



## Testing the utility of mitochondrial *cytochrome oxidase subunit 1* sequences for phylogenetic estimates of relationships between crane (*Grus*) species

D.B. Yu\*, R. Chen\*, H.A. Kaleri, B.C. Jiang, H.X. Xu and W.-X. Du

Department of Animal Science and Technology,  
Nanjing Agricultural University, Nanjing, China

\*These authors contributed equally to this study.  
Corresponding author: W.-X. Du  
E-mail: duwenxing@njau.edu.cn

Genet. Mol. Res. 10 (4): 4048-4062 (2011)  
Received November 1, 2011  
Accepted December 19, 2011  
Published December 21, 2011  
DOI <http://dx.doi.org/10.4238/2011.December.21.7>

**ABSTRACT.** Morphology and biogeography are widely used in animal taxonomy. Recent study has suggested that a DNA-based identification system, using a 648-bp portion of the mitochondrial gene *cytochrome oxidase subunit 1* (*COI*), also known as the barcoding gene, can aid in the resolution of inferences concerning phylogenetic relationships and for identification of species. However, the effectiveness of DNA barcoding for identifying crane species is unknown. We amplified and sequenced 894-bp DNA fragments of *COI* from *Grus japonensis* (Japanese crane), *G. grus* (Eurasian crane), *G. monacha* (hooded crane), *G. canadensis* (sandhill crane), *G. leucogeranus* (Siberian crane), and *Balearica pavonina* (crowned crane), along with those of 15 species obtained from GenBank and DNA barcoding, to construct four algorithms using *Tringa stagnatilis*, *Scolopax rusticola*, and *T. erythropus* as outgroups. The four phylum profiles showed good resolution of the major taxonomic groups. We concluded that reconstruction of the molecular phylogenetic tree can be helpful for classification and that

*COI* sequences are suitable for studying the molecular evolution of cranes. Although support for several deeper branches was limited, *COI* data gave remarkably good separations, especially considering that our analysis was based on just a fragment of the gene and that *COI* has generally been viewed as useful only for resolving shallow divergences.

**Key words:** Molecular phylogeny; Mitochondrial DNA; Cytochrome C oxidase 1

## INTRODUCTION

Cranes are large, long-legged, long-necked birds of the family Gruidae (order Gruiformes). They are divided into 15 species in four genera (*Balearica*, *Grus*, *Anthropoides*, and *Bucgeranus*) of two subfamilies, Balearicinae and Gruidae (Sibley and Monroe, 1990). Phylogenetic investigations have confirmed the monophyly of Gruidae using various methods, including morphometrics (Wood, 1979), allozyme analysis (Dessauer et al., 1992), DNA hybridization (Krajewski, 1989), mitochondrial gene sequence assessments (Krajewski et al., 1999), morphocladistics (Livezey, 1998), and tests that combine mitochondrial DNA sequences (Krajewski et al., 2010). All of these studies focused on species of the Gruidae family. However, classification of the avian order Gruiformes has been one of the least stable in the taxonomic history of birds (Fain et al., 2007; Krajewski et al., 2010; Yang et al., 2010b). Avian systematists have debated their phylogenetic relationships and even whether they constitute a monophyletic assemblage (Livezey, 1998; Sibley and Monroe, 1990). Livezey (1989) formerly divided Gruiformes into several families and found strong support for heliornithids and rails in the superfamily Ralloidea and for cranes, limpkins, and trumpeters in the superfamily Gruoidea. Recently, Fain (2001) studied the phylogeny of 'core Gruiformes', including Gruidae, Aramidae, Psophiidae, Heliornithidae, and Rallidae. However, cranes and limpkins are the more recent phyletic lineages based on DNA hybridization (Sibley and Monroe, 1990).

Nine species of 15 crane breeds are found in China. Unfortunately, destruction of wetlands sharply increased after the 1960s in northeast China and other regions of the Far East. Reclamation of wetlands not only leads to decreasing of the original habitat, but also blocks the connected water system. At the same time, the water covered areas in the Far East have been shrinking dramatically due to recent drought. Pollution from human activities also threatens the survival of the genus *Grus*. The Poyang Hu Lake area in China, an essential wintering site of *G. leucogeranus*, has been threatened by dramatic hydrological changes caused by the construction of the Three Gorges Dam in recent years, especially in 2011. Little rainfall caused the center of the lake to become almost a prairie, which will likely lead to an extremely rapid decline in Siberian cranes in the near future. In addition, cultivation methods, such as burning, can severely damage the nest/shelter material of species of the genus *Grus*, further narrowing its distribution.

Variation in mtDNA sequences has been widely used in intrapopulation and intraspecific phylogenetic analyses of the order Gruiformes (Fain et al., 2007; Krajewski et al., 2010; Yang et al., 2010b) and to assess genetic variation in species (Petersen et al., 2003; Ponomarev et al., 2004). These studies have provided new insights into vertebrate classification and affini-

ties, and have enabled vigorous testing of taxonomic schemes. More rapidly evolving mitochondrial genes have generally been employed to infer relationships among groups with more recent ancestry (Plejtel et al., 2008).

To develop effective conservation strategies for cranes, the relatedness of these species should be understood by detailed phylogenetic studies, requiring DNA markers. It is well known that DNA markers from mitochondrial genomes can only provide half of the story due to a uniparental (usually maternal) inherited pattern. So in the future, improving taxonomic standards and developing more sophisticated genetic barcoding tools will be required. In the mitochondrial genome, *cytochrome C oxidase subunit I (COI)* has been particularly popular for estimating relationships among closely allied taxa and efficient elucidation of biological diversity, especially the resolution of affinities at lower taxonomic levels. DNA barcode technology, using short *COI* sequences, has been proposed as a method for enabling rapid, accurate detection and identification of species. This method is accepted as a standard for identifying DNA of organisms and is now being proposed as a method for cataloguing life and developing a comprehensive species-specific sequence library for all eukaryotes (Marshall, 2005). This approach will significantly broaden the application of DNA barcoding in biodiversity studies. At present, a 5'-648 bp fragment of *COI* is considered a universal barcoding marker for the identification of animal species (Hebert et al., 2003a). This fragment is easily sequenced and provides greater than 97% species-level specificity for birds (Hebert et al., 2004) and mammals (Hajibabaei et al., 2007). DNA barcoding has been proposed as a standardized approach for the characterization of life forms in numerous groups of living organisms (Yoo et al., 2006; Hajibabaei et al., 2007) and the identification of new species (Yang et al., 2010a). In addition, for groups with numerous obstacles to taxonomic resolution, DNA barcodes can help in the routine identification of species in applied settings, in the detection of morphologically cryptic species, for host-specific lineages, and in the association of morphologically distinct life cycle forms within a species. DNA barcoding using mitochondrial *COI* (Hebert et al., 2003a) is now well established for animals. Previous molecular reports of cranes involved the *cytb*, *ND6*, and *D-Loop* regions and the mitochondrial genome (Krajewski et al., 2010). However, no independent information on crane mtDNA *COI* gene sequences was available. We collected samples from six species of Gruidae, analyzed their sequence variation in the 894 bp *COI* DNA barcode region, and constructed phylogenetic trees after aligning a library of DNA barcodes that includes 21 species of Gruiformes.

