The first set of universal nuclear protein-coding loci markers for avian phylogenetic and population genetic studies

Yang Liu[1§], Simin Liu[1§], Chia-Fen Yeh[2], Nan Zhang[1], Guoling Chen[1], Pinjia Que[3], Lu Dong[3]*, Shou-hsien Li[2*]

1 State Key Laboratory of Biocontrol, Department of Ecology/School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, Guangdong, China

2 Department of Life Sciences, National Taiwan Normal University, Taipei, 116, Taiwan, China

3 Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, 100875, China

*__Correspondence:__ Lu Dong, Email: donglu@bnu.edu.cn

Shou-hsien Li, Email: t43028@gmail.com

§ These authors equally contributed to this work.

Running header: Universal avian nuclear markers

Number of words in main body of paper: 4376

Number of references: 88

Abstract

Multiple nuclear markers provide genetic polymorphism data for molecular systematics and population genetic studies. They are especially required for the coalescent-based analyses that can be used to accurately estimate species trees and infer population demographic histories. However, in avian evolutionary studies, these powerful coalescent-based methods are hindered by the lack of a sufficient number of markers. In this study, we designed PCR primers to amplify 136 nuclear protein-coding loci (NPCLs) by scanning the published Red Junglefowl (*Gallus gallus*) and Zebra Finch (*Taeniopygia guttata*) genomes. To test their utility, we amplified these loci in 41 bird species representing 23 Aves orders. The sixty-three best-performing NPCLs, based on high PCR success rates, were selected which had various mutation rates and were evenly distributed across 17 avian autosomal chromosomes and the Z chromosome. To test phylogenetic resolving power of these markers, we conducted a Neoavian phylogenies analysis using 63 concatenated NPCL markers derived from 48 whole genomes of birds. The resulting phylogenetic topology, to a large extent, is congruence with results resolved by previous whole genome data. To test the level of intraspecific polymorphism in these makers, we examined the genetic diversity in four populations of the Kentish Plover (*Charadrius alexandrinus*) at 17 of NPCL markers chosen at random. Our results showed that these NPCL markers exhibited a level of polymorphism comparable with mitochondrial loci. Therefore, this set of pan-avian nuclear protein-coding loci has great potential to facilitate studies in avian phylogenetics and population genetics.

Although the next generation sequencing technologies have produced sequences data in the unprecedented quantity with relative low cost[1], traditional Sanger sequencing still has its niche in molecular evolutionary studies: pilot or small scale phylogenetic studies using PCR-based approach are cost-effectively and nearly available for every laboratory, beneficial to design sampling strategy and built an analysis scheme. By comparing molecular phylogenies based on different sizes of dataset, Rokas et al.[2] proposed that concatenation of a sufficient number of unlinked genes (>20) can overwhelm incongruent branches of the Tree of Life (TOL). Furthermore, tracing backwards from multiple genetic polymorphisms to find the most recent common ancestor (MRCA) of a group of individuals provides a sophisticated approach to clarify phylogenetic relationships among species (species tree approach) and to reconstruct the demographic history of populations[3,4]. However, the major drawback of this approach is the PCR performance of primers developed from one species is often unpredictable in the distantly related species; consequently, it is a time and cost consuming process to evaluate the performance of primers in a previously untested species. Therefore, a set of universal nuclear markers could provide an efficient way to ease this time consuming process. It should greatly facilitate the use of coalescent-based analyses to answer phylogenetic and population genetic questions[5].

Nuclear Protein-coding Loci (NPCLs) are exons without flanking introns[6], and are widely used in interspecific phylogenetic studies (e.g. *RAG1*[7], *c-myc* [8,9]). NPCL markers possess favorable properties including homogeneous base composition, varied evolutionary rates and easy alignment across species or populations[10,11]. Moreover, orthologous genes can be identified accurately using their annotations [12,13]. Several sets of universal NPCL markers had been developed specially for beetles[14], fish[15], reptiles[6], amphibian and vertebrates[16,17]. However, there is still no sufficient number of easily

amplifiable NPCL markers that can fulfill the needs of modern coalescent-based analysis for most of bird species. As the most common and species-rich group of terrestrial vertebrates, birds exhibit tremendous diversity in their phenotypes, ecology, habitats and behaviors[18]. So far, a considerable effort has been devoted to resolve the phylogenetic relationships from higher taxonomic categories[19–21] to sister species[22–26]. In addition to phylogenetics, modeling-based approaches using multiple nuclear genes have also shed light on population structure and demographic history and allowed inferences of selection pressures in non-model organisms[27–30]. The rapid advance in these sub-disciplines in evolutionary biology always hinges upon proper sampling design and a rigorous statistical approach, but it also requires data on multiple independent loci with an appropriate level of genetic polymorphism[31], which allows the application of sophisticated modeling and thus hypothesis testing.

Efforts of developing universal PCR primers have facilitated avian phylogenetic and population genetic studies[32–34]. For example, Dawson et al.[35] developed a set of microsatellite markers with high cross-species utility, suitable for paternity and population studies. Backström et al.[36] developed more than 200 exons flanking introns, which were evenly distributed throughout the avian genome. However, a variable number of indels (insertions and deletions) in the intron complicate the subsequent amplification, sequencing and alignment of these exons. Conserved and easily aligned exonic regions are ideal alternatives to compensate for resolving power for phylogenetic reconstruction[13]. Kimball et al.[37] tested the utility of 36 published markers on 42-199 bird species with only five exonic markers therein. Kerr et al.[38] developed 100 exonic markers from five avian genomes, and finally tested a subset of 25 markers in 12 avian orders. The quantity of NPCL markers is far from adequate as exon length should be longer than intron sequences to yield sufficient phylogenetic resolution[39]. Using a small

number of universal NPCL markers could increase the probability of error when estimating species relationships due to the conflict of gene tree topologies. To overcome the problem, it has been advocated to use more genes with longer sequences[40]. However, some obstacles have hindered the development of universal NPCL markers. Firstly, widespread flanking introns make the identification of the exon boundaries of a specific NPCL marker difficult[6]. Secondly, multiple nuclear loci are required to be distributed evenly and widely across the whole genome in order to indicate a variety of historical signals. And finally, low-cost and easy amplification are important requisites. The development of a universal set of NPCL markers for birds should significantly reduce the time required for future research as well as its cost, and facilitate the application of coalescent-based methods in avian evolutionary studies.

In this study, we aimed to develop a set of avian universal NPCL markers that can be widely utilized in avian phylogenetic and population genetic studies. By comparing the published genomes of the Red Junglefowl (*Gallus gallus*) and the Zebra Finch (*Taeniopygia guttata*), we designed 136 pairs of NPCL primers and amplified them in 41 species representing 23 avian orders to check their versatility. To test the resolving power of these markers, we further constructed a phylogenetic tree and estimated mutation rates by extracting universal NPCLs from 48 published avian genomes[41]. Moreover, samples from four populations of the Kentish Plover (*Charadrius alexandrinus*) were also amplified to estimate the intra-specific polymorphic level of these universal NPCLs.

Results

Pan-avian order amplifications of the novel NPCLs

The genome alignment and BLAST procedures resulted in 136 NPCL candidates, which were broadly distributed across 24 autosomal chromosomes and the Z chromosome of the Zebra Finch genome. Their original fragment length ranged from 815bp to 7176bp (Supplementary Table S1). We thus nominated each NPCL marker using abbreviation of the associated protein-coding regions according to gene annotation of Zebra Finch (Supplementary Table S1). More than one primer pairs were conducted for each NPCL marker candidate, and we finally chose the pair of PCR markers with the highest score denoting the level of conservatism between Zebra Finch and Red Junglefowl genomes.

In total, 5,146 PCRs were performed to amplify the 136 NPCLs in 41 species representing 23 avian orders (Fig. 1A). Among them, 2,875 (55.9%) of PCR performances produced a target band (Supplementary Table S3). For the 136 candidates, we successfully amplified 12 NPCLs in all 23 orders, with 100% PCR success rate (PSR). Sixty three of the 136 candidate NPCL markers had a relatively good overall PCR performance (PSR $\geq$ 80%) (Fig. 2A); all of them were successfully amplified in *Caprimulgiformes* and *Gruiformes*, and the PSR ranged from 65% to 97% in other orders (Fig. 1B, Supplementary Table S3). This set of 63 universal avian nuclear markers was distributed across 17 autosomal chromosomes and the Z chromosome (Fig. 3).

