

Determining the Position of Storks on the Phylogenetic Tree of Waterbirds by Retroposon Insertion Analysis

Tae Kuramoto^{1,†}, Hidenori Nishihara^{1,†}, Maiko Watanabe², and Norihiro Okada^{1,3,4,*}

¹Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama, Kanagawa, Japan

²Division of Microbiology, National Institute of Health Sciences, Setagaya, Tokyo, Japan

³Foundation for Advancement of International Science, Tsukuba, Ibaraki, Japan

⁴Department of Life Sciences, National Cheng Kung University, Tainan, Taiwan

*Corresponding author: E-mail: nokada@fais.or.jp.

†These authors contributed equally to this work.

Accepted: October 23, 2015

Data deposition: This project has been deposited at DDBJ under the accessions LC072269-LC072649.

Abstract

Despite many studies on avian phylogenetics in recent decades that used morphology, mitochondrial genomes, and/or nuclear genes, the phylogenetic positions of several birds (e.g., storks) remain unsettled. In addition to the aforementioned approaches, analysis of retroposon insertions, which are nearly homoplasy-free phylogenetic markers, has also been used in avian phylogenetics. However, the first step in the analysis of retroposon insertions, that is, isolation of retroposons from genomic libraries, is a costly and time-consuming procedure. Therefore, we developed a high-throughput and cost-effective protocol to collect retroposon insertion information based on next-generation sequencing technology, which we call here the STRONG (Screening of Transposons Obtained by Next Generation Sequencing) method, and applied it to 3 waterbird species, for which we identified 35,470 loci containing chicken repeat 1 retroposons (CR1). Our analysis of the presence/absence of 30 CR1 insertions demonstrated the intra- and interordinal phylogenetic relationships in the waterbird assemblage, namely 1) Loons diverged first among the waterbirds, 2) penguins (Sphenisciformes) and petrels (Procellariiformes) diverged next, and 3) among the remaining families of waterbirds traditionally classified in Ciconiiformes/Pelecaniformes, storks (Ciconiidae) diverged first. Furthermore, our genome-scale, in silico retroposon analysis based on published genome data uncovered a complex divergence history among pelican, heron, and ibis lineages, presumably involving ancient interspecies hybridization between the heron and ibis lineages. Thus, our retroposon-based waterbird phylogeny and the established phylogenetic position of storks will help to understand the evolutionary processes of aquatic adaptation and related morphological convergent evolution.

Key words: Aves, Ciconiiformes, chicken repeat 1, incomplete lineage sorting, next-generation sequencing.

Introduction

For over a century, the phylogeny of birds (*Aves*) has been studied based on morphology (Huxley 1867; Cracraft 1981; Livezey and Zusi 2007) and molecular analyses, for example, DNA–DNA hybridization (Sibley and Ahlquist 1990), mitochondrial genomics (Morgan-Richards et al. 2008; Pacheco et al. 2011; Gibb et al. 2013), and nuclear genomics (Fain and Houde 2004; Ericson et al. 2006; Hackett et al. 2008). These efforts, especially thanks to the recent availability of genomic data, have led to a general understanding of the evolutionary history of modern birds (Jarvis et al. 2014), although several questions remain unanswered.

One long-standing controversial issue is the phylogenetic relationships between the waterbird assemblage (*Aequornithes*; Mayr 2011), including Procellariiformes, Sphenisciformes, Gaviiformes, and the traditional orders of Ciconiiformes (Ciconiidae, Ardeidae, Threskiornithidae, Balaenicipitidae, Scopidae), and Pelecaniformes (Pelecanidae, Sulidae, Phalacrocoracidae, Anhingidae, Fregatidae). An evolutionary close relationship among these families has been supported by morphological and molecular studies (Sibley and Ahlquist 1990, Fain and Houde 2004; Ericson et al. 2006; Livezey and Zusi 2007; Hackett et al. 2008; Pacheco et al. 2011; Gibb et al. 2013). However, the phylogenetic

positions of certain waterbird families, for example, Ciconiidae (storks), have not been established and are still puzzling even when molecular analyses were used (Hackett et al. 2008, Gibb et al. 2013, Kimball et al. 2013; fig. 1).

The Ciconiidae family was first classified as a traditional Ciconiiformes along with Threskiornithidae and allies (Bonaparte 1854; Wetmore 1960), but thereafter the phylogenetic positions of Ciconiiformes members were found to vary depending on the study (Garrod 1874; Olson 1979; Cracraft et al. 2004). One study using DNA–DNA hybridization proposed that Ciconiidae, including condors, was a sister group to Pelecanidae (Sibley and Ahlquist 1990), whereas another suggested a sister relationship between Ciconiidae and the Procellariiformes + Sphenisciformes clade (Van Tuinen et al. 2001). Full-length mitochondrial genomic sequencing yielded different results owing possibly to taxon sampling and/or the analytical method employed (Watanabe, Nikaido, Tsuda, Kobayashi, et al. 2006; Pacheco et al. 2011; Gibb et al. 2013). For example, one study proposed that Ardeidae are the closest relatives of Ciconiidae (Gibb et al. 2013), whereas another suggested a sister relationship between Ciconiidae and Sphenisciformes (penguins) (Pacheco et al. 2011). Recent nuclear gene analyses (Hackett et al. 2008; Kimball et al. 2013; Jarvis et al. 2014) support the idea that the traditional orders of Pelecaniformes and Ciconiiformes are polyphyletic and that flamingos, which have been traditionally classified as Ciconiiformes, should be excluded from the waterbird assemblage (Hackett et al. 2008; Kimball et al. 2013; Jarvis et al. 2014). Accordingly, the latest avian classification defines Pelecanidae, Ardeidae, Threskiornithidae, Balaenicipitidae, and Scopidae as Pelecaniformes, and Ciconiidae as the only member of Ciconiiformes (Gill and Donsker 2015). Suloidea (Sulidae + Phalacrocoracidae + Anhingidae; Cracraft 1985) and Fregatidae were excluded from Pelecaniformes and designated as the order Suliformes (supplementary table S1, Supplementary Material online). However, the phylogenetic position of Ciconiidae has yet to be determined with a sufficiently great bootstrap probability (BP) even for large-scale studies, that is, those that used 19 or 50 nuclear gene sequences (fig. 1C and D; Hackett et al. 2008; Kimball et al. 2013). A recent genome-scale study that resolved many avian phylogenetic issues did not include Ciconiidae (Jarvis et al. 2014).

Certain locomotion-related morphological characteristics seem to have convergently evolved in marine birds (Livezey and Zusi 2007; Felice and O'Connor 2014). Also, in Ciconiiformes, some of their morphological features have been proposed to be a result of evolutionary convergence (Olson 1979). If the phylogenetic position of Ciconiidae could be unambiguously established, it would be possible to identify the characteristics that have been subjected to convergent evolution and to explain why waterbird phylogenetics has been perplexing for decades. Therefore, determining the

phylogenetic position of Ciconiidae is a very important yet long-standing puzzle in avian phylogenetics.

