

## Analysis of Selective Constraints on Mitochondrial DNA, Flight Ability and Physiological Index on Avian

Shanxin Zhang, Jiuqiang Han, *Member, IEEE*, Dexing Zhong\*, *Member, IEEE*, and Tuo Wang

**Abstract**—For most of the birds in the world, they can be divided into two main groups, i.e. resident birds and migratory ones. Most of the energy required for long-distance migration is supplied by mitochondria via oxidative phosphorylation. Therefore, the evolutionary constraints acted on the mitochondria DNA (mtDNA) are considered to vary with the locomotive abilities and flight speed. The flight speed is assumed to increase with mass and wing loading according to the fundamental aerodynamic theories, which is common between aves and aircrafts. We compared 148 avian mitochondrial genomes and main physiological parameters. More nonsynonymous nucleotide substitutions than synonymous ones are accumulated in low-speed and flightless birds rather than high-speed flying birds. No matter how the speed is obtained, directly measured or estimated through physiological index. Our results demonstrated that, besides artificial and environmental factors, selective constraints relevant to flight ability play an essential role in the evolution of mtDNA, even it might cause the extinction of avian species.

### I. INTRODUCTION

As is known, mitochondria have two important roles, i.e. energy metabolism and thermoregulation [1-3]. On one hand, mitochondrion supplies free energy adenosine triphosphate to support energy-consuming process by OXPHOS. On the other hand, it plays an essential role in thermoregulation process that is facilitated to adapt to various survival environment [4, 5]. More attention has been attracted on the role of selection in mtDNA evolution [1, 4-9]. Recent studies reveal that the 13 protein-encoding mtDNA in fish [10] are affected by dual functional constraints including energy production and heat generation. Related genes have different selective constraints depending on which factor exerts greater effect.

As one of the most successful groups of vertebrates, birds distribute all over the world with different morphology to accommodate to distinct circumstance. Obviously flight ability has a crucial role to the adaptability of birds in assisting them in feeding, perching, breeding and predator avoidance [11]. Flight needs very plenty of energy during whatever kind of flight modes. Bird's ecology and phylogeny, such as mass and wing loading, have significant impacts on the flight speed which reflects the flight adaptation [12]. Although during the long

evolution process some of the species remain the transitional locomotive models, some have degeneration of flight ability, and even some lost their ability to fly (e.g., kiwi, penguins and ratites), some of the locomotive styles of flightless species, such as diving or running, are still largely energy-consuming [13]. At the same time, about one fifth of the extant birds are migratory. These birds cruise a long distance between conventional destinations far beyond their breeding range. Compared with the resident birds, they need process more active metabolism to provide more energy. Moreover, resident birds may suffer more different thermoregulation for the changing climate, which influences heat generation efficiency. It seems that mtDNA of birds suffers dual functional selective constraints of genes related to energy production and heat generation.

In this paper, we focus on the different selective constraints of the bird mtDNA genes related to energy production and heat generation. We analyze all avian mitochondrial genomes (148 species) available in GenBank. The relationship between Ka/Ks ratio (nonsynonymous nucleotide substitutions relative to synonymous substitutions) and flight ability (directly measured flight speeds and indirectly inferred ones based on physiological index according to fundamental aerodynamic theories) has been analyzed. Classification by locomotion ability and whether migratory birds are taken into account to reveal the functional constraints in the evolution of mtDNA.

### II. MATERIALS AND METHODS

#### A. Subjects

Raw avian mitochondrial genomes are downloaded from Genbank, which contains 148 species available in July 2011. All of the 13 protein-coding mitochondrial genes are extracted from each genome and translated into amino acid sequences. Data for birds' morphology (masses, body length and wingspan) and speed are collected from previous studies [1, 12, 14] and some websites ("Animal Diversity Web" [15], IUCN [16] and "BirdLife"[17]).