## MATERIAL AND METHODS

### Animals

Twenty-one species of Gruiformes, including Aramidae, Gruidae, Rallidae, and Otididae, and three outgroups from the order Charadriiformes (*Tringa stagnatilis*, *T. erythropus*, and *Scolopax rusticola*) were used in our study (Table 1). New sequence data was obtained for six species of crane (*Balearica pavonina* (crowned crane), *G. japonensis* (Japanese crane), *G. grus* (Eurasian crane), *G. monacha* (hooded crane), *G. canadensis* (sandhill crane), and *G. leucogeranus* (Siberian crane)) belonging to the Gruidae family. The other 18 sequence samples were downloaded from the GenBank database (<http://www.ncbi.nlm.nih.gov>) and the Barcode of Life Database (BOLD, <http://www.barcodinglife.org>).

**Table 1.** Crane samples analyzed.

Order	Family	Genus	Species	Accession No.	Source		
Gruiformes	Aramidae	<i>Aramus</i>	<i>Aramus guarauna</i>	CDANS011-05[4705]	http://www.barcodinglife.org		
	Gruidae	<i>Anthropoides</i> <i>Balearica</i>	<i>Grus virgo</i> (demoiselle crane) <i>Balearica pavonina</i> (grey crowned crane) <i>Balearica regulorum</i> (black grey crowned crane)	FJ769845 In this study FJ769841	http://www.ncbi.nlm.nih.gov Nanjing Hongshan Zoo  http://www.ncbi.nlm.nih.gov		
Charadriiformes		<i>Grus</i>	<i>Grus americana</i> (whooping crane)	FJ769848	http://www.ncbi.nlm.nih.gov		
			<i>Grus antigone</i> (sarus crane)	FJ769854	http://www.ncbi.nlm.nih.gov		
			<i>Grus grus</i> (Eurasian crane)	In this study	Nanjing Hongshan Zoo		
			<i>Grus japonensis</i> (Japanese crane)	In this study	Nanjing Hongshan Zoo		
			<i>Grus nigricollis</i> (black-necked crane)	FJ769851	http://www.ncbi.nlm.nih.gov		
			<i>Grus paradisea</i> (blue crane)	FJ769844	http://www.ncbi.nlm.nih.gov		
			<i>Grus rubicunda</i> (brolga)	FJ769853	http://www.ncbi.nlm.nih.gov		
			<i>Grus vipio</i> (white-naped crane)	FJ769852	http://www.ncbi.nlm.nih.gov		
			<i>Grus monacha</i> (hooded crane)	In this study	Shanghai Zoo		
			<i>Grus canadensis</i> (sandhill crane)	In this study	Shanghai Zoo		
			<i>Grus carunculatus</i> (wattled crane)	FJ769843	http://www.ncbi.nlm.nih.gov		
			<i>Grus leucogeranus</i> (Siberian crane)	In this study	Nanjing Hongshan Zoo		
			<i>Aramides</i>	<i>Aramides</i>	<i>Aramides ypecaha</i>	KAARG156-07[MACN-Or-et 2122]	http://www.barcodinglife.org
			Rallidae		<i>Crex</i>	<i>Crex crex</i>	BONSC036-08[NHMO-BC553]
<i>Gallinix cinerea</i>	KBBI051-07[KRIBB 634]	http://www.barcodinglife.org					
Otididae		<i>Rallus</i>	<i>Rallus linicola</i>	BROMB213-06[IB-2599]	http://www.barcodinglife.org		
			<i>Tetrax tetrax</i>	KBPB128-06[UWBM 64702]	http://www.barcodinglife.org		
			<i>Tringa stagnatilis</i>	KBPB1246-06[UWBM 66737]	http://www.barcodinglife.org		
			<i>Tringa erythropus</i>	BON012-06[NHMO-BC12]	http://www.barcodinglife.org		
			<i>Scolopax rusticola</i>	KBBI073-07[KRIBB 998]	http://www.barcodinglife.org		

## DNA extraction and PCR amplification and sequencing

Three to five pieces of feather of each species were collected from the wing or tail. The quill coverts were then removed and stored in 2 ml Eppendorf tubes at  $-20^{\circ}\text{C}$ . Total DNA of each sample was extracted from the feather tissue using the standard phenol-chloroform method (Sambrook and Russel, 2001), and the quality of DNA was examined by a V spectrophotometer after electrophoresis on 0.5% agarose (Promega, Madison, WI, USA). The DNA barcodes were sequenced employing the method described below.

Oligonucleotide primers for amplification of the mitochondrial *COI* locus were designed using the Primer Premier 5.0 software (Premier Biosoft International, Palo Alto, CA, USA), and based on the mitochondrial genome sequence of *Anas platyrhynchos* (GenBank accession No. NC\_009684). The upstream primer P1 (5'-CCTCTGTAAAAAGGACTACAGCC-3') and the downstream primer 5'-GTTTCGATTCCCTTCTTCTGT-3' were used to amplify six overlapping fragments to give a total of 1534 bp of sequence from the 5' terminus of *COI*.

Each 15- $\mu\text{L}$  reaction mixture contained 1 U *Taq* DNA polymerase (TaKaRa, Dalian, China), approximately 100 ng genomic DNA, 2.5 mM  $\text{MgCl}_2$ , 10X PCR buffer, 200  $\mu\text{M}$  dNTPs, and 0.2  $\mu\text{M}$  of each primer. The PCR reaction was performed in a GeneAmp 9700 PCR System (Applied Biosystems) under the following conditions: initial denaturation for 3 min at  $94^{\circ}\text{C}$ ; 35 cycles of denaturation for 45 s at  $94^{\circ}\text{C}$ , annealing temperature for 60 s at  $53^{\circ}\text{C}$ , and extension for 90 s at  $72^{\circ}\text{C}$ ; followed by a final elongation at  $72^{\circ}\text{C}$  for 10 min.

The PCR products were separated on 1.0% agarose gels and bands were visualized by ethidium bromide staining and viewed with an ultraviolet light source. The PCR products were purified using a GeneClean Kit (Qbiogene, Carlsbad, CA, USA), ligated into the pMD18-T cloning vector (TaKaRa, Dalian, China) according to manufacturer instructions, and transformed into competent *E. coli* DH5 $\alpha$  cells. The plasmid DNA was extracted using a Bio Dev Plasmid Rapid Isolation Kit (Bio Dev, Beijing, China) and sequenced on an ABI 377 DNA sequencer (ABI, Foster, CA, USA).