Another important issue concerning waterbirds is the phylogenetic relationship among herons (Ardeidae), ibises (Threskiornithidae), and the pelican-related group (Pelecanidae + Balaenicipitidae + Scopidae). A 19-loci nuclear gene analysis proposed that ibises are evolutionarily closer to herons than pelicans (BP = 72; fig. 1C; Hackett et al. 2008), whereas a genome-scale study concluded that pelicans and herons form a monophyletic group (BP = 100) (Jarvis et al. 2014). To elucidate the evolutionary history of these birds, other molecular phylogenetic markers (e.g., retroposon insertions) should be characterized.

Retroposons, such as short and long interspersed elements (SINEs and LINEs), are transposable elements that propagate via reverse transcription of their RNA intermediates and subsequent integration into host genomes (Okada et al. 1997; Deininger and Batzer 2002). Because there is exceedingly small likelihood that retroposons would be precisely excised or independently inserted into exactly the same orthologous site in different lineages, a retroposon insertion is a unique, powerful, nearly homoplasy-free phylogenetic molecular marker (Shedlock and Okada 2000). These features of retroposons have allowed researchers to resolve difficult phylogenetic issues for various species, for example, fish (Murata et al. 1993), turtles (Sasaki et al. 2004), mammals (Shimamura et al. 1997; Nikaido et al. 1999, 2001; Kriegs et al. 2006; Nishihara et al. 2006, 2009; Churakov et al. 2009), and birds (Watanabe, Nikaido, Tsuda, Inoko, et al. 2006; Kaiser et al. 2007; Kriegs et al. 2007; Suh et al. 2011, 2012; Haddrath and Baker 2012).

The chicken repeat 1 retroposon (CR1) family belongs to LINEs (Stumph et al. 1981; Burch et al. 1993; Suh et al. 2014) and represents the most active transposable element in avian genomes, for example, it accounts for ~85% of all transposable elements in chicken (Hillier et al. 2004). Although a full-length CR1 is ~4.5 kb, most are truncated at the 5' end owing to incomplete integration that starts at the 3' end (Vandergon and Reitman 1994; Hillier et al. 2004). For example, >90% of CR1s are found to be <1 kb in length in the chicken genome (unpublished data). The abundance of CR1s has allowed researchers to use them as avian phylogenetic markers (Watanabe, Nikaido, Tsuda, Inoko, et al. 2006; Kaiser et al. 2007; Kriegs et al. 2007; Suh et al. 2011, 2012; Haddrath and Baker 2012) and more detailed in the analysis of the waterbirds (Watanabe, Nikaido, Tsuda, Inoko, et al. 2006), although, in that study, the position of Ciconiidae was not resolved. In general, retroposon insertion analysis requires construction of a genomic library to screen for CR1-containing clones by Southern hybridization (Watanabe, Nikaido, Tsuda, Inoko, et al. 2006; Nishihara and Okada 2008), except when genomic sequence data of a reference species are available in a public database (Kaiser et al. 2007; Kriegs et al. 2007; Suh et al. 2011). Moreover, it

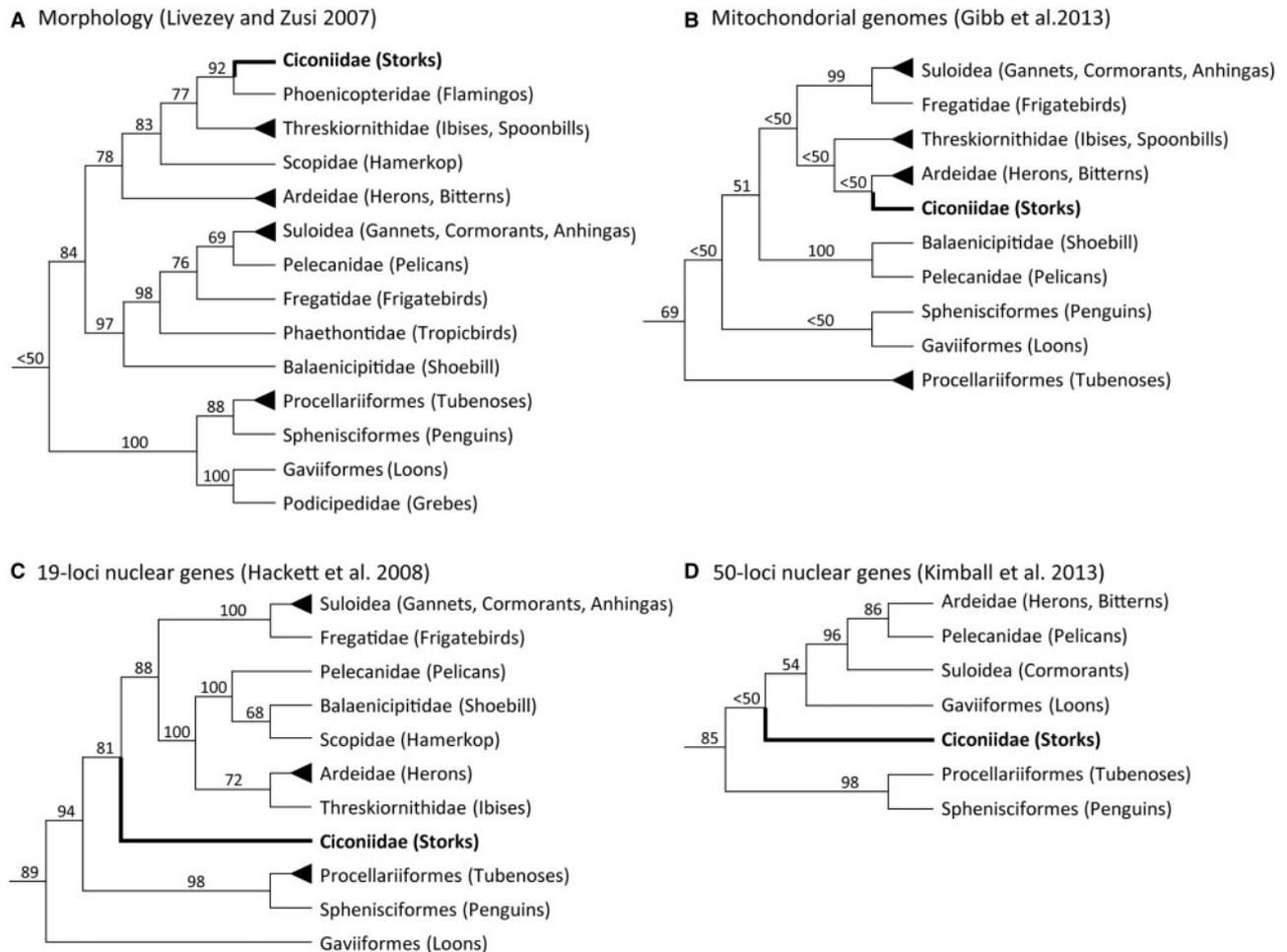


FIG. 1.—Four representative waterbird phylogenetic trees that include storks based on (A) morphology (Livezey and Zusi 2007), (B) mitochondrial genomic analysis (Gibb et al. 2013), (C) 19 nuclear loci analysis (Hackett et al. 2008), and (D) 50 nuclear loci analysis (Kimball et al. 2013). Values above each node are bootstrap probabilities. The Ciconiidae lineages are highlighted.