We first divided 148 avian species into two groups: strongly locomotive birds and weakly locomotive ones. The ability of fast locomotion includes almost all high energy consuming movement styles, no matter the species are flying, running or swimming. Then these taxa are divided into two main classes according to whether they are migration with the criterion used in "BirdLife" website.

#### B. Construct Phylogenetic Tree

The amino acid sequences of all 13 protein-coding genes of each species are concatenated. Sequences were aligned with a combination of ClustalX with default setting and manual alignment, and the results were verified by visual inspection.

This work was supported by grants from National Natural Science Foundation of China (No. 61105021, No. 61071217), Ph.D. Program Foundation of the Ministry of Education of China (No. 20110201110010), China Postdoctoral Science Foundation (No. 2011M501442) and the Fundamental Research Funds for the Central Universities.

S. Zhang (e-mail: [shanxin.zh@gmail.com](mailto:shanxin.zh@gmail.com)), J. Han (email: [jqhan@xjtu.edu.cn](mailto:jqhan@xjtu.edu.cn)), D. Zhong (e-mail: [bell@xjtu.edu.cn](mailto:bell@xjtu.edu.cn)), and T. Wang (e-mail: [tuowang@xjtu.edu.cn](mailto:tuowang@xjtu.edu.cn)) are with the Ministry of Education Key Lab for Intelligent Networks and Network Security, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China.

\*Corresponding author, phone: +86-029-82668665-175, [bell@xjtu.edu.cn](mailto:bell@xjtu.edu.cn).

PAUP\* [18] was used for MP analyses. We analyzed the sequences by conducting a heuristic search of 1000 random-additional replicates with TBR branch-swapping. Bootstrap support was evaluated with 500 pseudo-replicates of 5 random-additions each, as mentioned by Hackett [14].

ML analyses were conducted with phyML [19], using the dataset based on the 13 protein-coding genes. ML model support was calculated by analyzing 100 bootstrap replicates.

Recently, Hackett construct a robust avian nuclear tree based on 19 nuclear genes [14]. We used their tree to guide our mitochondrial phylogeny. At the same time, we compared the trees we obtained by the two methods and we chose the ML tree as our synthetical phylogenetic tree. The tree was edited with Dendroscope [20].

### C. Ratio of Nonsynonymous to Synonymous Nucleotide Substitutions (Ka/Ks)

PAML [21, 22] was used to analysis the ratio of nonsynonymous to synonymous nucleotide substitutions (Ka/Ks) of all individual data sets estimated for each branch of the phylogenetic tree. The avian tree was divided into five monophyletic subtrees to reduce the time consumed to estimate the data, and each of those was evaluated independently. Model 1 was utilized to estimate a free Ka/Ks ratio for each branch. In other words, we focused on the Ka/Ks values associated with terminal branches, *i.e.* the rate between modern species and their most recent reconstructed ancestors, as mentioned by Shen *et al.* [1].

Species with missing molecular data are discarded because their most recent ancestors on the tree are smaller than 10, as a result they contain a number of nonsynonymous substitutions, as mentioned by Sun *et al.* [10].

### D. Statistic Treatments

We conducted a linear regression analysis comparing Ka/Ks ratios to the maximum speed for 38 avian species. Some birds without speed records can be indirectly inferred by their physiological index according to fundamental aerodynamics formula, the lift force ( $L$ ) generated on a wing is related to flight speed ( $U$ ) as following:

$$L = \frac{1}{2} \rho \cdot C_L \cdot S \cdot U^2 \quad (1)$$

Where  $\rho$  is air density,  $S$  is wing area, and  $C_L$  is the lift coefficient [7, 16]. In horizontal cruising flight  $L$  balances the weight ( $m \times g$ ), therefore the flight speed  $U$  can be rewritten as following:

$$U = \sqrt{\frac{2g}{\rho \cdot C_L}} \cdot \sqrt{m/S} = k \sqrt{m/S} \quad (2)$$

Where  $k$  is a substituted constant, which provides that the three parameters are about equal among bird species and is inferred from the parameters in [23]. This substituted constant  $k$  can be used to infer the flight speed when the value of maximum speed is out of record and hard to measure under the existing conditions whereas the average mass and wing area are available or easy to get.