Interspecific mutation rate and phylogenic construction of the 63 universal NPCLs

The genome-based BLAST results showed that the widely used genetic markers, cytochrome *b* (*cyt b*) of mitochondrial DNA(mtDNA) and *RAG1*, an extensively used nuclear gene[42,43] were located in all 48 published genomes[41]. For the newly developed NPCLs, we located 56 loci across all 48 avian genomes. Among the remaining seven NPCLs, six of them were located in 47 genomes and two missing data recorded at the locus *FUT10*. Combined, BLAST results confirmed the set of 63 universal NPCL

132 markers were orthologous among these 48 species (Supplementary Table S4) and the resulting concatenated matrix with sequences of approximately 96 kb was obtained.

134 The range of the estimated mutation rates for the universal avian NPCLs is broad; it ranged from 0.0997 to 0.7317 $\times 10^{-8}$ per site per million years (Fig. 2B). Among these 63 NPCLs, the mutation rates 136 of 27 were slower than the mutation rate of *RAG1*, whilst the other 36 NPCLs were faster. All NPCLs showed a slower mutation rate than that of the mitochondrial *cyt b*.

138 We constructed a Maximum Likelihood (ML) tree based on 63 concatenated NPCLs from 48 species, representing 34 orders of extant birds (Fig. 4). The resulting topology is largely similar with 140 the recent phylogenomic studies[41,44,45]. Neoaves and Galloanseres, which united in the infraclass Neognathae, as well as Palaeognathae were three major groups with highest bootstrap support (100%). 142 Among Neoaves group, two major clades, core landbirds (Telluraves) and core waterbirds (Aequornithia) were strongly supported by whole-genome data[41] and 259 independent nuclear loci[44]. 144 Within core landbirds, the clade containing Passerimorphae (Passeriformes + parrots), Falconiformes (falcons), Cariamiformes (seriemas) is sister to Craciimorphae (bee-eaters + woodpeckers + hornbills + 146 trogons + cuckoo-roller + mousebirds), which is paraphyletic to Strigiformes (owls) and its sister clade Accipitrimorphae (eagles + New World vultures). Within core waterbirds, Pelecanlmorphae (pelicans + 148 herons + ibises+cormornts) and Procellariimorphae (fulmars + penguins) are two monophyletic groups, sister to Gaviimorphae (loons). Other clades such as Phoenicopterimorphae (flamingos + grebes), 150 Otidimorphae (bustards + turacos + cuckoos), Caprimulgimorphae (hummingbirds + swifts + nightjars) and Phaethontimorphae (tropicbirds + sunbitterns) are identical with previous studies[41,45,46]. However, 152 we also found discordances between this phylogenetic tree and previous results[41,44–46] , specifically in some branches with conflict placements with low support. For example, Columbiformes (doves),

154 Pterocliformes (sandgrouses) and Mesitornithiformes (mesites) are not clustered into Columbimorphae. The placement of Charadriiformes (plovers), Gruiformes (cranes) and Opisthocomiformes (hoatzins) 156 are incongruence with Jarvis et al[41], respectively.

158 *Intraspecific polymorphism of 17 randomly selected NPCL markers*

A total of 12,420bp DNA sequences, including 11,196bp of 17 NPCLs and 1,224bp of two 160 mitochondrial loci were sequenced in 40 samples representing four populations of the Kentish Plover. The NPCL markers showed varied degrees of polymorphism, with the exception of locus *KBTBD8* 162 (Fig. 5). There were 10 polymorphic sites in loci *BIRC2* and *FMN2*, while there were only 1-6 polymorphic sites in other loci. Correspondingly, *BIRC2* and *FMN2* possessed the highest values of 164 haplotype and nucleotide diversity (mean *Hd*= 0.84 and 0.92, mean π= 0.0047 and 0.0040, respectively). In contrast, a mitochondrial gene *ND3* had only one polymorphic site, yielding a low 166 haplotype (mean *Hd*= 0.36) and nucleotide diversity (mean π= 0.0009). When we combined the results on interspecific mutation rates and intraspecific polymorphism, we found that the inter- and 168 intraspecific genetic diversity of our gene markers were incongruent; although the estimated mutation rates at the study NPCLs were all much lower than that of mitochondrial gene *cyt b*, the intraspecific 170 polymorphism at nine NPCL markers was higher than that of two mitochondrial genes.

The genetic polymorphism parameters varied greatly, not only among genes, but among 172 populations as well. For example, the *Hd* value of the *MAML3* gene was lowest in the Taiwan population (*Hd*= 0.51) and highest in the Qinghai population (*Hd*= 0.88). The measure for nucleotide 174 diversity, π of the *NCOA6* gene was lowest in the Taiwan population (π= 1.59) and highest in the Guangxi population (π= 3.81). Detailed information on measures from each population is available in

Supplementary Table S5. The HKA test suggested no departure from the neutral expectation for any of the 17 NPCL markers. Similarly, the test of Tajima's *D* showed that none of the 17 NPCL markers deviated significantly from neutrality (Supplementary Table S5).

Discussion

We developed a set of 63 avian universal NPCL markers with diverse mutation rates and levels of intraspecific polymorphism. Our results showed that the 63 NPCL markers were successfully amplified in most of the species tested, representing 23 extant orders across major lineages of the avian tree of life (PCR success rate >80%), and denoted different levels of inter- and intraspecific polymorphism. Therefore, our NPCL set will provide a highly versatile genetic toolkit for a broad range of molecular phylogenetic and ecological applications. Moreover, the genetic marker system we provide here is cheap and easy to apply. Any molecular laboratories that are capable of performing PCRs can adopt our marker system effortlessly. Hence, this novel set of universal NPCL markers has great potential to be widely applied in evolutionary biology studies in birds.

Inherited from different chromosomes, concatenation of nuclear markers contribute multiple independent estimates to species trees[47,48], in order to alleviate the node conflicts of gene trees caused by incomplete lineage sorting, horizontal gene transfer, inconsistent evolutionary rates, gene duplication and/or gene loss and so on[49,50]. We constructed an avian phylogenetic tree using concatenation of 63 NPCLs across 18 chromosomes from 48 genomes. The result is largely similar with previous phylogenomic works using different data types like multiple nuclear loci[41,44], introns[20], ultraconserved elements (UCEs)[46,51] and retroposon presence/absence matrix[45]. The congruent parts of topology reveal multiple cluster clades, such as Telluraves (core landbirds), Aequornithes (core

198 waterbirds) and Phoenicopterimorphae, Otidimorphae, Caprimulgimorphae and Phaethontimorphae. We also find some unresolved placements comparing with Jarvis et al[41]. These include Columbiformes
200 (doves), Pterocliformes (sandgrouses), Mesitornithiformes (mesites), Charadriiformes (plovers) and Gruiformes (cranes), which exhibits hard polytomies in the avian tree of life. Though recent efforts in
202 avian phylogenomic studies using whole-genome[41] or genome-level data[44], irresolvable relationships have been found in some clades[41,44-46]. Suh et al. [45] investigated the causes of phylogenetic
204 irresolvabilities and concluded that such phylogenetic discordances were originated from prevalent ancestral polymorphism denoted by incomplete lineage sorting (ILS)[52] , which is probably associated
206 with an initial near-K-Pg super-radiation[41] in Neoaves. Unlike the two other main radiations that gave rise to the core waterbirds and core landbirds clades, the massive near-K-Pg super-radiation in
208 Neoaves, containing several unresolved lineages, leads extreme ILS and associated network-like phylogenetic relationships[45-46]. On one hand, again, the topology reconstructed by the present set of
210 universal NPCL markers captures these patterns, and suggests hard polytomies due to biological limitation of phylogenetic methods. On the other hand, it implies that our NPCL markers have
212 sufficient polymorphism to resolve phylogenetic relationships among lineages with less ILS in Neoaves.

214 We also found that this set of novel NPCL markers has the potential to be applied in population genetic studies, in which researchers usually prefer to use abundant markers with high mutation rates.
216 For example, microsatellites were developed for specific species or orders to detect differences in genotypes and further to quantify intraspecific genetic diversity[35,53,54]. But introns like microsatellites
218 have high levels of length homoplasy[55]. It is commonly assumed that NPCLs are conservative loci, highly suitable to address questions concerning high-level systematics[40]. However, some population

genetic studies highlight the importance of using functional exonic SNPs in population genetic studies[11,56], comparing to neutral markers (such as microsatellite and mitochondrial DNA). Datasets that contain numbers of several to more than 100 exon genes[51,57] can support accurate and reliable estimates of population genetic parameters[55,58], and have a substantial power in population genetic analysis[5,40,50,59]. Studying 17 NPCL markers in the Kentish Plover, we found that sixteen had low to moderate levels of intraspecific polymorphism and nine of them showed higher genetic diversity than mitochondrial genes in this study. Although a previous study showed a low level of genetic differentiation across Eurasian populations[60], this species exhibits variability in morphology and behavior among and within populations in East Asia[61], warranting further coalescent-based analysis of their evolutionary history. This dataset provides sufficient information to study population genetics in the Kentish Plover in East Asia.