is difficult to determine early divergent branching points using retrotransposons because it is difficult to isolate older CR1s with intense accumulated mutations and consequently because of a biased isolation of recently inserted ones. In addition, screening by hybridization requires a huge effort and its cost-effectiveness is low. A more cost-effective polymerase chain reaction (PCR)-based protocol for loci with inserted CR1s has been proposed (Suh et al. 2012). Also, another extensive protocol using PCR and magnetic beads for isolation of a certain SINE has been proposed, but it still requires substantial experimental effort (Platt et al. 2015).

To overcome the inconvenience of traditional hybridization-based methods used for avian phylogenetics, we developed a cost-effective protocol for comprehensive CR1 searching based on next-generation sequencing (NGS) technology as reported herein. We used this protocol, designated as the STRONG (Screening of Transposons Obtained by Next Generation Sequencing) method, to analyze retrotransposon insertions in waterbird genomes, which allowed us to clarify their

order/family-level phylogeny and the phylogenetic position of Ciconiidae. In addition, a comprehensive CR1 analysis using a recently compiled genomic data set allowed us to uncover the ancient, complex evolutionary history among Ardeidae, Threskiornithidae, and Pelecanidae lineages, which appears to involve the past interspecies hybridization.

Materials and Methods

The STRONG Method; NGS-based Identification of CR1s

The STRONG method was schematically shown in figure 2. We collected 35,470 loci with CR1 insertions from the genomes of the white-chinned petrel (*Procellaria aequinoctialis*), little egret (*Egretta garzetta*), and brown pelican (*Pelecanus occidentalis*) in total. First, genomic DNA from each of the three species was extracted from muscle or liver tissue according to Sambrook et al. (1989). DNA fragments were used to construct paired-end libraries containing ~170 bp inserts for NGS. High-throughput paired-end

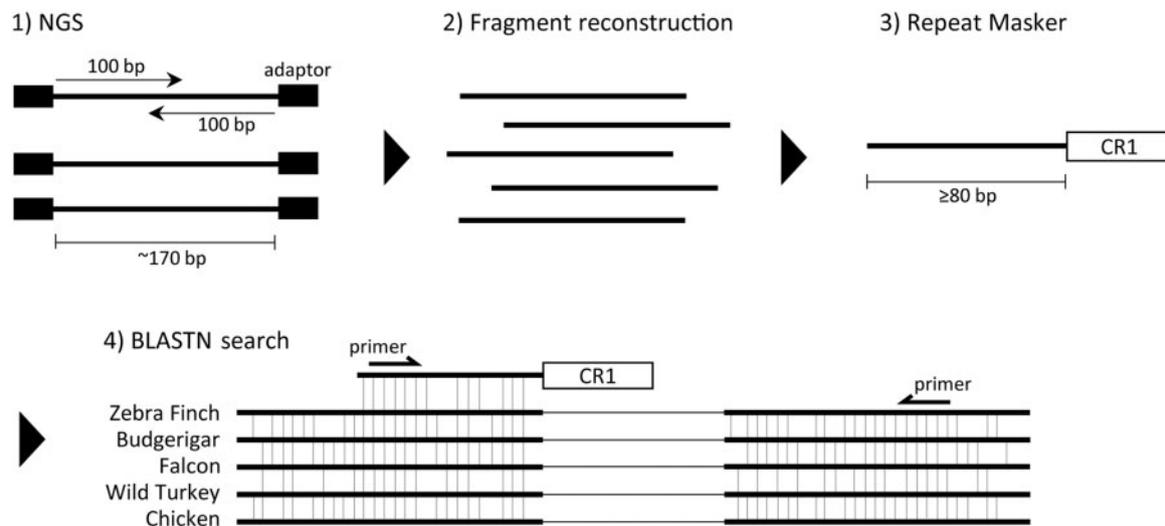


FIG. 2.—The STRONG method used in this study. The procedure involves 1) paired-end sequencing, 2) reconstruction of sequences of ~170 bp, 3) CR1 search by RepeatMasker, and 4) ortholog searches for the retroposon-flanking sequences.

sequencing was conducted with an Illumina HiSeq2000 instrument at Beijing Genomics Institute (Shenzhen, China), and 5.9, 8.4, and 9.9 M pairs of 100 nt in length were obtained for the petrel, egret, and pelican fragments, respectively (table 1). For each read pair, the original insert sequence was reconstructed by finding sequences containing ≥ 15 identical nucleotides for the forward and reverse reads, which resulted in 5.8 M (97.1%), 8.1 M (96.0%), and 9.7 M (97.7%) sequences for the petrel, egret and pelican, respectively. Next, RepeatMasker (<http://www.repeatmasker.org>, last accessed October 31, 2015) in conjunction with the Aves repeat library (version 20140131) was used to find all CR1s in the sequences. Fragments containing a CR1 and a 5'- or 3'-flanking sequence of ≥ 80 nucleotides accounted for 15,168, 62,781, and 65,853 sequences for the petrel, egret, and pelican, respectively. All flanking sequences were searched with BLASTN against the genome of zebra finch (taeGut1 available at the UCSC Genome Browser; <http://genome.ucsc.edu/>, last accessed October 31, 2015), whose genomic data and retroposon annotation were best established among Neoaves. The BLASTN searches were conducted with the parameters $r = 2$, $G = 5$, $E = 2$, and an e-value of $< 1 \times 10^{-10}$. The numbers of orthologs, found as single-hit sequences by BLASTN, were 4,098, 15,114, and 16,258 for the petrel, egret, and pelican, respectively. Based on the zebra finch ortholog information, the corresponding orthologous sequences in the genomes of the medium ground finch (*Geospiza fortis*), budgerigar (*Melopsittacus undulatus*), wild turkey (*Meleagris gallopavo*), and chicken (*Gallus gallus*) were retrieved from the multispecies alignment data (geoFor1.7way.maf available in the UCSC database). In addition, the orthologous sequences in the genomes of the peregrine falcon (*Falco peregrinus*; GenBank: AKMT000000000) and saker falcon (*Falco cherrug*;

GenBank: AKMU000000000) were obtained by BLASTN searches using the zebra finch sequences as queries as described above. The collected sequences were aligned by MAFFT (Kato and Standley 2013). Among the collected loci with CR1 insertions, RepeatMasker analysis revealed that CR1 insertions were absent in zebra finch for 2,684, 11,615, and 11,717 loci from the petrel, egret, and pelican, respectively, which are ready for the following PCR analysis.