## Alignment and sequence properties

New sequences were compared to those in GenBank (*B. pavonina*, FJ769842; *G. grus*, FJ769849; *G. japonensis*, FJ769847; *G. monacha*, FJ769850; *G. canadensis*, FJ769855; *G. carunculatus*, FJ769843; and *G. leucogeranus*, FJ769846). We performed phylogenetic analyses using the newly obtained *COI* sequences and the *COI* sequences from 18 other species.

We manually aligned the sequences using the BioEdit sequence alignment editor (Hall, 1999). Multiple alignments were carried out with Clustal X (Thompson et al., 1997), and the alignment was subsequently pruned to 894 bp. We confirmed the accuracy of the sequences by checking whether they could be translated into proteins with EditSeq (DNASTAR package; DNASTAR Inc., 1996). We calculated sequence divergence for several groups of *COI* sequences. Average pairwise genetic distances were calculated using the Kimura two-parameter model (Kimura, 1980), as implemented in MEGA 4 (Tamura et al., 2007). Kimura's two-parameter model was used for two reasons. First, it makes fewer assumptions about the nature of sequence change than more heavily parameterized models (Herbert et al., 2003a). This is an important attribute when transition rules governing nucleotide substitution vary among taxa. Second, it provides a conservative estimate of long branches

because it underestimates the number of multiple hits compared to more complex models (Nei and Kumar, 2000).

Nucleotide saturation was analysed by plotting the number of transitions and transversions on each codon position against the Tamura and Nei (1993) (TN93) genetic distance using the DAMBE program (Xia and Xie 2001). Saturation was considered to have happened if the scatter of points showed leveling-off of mutations as sequence divergence increased.

### Phylogenetic analysis

The algorithms used for phylogenetic analyses included the neighbor-joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian analysis methods. We employed NJ analysis as implemented in PAUP \*4.0b (Wilgenbusch and Swofford, 2003) to examine relationships among taxa. The NJ method has been shown to be computationally efficient, with a record of recovering trees that are as good as those generated by alternate methods. We used the NJ algorithm first because it provides the necessary speed of analysis for large data sets. We measured the success of classification by determining whether each test sequence grouped most closely with the single representative of its species in the profile or with another species. After conducting all classification tests, we added all test sequences to the profile to allow for a more detailed examination of the factors that enable successful identification.

MP and ML analyses were performed with PAUP \*4.0b. An MP analysis was first carried out on each gene region separately, and the data set was combined under a heuristic search strategy, with all sites weighted equally. Gaps were treated as missing data, and 1000 random-addition sequences and TBR (the tree-bisection-reconnection) branch-swapping were used. To assess the support for branching events, non-parametric bootstrapping was performed with 1000 pseudoreplicates, under a heuristic search strategy, and 100 random-addition sequences in each pseudoreplicate (Pedersen et al., 2007).

A model of molecular evolution was first assigned to the data using the log-likelihood function of MODELTEST 3.0 (Posada and Crandall, 2001). MODELTEST compares 56 different nested substitutional models and uses log-likelihood scores to determine which model best fits the data. The best-fit nucleotide substitution models for our data were GTR+I+G. ML analysis was performed in PAUP \*4.0b under the selected optimal model under the heuristic search strategy with 10 random-addition sequences and TBR branch-swapping. With a starting tree obtained by 100 random taxa additions, 1000 bootstrap replicates were performed.

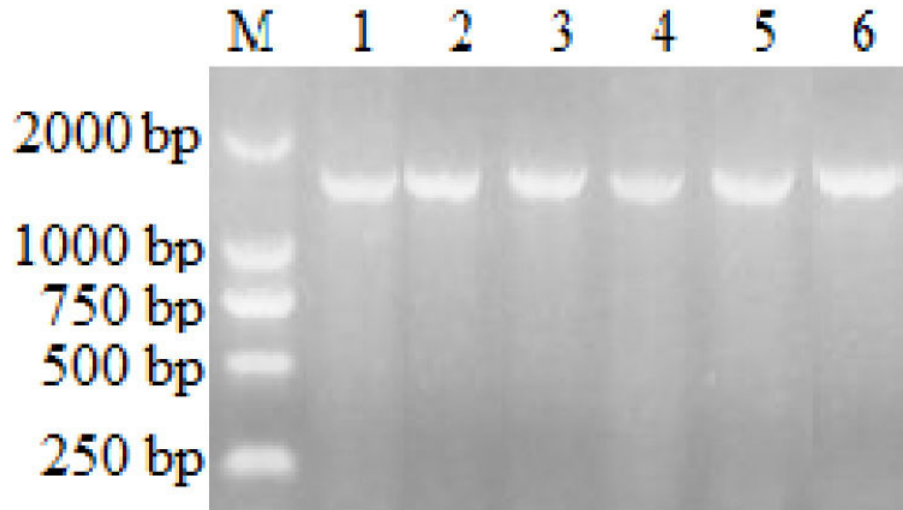
Bayesian analysis was conducted on our data set using MrBayes3.1.1 (Ronquist and Huelsenbeck, 2003) and the GTR+I+G models. Model parameter values were treated as unknown variables with uniform prior probabilities and were estimated during the analysis. Four chains (three heated and one cold) were run, starting from a random tree and proceeding for 2.4 million Markov chain Monte Carlo generations, sampling the chains every 200 generations. The point of stationarity was then determined (when the distribution of likelihoods was seen to have reached a plateau). The trees prior to stationarity were discarded (for all runs, 3000 trees were discarded), and the remaining trees were used to compute 50% majority rule consensus trees. The posterior probabilities were summarized accordingly.