Furthermore, we would like to pay more attention on the strongly locomotive birds about the selective constraints on

mtDNA, flight ability and physiological index of them, especially on crested ibis for the potential reason of extinction.

## III. RESULTS

The maximum likelihood phylogenetic tree of aves based on 148 complete avian mitochondrial genomes is presented in Fig. 1. As shown in Fig.1, residential birds (marked in purple), most of which are weekly locomotive or flightless birds, do not cluster together.

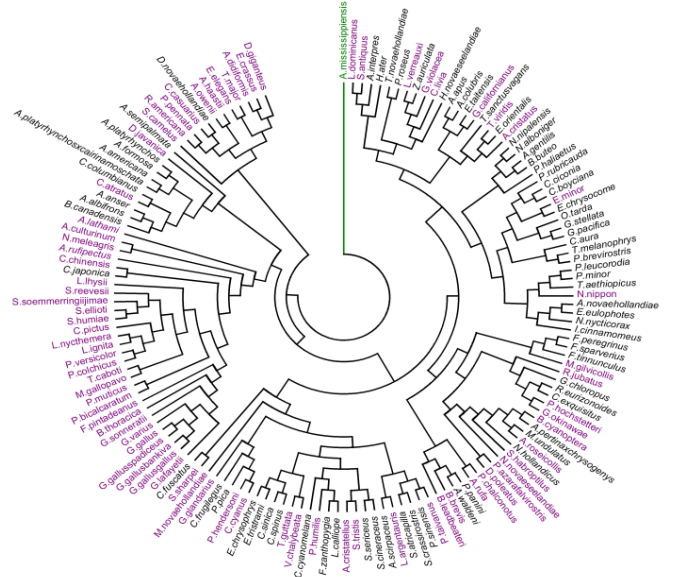


Figure 1. Avian phylogenetic tree reconstructed from mtDNA data. The tree of 148 avian species was constructed from all 13 protein-coding genes from mtDNA genomes using the maximum parsimony (MP) and maximum likelihood (ML) synthetical methods. The residential birds are marked in purple and the outgroup (*A. mississippiensis*) is marked in green.

The mean Ka/Ks ratios of all mtDNA genes is 0.0457 in the weakly locomotive group, which is distinct greater (19.32% larger) than that of the strongly locomotive group. The mean Ka/Ks ratios of the residential birds is 0.0433, which is also much greater (19.28% larger) than that of the migratory birds.

The linear regression comparing Ka/Ks to the speeds is demonstrated in Fig. 2A, and the one comparing Ka/Ks to the derived maximum speeds are shown in Fig. 2B. In general, the Ka/Ks ratios is significantly negatively correlated with increasing speed, revealing a relaxation of selective constraints following the degeneration of locomotive ability.

The distribution of 38 strongly locomotive avian species is plotted in a three-dimensional space based on three feature vectors (*i.e.* Ka, Ks and Speed), shown in Fig. 3. The red plane is the hyperplane with Ka/Ks=0.04 which is the approximate value dividing the strongly locomotive birds from the weakly ones and the migratory birds from the residential ones (average values of Ka/Ks ratios, strongly locomotive birds: 0.0383, weakly ones: 0.0457; migratory birds: 0.0363, residential ones: 0.0433). Fifteen representative birds are selected and linked to demonstrate their location and relation in the specific 3D space.

The value of Ka/Ks ratio varied among these 13 genes, demonstrating that different genes accumulated different amounts of deleterious mutations between the strongly and

weakly locomotive groups (Fig. 4), also between the migratory and residential birds (Fig. 5).