Analysis of Inserted Retroposons

PCR primers were designed using relatively conserved flanking regions of each CR1 locus where the CR1 insertion was not present in two distantly related birds, i.e., chicken and zebra finch. For CR1 insertion analysis, we used the genomes of eight waterbird families as well as those of four nonwaterbird species (supplementary table S1, Supplementary Material online). These four species may be closely related to waterbirds (Fain and Houde 2004; Ericson et al. 2006; Livezey and Zusi 2007; Hackett et al. 2008; Jarvis et al. 2014). The presence or absence of each CR1 was characterized for randomly selected 49, 119, and 96 loci from petrel, egret, and pelican DNA, respectively (supplementary table S2, Supplementary Material online). All PCR experiments were performed with an initial denaturation step at 95°C for 3 min, followed by 35 cycles of denaturation at 95°C for 1 min, annealing at 50 or 55°C for 30 sec, and extension at 72°C for 2 min with the use of Ex Taq kit reagents (TaKaRa BIO, Shiga, Japan). After knowing that PCR products were well electrophoretically separated, we sequenced them to confirm that they represented bona fide informative loci. The presence or absence of a CR1 was determined by RepeatMasker after aligning the sequences using MEGA (Tamura et al. 2013). All

Table 1

Number of Read Pairs/Sequences Obtained after Each Step in the Construction of NGS-based CR1 Libraries for the White-chinned Petrel, Little Egret, and Brown Pelican

Species	Read-pairs (step 1)	Reconstructed fragments (step 2)	CR1s (step 3)	Orthologs corresponding to zebra finch loci (step 4)
White-chinned petrel	5,939,158	5,768,938	15,168	4,098
Little egret	8,389,771	8,056,647	62,781	15,114
Brown pelican	9,918,513	9,693,328	65,853	16,258

NOTE.—Data and step number are given in figure 2.

sequences were deposited in DNA Data Bank of Japan (DDBJ) (accession numbers: LC072269–LC072649). In the PCR-based CR1 insertion analysis, the significance for the number of CR1 loci to obtain the confidence of a node was evaluated based on the likelihood ratio test (Waddell et al. 2001).

Comparison of Loci with Inserted CR1s based on Genomic Assemblies

We also performed a genome-scale CR1 insertion analysis using the information reported by Jarvis et al. (2014) concerning the genomic assemblies of the little egret, crested ibis (*Nipponia nippon*), Dalmatian pelican (*Pelecanus crispus*), and great cormorant (*Phalacrocorax carbo*) to complement their evolutionary history. All CR1s in these four genomic assemblies were identified by RepeatMasker using the Aves repeat library (version 20140131). To find orthologs of loci with CR1 insertions, 150 bp upstream and downstream flanking sequences of all CR1s were collected, and after removal of repetitive sequences each set of the flanking sequences was used as a query in a BLASTN search against the other three genomes. The BLASTN search was conducted with the parameters described above. Sequences were considered orthologous and retained if two hits from two independent searches using upstream and downstream sequences of CR1, respectively, were located in approximately the same region, that is, <2 kbp apart, so that most of CR1 insertions can be tested, based on the fact that >98% of all CR1 are <2 kbp due to 5'-truncation in chicken (unpublished data). A sequence alignment of the four orthologous sequences was performed with MAFFT (Katoh and Standley 2013). After removing overlapped data, the presence or absence of a CR1 in each locus was determined by RepeatMasker followed by manual validation. The absence of CR1s in the zebra finch, *Fulmarus*, and *Gavia* was confirmed for all phylogenetically informative loci obtained by examining the orthologs in their genomes (Jarvis et al. 2014). The significance of the number of CR1 loci to obtain the confidence of a node involving the four waterbird lineages was evaluated by chi-square test. The interspecies hybridization was tested based on the KKSC significance test (Doronina et al. 2015; http://retrogenomics.uni-muenster.de:3838/KKSC_significance_test/, last accessed October 31, 2015).

http://retrogenomics.uni-muenster.de:3838/KKSC_significance_test/, last accessed October 31, 2015).

Results

Phylogenetic Position of Ciconiidae on the Waterbird Phylogenetic Tree

The presence or absence of a CR1 was characterized for 49, 119, and 96 loci obtained from the genomes of the white-chinned petrel, little egret, and brown pelican, respectively. Among them, 10, 12, and 8 loci, respectively, were found to be phylogenetically informative (fig. 3 and [supplementary table S3, Supplementary Material](#) online). Five CR1 insertions suggested the monophyly of the waterbird assemblage (Aequornithes) (branch 1 in fig. 3B). Because we did not find a CR1 insertion that was inconsistent with the presence/absence pattern, this conclusion was statistically significant for the [5 0 0] retroposon insertion pattern as shown by the likelihood ratio test ($P=0.0041$; Waddell et al. 2001). The monophyly of Pelecaniformes, Ciconiiformes, Procellariiformes, and Sphenisciformes was strongly indicated by the eight additional CR1 insertions found for branch 2 ($P=0.00015$ [8 0 0]; fig. 3B), suggesting the first divergence of Gaviiformes. From morphological, mitochondrial, and nuclear genetic studies, it has not been fully resolved what species diverged first in the waterbird assemblage (fig. 1). In addition to our previous evidence (Watanabe, Nikaido, Tsuda, Inoko, et al. 2006), this study clearly demonstrated the early divergence history of waterbirds. Our tree is consistent with those produced by nuclear gene analyses (Hackett et al. 2008; Jarvis et al. 2014) but not with that of the 50-loci gene analysis of Kimball et al. (2013) (fig. 1).