## RESULTS

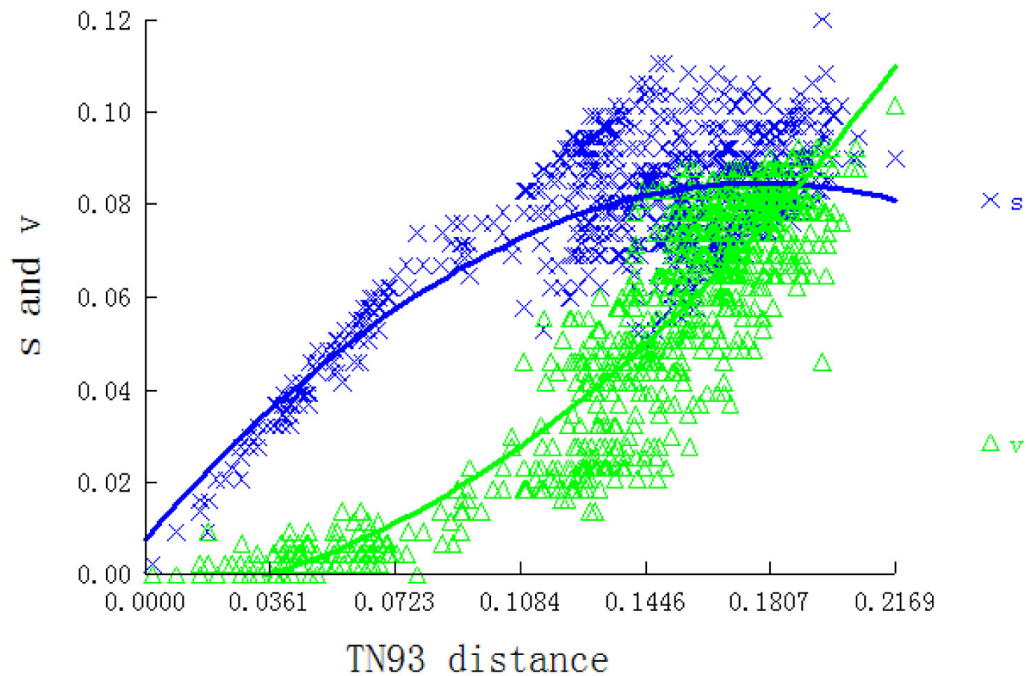
A 1534 bp segment was amplified from DNA of six species by employing the *COI* primer pairs (Figure 1). No insertions/deletions or stop codons were observed. The six species represented only by feather samples had conserved 894 bp of 5' regions of *COI* with high sequence similarity (100%) to their respective species data in GenBank (Nos. FJ769842, FJ769849, FJ769847, FJ769850, FJ769855, FJ769843, and FJ769846). Thus, the correct identity of these species was confirmed. After mtDNA data matrix alignment, which included the outgroup taxa with BioEdit, analysis of the 894 bp sequences showed that 603 (67%) positions were conserved sites, 291 (33%) were polymorphic sites, 237 (27%) were parsimony informative, and 54 (6%) were single-variable sites (for the crane taxa only, 701 (78%) sites were constant, 193 (22%) were variable, 140 (16%) were parsimony-informative, and 53 (6%) were single-variable). The average nucleotide frequencies of T, C, A, and G were 26.4, 30.2, 27.2, and 16.3%, respectively. The content of pyrimidine (56.8%) was higher than that of purine (43.2%). These values showed a strong A + T (53.5%) to G + C (46.5%) asymmetry in nucleotide composition (Table 2). The model found to be optimal for our data by Modeltest3.7 was GTR+I+G. The transition/transversion ratio, base frequencies, and gamma distribution shape parameter as estimated by Modeltest were  $Ti/Tv = 8.7879$ ;  $A = 0.2885$ ,  $C = 0.3589$ ,  $G = 0.1481$ , and  $T = 0.2044$ ; and  $\alpha = 2.5876$ , respectively. The transition/transversion ratio of *COI* in cranes in bias was  $R = 2.21$ , which was greater than the critical value (2.0) of  $Ts/Tv$  (Knight and Mindell, 1993). Because transitions and transversions in *COI* were accumulated linearly and showed no saturation patterns at any position (Figure 2), all nucleotide positions could be employed in the subsequent analysis.

**Table 2.** Nucleotide composition of 24 species from five bird families.

	T(U)	C	A	G	T+A	C+G
<i>Grus americana</i>	26.2	30.0	27.7	16.1	53.9	46.1
<i>Grus antigone</i>	26.8	29.1	28.1	16.0	54.9	45.1
<i>Grus carunculatus</i>	26.5	29.2	27.7	16.6	54.3	45.7
<i>Grus grus</i>	26.7	29.5	27.5	16.2	54.3	45.7
<i>Grus japonensis</i>	25.6	30.4	27.5	16.4	53.1	46.9
<i>Grus leucogeranus</i>	26.3	29.6	27.7	16.3	54.0	46.0
<i>Grus nigricollis</i>	26.5	29.6	27.5	16.3	54.0	46.0
<i>Grus paradisea</i>	26.2	29.9	27.1	16.9	53.2	46.8
<i>Grus rubicunda</i>	26.8	29.1	28.2	15.9	55.0	45.0
<i>Grus vipio</i>	26.3	29.5	28.2	16.0	54.5	45.5
<i>Grus virgo</i>	26.0	30.2	27.1	16.8	53.0	47.0
<i>Grus monacha</i>	26.3	29.9	27.7	16.1	54.0	46.0
<i>Grus canadensis</i>	26.1	29.8	27.7	16.4	53.8	46.2
<i>Balearica regulorum</i>	25.7	30.3	26.8	17.1	52.6	47.4
<i>Balearica pavonina</i>	25.1	30.9	27.1	17.0	52.1	47.9
<i>Aramus guarana</i>	25.1	32.2	26.0	16.7	51.1	48.9
<i>Aramides ypecaha</i>	26.8	31.2	26.7	15.3	53.5	46.5
<i>Crex crex</i>	27.2	29.5	27.9	15.4	55.1	44.9
<i>Rallus limicola</i>	25.0	31.7	26.4	16.8	51.5	48.5
<i>Gallix rex cinerea</i>	27.7	29.2	26.9	16.2	54.6	45.4
<i>Tetrax tetrax</i>	26.0	31.9	25.7	16.5	51.7	48.3
<i>Tringa stagnatilis</i>	28.0	30.7	25.1	16.2	53.1	46.9
<i>Scolopax rusticola</i>	27.0	30.7	26.0	16.3	53.0	47.0
<i>Tringa erythropus</i>	27.6	31.2	25.5	15.7	53.1	46.9
Avg.	26.4	30.2	27.2	16.3	53.5	46.5



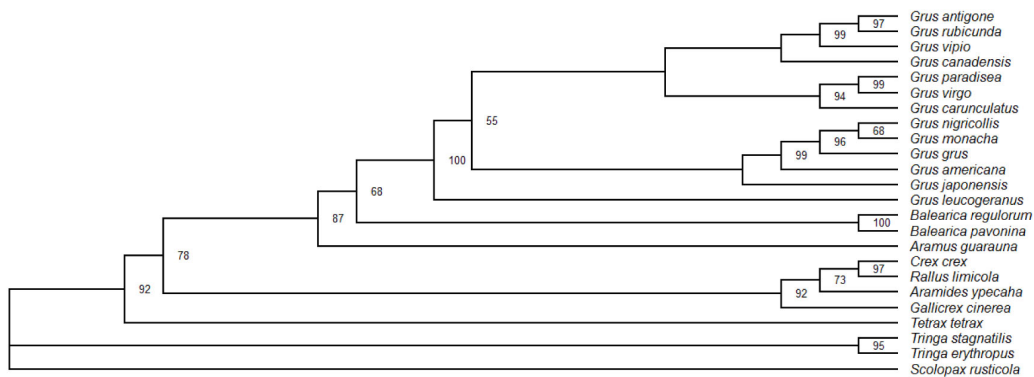
**Figure 1.** Agarose gel electrophoresis of PCR products for six species. Lane *M* = DNA marker, DL2000 (Takara Co.). Lane 1 = crowned crane; lane 2 = Japanese crane; lane 3 = Eurasian crane; lane 4 = hooded crane; lane 5 = sandhill crane; lane 6 = Siberian crane.