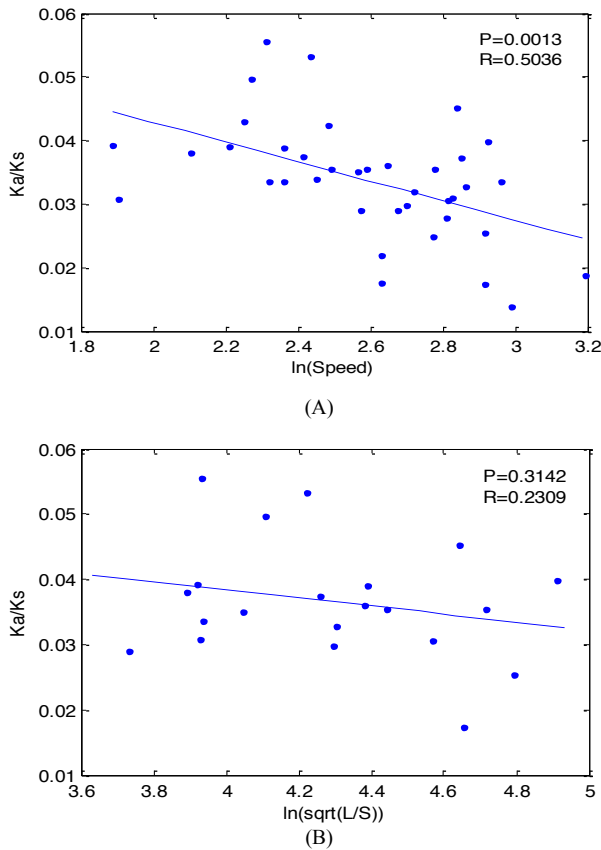


Figure 2. Negative correlation between Ka/Ks ratios of mtDNA and locomotive speed. (A) Ordinary linear regression of Ka/Ks ratios to loge speed for 38 avian species,  $P=0.0013$ ,  $R=0.5036$ . (B) The linear regression of Ka/Ks ratios to loge (L/S), which is the indirectly inferred speed according to fundamental aerodynamics formula, for 38 avian species,  $P=0.3142$ ,  $R=0.2309$ .

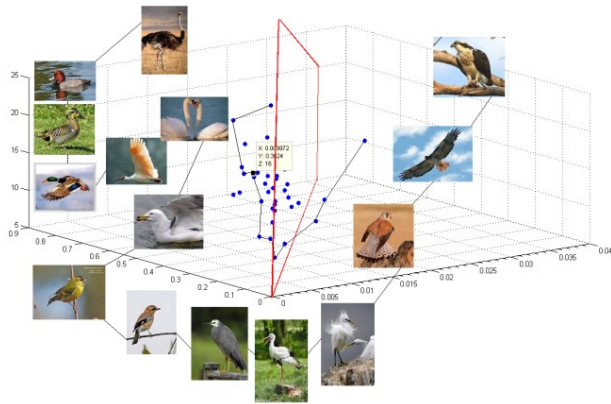


Figure 3. 3D schematic diagram based on three feature vector (i.e. Ka, Ks and Speed) for 38 strongly locomotive avian species. The red plane is the hyperplane with  $Ka/Ks=0.04$  which is the approximate value dividing the strongly locomotive birds from the weakly ones and the migratory birds from the residential ones. Fifteen representative birds are selected and linked to demonstrate their location and relation in the specific 3D space. From the left part to the right, there are *Struthio camelus*, *Aythya Americana*, *Anas Formosa*, *Anas platyrhynchos*, *Nipponia Nippon* (Crested ibis, the black square in the graph), *Cygnus columbianus*, *Larus dominicanus*, *Carduelis spinus*, *Garrulus glandarius*, *Ardea novaehollandiae*, *Ciconia ciconia*, *Egretta eulophotes*, *Falco tinnunculus*, *Accipiter gentiles* and *Pandion haliaetus*.