In order to infer correct phylogenetic relationships in different taxonomic levels, it is essential to choose unlinked genes with different mutation rates[62]. Our novel set of NPCL markers offer a wide range of mutation rates. The comparisons of mutation rates between the new NPCL markers with commonly used nuclear loci *RAG1*[7] and some loci at mtDNA[63] provide a reference for marker choice (Figure 2B, 5). In principal, it is advisable to use markers with slow mutation rates to resolve deep nodes and fast mutation rates to population genetic studies. Moreover, coalescent theory is widely used to estimate species tree and population demographic parameters, such as divergence times[64,65] and effective population sizes (N_e) [59]. The associated analyses, such as species-tree estimation, e.g. MP-EST[66], *BEAST[67], BP&P[68], and demographic analysis, such as Isolation with Migration (IM) model[69,70] and Approximate Bayesian Computation (ABC) simulations[59] require multiple independent loci with different demographic histories and mutation rates. In this regard, markers from different

242 genomic segments, such as introns (developed previously[35–37]) and exons we developed are preferred to combine and to be used. It is no doubt that the present marker set is a useful resource to generate

244 multilocus datasets for avian evolutionary studies in different taxonomic levels. In fact some studies have used these novel NPCL markers to apply the aforementioned analyses[25,30].

246 Compared with traditional Sanger sequencing, the fast development of next generation sequencing (NGS) techniques has enabled researchers to obtain genetic polymorphisms easily[71]. For

248 example, Jarvis et al.[41] performed a highly resolved phylogenetic tree of 48 species using phylogenomic methods, and Prum et al.[44] conducted a comprehensive phylogeny of 198 species within

250 the Neoaves, which diversified very quickly, using genome-scale data by targeted NGS. Multilocus methods do not use as much as genomic data. However, we consider that this set of universal NPCL

252 markers has its niche in avian molecular studies. Sanger sequencing technique of NPCL markers is less sensitive to the quality of template DNA like sequence capture approach than other genomic

254 approaches. Degraded DNA or a small quantity of DNA is also workable, like feather and museum samples. It is always a tradeoff between template DNA quality and PCR product length. With the novel

256 NPCL markers, we aimed to amplify a fragment of 700-1200 bp sequence of each locus. Hence they should be applicable to avian blood, tissue and feathers. Moreover, a thorough analysis pipeline for

258 traditional PCR-based method is available, supported by a series of visualized operating software, e.g. MEGA, DNASTAR, DnaSP, BEAST and etc., which are widely used in molecular phylogenetic

260 analysis. Processing genomic data always places high demands on bioinformatics and computational power[72]. High-quality samples for NGS, project budget, bioinformatic facilities are not available to all

262 laboratory. It is still useful and necessary to align orthologous sequences across multiple hierarchical levels using NPCL markers, especially for a pilot or small scale study.

However, there are some limitations when using this set of universal NPCL markers. Firstly, PCR performances were simultaneously tested under a unified protocol (e.g. Tm=50℃), so that the PSR of each NPCL marker might be underestimated. Reducing the annealing temperature by 1~2℃ would improve the success rate in practice. There is also the possibility that PCR produced target sequences but also non-specific amplicons. We could slightly raise the annealing temperature to increase specificity or perform extra steps including gel purification and cloning. Furthermore, the interspecific polymorphic parameters of the 17 NPCL markers are reference values for Kentish plovers. Different evolutionary forces, such as genetic drift or natural selection, can act on different regions of the genome, causing a various evolutionary rates and demographic histories in different species[73]. Thus, different combinations of markers are important for specific questions. For example, NPCL markers on the Z chromosome could be selected to solve questions involving sexual selection and mate choice. There is also a trade-off between the number of markers and time- and cost-efficiency. In avian phylogenetic analysis, random errors can be reduced by employing more markers, whilst, as a consequence of this procedure, systemic errors would increase due to differences in nucleotide composition and various mutation rates[74]. Kimball et al. proposed that adopting various analytical methods might overcome these adverse effects[75].

In conclusion, we have developed 63 avian universal NPCL markers, evenly distributed across 17 autosome chromosomes and the Z chromosome. This set of universal NPCL markers had high PCR success rates (PSR>80%) in 23 avian orders. Its wide range of mutation rates are suitable to resolve phylogenetic relationships at both low and high-level. Furthermore, various intraspecific polymorphisms are potentially useful to provide deep-level divergence and demographic information for population genetics. Though high-throughput genetic polymorphism data from next generation sequencing

286 undoubtedly provide a more comprehensive vision for avian evolutionary history and genomic patterns, we believe that this set of exonic markers provides a relatively reliable and repeatable solution and could
288 have widespread application in phylogenetic and population genetics studies.

290 **Methods**

Development of NPCL markers and primer design

292 To screen NPCL marker candidates, we aligned parts of the genome of two species with a distant phylogenetic relationship[76], the Red Junglefowl (GCA_000002315) and the Zebra Finch
294 (GCA_000151805). Firstly we identified long (>600bp) single-copy exons within the genome of the Zebra Finch and took these exons as templates. Then we aligned them with the genome of the Red
296 Junglefowl using BLAST (Basic Local Alignment Search Tool). We assumed that query sequences of the Red Junglefowl with identity more than 80% and length more than 50% of templates length were
298 orthologous exons and employed them as NPCL marker candidates.

We used the program Primer3[77] to design the primers for NPCL marker candidates. We selected
300 exon sequences of Zebra Finch as templates, focusing on High-scoring Segment Pairs region (700-1200bp). For each primer pair, the oligomer ranged from 18bp to 25bp and GC content ranged from
302 20% to 80%. Furthermore, we tested a single primer for self-complementarity by setting complementarity score to less than 6.00, so as to predict the tendency of primers to anneal to each other
304 without necessarily causing self-priming in the PCR. Complementarity 3' score was set as default (< 3.00) to test the complementarity between left and right primers.

306

Tests on the universality of the NPCL markers

To test the amplification performance of these new NPCL markers, we selected 41 species of 23 representative Aves orders (Supplementary Table S2). We used a set of 10000 trees with 9993 operational taxonomic unites (OTUs) downloaded from http://birdtree.org/ to demonstrate the phylogenetic relationships among selected species[18].

Total genomic DNA was extracted from ethanol-preserved muscle tissue or blood using a TIANamp Genomic DNA kit (TIANGEN, China) and stored at 4℃. DNA concentration and purity were estimated by NanoDrop 2000 (Thermo Scientific, USA). We used Touchdown PCR (TD-PCR)[78], an improved standard PCR to test the utility of primers sensitively, by decreasing the annealing temperature 1℃/cycle from Tm+10℃ to Tm (Melting Temperature). The Touchdown PCR was performed in a Veriti96 PCR thermal cycler system (ABI, USA) using a 10μl reaction containing 2μl template DNA (10-70ng totally), with mixed concentrations of 10×PCR buffer, 20μM dNTP, 10mM of each forward and reverse primer, and 5U Taq polymerase (Takara, China). The initial temperature profile was 2 min at 94℃, 10 cycles at 94℃ for 30s, 60-50℃ (decreasing the annealing temperature by 1℃ per cycle) for 30s and 72℃ for 90s followed by 30 similar cycles but with a constant temperature of 50℃. This process was concluded with an extra elongation step at 72℃ for 10 min. A successful amplification was recorded if a single clear band (target locus) was observable under ultraviolet light after being isolated on a 1% TAE agarose gel at 120V for 30 min.

Estimation of inter-species mutation rates and Construction of Neoavian phylogeny

We downloaded 48 avian genomes covering all orders of Neoaves[41]. Sixty-five gene sequences, including the 63 universal NPCLs and two frequently-used genes (*RAG1*, and mitochondrial cytochrome *b* (*cyt b*)) as control, were retrieved and aligned in these genomes against genes in Zebra

330 Finch (abbreviation as *Tgu1*) using BLAST. These sequences were extracted and filtered in batches by own-developed Perl script in the Tianhe-2 server (School of Advanced Computing, Sun Yat-sen

332 University). Because of different genome sequence format, the script was unable to retrieve sequences in four species, i.e. *Anas platyrhynchos*, *Gallus gallus*, *Meleagris gallopavo* and *Melopsittacus*

334 *undulates*[41]. Hence, we manually searched and obtained orthologues of the four species using BLAST tool on the website https://blast.ncbi.nlm.nih.gov/Blast.cgi. We further aligned the obtained NPCL

336 orthologous sequences using MEGA v6.0[79].