**Figure 2.** Saturation plots for the mitochondrial *COI* sequences of 894 bp. The number of transitions and transversions of each pairwise comparison of taxa are plotted against the TN93 model-corrected distance. The broken lines show the mean values of transition and transversion, respectively.

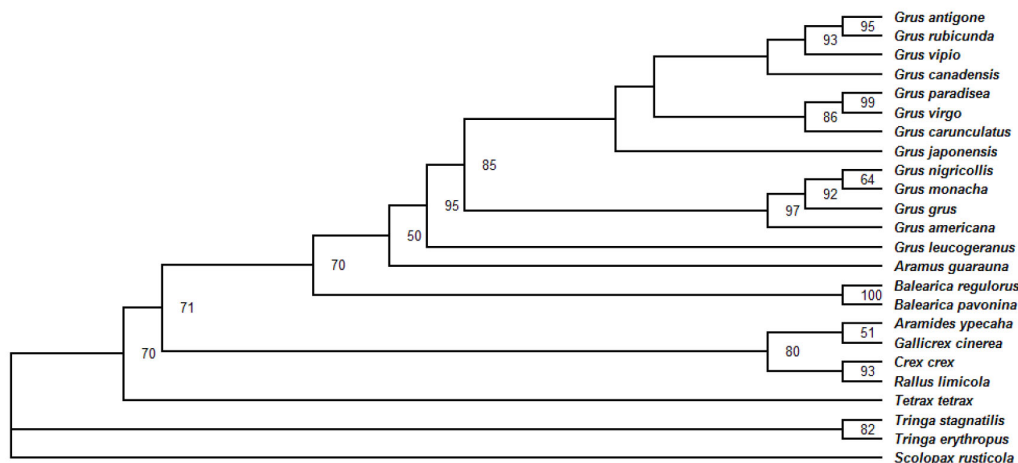


The phylogenetic trees of Gruidae and other species from the Aramidae, Rallidae, and Otidae families were constructed using the MJ, MP, ML, and Bayesian methods based on the 894 bp of the *COI* sequence (Figures 3-6). MP and ML analyses were performed on the gene data set for the 15 species of crane and the three species of sandpiper (Scolopacidae) with PAUP\*. ML analysis resulted in only one tree (-Ln likelihood = 4891.3467, Akaike Information Criterion (AIC) = 9802.6934), which was similar to the most parsimonious tree (Figure 4). These results show that the topological structures of the two trees were basically identical. The three outgroup taxa, *T. stagnatilis*, *T. erythropus*, and *S. rusticola*, were at the base of both trees. Within the *Grus* genus, however, the group was monophyletic and could be separated into two distinct subclades. The African crowned cranes (*Balearica*) were widely divergent from all other cranes. They showed relatively close relationships among the *Grus* species groups and were strongly supported by bootstrap values. With the exception of the MP tree, the trees strongly supported *G. monachus* and *G. nigricollis* as sister species, with *G. grus*, *G. americana*, and *G. japonensis* as successively earlier branches within the *americana* group. In the MP tree, however, clades Antigone and Anthropoids clustered with *G. japonensis*, and had low bootstrap values (BP = 30, <50%). In all four trees, *G. leucogeranus* was sister to all other gruines, and *G. antigone* and *G. rubicunda* clustered together; these two later species grouped with *G. vipio* in the Antigone group. In addition, *G. canadensis* clustered with the Antigone group, but with unreliable support in the NJ, MP, and ML trees (BP <50%) so that its position with respect to Antigone is not resolved in this study. In short, these relationships defined among the Gruidae were consistent with previous reports, thus providing additional support for the validity of the sequence described here (Krajewski et al., 2010).



**Figure 3.** Neighbor-joining tree of partial *COI* sequences. Bootstrap values are shown at the nodes. Numbers above the branches denote the posterior probabilities (only those  $\geq 0.50$  are shown).

In addition, *Aramus guarauna* from the *Aramus* genus was included as a member of Gruidae, apart from *Aramides*, *Crex*, *Gallicrex*, *Rallus*, and *Tetrax*, in the four trees, and the bootstrap support for the monophyly of the clade as a whole remained high (87%, NJ; 50%, MP; 87%, ML; 0.51, Bayesian).



**Figure 4.** Strict consensus tree from maximum parsimony analyses based on partial *COI* sequences (tree length = 885, consistency index [CI] = 0.4678, homoplasy index [HI] = 0.5322). Bootstrap values are shown at the branches. MP analyses employed heuristic searches (1000 random sequence addition, TBR branch-swapping, unordered and unweighted, character states; bootstrap analyses utilised full heuristic searches). Numbers above the branches denote the posterior probabilities (only those  $\geq 50$  are shown).

## DISCUSSION

An ideal barcode should possess sufficient variation among the sequences to discriminate species; however, it also needs to be sufficiently conserved so that there is less variability within species than between species (Kress and Erickson, 2008). DNA barcoding employs sequences from a short standardised gene region to identify species (Hebert et al., 2010). Now, *COI* can serve as the core of a global bioidentification system for animals, because it is suitable to diagnose diversity among DNA sequences to identify organisms. As an effective DNA-based identification system, (Hebert et al., 2010) confirmed that *COI* can be developed with modest effort. Recent studies have shown that sequence divergences are much larger among species than within species (Hebert et al., 2003a); thus, analysis of DNA divergences could be used for species identification (Hebert et al., 2004). Barcoding, together with related developments in sequencing technology, is likely to provide an efficient approach for the assembly of such genetic data. Kerr et al. (2004) confirmed that DNA barcoding can be effectively applied across the geographical and taxonomic expanse of North American birds. The tight clustering of mtDNA sequences within species observed in our study not only bolsters the view that species are fundamental biological units, but also reveals that their identification is usually uncomplicated. Hebert et al. (2003a) and Kerr et al. (2009b) found that all bird species had a different barcode and none was shared between species. Collectively, these phylogeographic studies establish that intraspecific divergences are ordinarily well below those that separate congeneric species pairs (Hebert et al., 2003b).