Regarding the phylogenetic position of Ciconiidae, we found that seven CR1 loci represent a monophyletic clade for the families traditionally classified as Ciconiiformes and Pelecaniformes (branch 3 in fig. 3B). This clade is a sister group of Austrodyptornithes (the Procellariiformes + Sphenisciformes clade). We found five additional retroposon insertions, suggesting that Ciconiidae had diverged from the traditional group of Ciconiiformes and Pelecaniformes (branch 4 in fig. 3B). We found only

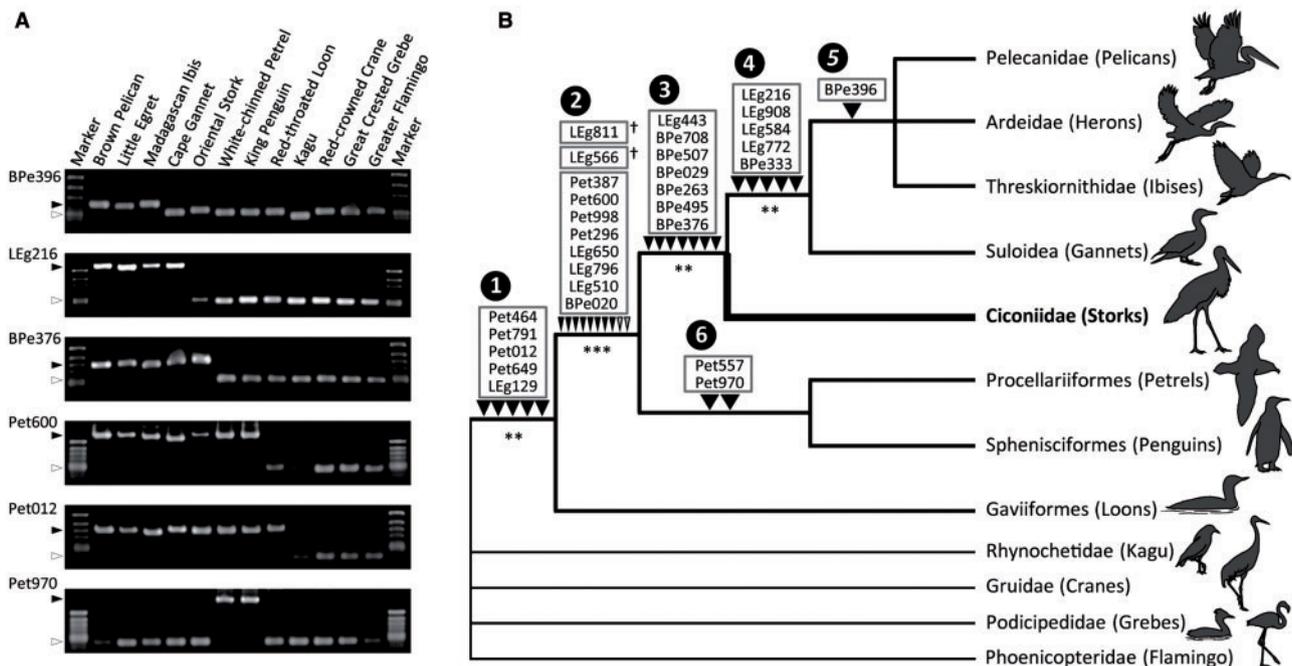


FIG. 3.—Waterbird phylogeny reconstructed by retroposon insertion analysis. (A) Electrophoretic analysis of the presence/absence of CR1 for 14 representative loci. The solid and open arrowheads identify PCR products that do or do not contain CR1, respectively. Left and right lanes labeled “Marker,” *HincII* digest of Φ X174. Sequence alignments are shown in [supplementary figure S1, Supplementary Material](#) online. (B) Waterbird phylogenetic tree based on our CR1 insertion analysis. The numbers of CR1 insertions are shown as arrowheads, and their associated loci are named above the branches. The names of the branches 1–6 supported by the CR1 insertions are shown above the corresponding branch. The highlighted line is leading to storks. Asterisks represent statistically significant support for each clade (*** $P < 0.001$; ** $P < 0.01$). At the two loci of LEg811 and LEg566, identified by grey arrowheads and the names marked with daggers, it is likely that CR1s were inserted at branch 2 and were subsequently subjected to incomplete lineage sorting, resulting in showing conflict insertion patterns of LEg566 with branch 3 and 4 and that of LEg811 with branch 6.

two CR1s whose insertion patterns were inconsistent with any of the other 28 loci (fig. 3 and [supplementary table S3, Supplementary Material](#) online). One locus, LEg811, had a CR1 insertion in the genomes of all the waterbirds except for the loon and petrel. For locus LEg566, CR1 was observed for all waterbirds except those of the loon and gannet. The two incongruent CR1 patterns may have arisen from incomplete sorting of CR1 presence/absence dimorphisms in the ancestral lineages (Shedlock et al. 2004), as shown by other retroposon studies in fish (Takahashi et al. 2001; Terai et al. 2003), birds (Suh et al. 2011; Matzke et al. 2012), and mammals (Nikaido et al. 2006; Churakov et al. 2009; Nishihara et al. 2009). Given that the CR1s were absent from LEg811 and LEg566 in the loon genome, these CR1 insertions probably occurred after divergence of Gaviiformes (branch 2 in fig. 3B). Subsequently, the CR1 presence/absence dimorphisms were not fixed in the ancestral populations of the petrel and gannet lineages during the relatively rapid divergences of these waterbird families/orders. Even though the CR1 in LEg566 showed an incongruent insertion pattern, the two clades defined by branches 3 and 4 were significantly supported ($P = 0.0014$ for [7 1 0] of branch 3 and $P = 0.0096$ for [5 1 0] of branch 4; Waddell et al. 2001).

Thus, we believe that our results represent conclusive evidence for the phylogenetic position of Ciconiidae. This result also indicates that the traditional Ciconiiformes (Ciconiidae, Ardeidae, Threskiornithidae, Balaenicipitidae, and Scopidae) is polyphyletic, a finding that is in agreement with recent nuclear gene analyses (Hackett et al. 2008; Kimball et al. 2013), although the relatively close relationship of Ciconiidae with Ardeidae and Threskiornithidae is suggested by morphological (Cracraft 1981; Livezey and Zusi 2007) and, in part, by a mitochondrial genomic analysis (Gibb et al. 2013).

Regarding other phylogenetic relationships, we characterized one locus for which a CR1 insertion was found only for Pelecanidae + Ardeidae + Threskiornithidae (branch 5 in fig. 3B). In addition, two CR1 insertions were observed exclusively in Procellariiformes and Sphenisciformes (branch 6 in fig. 3B) although one (LEg811) had an inconsistent pattern ($P = 0.15$ [2 1 0]; Waddell et al. 2001). The two CR1-containing loci supporting the formation of the Procellariiformes + Sphenisciformes clade agree with recent nuclear gene analyses (Hackett et al. 2008; Kimball et al. 2013; Jarvis et al. 2014) and, in part, with morphological studies (Livezey and Zusi 2007) (fig. 1).

