

# SCIENTIFIC REPORTS



OPEN

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

Yang Liu<sup>1</sup>, Simin Liu<sup>1</sup>, Chia-Fen Yeh<sup>2</sup>, Nan Zhang<sup>1</sup>, Guoling Chen<sup>1</sup>, Pinjia Que<sup>3</sup>, Lu Dong<sup>3</sup> & Shou-hsien Li<sup>2</sup>

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 markers, 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<sup>1</sup>, traditional Sanger sequencing still has its niche in molecular evolutionary studies: pilot or small scale phylogenetic studies using PCR-based approach are cost-effective 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.*<sup>2</sup> 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<sup>3,4</sup>. However, the major drawback of this approach is that 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<sup>5</sup>.

<sup>1</sup>State Key Laboratory of Biocontrol, Department of Ecology/School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, Guangdong, China. <sup>2</sup>Department of Life Sciences, National Taiwan Normal University, Taipei, 116, Taiwan, China. <sup>3</sup>Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, 100875, China. Yang Liu and Simin Liu contributed equally. Correspondence and requests for materials should be addressed to L.D. (email: [donglu@bnu.edu.cn](mailto:donglu@bnu.edu.cn)) or S.-h.L. (email: [t43028@gmail.com](mailto:t43028@gmail.com))

Nuclear Protein-coding Loci (NPCLs) are exons without flanking introns<sup>6</sup>, and are widely used in interspecific phylogenetic studies (e.g. *RAG1*<sup>7</sup>, *c-myc*<sup>8,9</sup>). NPCL markers possess favorable properties including homogeneous base composition, varied evolutionary rates and easy alignment across species or populations<sup>10,11</sup>. Moreover, orthologous genes can be identified accurately using their annotations<sup>12,13</sup>. Several sets of universal NPCL markers had been developed specially for beetles<sup>14</sup>, fish<sup>15</sup>, reptiles<sup>6</sup>, amphibian and vertebrates<sup>16,17</sup>. 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<sup>18</sup>. So far, a considerable effort has been devoted to resolve the phylogenetic relationships from higher taxonomic categories<sup>19–21</sup> to sister species<sup>22–26</sup>. 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<sup>27–30</sup>. 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<sup>31</sup>, 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<sup>32–34</sup>. For example, Dawson *et al.*<sup>35</sup> developed a set of microsatellite markers with high cross-species utility, suitable for paternity and population studies. Backström *et al.*<sup>36</sup> 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<sup>13</sup>. Kimball *et al.*<sup>37</sup> tested the utility of 36 published markers on 42–199 bird species with only five exonic markers therein. Kerr *et al.*<sup>38</sup> 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<sup>39</sup>. 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<sup>40</sup>. 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<sup>6</sup>. 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 set of universal 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<sup>41</sup>. 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