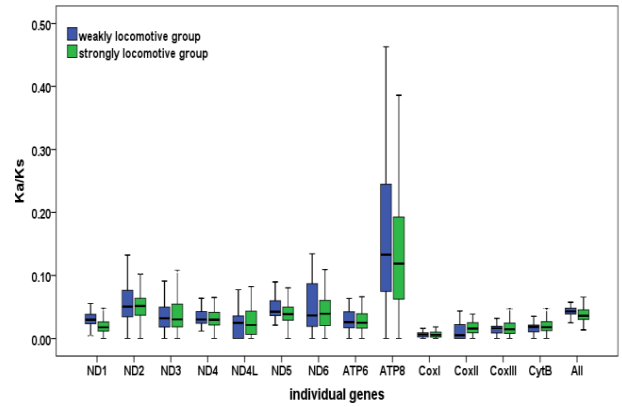


Figure 4. Comparisons of average Ka/Ks ratios for 13 individual genes and all protein-coding genes set between weakly and strongly locomotive groups. Boxes include 50% of the distributions. ND1,  $P<0.001$ ; ND2,  $P=0.361$ ; ND3,  $P=0.591$ ; ND4,  $P=0.207$ ; ND4L,  $P=0.541$ ; ND5,  $P=0.004$ ; ND6,  $P=0.185$ ; ATP6,  $P=0.973$ ; ATP8,  $P=0.376$ ; CoxI,  $P=0.823$ ; CoxII,  $P=0.011$ ; CoxIII,  $P=0.420$ ; CytB,  $P=0.070$ ; All genes,  $P=0.002$ . Note: \* $P=0.001\sim0.05$ , \*\* $P<0.001$ .

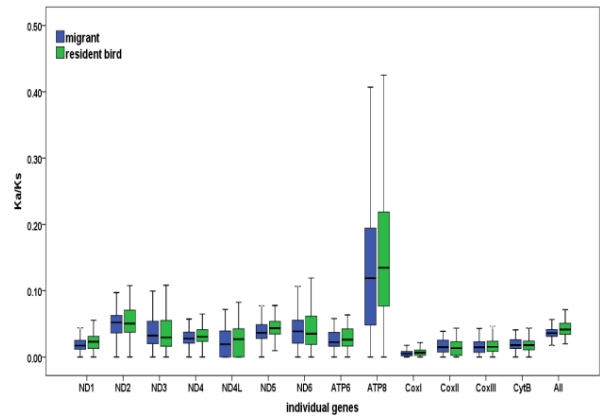


Figure 5. Comparisons of average Ka/Ks ratios for 13 individual genes and all protein-coding genes set between migratory and residential birds groups. Boxes include 50% of the distributions. ND1,  $P=0.032$ ; ND2,  $P=0.097$ ; ND3,  $P=0.511$ ; ND4,  $P=0.175$ ; ND4L,  $P=0.333$ ; ND5,  $P=0.018$ ; ND6,  $P=0.734$ ; ATP6,  $P=0.313$ ; ATP8,  $P=0.333$ ; CoxI,  $P=0.152$ ; CoxII,  $P=0.570$ ; CoxIII,  $P=0.542$ ; CytB,  $P=0.394$ ; All genes,  $P=0.001$ . Note: \* $P=0.001\sim0.05$ .

#### IV. DISCUSSION

As birds evolved from a single common ancestor [15], in this paper, we can determine whether the mtDNA of different groups of birds experienced different strength of selective pressures during the evolution through calculating the ratio of rates of nonsynonymous (change in amino acid) substitution over synonymous (silent) substitutions (Ka/Ks). The result demonstrates that a relatively stronger negative selection occurred in the strongly locomotive group compared to the weakly locomotive group. The result also indicates that relaxation of selective constraints occurred in the residential birds, since the migratory birds need more energy consumption during their long-distance flight.