Hybridization during the Divergence History of Pelecaniformes

Because the phylogenetic relationship for the pelican, herons (egret), and ibis was not resolved by the analysis described above, we performed a genome-scale CR1 insertion analysis using the large-scale sequence data reported by Jarvis et al. (2014). The sequence data for the little egret, crested ibis, Dalmatian pelican, and great cormorant served as representatives of Ardeidae, Threskiornithidae, Pelecanidae, and Suloidea, respectively. The presence/absence patterns for all CR1 elements in the genomes of the 4 species were investigated, and 97 phylogenetically informative CR1 insertions were found (fig. 4). Among them, insertions in 63 loci were indicative of Suloidea diverging first ($P < 0.001$, chi-square test; fig. 4A and B), although several incongruent patterns were also found. That Suloidea diverged first was reinforced by the observation that the numbers of isolated CR1 loci were not biased among the species, that is, 26, 19, and 23 loci for the egret, ibis, and pelican genomes, respectively (5 were isolated from 2 species). Therefore, our study provides conclusive evidence for the monophyly of Ardeidae + Threskiornithidae + Pelecanidae and confirms recent reports that used nuclear gene analyses (Hackett et al. 2008; Jarvis et al. 2014). In addition, characterization of the 19 incongruent loci suggested that a rapid ancestral radiation of these families occurred. This conclusion also supports the recently proposed waterbird classification (IOC World Bird List Version 5.2) where Ardeidae and Threskiornithidae are included within Pelecaniformes.

Regarding the relationship among Pelecanidae (pelicans), Ardeidae (herons), and Threskiornithidae (ibises), recent nuclear gene analyses suggest monophyly for Pelecanidae + Ardeidae (Jarvis et al. 2014) or Ardeidae + Threskiornithidae (Hackett et al. 2008). In our retroposon analysis, nine CR1 insertions were found to be shared by Pelecanidae and Ardeidae, whereas six insertions were found in Ardeidae and Threskiornithidae (fig. 4C). These 15 CR1 elements were evenly isolated from the genomes of the 3 species. Namely, seven, four, and five loci were found from the egret, ibis, and pelican, respectively (one was found from two species). Accordingly, no ascertainment bias was apparent in our study. Notably, no CR1 insertion was shared by Pelecanidae and Threskiornithidae ($P = 0.031$, KKSC significance test; Doronina et al. 2015), which is important because if the three species diverged simultaneously, an equal number of inconsistent insertions would theoretically be observed owing to incomplete lineage sorting (Waddell et al. 2001; Churakov et al. 2009; Nishihara et al. 2009). We, therefore, interpreted these results as follows. Regarding the phylogenetic relationship among Pelecanidae (pelicans), Ardeidae (herons), and Threskiornithidae (ibises), the monophyly of Pelecanidae and Ardeidae seemed most likely because the nine CR1-containing loci described above supported the

formation of the clade as does the genome-scale sequence analysis of Jarvis et al. (2014), which had 100% BP. If this is the case, our result that no CR1-containing locus was found to be shared by Pelecanidae and Threskiornithidae can be explained as follows. After divergence of Suloidea, followed by the establishment of the three species, namely, Threskiornithidae, Pelecanidae, and Ardeidae, interspecies hybridization between the ancestral populations of Ardeidae and Threskiornithidae lineages occurred. This is the most plausible explanation for the current retroposon data (fig. 4).

Discussion

Advantages of a Genome-Scale Retroposon Screen for Retroposon-based Phylogeny as well as Resolution of Rapid Radiation

This study demonstrates the usefulness and effectiveness of our NGS-based protocol and the genome-scale CR1 insertion analysis, which we call the STRONG method, to resolve waterbird phylogeny. The greatest advantage of our method is that a huge number of retroposons can be isolated. For example, from the analysis of the brown pelican genome produced by $\sim 1 \times$ coverage of its genome with the assumption that the size of the genome is 1.4 Gb (Gregory et al. 2007), we obtained $\sim 16,000$ CR1-containing sequences for which orthologous loci are also found in the zebra finch genome. The second advantage of the method is that sequencing is not inherently biased, which enabled us to find both anciently and recently inserted retroposons. Although the protocol requires a reference genomic sequence for subsequent PCR analysis, this constraint may be overcome as the amount of whole-genome data increases. We expect the STRONG method to become a standard approach for retroposon insertion analysis.

The availability of a number of retroposons should allow us to overcome phylogenetic problems related to incomplete lineage sorting and introgression. For example, rapid radiation from the common ancestor of Neoaves to the major avian orders is thought to have occurred within a very short evolutionary time (Feduccia 1995; Jetz et al. 2012; Jarvis et al. 2014), and retroposon insertion analysis has shed light on incomplete lineage sorting (Suh et al. 2011; Matzke et al. 2012). On the other hand, inconsistent retroposon insertions have not been reported for the cases of divergences in Galliformes (Kaiser et al. 2007; Kriegs et al. 2007), paleognaths (Haddrath and Baker 2012), and grebes (Suh et al. 2012), suggesting that radiation did not occur during their evolution. In this study, the phylogenetic relationships among waterbirds were established using statistically significant numbers of retroposon insertions ($P < 0.001$ for two branches and $P < 0.01$ for three branches, fig. 3) despite the presence of CR1 loci with incomplete lineage sorting. Specifically, the monophyly of Pelecanidae, Ardeidae, and Threskiornithidae is significantly supported by 63 CR1 loci

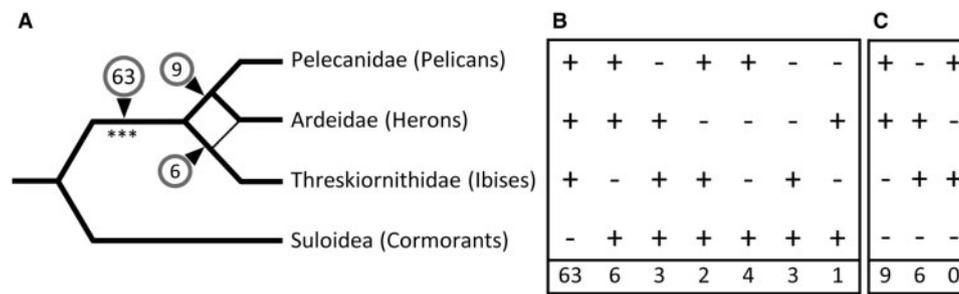


Fig. 4.—CR1 insertion analysis for Pelecanidae, Ardeidae, Threskiornithidae, and Phalacrocoracidae based on their genomic assemblies. (A) Relationship among the four families (***P* < 0.001). The circled numbers represent the numbers of supporting CR1 insertions. (B) The monophyly of Pelecanidae, Ardeidae, and Threskiornithidae is supported by 63 CR1 loci, but 6 incongruent patterns involving 19 markers are also present. The presence or absence of CR1 is represented by “+” and “-”, respectively, and each pattern is supported by the number of CR1 insertions given at the bottom of the figure. (C) Three insertion patterns for the relationship among Pelecanidae, Ardeidae, and Threskiornithidae. The numbers of supporting CR1 insertions are given at the bottom of the figure.

despite the presence of 19 insertions showing inconsistent patterns. These observations suggest that moderately rapid radiation probably occurred during the divergences of waterbird orders/families. We demonstrated that, even in the presence of incomplete lineage sorting, inclusion of a relatively large number of retroposons can overcome the uncertainty introduced by incomplete lineage sorting (Doronina et al. 2015).