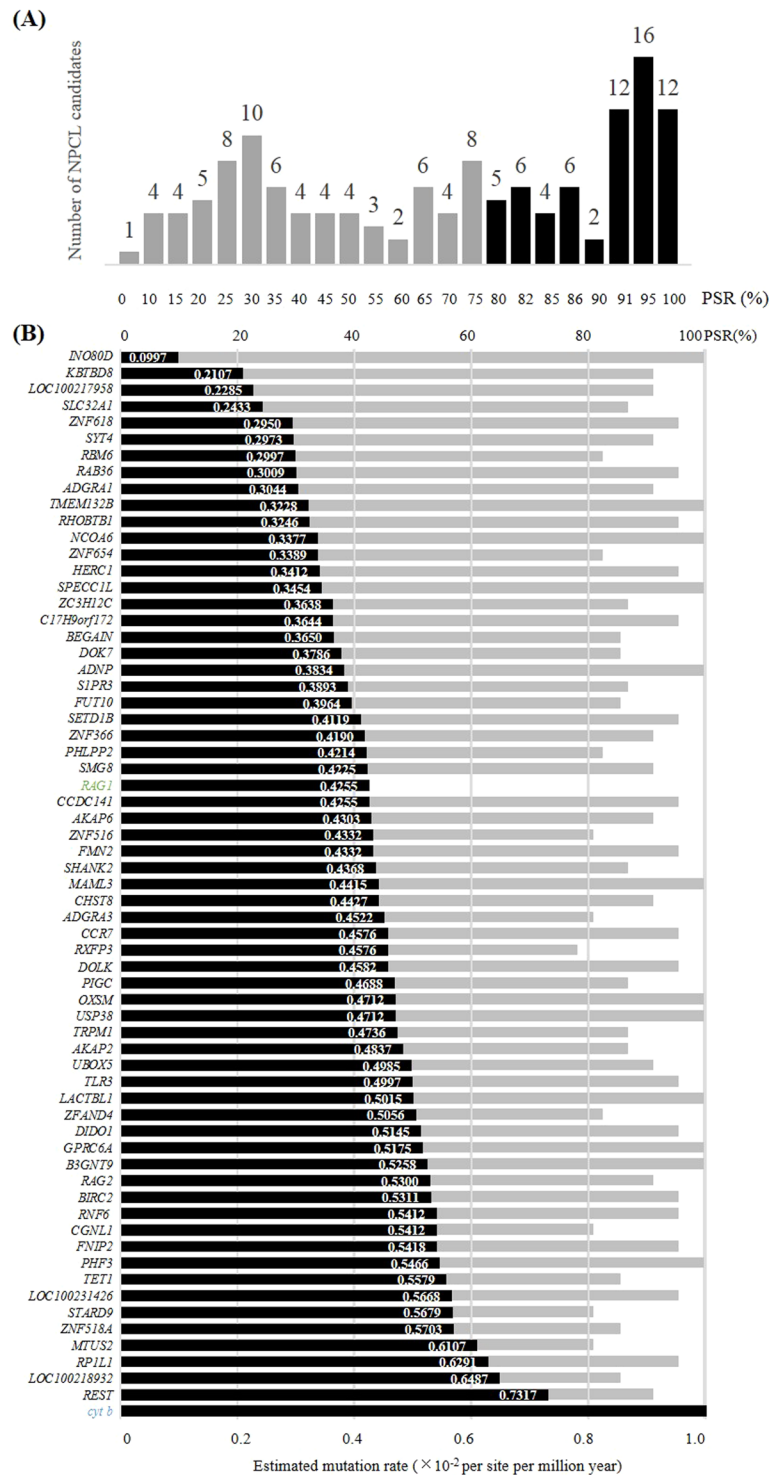
**Pan-avian order amplifications of the novel NPCLs.** The genome alignment and BLAST procedures resulted in 136 NPCL candidates, which were broadly distributed across 24 autosomal chromosomes and the Z chromosome of the Zebra Finch genome. Their original fragment length ranged from 815 bp to 7176 bp (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).

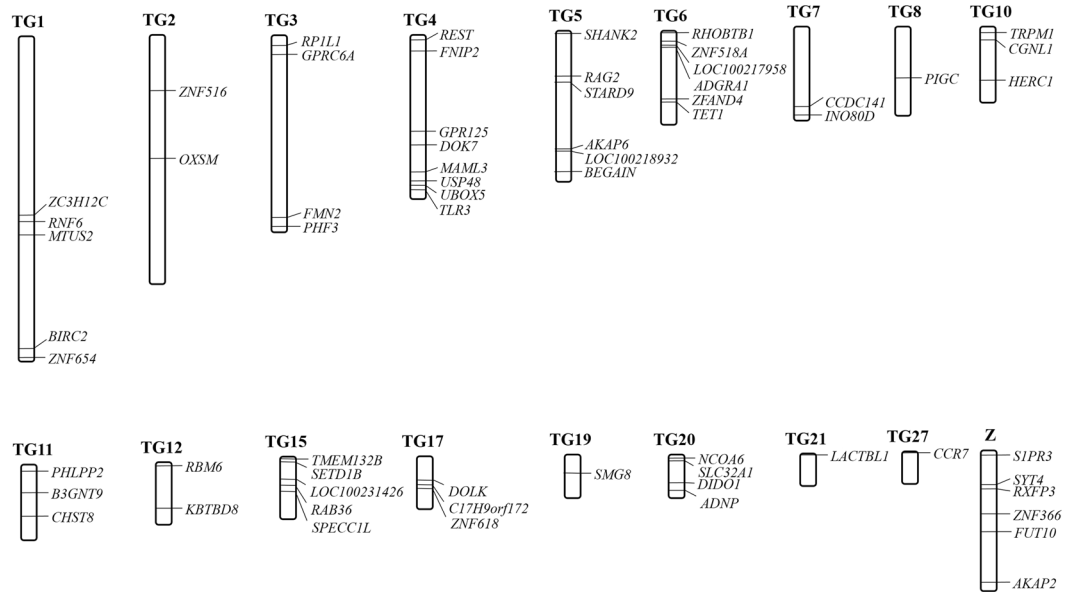
**Interspecific mutation rate and phylogenic construction of the 63 universal NPCLs.** The genome-based BLAST results showed that the widely used genetic markers, cytochrome *b* (*cyt b*) of mitochondrial DNA (mtDNA) and *RAG1*, an extensively used nuclear gene<sup>42,43</sup> were located in all 48 published genomes<sup>41</sup>. 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 this set of 63 universal NPCL markers were orthologous among these 48 species (Supplementary Table S4) and the resulting concatenated matrix with sequences of approximately 96 kb was obtained (alignment available at: DRYAD <https://doi.org/10.5061/dryad.ht3823d>).

