Steinkamp R, Waack S, Morgenstern B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32, W309–312, 2004
 - 29) Stanke M, Tzvetkova A, Morgenstern B. AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biol* 7 Suppl 1, S11.1–8, 2006
 - 30) Zhang Q, Edwards SV. The evolution of intron size in amniotes: a role for powered flight? *Genome biology and evolution* 4, 1033–1043, 2012
 - 31) Zianni M, Tessanne K, Merighi M, Laguna R, Tabita FR. Identification of the DNA bases of a DNase I footprint by the use of dye primer sequencing on an automated capillary DNA analysis instrument. *J Biomol Tech* 17, 103–113, 2006