Very recently Suh et al. (2015) performed a genome-scale analysis of >2,000 long terminal repeat (LTR) retrotransposon insertions based on the genomic data of Jarvis et al. (2014). They reported that the waterbird phylogeny proposed by Jarvis et al. (2014) (excluding storks) was well supported by retroposon insertions, and waterbirds showed less inconsistent insertion patterns than deeper interordinal relationships, also providing the support to our present conclusion.

Phylogenetic Position of Storks (Ciconiidae) in the Waterbird Assemblage

Our retroposon analysis strongly supports the monophyly of the families traditionally classified as Pelecaniformes and Ciconiiformes (*P* = 0.0014 for branch 3 in fig. 3B; Waddell et al. 2001). Most birds in this group live in semiaquatic environments, whereas other waterbirds (e.g., Gaviiformes, Procellariiformes, and Sphenisciformes) are more adapted to aquatic life. Regarding the phylogenetic position of storks (Ciconiidae), various birds, for example, the condor (Sibley and Ahlquist 1990), penguin (Pacheco et al. 2011), Ardeidae (Fain and Houde 2004; Gibb et al. 2013), and flamingo (Cracraft 1981; Livezey and Zusi 2007), have been proposed as a sister group. Our retroposon insertion analysis provides conclusive evidence that Ciconiidae was the first to diverge from the traditional Pelecaniformes and Ciconiiformes group, which is consistent with the results for the 19 nuclear gene analysis of Hackett et al. (2008) (fig. 1C) but not with the 50 nuclear gene analysis of Kimball et al. (2013) (fig. 1D).

Having established a conclusive retroposon-based waterbird phylogeny, it can now be used to reconsider and discuss their evolution in terms of morphology and ecology. For example, Livezey and Zusi (2006) proposed a diagnostic, synapomorphic feature of the radius for traditional Ciconiimorphae (storks, herons, ibis, hamerkop, and flamingos). However, among the Ciconiimorphae, herons and ibises (fig. 3) and hamerkops (Hackett et al. 2008) are more closely related to pelicans and Suloidea, which do not share this feature, than are storks and flamingos. The flamingo was excluded from the waterbird assemblage (fig. 3), which is consistent with other molecular studies (Hackett et al. 2008; Jarvis et al. 2014). Accordingly, it is likely that this osteological feature convergently evolved for storks, flamingos, and other birds on our tree. In addition, Cracraft (1981) reported that the legs and skulls of storks, flamingos, and ibises share very similar osteological features, which also suggests that many morphological characters have been subjected to convergent evolution, presumably owing to their similar means of locomotion (Livezey and Zusi 2007; Felice and O’Connor 2014).

Livezey and Zusi (2006) characterized the sterna as a distinctive feature that is shared by pelicans, herons, ibises, Suloidea, storks, and flamingos. Our retroposon analysis demonstrates that all these birds, except flamingos, form a monophyletic group (branch 3 in fig. 3B). Therefore, it is likely that this osteological feature has evolved once in the waterbird assemblage and once in the flamingo lineage. Thus, based on our waterbird phylogeny, re-examination of morphological characteristics should reveal conclusive diagnostic synapomorphies and features that tend to be subjected to convergent evolution, as recently shown by two studies (Smith 2012; Felice and O’Connor 2014).

Ancient Interspecies Hybridization Detected by Our Retroposon Analysis

This study provides the most likely scenario for the evolution of Pelecanidae (pelicans), Ardeidae (herons), and

Threskiornithidae (ibises) using genome-scale retroposon insertion analysis. Namely, we suggest that an interspecies hybridization occurred involving the ancestral populations of Ardeidae and Threskiornithidae lineages (fig. 4). This suggestion has not been proposed before, which indicates the advantage achieved by using the nearly homoplasy-free nature of retroposon insertion analysis.

Jarvis et al. (2014) performed a genome-scale comparative analysis of long terminal repeat retroposons in an attempt to resolve the phylogenetic relationships of land birds. They found 22 loci consistent with the monophyly of owls and Eucavitaves (e.g., cuckoo-rollers and woodpeckers), whereas 15 other loci are consistent with the monophyly of owls and Accipitrimorphae (e.g., hawks and condors). However, no retroposon locus has been found that supports the grouping of Eucavitaves and Accipitrimorphae (Jarvis et al. 2014). We consider these findings to be the support for the occurrence of an interspecies hybridization event involving the owl lineage.

Thus, species divergence did not always occur with constant bifurcations during and after the radiation of Neoaves. To characterize the evolution of a species, sequence analyses as well as rare genomic changes, for example, retroposon insertions, should be used (Rokas and Holland 2000). Researchers must be cognizant that there is always the possibility that ancient interspecies hybridization occurred during evolution and that it can be assessed by retroposon insertion analysis.

Supplementary Material

Supplementary figure S1 and tables S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

Acknowledgments

We thank the National Research Institute of Far Seas Fisheries, Fisheries Research Agency, the National Research Institute of Aquaculture, Fisheries Research Agency, the National Institute for Environmental Studies, the Zoorasia Yokohama Zoological Gardens, and the Ueno Zoological Gardens for providing tissue samples from the petrel and gannet, the penguin, the loon and grabe, the kagu, and all other birds, not specifically mentioned, respectively. We also thank the Biomaterial Analysis Center in Tokyo Institute of Technology for DNA sequencing. The computational analysis was conducted using the super computer system of the Institute of Statistical Mathematics. Grants from the JSPS KAKENHI (numbers 21227002 and 26291075 to N.O.) and JSPS KAKENHI (numbers 22770080, 26840117, 26106004, and 23123504 to H.N.) supported this work.

Literature Cited

Bonaparte CL. 1854. *Conspectus systematis ornithologiae*. *Ann Sci Nat Zool Ser 4*. 1:1–48.