To estimate the mutation rate of each gene, we firstly computed the overall mean genetic distance

338 at each NPCL marker in MEGA v6.0 with 1000 bootstrap replicates. Then we calculated the ratio of genetic distance between each NPCL and *cyt b*. Finally, we multiplied the ratio by the average mutation

340 rate in the *cyt b* (0.01035 mutations per site per million years)[80] to get the average mutation rate of each NPCL[81].

342 To construct the phylogenetic relationship of Neoavian birds in order level, we concatenated all NPCL sequences and reconstructed the maximum likelihood tree by RAxML v8.2.1[82], with GTRCAT

344 model and 1,000 bootstrap runs. Maximum-likelihood-bootstrap proportions ≥ 70% were considered strong support[83].

346

Intra-species polymorphism measurements

348 We amplified 17 random NPCLs in 40 Kentish Plover (*Charadrius alexandrinus*) blood samples from live-trapped birds in a noninvasive manner. To compare our data with previous genetic analyses on

350 European populations of the Kentish Plover[84] and compare the degree of polymorphism between nuclear and mtDNA, we added two mtDNA loci, ATPase subunit six concatenated with partial ATPase

352 subunit 8 (*ATPase6/8*) and NADH dehydrogenase subunit 3 (*ND3*). Blood samples were collected from four breeding populations of plovers, Guangxi (GX), Qinghai (QH), Hebei (HB) and Taiwan (TW)
354 (Table S2). The same protocol for DNA extraction and PCR amplification was followed as above, and the products were sequenced on ABI3730XL (Applied Biosystems, USA) by Beijing Genomics
356 Institute (BGI, China).

Both strands of the amplicons were assembled, and the heterozygosity of nuclear genes was
358 detected using SeqMan v7.1.0.44[85]. Some parameters of the DNA polymorphism, the number of polymorphic sites (*S*) and haplotypes (*H*), haplotype diversity (*Hd*), and nucleotide diversity (π) were
360 calculated using DnaSP v5.0[86] . The neutrality of each locus was tested using the Hudson-Kreitman-Aguade (HKA) test[87] and Tajima's *D*[88] implemented in DnaSP v5.0.

362

364 **Acknowledgements**

We are grateful to Per Alström, Fuming Lei, Fasheng Zou, Xiaojun Yang, Ulf Johansson, Chung-Yu
366 Chiang, Jonathan Reeves, Yingyong Wang, Menxiu Tong, Qin Huang, Zhechun Zhang, Xuejing Wang, Xin Lin, Jian Zhao for supplying tissue, blood or DNA samples used in this study, and Zhenhao Luo for
368 providing technical support to BLAST script methods, and Alan Watson for editing the text. This study was supported by the National Science Foundation of China (No. 31301875 & No. 31572251 to Yang
370 Liu; No. 31471987 to Lu Dong, No. 31600297 to Pinjia Que), and the National Key Program of Research and Development, Ministry of Science and Technology Grant 2016YFC0503200 to Lu Dong, and some
372 DNA samples of birds were collected during 'The Comprehensive Scientific Survey of Biodiversity from Luoxiao Range Region in China (2013FY111500)'. Computational work was funded by Special Program

for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) under Grant No. U1501501 to Yang Liu.

Author contributions statement

Y.L., SH.L. and D.L. designed this study. CF. Y. and N.Z. carried out the primer design and BLAST procedures. GL.C. and PJ.Q. provided materials and technical support in the lab. SM.L. completed wet lab experiments, analyzed the data and wrote the manuscript with Y.L.

Competing interests

The author declare no competing interests.

386 **References**

1. Ansorge, W. J. Next-generation DNA sequencing techniques. *New Biotechnol.* **25,** 195–203
388 (2009).
2. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving
390 incongruence in molecular phylogenies. *Nature* **425,** 798 (2003).
3. Kingman, J. F. C. On the genealogy of large populations. *J. Appl. Probab.* **19,** 27–43 (1982).
392 4. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105,**
437–460 (1983).
394 5. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of
life. *Nat. Rev. Genet.* **6,** 361–375 (2005).
396 6. Townsend, T. M., Alegre, R. E., Kelley, S. T., Wiens, J. J. & Reeder, T. W. Rapid development
of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from
398 squamate reptiles. *Mol. Phylogenet. Evol.* **47,** 129–142 (2008).
7. Groth, J. G. & Barrowclough, G. F. Basal divergences in birds and the phylogenetic utility of
400 the nuclear RAG-1 gene. *Mol. Phylogenet. Evol.* **12,** 115–123 (1999).
8. Ericson, P. G. P., Johansson, U. S. & Parsons, T. J. Major divisions in oscines revealed by
402 insertions in the nuclear gene c-myc: a novel gene in avian phylogenetics. *The Auk* **117,** 1069–
1178 (2000).
404 9. Johansson, U. S., Irestedt, M., Parsons, T. J. & Ericson, P. G. P. Basal phylogeny of the
tyrannoidea based on comparisons of cytochrome b and exons of nuclear c-myc and RAG-1
406 genes. *The Auk* **119,** 984 (2002).
10. Boekhorst, J. & Snel, B. Identification of homologs in insignificant blast hits by exploiting
408 extrinsic gene properties. *BMC Bioinformatics* **8,** 356 (2007).
11. Zhan, X. *et al.* Exonic versus intronic SNPs: contrasting roles in revealing the population
410 genetic differentiation of a widespread bird species. *Heredity* **114,** 1–9 (2015).
12. Remm, Maido, Christian, E. S. & Erik, L. S. Automatic clustering of orthologs and in-paralogs
412 from pairwise species comparisons. *J. Mol. Biol.* **314,** 1041–1052 (2001).
13. Thomson, R. C., Wang, I. J. & Johnson, J. R. Genome-enabled development of DNA markers
414 for ecology, evolution and conservation. *Mol. Ecol.* **19,** 2184–2195 (2010).
14. Che, L.-H. *et al.* Genome-wide survey of nuclear protein-coding markers for beetle

phylogenetics and their application in resolving both deep and shallow-level divergences. *Mol. Ecol. Resour.* (2017).

15. Li, C., Ortí, G., Zhang, G. & Lu, G. A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol. Biol.* **7,** 44 (2007).
16. Shen, X. X., Liang, D., Feng, Y. J., Chen, M. Y. & Zhang, P. A versatile and highly efficient toolkit including 102 nuclear markers for vertebrate phylogenomics, tested by resolving the higher level relationships of the Caudata. *Mol. Biol. Evol.* **30,** 2235–2248 (2013).
17. Fong, J. J. & Fujita, M. K. Evaluating phylogenetic informativeness and data-type usage for new protein-coding genes across Vertebrata. *Mol. Phylogenet. Evol.* **61,** 300–307 (2011).
18. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491,** 444–448 (2012).
19. Berv, J. S. & Prum, R. O. A comprehensive multilocus phylogeny of the Neotropical cotingas (Cotingidae, Aves) with a comparative evolutionary analysis of breeding system and plumage dimorphism and a revised phylogenetic classification. *Mol. Phylogenet. Evol.* **81,** 120–136 (2014).
20. Hackett, S. J. *et al.* A phylogenomic study of birds reveals their evolutionary history. *Science* **320,** 1763–1768 (2008).
21. Jønsson, K. A. *et al.* A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). *Mol. Phylogenet. Evol.* **94,** 87–94 (2016).
22. Backström, N., Sætre, G.-P. & Ellegren, H. Inferring the demographic history of european ficedula, flycatcher populations. *BMC Evol. Biol.* **13,** 2 (2013).
23. Chu, J.-H. *et al.* Inferring the geographic mode of speciation by contrasting autosomal and sex-linked genetic diversity. *Mol. Biol. Evol.* **30,** 2519–2530 (2013).
24. Dong, F. *et al.* Molecular systematics and plumage coloration evolution of an enigmatic babbler (Pomatorhinus ruficollis) in East Asia. *Mol. Phylogenet. Evol.* **70,** 76–83 (2014).
25. Wang, N. *et al.* Incipient speciation with gene flow on a continental island: Species delimitation of the Hainan Hwamei (Leucodioptron canorum owstoni, Passeriformes, Aves). *Mol. Phylogenet. Evol.* **102,** 62–73 (2016).
26. Yeung, C. K. L. *et al.* Beyond a morphological paradox: Complicated phylogenetic relationships of the parrotbills (Paradoxornithidae, Aves). *Mol. Phylogenet. Evol.* **61,** 192–202

(2011).