The reasons for the phenomenon can be summarized as following: the first hypothesis is due to the impact of effective population sizes ( $N_e$ ). Weakly locomotive birds generally have lower  $N_e$ , since their lower dispersal ability. As  $N_e$  declines, slightly deleterious mutations have a greater probability of reaching fixation [24]. Therefore, weakly locomotive birds tend to accumulate more nonsynonymous mutations and greater

value of Ka/Ks. The metabolic power required for locomotion is linearly related to speed [25]. Up to 95% of the cell's energy requirements are provided by OXPHOS, in which proteins encoded by mtDNA play a critical role. The maximum speeds would be expected to strongly correlate with the Ka/Ks ratio of mtDNA genes.

The values of Ka/Ks ratios of *Pandion haliaetus*, *Accipiter gentiles* and *Falco tinnunculus* are close to or even a little greater than the average value of the weakly locomotive birds, because their synonymous substitutions are relatively greater than their nonsynonymous substitutions. This feature is even significant in the performance of *Falco tinnunculus*, whose nonsynonymous substitutions and synonymous substitutions are both considerable smaller. While the Ka/Ks ratios of *Accipiter gentiles* and *Pandion haliaetus* are almost equivalent, and both of them are in the same seventh level branch. These species are all experts in aerodynamics (aerodynamics need less energy consumption than flapping wings to keep flight speed), therefore although they accumulated more nonsynonymous nucleotide substitutions relative to synonymous substitutions in mtDNA genes; it doesn't seriously affect their flight speeds to make them become weakly locomotive birds.

From the avian phylogenetic tree, crested ibis, *Nipponia nippon*, is close neighbor of *Threskiornis aethiopicus*, *Platalea minor* and *Platalea leucorodia* (same sub-branch); *Egretta eulophotes*, *Ardea novaehollandiae* (different sub-branch) and *Ciconia ciconia* (relatively further). However, in Fig. 3 crested ibis (marked with Black Square) is closed to *Anas platyrhynchos* and *Cygnus columbianus* instead of its phylogenetic neighbors (*Ardea novaehollandiae*, *Ciconia ciconia*, *Egretta eulophotes*). Crested ibis is further away from its phylogenetic neighbors, whose values of Ka and Ks are relatively small. Specifically, the value of Ka of crested ibis is about 6 times greater than the average one of *Platalea minor* and *Platalea leucorodia*; while the value of Ks of crested ibis is about 9 times greater than the average one of them, respectively.

Why does not crested ibis locate close to its phylogenetic neighbors in the 3D feature space? Although the value of Ka/Ks ratio of crested ibis is smaller than its phylogenetic neighbors (*Platalea minor* and *Platalea leucorodia*), the individual value of Ka and Ks are significant greater than the average values of its neighbors (much more synonymous substitutions in mtDNA genes). A relatively acknowledged reason is that the groups of wild crested ibis have been disappeared at the end of 1970s due to environmental degradation and other factors, then until May of 1981 scientist have found a small group of them in Yang County of Shaanxi Province, China. Therefore, its *Ne* was extremely small at that time and prior, so that slightly deleterious mutations have a greater probability of reaching fixation to accumulate more nonsynonymous and synonymous mutations in mtDNA genes. The nonsynonymous mutations in mtDNA may be another important reason for the almost extinction of crested ibis.

For the locomotive ability of avian species, some genes may play more important roles than other genes, which have undergone much stronger selective constraints to eliminate deleterious mutations to maintain their function than others [1]. For whether migratory or residential the bird is, flight ability is just one of the factors determining this natural behavior besides

climate, environment *etc.*, therefore less significant differences (only 2 genes) are found among the 13 individual genes from the mtDNA. The functional constraints can influence the evolution of mtDNA in avian species, which may influence their locomotive speed, change their locomotive type (from flight to running or swimming), even change them from migratory birds to residential ones (*e.g.* Crested ibis).