- Burch JB, Davis DL, Haas NB. 1993. Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. *Proc Natl Acad Sci U S A*. 90:8199–8203.
- Churakov G, et al. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res*. 19:868–875.
- Cracraft J. 1981. Toward a phylogenetic classification of the recent birds of the world (Class Aves). *Auk* 98:681–714.
- Cracraft J. 1985. Monophyly and phylogenetic relationships of the Pelecaniformes: a numerical cladistic analysis. *Auk* 102:834–853.
- Cracraft J, et al. 2004. Phylogenetic relationships among modern birds (Neornithes): toward an avian tree of life. In: Cracraft J, Donogue MJ, editors. *Assembling the tree of life*. New York: Oxford University Press. p. 468–489.
- Deininger PL, Batzer MA. 2002. Mammalian retroelements. *Genome Res*. 12:1455–1465.
- Doronina L, et al. 2015. Exploring Massive Incomplete Lineage Sorting in Arctoids (Laurasiatheria, Carnivora). *Mol Biol Evol*, forthcoming.
- Ericson PG, et al. 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol Lett*. 2:543–547.
- Fain MG, Houde P. 2004. Parallel radiations in the primary clades of birds. *Evolution* 58:2558–2573.
- Feduccia A. 1995. Explosive evolution in tertiary birds and mammals. *Science* 267:637–638.
- Felice RN, O'Connor PM. 2014. Ecology and caudal skeletal morphology in birds: the convergent evolution of pygostyle shape in underwater foraging taxa. *PLoS One* 9:e89737.
- Garrod AH. 1874. On certain muscles of birds and their value in classification. Part II. *Proc Zool Soc Lond*. 42:111–123.
- Gibb GC, Kennedy M, Penny D. 2013. Beyond phylogeny: pelecaniform and ciconiiform birds, and long-term niche stability. *Mol Phylogenet Evol*. 68:229–238.
- Gill F, Donsker D, editors. 2015. IOC world bird list (v 5.2), <http://www.worldbirdnames.org/>, doi:10.14344/IOC.ML.5.2.
- Gregory TR, et al. 2007. Eukaryotic genome size databases. *Nucleic Acids Res*. 35:D332–D338.
- Hackett SJ, et al. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Haddrath O, Baker AJ. 2012. Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. *Proc Biol Sci*. 279:4617–4625.
- Hillier LW, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.
- Huxley TH. 1867. On the classification of birds; and on the taxonomic value of the modifications of certain of the cranial bones observable in that class. *Proc Zool Soc Lond*. 145:415–472.
- Jarvis ED, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491:444–448.
- Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol*. 24:338–347.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.
- Kimball RT, Wang N, Heimer-McGinn V, Ferguson C, Braun EL. 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. *Mol Phylogenet Evol*. 69:1021–1032.
- Kriegs JO, et al. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol*. 4:e91.

- Kriegs JO, et al. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol.* 7:190.
- Livezey BC, Zusi RL. 2006. Phylogeny of Neornithes. *Bull Carnegie Mus Nat Hist.* 37:1–544.
- Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool J Linn Soc.* 149:1–95.
- Matzke A, et al. 2012. Retroposon insertion patterns of neoavian birds: strong evidence for an extensive incomplete lineage sorting era. *Mol Biol Evol.* 29:1497–1501.
- Mayr G. 2011. Metaves, Mirandornithes, Strisores and other novelties—a critical review of the higher-level phylogeny of neornithine birds. *J Zool Syst Evol Res.* 49:58–76.
- Morgan-Richards M, et al. 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. *BMC Evol Biol.* 8:20.
- Murata S, et al. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc Natl Acad Sci U S A.* 90:6995–6999.
- Nikaido M, Rooney AP, Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci U S A.* 96:10261–10266.
- Nikaido M, et al. 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc Natl Acad Sci U S A.* 98:7384–7389.
- Nikaido M, et al. 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. *Mol Biol Evol.* 23:866–873.
- Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci U S A.* 103:9929–9934.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A.* 106:5235–5240.
- Nishihara H, Okada N. 2008. Retroposons: genetic footprints on the evolutionary paths of life. *Methods Mol Biol.* 422:201–225.
- Okada N, Hamada M, Ogiwara I, Ohshima K. 1997. SINEs and LINEs share common 3' sequences: a review. *Gene* 205:229–243.
- Olson SL. 1979. Multiple origins of the Ciconiiformes. *Proc Colonial Waterbird Group.* 2:165–170.
- Pacheco MA, et al. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol Biol Evol.* 28:1927–1942.
- Platt RN II, et al. 2015. Targeted capture of phylogenetically informative Ves SINE insertions in genus *Myotis*. *Genome Biol Evol.* 7:1664–1675.
- Rokas A, Holland PWH. 2000. Rare genomic changes as a tool for phylogenetics. *Trends Ecol Evol.* 15:454–459.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: laboratory manual*. 2nd ed. New York: Cold Spring Harbor Laboratory Press.
- Sasaki T, et al. 2004. First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. *Mol Biol Evol.* 21(4):705–715.
- Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays* 22:148–160.
- Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol.* 19:545–553.
- Shimamura M, et al. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388:666–670.
- Sibley CG, Ahlquist JE. 1990. *Phylogeny and classification of birds: a study in molecular evolution*. New Haven (CT): Yale University Press.
- Smith ND. 2012. Body mass and foraging ecology predict evolutionary patterns of skeletal pneumaticity in the diverse “waterbird” clade. *Evolution* 66:1059–1078.
- Stumph WE, Kristo P, Tsai MJ, O'Malley BW. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Res.* 9:5383–5397.
- Suh A, Kriegs JO, Donnellan S, Brosius J, Schmitz J. 2012. A universal method for the study of CR1 retroposons in nonmodel bird genomes. *Mol Biol Evol.* 29:2899–2903.
- Suh A, Smeds L, Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of Neoavian birds. *PLoS Biol.* 13:e1002224.
- Suh A, et al. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.
- Suh A, et al. 2014. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol Evol.* 7:205–217.
- Takahashi K, Terai Y, Nishida M, Okada N. 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Mol Biol Evol.* 18:2057–2066.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 30:2725–2729.
- Terai Y, Takahashi K, Nishida M, Sato T, Okada N. 2003. Using SINEs to probe ancient explosive speciation: “hidden” radiation of African cichlids? *Mol Biol Evol* 20:924–930.
- Van Tuinen M, Butvill DB, Kirsch JA, Hedges SB. 2001. Convergence and divergence in the evolution of aquatic birds. *Proc Biol Sci.* 268:1345–1350.
- Vandergon TL, Reitman M. 1994. Evolution of chicken repeat 1 (CR1) elements: evidence for ancient subfamilies and multiple progenitors. *Mol Biol Evol.* 11:886–898.
- Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.
- Watanabe M, Nikaido M, Tsuda TT, Kobayashi T, et al. 2006. New candidate species most closely related to penguins. *Gene* 378:65–73.
- Watanabe M, Nikaido M, Tsuda TT, Inoko H, et al. 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. *Gene* 365:57–66.
- Wetmore A. 1960. *A classification for the birds of the world*. Smithsonian Miscellaneous Collections. Vol. 139. Washington (DC): Smithsonian Institution. p. 1–37.

Associate editor: Wen-Hsiung Li