To date, the *COI* region has also been widely used to firmly establish the core barcode region for animals (Frezal and Leblois, 2008), and its performance for classification and identification have been evaluated in birds from several regions, including North America (Kerr et

al., 2007), Brazil (Chaves et al., 2008), Argentina (Kerr et al., 2009b), Korea (Yoo et al., 2006), and the eastern Palearctic (Kerr et al., 2009a). Although our samples were not collected within breeding habitat, our results did not suggest a relationship between the sequence differences within species and the geographic distances between specimen collection points, as found by Yoo et al. (2006) and Hebert et al. (2003a).

Most researchers agree that species are key units of biological systems, but disagree about how best to define them. Species-level paraphyly or even polyphyly in mtDNA is not common in the animal kingdom. According to Funk and Omland (2003), at least 702 arthropod species are polyphyletic (26.5%), and 23% of 2319 animal species show species-level paraphyly or polyphyly in their mtDNA gene trees, including 16.7% of 331 bird species. Several explanations have been offered for this: the analyses may have consisted of inadequate sequence data or inaccurate taxonomy and/or may have included interspecific hybridisation or introgression, and incomplete sorting (Park et al., 2009). For example, the mtDNA inherited marker could be a result of hybridisation events, which results in morphology that differs from the mtDNA make-up of the individual (Petersen et al., 2003). McKay and Zink (2010) examined 151 studies published between 1990 and 2008, and found that of the 856 potentially paraphyletic species, 122 (14.3%) showed mtDNA gene tree paraphyly. Most instances (86%) of paraphyly due to incorrect taxonomy can be eliminated by elevating a subspecies (or group of subspecies) to species status. mtDNA is rarely misleading and is usually a helpful tool for species delimitation. It has proven effective for exposing these taxonomic problems. Indeed, for DNA barcoding to work, sequence variation must be high enough between species so that they can be discriminated from one another; however, it must be low enough within species that a clear threshold between intra- and interspecific genetic variation can be defined. Sequence analysis of the *COI* amplicon has established that members of a species usually show low sequence variation, averaging 0.43% (SE = 0.017%), while congeneric species possess 18-fold higher mean divergences (7.70%, SE = 0.033%). The average interspecies genetic distance among cranes was 0.0406. We found that levels of sequence variation between species were slightly lower than those in most other groups. For example, interspecific variation in cranes averaged 4.06 vs 7.93% in North American birds (Hebert et al., 2004). We found that pairwise distance analysis can be used to precisely distinguish between genera and that pairwise distances of <0.1 indicate species belonging to the same genus.

Many studies have found that reduced coding, such as amino acid recoding of the nucleotide sequences, negatively affects phylogenetic reconstruction. Pleijel et al. (2008) found that converting nucleotide sequences into amino acid sequences or purine/pyrimidine coding reduced the phylogenetic signal and could introduce artificial relationships. Other research has found that the inclusion of third codon positions does not negatively affect phylogenetic reconstruction when all available data are analysed together. The best phylogenetic inference results are obtained when all available data are included and analysed together. Therefore, we conservatively tested the effectiveness of *COI* barcodes as an identification tool in phylogenetic analysis based on four methods (NJ, MP, ML, and Bayesian). All four methods yielded an identical ingroup topology with high support values. In short, the four phylum profiles showed good resolution of the major taxonomic groups (Figures 3-6), and the same topologic trees showed that different optimisation criteria produced invariable relationships among the three outgroups and the 15 species of crane. The three outgroup taxa clustered together to form a monophyletic group with robust support (>50%). In Gruidae, our *COI* results confirm

Krajewski's complete mtDNA phylogenetic relationships (Krajewski et al., 2010). Previous studies (Krajewski, 1989; Krajewski et al., 1999) on the subject have used a variety of molecular markers (Cytochrome b, ND6), which have led to results with limited resolution (Krajewski et al., 2010).

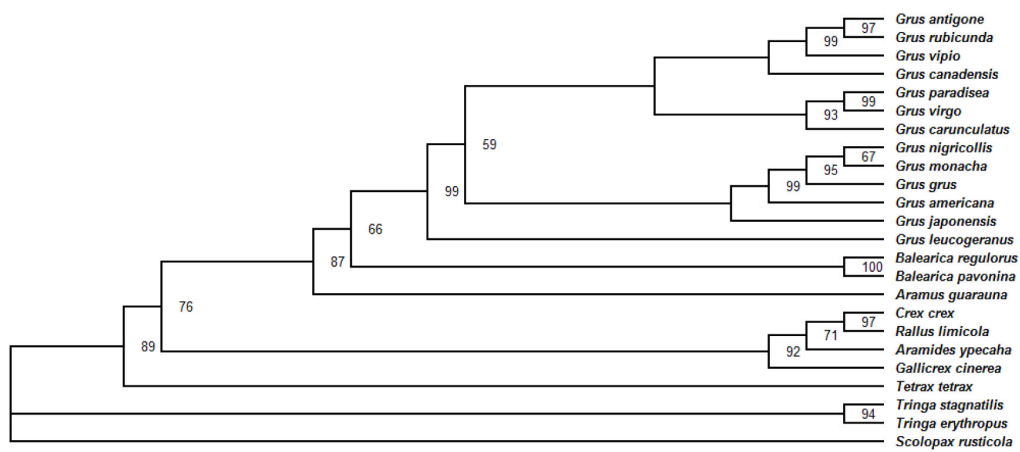
We found that the three trees supported a basal position for the Siberian crane (*G. leucogeranus*) within the Gruidae clade (NJ, 100%; MP, 50%; ML, 99%) as found by Krajewski et al. (2010). Crane classification was reasonably stable throughout the 20th century until Krajewski (1989) suggested that *Grus*, as constituted by Peters (1934), was not monophyletic because of the early separation of the Siberian crane (then *G. leucogeranus*) from other gruines. Indeed, because Livezey (1998) transferred the Siberian crane from *Grus* to the monotypic *Leucogeranus* and Krajewski (1989) placed *Anthropoides* as sister to *Bugeranus* based on molecular data, five well-supported species groups have been identified within the Gruidae: *Leucogeranus*, *Anthropoides*, *Canadensis*, *Antigone*, and *Americana* (Krajewski et al., 2010; Livezey, 1998), which our present research further validates. In our current study, the monophyly of the Gruidae clade, except for *A. guarauna*, was confirmed based on a short part of *COI*, as has been found using other methods, including morphometrics (Wood, 1979), allozyme analysis (Dessauer et al., 1992), DNA hybridization (Krajewski, 1989), and mitochondrial gene sequence assessments (Krajewski et al., 1999).