## REFERENCES

- [1] Y. Y. Shen, *et al.*, "Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability," *Genome research*, vol. 19, pp. 1760-1765, 2009.
- [2] J. W. O. Ballard and M. C. Whitlock, "The incomplete natural history of mitochondria," *Molecular Ecology*, vol. 13, pp. 729-744, 2003.
- [3] D. M. Rand, "Mitigating mutational meltdown in mammalian mitochondria," *PLoS biology*, vol. 6, p. e35, 2008.
- [4] D. Mishmar, *et al.*, "Natural selection shaped regional mtDNA variation in humans," *Proceedings of the National Academy of Sciences*, vol. 100, pp. 171-176, 2003.
- [5] E. Ruiz - Pesini and D. C. Wallace, "Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA," *Human mutation*, vol. 27, pp. 1072-1081, 2006.
- [6] J. Elson, *et al.*, "Comparative genomics and the evolution of human mitochondrial DNA: assessing the effects of selection," *The American Journal of Human Genetics*, vol. 74, pp. 229-238, 2004.
- [7] C. Sun, *et al.*, "The role of climate in human mitochondrial DNA evolution: a reappraisal," *Genomics*, vol. 89, pp. 338-342, 2007.
- [8] T. A. Castoe, *et al.*, "Adaptive evolution and functional redesign of core metabolic proteins in snakes," *PLoS One*, vol. 3, p. e2201, 2008.
- [9] R. R. Da Fonseca, *et al.*, "The adaptive evolution of the mammalian mitochondrial genome," *Bmc Genomics*, vol. 9, p. 119, 2008.
- [10] Y. B. Sun, *et al.*, "Evaluating the roles of energetic functional constraints on teleost mitochondrial-encoded protein evolution," *Molecular biology and evolution*, vol. 28, pp. 39-44, 2011.
- [11] C. Berg and J. Rayner, "The moment of inertia of bird wings and the inertial power requirement for flapping flight," *Journal of experimental biology*, vol. 198, pp. 1655-1664, 1995.
- [12] T. Alerstam, *et al.*, "Flight speeds among bird species: allometric and phylogenetic effects," *PLoS biology*, vol. 5, p. e197, 2007.
- [13] H. F. James and S. L. Olson, "Flightless birds," *Natural History*, vol. 92, pp. 30-40, 1983.
- [14] S. J. Hackett, *et al.*, "A phylogenomic study of birds reveals their evolutionary history," *science*, vol. 320, pp. 1763-1768, 2008.
- [15] P. Myers, *et al.*, "The animal diversity web," Accessed October, vol. 12, p. 2, 2006.
- [16] S. C. IUCN, "IUCN red list categories and criteria: version 3.1," Prepared by the IUCN Species Survival Commission, 2001.
- [17] B. International, "The BirdLife Checklist of the Birds of the World with Conservation Status and Taxonomic Sources," 2007.
- [18] D. L. Swofford, "PAUP\*. Phylogenetic Analysis Using Parsimony (\* and Other Methods)". Version 4, 2003.
- [19] S. Guindon and O. Gascuel, "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood," *Systematic biology*, vol. 52, pp. 696-704, 2003.
- [20] D. Huson, *et al.*, "Dendroscope: An interactive viewer for large phylogenetic trees," *BMC bioinformatics*, vol. 8, p. 460, 2007.
- [21] Z. Yang, "PAML: a program package for phylogenetic analysis by maximum likelihood," *Computer applications in the biosciences: CABIOS*, vol. 13, pp. 555-556, 1997.
- [22] Z. Yang, "PAML 4: phylogenetic analysis by maximum likelihood," *Molecular biology and evolution*, vol. 24, pp. 1586-1591, 2007.
- [23] G. Spedding, "The aerodynamics of flight," *Advances in comparative and environmental physiology*, vol. 11, pp. 51-111, 1992.
- [24] M. Kimura, "On the probability of fixation of mutant genes in a population," *Genetics*, vol. 47, p. 713, 1962.
- [25] C. R. Taylor, *et al.*, "Energetics and mechanics of terrestrial locomotion. I. Metabolic energy consumption as a function of speed and body size in birds and mammals," *Journal of experimental biology*, vol. 97, pp. 1-21, 1982.