27. Hung, C.-M., Drovetski, S. V. & Zink, R. M. Matching loci surveyed to questions asked in phylogeography. *Proc. R. Soc. B Biol. Sci.* **283,** 20152340 (2016).
28. Lim, H. C. & Sheldon, F. H. Multilocus analysis of the evolutionary dynamics of rainforest bird populations in Southeast Asia: population history of Sundaland birds. *Mol. Ecol.* **20,** 3414–3438 (2011).
29. Shaner, P.-J. L. *et al.* Climate niche differentiation between two passerines despite ongoing gene flow. *J. Anim. Ecol.* **84,** 829–839 (2015).
30. Wang, P. *et al.* The role of niche divergence and geographic arrangement in the speciation of Eared Pheasants (Crossoptilon, Hodgson 1938). *Mol. Phylogenet. Evol.* **113,** 1–8 (2017).
31. Burleigh, J. G., Kimball, R. T. & Braun, E. L. Building the avian tree of life using a large-scale, sparse supermatrix. *Mol. Phylogenet. Evol.* **84,** 53–63 (2015).
32. Irestedt, M., Fjeldså, J., Johansson, U. S. & Ericson, P. G. . Systematic relationships and biogeography of the tracheophone suboscines (Aves: Passeriformes). *Mol. Phylogenet. Evol.* **23,** 499–512 (2002).
33. Helbig, A. J., Kocum, A., Seibold, I. & Braun, M. J. A multi-gene phylogeny of aquiline eagles (Aves: Accipitriformes) reveals extensive paraphyly at the genus level. *Mol. Phylogenet. Evol.* **35,** 147–164 (2005).
34. Ericson, P. G. P. *et al.* Higher-level phylogeny and morphological evolution of tyrant flycatchers, cotingas, manakins, and their allies (Aves: Tyrannida). *Mol. Phylogenet. Evol.* **40,** 471–483 (2006).
35. Dawson, D. A. *et al.* New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Mol. Ecol. Resour.* **10,** 475–494 (2010).
36. Backström, N., Fagerberg, S. & Ellegren, H. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol. Ecol.* **17,** 964–980 (2007).
37. Kimball, R. T. *et al.* A well-tested set of primers to amplify regions spread across the avian genome. *Mol. Phylogenet. Evol.* **50,** 654–660 (2009).
38. Kerr, K. C. R., Cloutier, A. & Baker, A. J. One hundred new universal exonic markers for birds

476 developed from a genomic pipeline. *J. Ornithol.* **155,** 561–569 (2014).

39. Chojnowski, J. L., Kimball, R. T. & Braun, E. L. Introns outperform exons in analyses of basal
478 avian phylogeny using clathrin heavy chain genes. *Gene* **410,** 89–96 (2008).

40. Brito, P. H. & Edwards, S. V. Multilocus phylogeography and phylogenetics using sequence-
480 based markers. *Genetica* **135,** 439–455 (2009).

41. Jarvis, E. D. *et al.* Whole-genome analyses resolve early branches in the tree of life of modern
482 birds. *science* **346,** 1320–1331 (2014).

42. Cibois, A. & Cracraft, J. Assessing the passerine "Tapestry": phylogenetic relationships of the
484 Muscicapoidea inferred from nuclear DNA sequences. *Mol. Phylogenet. Evol.* **32,** 264–273 (2004).

486 43. Paton, T. A., Baker, A. J., Groth, J. G. & Barrowclough, G. F. RAG-1 sequences resolve phylogenetic relationships within Charadriiform birds. *Mol. Phylogenet. Evol.* **29,** 268–278
488 (2003).

44. Prum, R. O. *et al.* A comprehensive phylogeny of birds (Aves) using targeted next-generation
490 DNA sequencing. *Nature* **526,** 569–573 (2015).

45. Suh, A., Smeds, L. & Ellegren, H. The Dynamics of Incomplete Lineage Sorting across the
492 Ancient Adaptive Radiation of Neoavian Birds. *PLOS Biol.* **13,** e1002224 (2015).

46. Suh, A. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves.
494 *Zool. Scr.* **45,** 50–62 (2016).

47. Salas-Leiva, D. E. *et al.* Conserved genetic regions across angiosperms as tools to develop
496 single-copy nuclear markers in gymnosperms: an example using cycads. *Mol. Ecol. Resour.* **14,** 831–845 (2014).

498 48. Waters, J. M., Rowe, D. L., Burridge, C. P. & Wallis, G. P. Gene Trees versus Species Trees: Reassessing life-History evolution in a freshwater fish radiation. *Syst. Biol.* **59,** 504–517 (2010).

500 49. Edwards, S. V. Is a new and general theory of molecular systematics emerging? *Evolution* **63,** 1–19 (2009).

502 50. Szöllősi, G. J., Tannier, E., Daubin, V. & Boussau, B. The Inference of Gene Trees with Species Trees. *Syst. Biol.* **64,** e42–e62 (2015).

504 51. McCormack, J. E. *et al.* A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE* **8,** e54848 (2013).

506 52. Maddison, W. P. & Knowles, L. L. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Syst. Biol.* **55,** 21–30 (2006).

508 53. Galbusera, P., Dongen, S. van & Matthysen, E. Cross-species amplification of microsatellite primers in passerine birds. *Conserv. Genet.* 163–168 (2000).

510 54. Wang, B. *et al.* Development and characterization of novel microsatellite markers for the Common Pheasant (Phasianus colchicus) using RAD-seq. *Avian Res.* **8,** (2017).

512 55. Sunnucks, P. Efficient genetic markers for population biology. *Trends Ecol. Evol.* **15,** 199–203 (2000).

514 56. Freamo, H., O'Reilly, P., Berg, P. R., Lien, S. & Boulding, E. G. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral

516 SNPs. *Mol. Ecol. Resour.* **11,** 254–267 (2015).

57. Bapteste, E. *et al.* The analysis of 100 genes supports the grouping of three highly divergent

518 amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc. Natl. Acad. Sci.* **99,** 1414–1419 (2002).

520 58. Edwards, S. & Bensch, S. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Mol. Ecol.* **18,** 2930–2933 (2009).

522 59. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162,** 2025–2035 (2002).

524 60. Küpper, C. *et al.* High gene flow on a continental scale in the polyandrous Kentish plover *Charadrius alexandrinus*. *Mol. Ecol.* **21,** 5864–5879 (2012).

526 61. Rheindt, F. E. *et al.* Conflict between Genetic and Phenotypic Differentiation: The Evolutionary History of a 'Lost and Rediscovered' Shorebird. *PLoS ONE* **6,** e26995 (2011).

528 62. Nosenko, T. *et al.* Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogenet. Evol.* **67,** 223–233 (2013).

530 63. Zink, R. M. & Barrowclough, G. F. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* **17,** 2107–2121 (2008).

532 64. Smith, B. T. & Klicka, J. Examining the role of Effective population size on mitochondrial and multilocus divergence time discordance in a songbird. *PLoS ONE* **8,** e55161 (2013).

534 65. Thorne, J. L. & Kishino, H. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* **51,** 689–702 (2002).

66. Liu, L., Yu, L. & Edwards, S. V. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10,** 302 (2010).

67. Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29,** 1969–1973 (2012).

68. Yang, Z. & Rannala, B. Bayesian species delimitation using multilocus sequence data. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 9264–9269 (2010).

69. Hey, J. Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of Drosophila pseudoobscura and D. persimilis. *Genetics* **167,** 747–760 (2004).

70. Hey, J. & Nielsen, R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl. Acad. Sci.* **104,** 2785–2790 (2007).

71. McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C. & Brumfield, R. T. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol. Phylogenet. Evol.* **66,** 526–538 (2013).

72. Roure, B., Baurain, D. & Philippe, H. Impact of Missing Data on Phylogenies Inferred from Empirical Phylogenomic Data Sets. *Mol. Biol. Evol.* **30,** 197–214 (2013).

73. Lande, R. Natural selection and random genetic drift in phenotypic evolution. *Evolution* **30,** 314–334 (1976).

74. Felsenstein, J. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Biol.* **27,** 401–410 (1978).

75. Kimball, R. T., Wang, N., Heimer-McGinn, V., Ferguson, C. & Braun, E. L. Identifying localized biases in large datasets: A case study using the avian tree of life. *Mol. Phylogenet. Evol.* **69,** 1021–1032 (2013).