*Grus canadensis* clustered with the *Antigone* species group, but with low support in the NJ, MP, and ML trees (BP < 50%); it formed a single species group in the BA tree (0.51). Thus, its position with respect to *Antigone* was not resolved in our study. This could be because the different models and methods produced different results, the rates of evolution of the three molecular markers were different, or the mtDNA did not improve the resolution of the deep nodes of the phylogenetic tree. Other, deeper reasons for this include the following: i) the evolutionary relationships could be oversimplified by the data, ii) the genetic diversity may be underestimated using markers, and iii) uncertainty in genealogical analysis might increase because of the increased probability of more missing links in mitochondrial haplotypes.

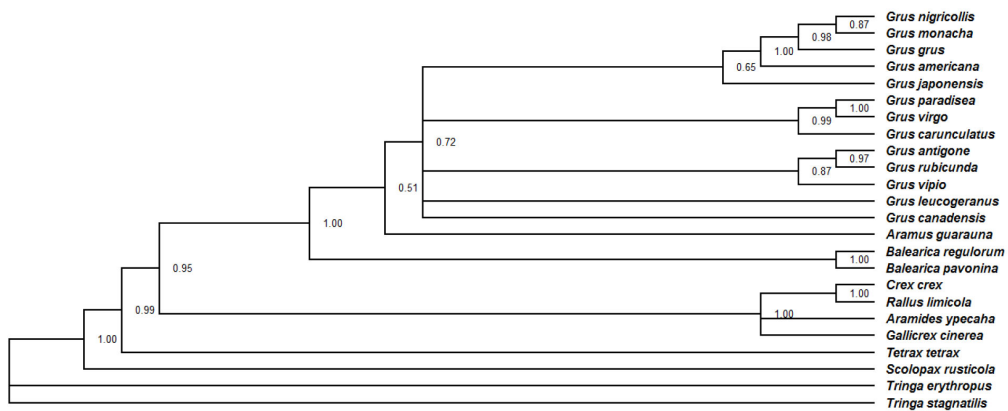
Indeed, the argued taxonomic category of *G. canadensis* emerged in many reports with various methods. Based on allozyme evidence, Dessauer et al. (1992) reported a sandhill crane group with two species clusters: one composed of the three sandhill crane subspecies plus the Siberian and sarus cranes, and one including the broilga crane of the genus *Grus*, the wattled crane of the genus *Bugeranus*, and the demoiselle and blue cranes of the genus *Anthropoides*. The midcontinental population of sandhill cranes, *G. canadensis*, in North America consists of approximately 500,000 individuals of three migratory subspecies: lesser (*G. c. canadensis*), Canadian (*G. c. rowani*), and greater (*G. c. tabida*). Limited morphometric and geographic investigations, along with DNA hybridization and biochemical allozyme analyses, have not clarified the relationships among *G. canadensis* subspecies (Krajewski, 1999). Petersen et al. (2003) found that genetic divergence ranged from 6.5 to 14.5% between *G. c. canadensis* and *G. c. tabida*, 0.5 to 6.6% within *G. c. canadensis*, and 0.1 to 6.0% within *G. c. tabida*. Krajewski et al. (2010) placed the sandhill crane (*G. canadensis*) as sister to the *Antigone* species group based on complete mitochondrial genomic sequences, resulting in a Pacific Rim clade that had not previously been suggested.

These resultant MP and BA trees have the same topology structure as the NJ and ML trees. In our current study, the family Aramididae (e.g., *A. guarauna*) was embedded in the clade of traditional Gruidae with strong support in the NJ (Figure 3), MP (Figure 4), ML (Figure 5),

and BA (Figure 6) trees (87% bootstrap values, 50% bootstrap values, 87% bootstrap values, and 0.51 posterior probability values, respectively). The same topology (with Aramididae as a close sister to Gruidae) has been reported in several studies (Fain et al., 2007; Livezey, 1989; Yang et al., 2010b). Livezey (1998) showed that the families Aramididae and Gruidae both belong to the superfamily Gruoidea, but the family Rallidae belongs to the order Ralliformes as opposed to the order Gruiformes. Opinions on the systematic relationships of Gruiformes are as diverse as the families included within it.



**Figure 5.** Maximum likelihood phylogram for *cytochrome C oxidase I (COI)* sequences based on a GTR+I+G pattern (Lset Base = [0.2885 0.3589 0.1482] Nst = 6 Rmat = [2.2027 10.8487 1.9942 1.0000 31.8922] Rates = gamma Shape = 2.5883 Pinvar = 0.6302). Numbers above the branches denote the posterior probabilities (only those  $\geq 50$  are shown).



**Figure 6.** Bayesian tree reconstructed from *cytochrome C oxidase I (COI)* sequences. Numbers above the branches denote the posterior probabilities (only those  $\geq 0.50$  are shown) with gall types and primary host mapped on top.

Despite ongoing debate over monophyly of the order and relationships among its various members, recent opinions have converged on the monophyly of a 'core' group of five families classified as the suborder Grues: the rails (Rallidae), the cranes (Gruidae), the limpkins (Aramidae), the trumpeters (Psophiidae), and the finfoots (Heliornithidae) (Hackett et al., 2008; Yang et al., 2010b). A consensus has begun to emerge that there is a monophyletic 'core' consisting of Gruidae, Aramidae, Psophiidae, Heliornithidae, and Rallidae (Fain et al., 2007). Based on the mtDNA sequences of 12S rRNA, 16S rRNA, and tRNA-Val, (Yang et al., 2010b) reported that Otididae was a sister group to 'core Gruiformes' and Charadriiformes with strong support (97% posterior probability values) in Bayesian analysis. However, Hackett et al. (2008) reported that Otididae was outside of the core Gruiformes and included bustards in the order 'Otidiformes'. Fain et al. (2007) indicated that the genus *Tetrax* (Gruiformes: Otididae) is located at a basal position of 'core Gruiformes' (Aves: Grues) with strong support, suggesting that Otididae belongs to the order Gruiformes. Houde et al. (1997) addressed the problem of Gruiform relationships with mitochondrial 12S rRNA sequences and found that limpkins were supported as a sister group to cranes, indicating that *G. japonensis* had a prior evolutionary origin. In short, the phylum profile showed good resolution of the major taxonomic groups.

## ACKNOWLEDGMENTS

Research supported by the National Natural Science Foundation of China (#31101705 and #31101032), the Natural Science Foundation of Jiangsu Province (#SBK201021501), the Research Fund for the Doctoral Program of Higher Education of China (#200803071048), a Youth Science and Technology Innovation Fund Project (#KJ09013) and the Agricultural Science and Technology Independent Innovation Fund of Jiangsu Province (#09CX116).

## REFERENCES

- Chaves AV, Clozato CL, Lacerda DR, Sari EH, et al. (2008). Molecular taxonomy of Brazilian tyrant-flycatchers (Passeriformes: Tyrannidae). *Mol. Ecol. Resour.* 8: 1169-1177.
- Dessauer HC, Gee GF and Rogers JS (1992). Allozyme evidence for crane systematics and polymorphisms within populations of Sandhill, Sarus, Siberian, and whooping cranes. *Mol. Phylogenet. Evol.* 1: 279-288.
- Fain MG (2001). Phylogeny and evolution of cranes (Aves: Gruidae) inferred from DNA sequences of multiple genes. Ph.D. dissertation.
- Fain MG, Krajewski C and Houde P (2007). Phylogeny of "core Gruiformes" (Aves: Grues) and resolution of the Limpkin-Sungrebe problem. *Mol. Phylogenet. Evol.* 43: 515-529.
- Frezal L and Leblois R (2008). Four years of DNA barcoding: current advances and prospects. *Infect. Genet. Evol.* 8: 727-736.
- Funk DJ and Omland KE (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Systemat.* 34: 397-423.
- Hackett SJ, Kimball RT, Reddy S, Bowie RC, et al. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science* 320: 1763-1768.
- Hajibabaei M, Singer GA, Clare EL and Hebert PD (2007). Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol.* 5: 24.
- Hebert PD, Cywinska A, Ball SL and deWaard JR (2003a). Biological identifications through DNA barcodes. *Proc. Biol. Sci.* 270: 313-321.
- Hebert PD, Ratnasingham S and deWaard JR (2003b). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.* 270 (Suppl 1): S96-S99.
- Hebert PD, Stoeckle MY, Zemlak TS and Francis CM (2004). Identification of Birds through DNA Barcodes. *PLoS Biol.* 2: e312.

- Hebert PD, deWaard JR and Landry JF (2010). DNA barcodes for 1/1000 of the animal kingdom. *Biol. Lett.* 6: 359-362.
- Houde P, Cooper A, Leslie E, Strand AE, et al (1997). Phylogeny and Evolution of 12S rDNA in Gruiformes. In: Avian Molecular Evolution and Systematics (Mindell DP, ed.). Academic Press, San Diego, 121-158.
- Kerr KC, Stoeckle MY and Dove CJ (2007). Comprehensive DNA barcode coverage of North American birds. *Mol. Ecol. Notes* 7: 535-543.
- Kerr KC, Birks SM and Kalyakin MV (2009a). Filling the gap - COI barcode resolution in eastern Palearctic birds. *Front. Zool.* 6: 29.
- Kerr KC, Lijtmaer DA, Barreira AS, Hebert PD, et al. (2009b). Probing evolutionary patterns in neotropical birds through DNA barcodes. *PLoS One* 4: e4379.
- Kimura M (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111-120.
- Krajewski C (1989). Phylogenetic relationships among cranes (Gruiformes: Gruidae) based on DNA hybridization. *Auk* 106: 603-618.
- Krajewski C, Fain MG, Buckley L and King DG (1999). Dynamically heterogeneous partitions and phylogenetic inference: an evaluation of analytical strategies with cytochrome b and ND6 gene sequences in cranes. *Mol. Phylogenet. Evol.* 13: 302-313.
- Krajewski C, Sipiorski JT and Anderson FE (2010). Complete mitochondrial genomes and the phylogeny cranes (Gruiformes: Gruidae). *Auk* 127: 440-452.
- Kress WJ and Erickson DL (2008). DNA barcodes: genes, genomics, and bioinformatics. *Proc. Natl. Acad. Sci. U. S. A.* 105: 2761-2762.
- Livezey BC (1989). Flightlessness in grebes (Aves, Podicipedidae): its independent evolution in three genera. *Evolution* 43: 29-54.
- Livezey BC (1998). A phylogenetic analysis of the Gruiformes (Aves) based on morphological characters, with an emphasis on the rails (Rallidae). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 353: 2077-2151.
- Marshall E (2005). Taxonomy. Will DNA bar codes breathe life into classification? *Science* 307: 1037.
- McKay BD and Zink RM (2010). The causes of mitochondrial DNA gene tree paraphyly in birds. *Mol. Phylogenet. Evol.* 54: 647-650.
- Park SH, Zhang Y, Piao H, Yu DH, et al. (2009). Use of cytochrome c oxidase subunit i (COI) nucleotide sequences for identification of the Korean Luciliinae fly species (Diptera: Calliphoridae) in forensic investigations. *J. Korean Med. Sci.* 24: 1058-1063.
- Pedersen N, Holyoak DT and Newton AE (2007). Systematics and morphological evolution within the moss family Bryaceae: a comparison between parsimony and Bayesian methods for reconstruction of ancestral character states. *Mol. Phylogenet. Evol.* 43: 891-907.
- Petersen JL, Bischof R, Krapu GL and Szalanski AL (2003). Genetic variation in the midcontinental population of sandhill cranes, *Grus canadensis*. *Biochem. Genet.* 41: 1-12.
- Pleijel F, Jondelius U, Norlinder E, Nygren A, et al. (2008). Phylogenies without roots? A plea for the use of vouchers in molecular phylogenetic studies. *Mol. Phylogenet. Evol.* 48: 369-371.
- Ronquist F and Huelsenbeck JP (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
- Sambrook J and Russell DW (2001). *Molecular Cloning: A Laboratory Manual*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Sibley CG and Monroe BL (1990). *Infraorder Gruidae*. In: *Distribution and Taxonomy of Birds of the World* Yale University Press, New Haven, London, 217-219.
- Tamura K and Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10: 512-526.
- Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24: 1596-1599.
- Wilgenbusch JC and Swofford D (2003). Inferring Evolutionary Trees with PAUP\*. *Curr. Protoc. Bioinformatics* Chapter 6, Unit 6.4.
- Wood DS (1979). Phenetic relationships within the family Gruidae. *Wilson Bull.* 91: 384-399.
- Xia X and Xie Z (2001). DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.* 92: 371-373.
- Yang R, Wu X, Yan P and Li X (2010a). Using DNA barcodes to identify a bird involved in a birdstrike at a Chinese airport. *Mol. Biol. Rep.* 37: 3517-3523.
- Yang R, Wu X, Yan P, Su X, et al. (2010b). Complete mitochondrial genome of *Otis tarda* (Gruiformes: Otidae) and phylogeny of Gruiformes inferred from mitochondrial DNA sequences. *Mol. Biol. Rep.* 37: 3057-3066.
- Yoo HS, Eah JY, Kim JS, Kim YJ, et al. (2006). DNA barcoding Korean birds. *Mol. Cells* 22: 323-327.