76. Nam, K. *et al.* Molecular evolution of genes in avian genomes. *Genome Biol.* **11,** R68 (2010).

77. Rozen, S. & Skaletsky, H. Primer3 on the WWW for general users and for biologist programmers. in *Bioinfrormatics methods and protocols* **132,** 365–386 (Humana Press, 2000).

78. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19,** 4008–4008 (1991).

79. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular

566 Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30,** 2725–2729 (2013).

80. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17,** 2321–2328
568 (2008).

81. Li, J. W. *et al.* Rejecting strictly allopatric speciation on a continental island: prolonged
570 postdivergence gene flow between Taiwan (*Leucodioptron taewanus* , Passeriformes Timaliidae) and Chinese (*L. canorum canorum*) hwameis. *Mol. Ecol.* **19,** 494–507 (2010).

572 82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 (2014).

574 83. Hillis, D. M. & Bull, J. J. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42,** 182–192 (1993).

576 84. Küpper, C. *et al.* Kentish versus Snowy Plover: Phenotypic and Genetic Analyses of *Charadrius alexandrinus* Reveal Divergence of Eurasian and American Subspecies. *The Auk*
578 **126,** 839–852 (2009).

85. Swindell, S. R. & Plasterer, T. N. SEQMAN. *Seq. Data Anal. Guideb.* 75–89 (1997).

580 86. Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25,** 1451–1452 (2009).

582 87. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116,** 153–159 (1987).

584 88. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–595 (1989).

586

588 **Figure caption**

590 **Figure 1. PCR performance for the 136 NPCL marker candidates in 23 avian orders. (A)** Genetic relationships among our experimental samples. 41 species are highlighted in different colors
592 representing 23 avian orders widely distributed in the avian phylogenetic tree. **(B)** PCR performance for 136 NPCL marker candidates. Each square represents a PCR result. Success is shown in black and
594 failure in white. 430 of 5146 reactions that could not be produced due to a paucity of DNA are shown in grey. The gene name and PCR success rate of each NPCL marker are indicated to the left. The
596 success rate of each avian order is indicated at the bottom of the matrix of 63 universal NPCL markers.

598 **Figure 2. PCR success rate distribution for the 136 NPCL marker candidates and mutation rates for the 63 universal NPCL markers. (A)** PCR success rate distribution for 136 candidates in 23 avian
600 orders. The 63 NPCL markers with PCR success rates higher than 80% are shown in black; other loci with PSR success rates below 80% (in grey) were excluded from the subsequent analysis. The number
602 above each bar shows the number of NPCL marker candidates. **(B)** Mutation and success rates for the 63 universal NPCL markers. The markers were sorted according to estimated mutation rates (bars in
604 black) from low to high. The number on the right of each bar is the mutation rate of each NPCL marker. PCR success rates are shown underneath (bars in grey). The mutation rates of widely-used
606 NPCL *RAG1* (in green) and mitochondrial gene *cyt b* (in blue) were selected as references.

608 **Figure 3. Chromosome mapping of the 63 avian universal NPCL markers in the genome of Zebra Finch (*Taeniopygia guttata*).** The 63 universal NPCL markers with more than 80% PCR success rate
610 were widely distributed in 17 autosomal chromosomes and the Z chromosome.

612 **Figure 4. Phylogenetic analysis of Neoaves using 63 NPCLs from 48 bird species.** The names of species are abbreviated (refer to supplementary Table S4), representing 30 orders of Neoavian birds.
614 Superorders are labelled on the nodes as classification in previous study[41] using genomic data. Bootstrap support over 70% are indicated above nodes.

616

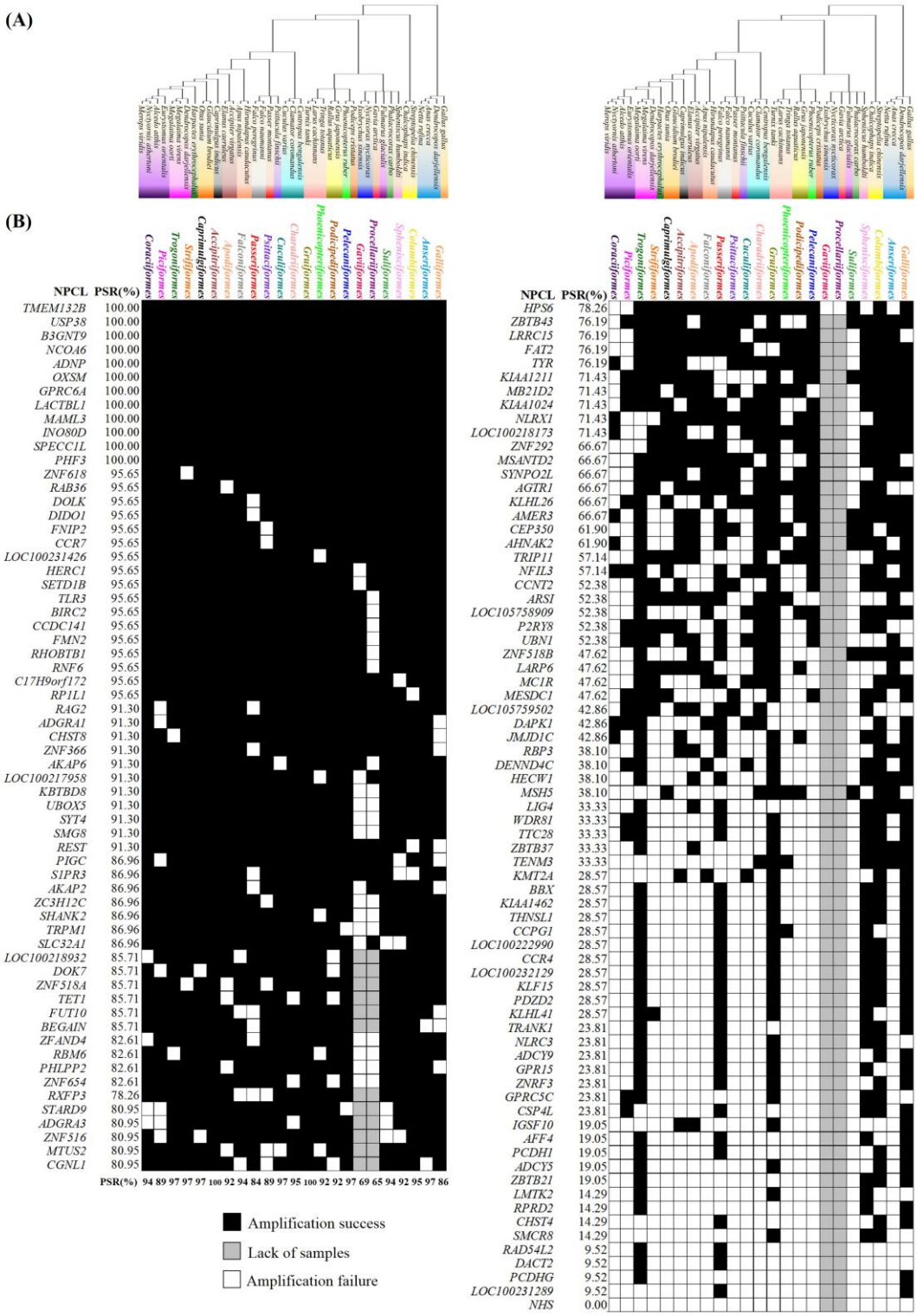
Figure 5. Polymorphism at 17 avian universal NPCL markers and two mitochondrial loci in four

618 **populations of Kentish Plover (*Charadrius alexandrinus*).** Markers were sorted according to their level of polymorphisms, from low to high. Gene names can be found at the bottom of the three box

620 plots. Mitochondrial genes (in blue) were selected as references **(A)** Average number of nucleotide sites ranging from 0 to 10. **(B)** Haplotype diversity ranging from 0 to 0.97. **(C)** Nucleotide diversity

622 ranging from 0 to 5.59.

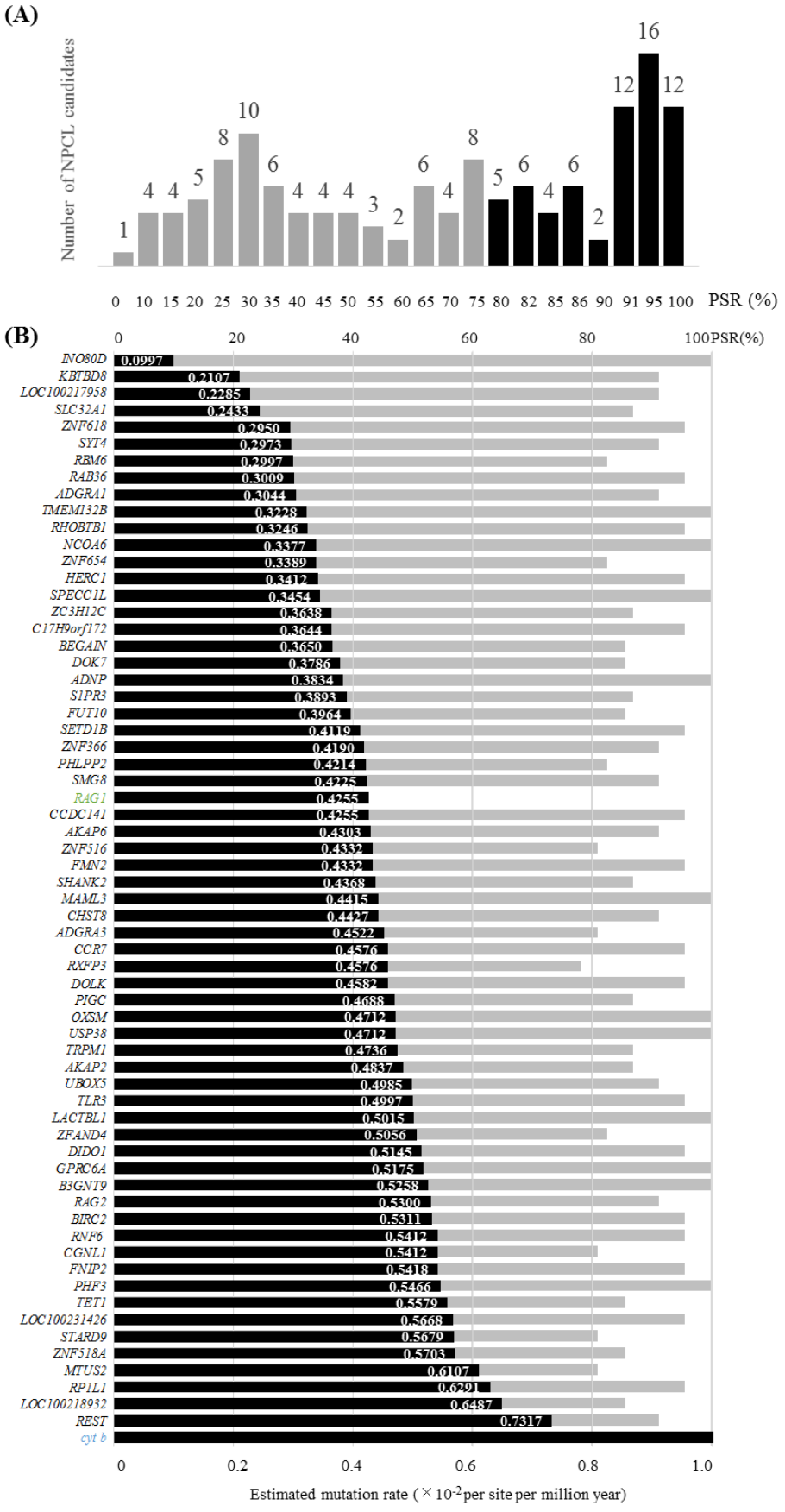
624 **Figure 1**



626

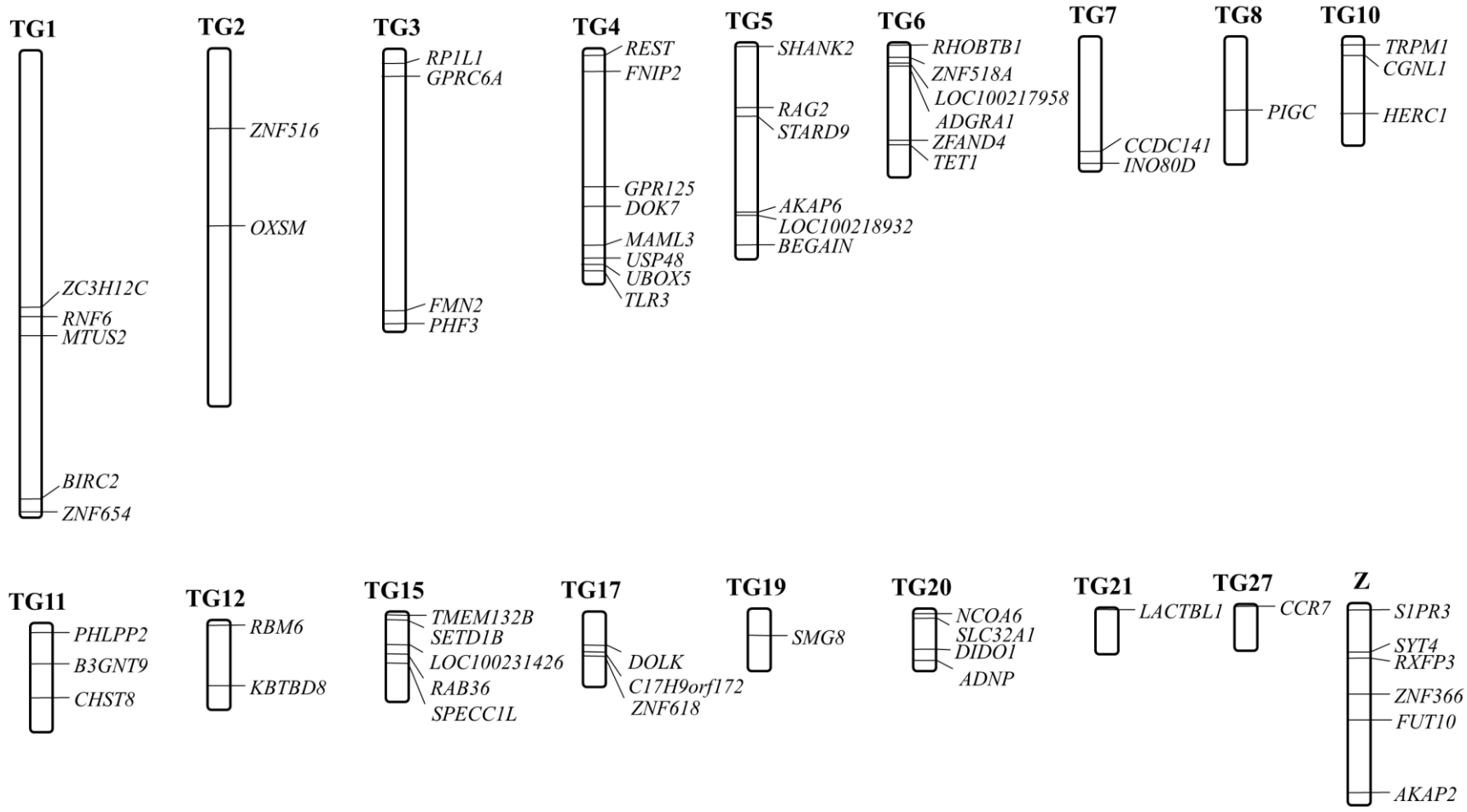
628

Figure 2



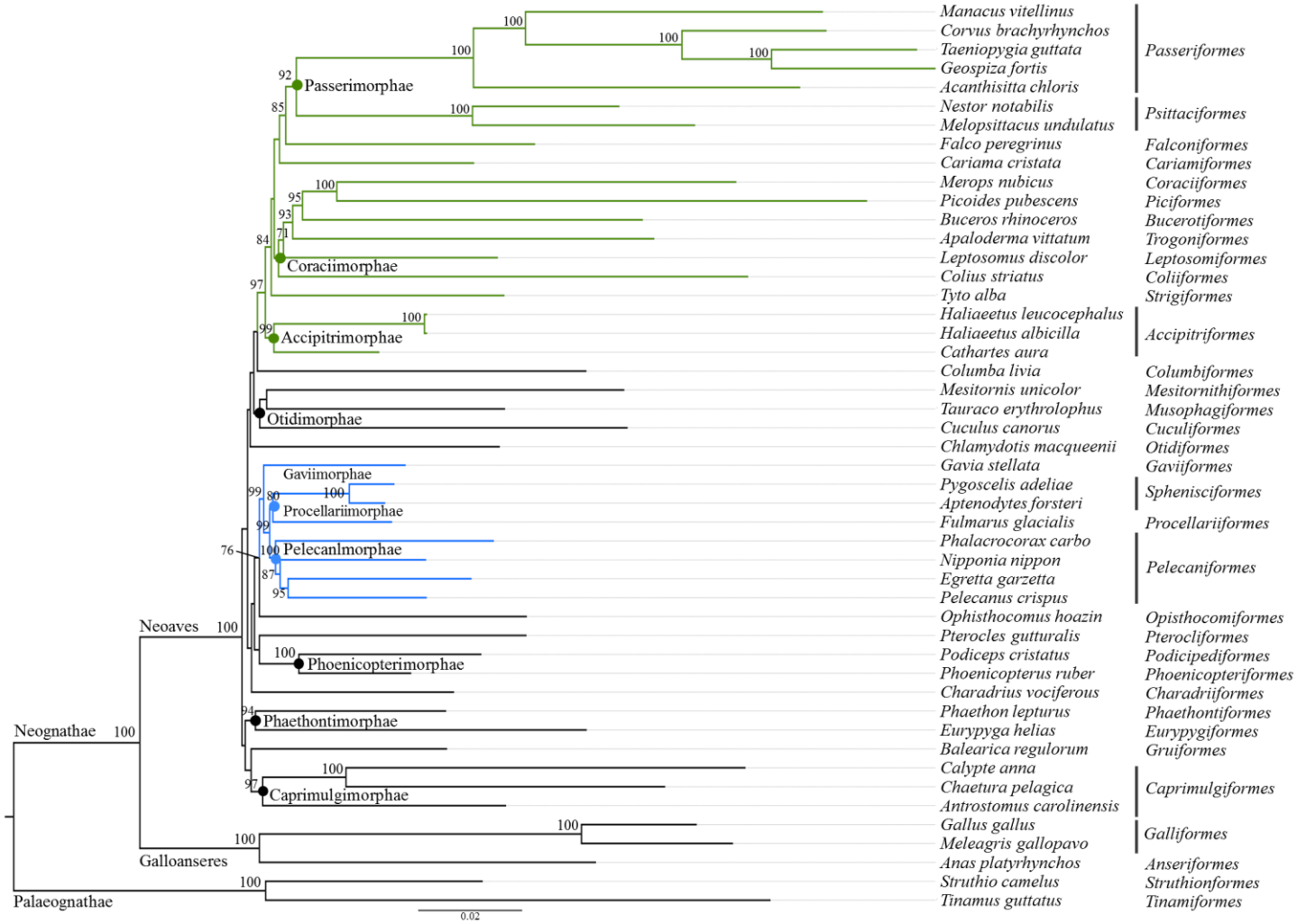
630

Figure 3


TG1
ZC3H12C
RNF6
MTUS2
BIRC2
ZNF654
TG2
ZNF516
OXSM
TG3
RP1L1
GPRC6A
FMN2
PHF3
TG4
REST
FNIP2
GPR125
DOK7
MAML3
USP48
UBOX5
TLR3
TG5
SHANK2
RAG2
STARD9
AKAP6
LOC100218932
BEGAIN
TG6
RHOBTB1
ZNF518A
LOC100217958
ADGRA1
ZFAND4
TET1
TG7
CCDC141
INO80D
TG8
PIGC
TG10
TRPM1
CGNL1
HERC1
TG11
PHLPP2
B3GNT9
CHST8
TG12
RBM6
KBTBD8
TG15
TMEM132B
SETD1B
LOC100231426
RAB36
SPECC1L
TG17
DOLK
C17H9orf172
ZNF618
TG19
SMG8
TG20
NCOA6
SLC32A1
DIDO1
ADNP
TG21
LACTBL1
TG27
CCR7
Z
S1PR3
SYT4
RXFP3
ZNF366
FUT10
AKAP2

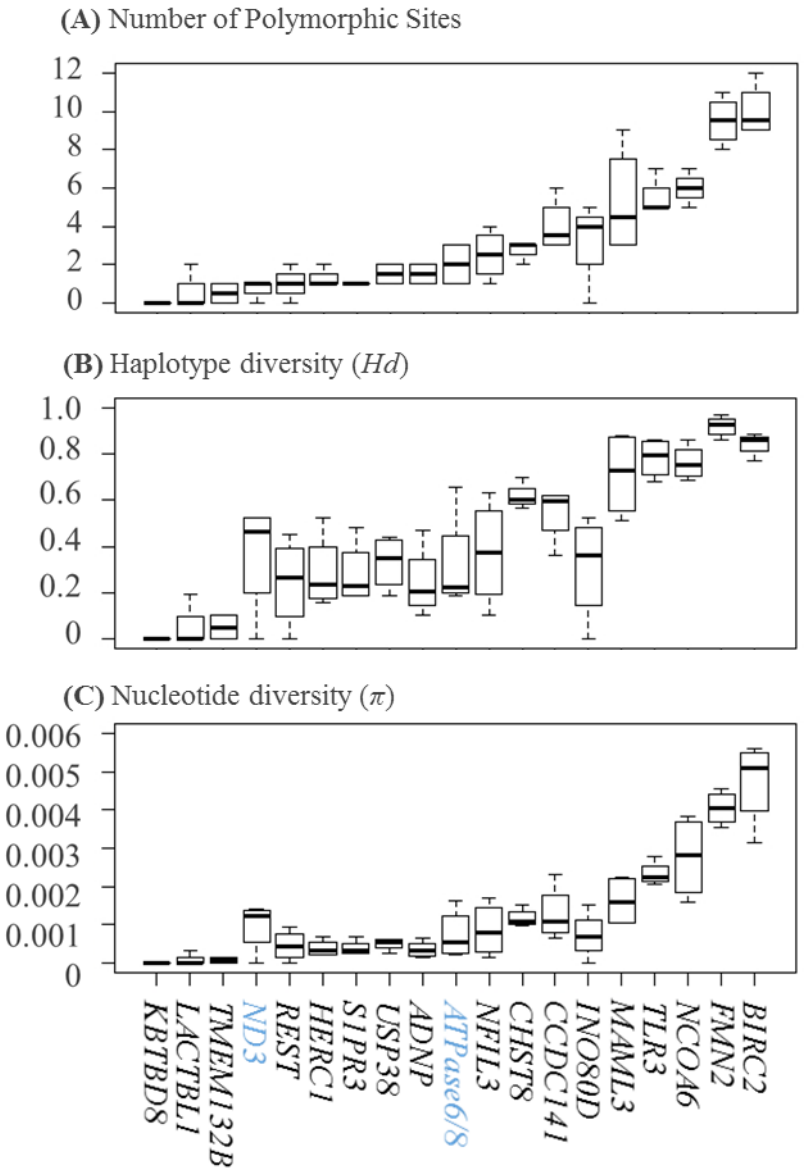
632

Figure 4


Manacus vitellinus
Corvus brachyrhynchos
Taeniopygia guttata
Geospiza fortis
Acanthisitta chloris
Passeriformes
Passerimorphae
Nestor notabilis
Melopsittacus undulatus
Psittaciformes
Falco peregrinus
Falconiformes
Cariama cristata
Cariamiformes
Merops nubicus
Coraciiformes
Picoides pubescens
Piciformes
Buceros rhinoceros
Bucerotiformes
Apaloderma vittatum
Trogoniformes
Coraciimorphae
Leptosomus discolor
Leptosomiformes
Colius striatus
Coliiformes
Tyto alba
Strigiformes
Haliaeetus leucocephalus
Haliaeetus albicilla
Accipitrimorphae
Accipitriformes
Cathartes aura
Columba livia
Columbiformes
Mesitornis unicolor
Mesitornithiformes
Tauraco erythrolophus
Musophagiformes
Otidimorphae
Cuculus canorus
Cuculiformes
Chlamydotis macqueenii
Otidiformes
Gavia stellata
Gaviiformes
Gaviimorphae
Pygoscelis adeliae
Aptenodytes forsteri
Spheniscidae
Procellariimorphae
Fulmarus glacialis
Procellariiformes
Phalacrocorax carbo
Nipponia nippon
Pelecanlmorphae
Egretta garzetta
Pelecaniformes
Pelecanus crispus
Ophisthocomus hoazin
Opisthocomiformes
Neoaves
Pterocles gutturalis
Pterocliformes
Podiceps cristatus
Podicipediformes
Phoenicopterimorphae
Phoenicopterus ruber
Phoenicopteriformes
Charadrius vociferous
Charadriiformes
Phaethontimorphae
Phaethon lepturus
Phaethontiformes
Neognathae
Eurypyga helias
Eurypygiformes
Balearica regulorum
Gruiformes
Calypte anna
Chaetura pelagica
Caprimulgimorphae
Caprimulgiformes
Antrostomus carolinensis
Gallus gallus
Meleagris gallopavo
Galliformes
Galloanseres
Anas platyrhynchos
Anseriformes
Struthio camelus
Struthionformes
Palaeognathae
Tinamus guttatus
Tinamiformes
0.02

634

Figure 5



636