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 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*.

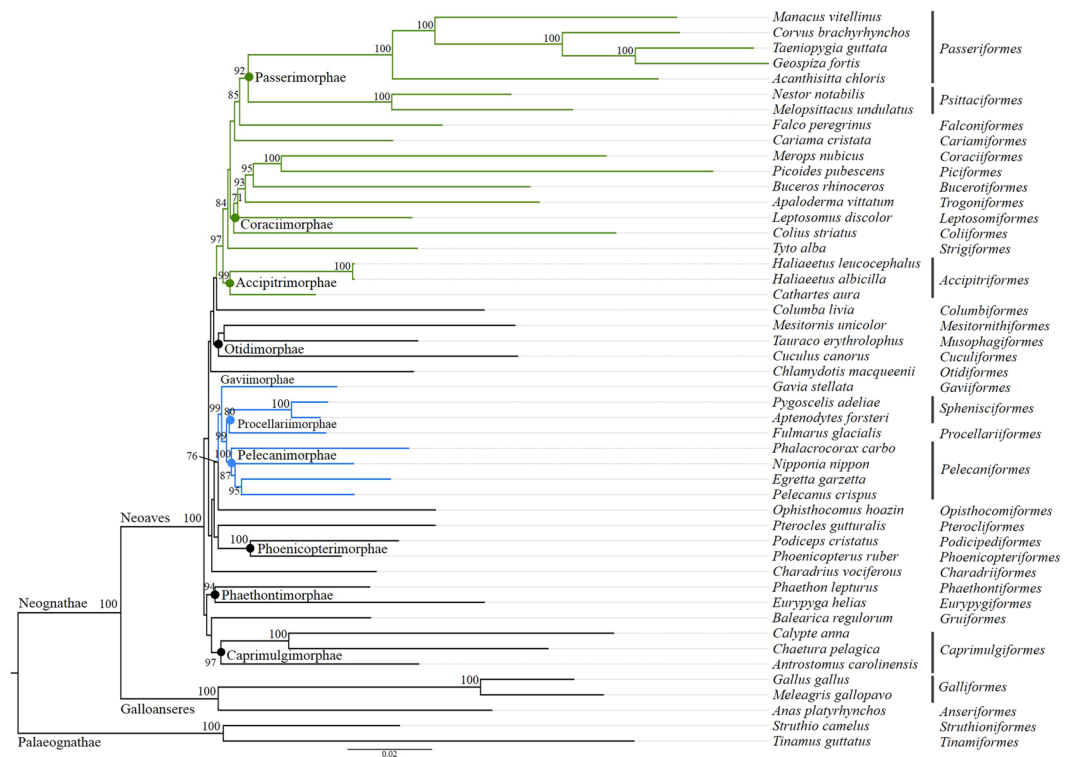




**Figure 2.** PCR success rate distribution for the 136 NPCL marker candidates and mutation rates for the 63 universal NPCL markers. **(A)** PCR success rate distribution for 136 candidates in 23 avian orders. The 63 NPCL markers with PCR success rates higher than 80% are shown in black; other loci with PSR success rates below 80% (in grey) were excluded from the subsequent analysis. The number 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 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 NPCL *RAG1* (in green) and mitochondrial gene *cyt b* (in blue) were selected as references.



**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 were widely distributed in 17 autosomal chromosomes and the Z chromosome.



**Figure 4.** Phylogenetic analysis of Neoaves using 63 NPCLs (96,000 bp) from 48 bird genomes in RAxML. The scientific names of species and corresponding orders are indicated to the right. Superorders are labelled on the nodes and two main clades representing core landbirds and waterbirds are colored in green and blue as classification in previous study<sup>41</sup> using genomic data. Bootstrap support over 70% are indicated above nodes.

three major groups with highest bootstrap support (100%). Among Neoaves group, two major clades, core landbirds (Telluraves) and core waterbirds (Aequornithia) were strongly supported by whole-genome data<sup>41</sup> and 259 independent nuclear loci<sup>44</sup>. Within core landbirds, the clade containing Passerimorphae (Passeriformes + parrots), Falconiformes (falcons), Cariamiformes (seriemas) is sister to Coraciimorphae (bee-eaters + woodpeckers



+ hornbills + trogons + cuckoo-roller + mousebirds), which is paraphyletic to Strigiformes (owls) and its sister clade Accipitriformes (eagles + New World vultures). Within core waterbirds, Pelecaniformes (pelicans + herons + ibises + cormorants) and Procellariiformes (fulmars + penguins) are two monophyletic groups, sister to Gaviiformes (loons). Other clades such as Phoenicopteriformes (flamingos + grebes), Otidiformes (bustards + turacos + cuckoos), Caprimulgiformes (hummingbirds + swifts + nightjars) and Phaethontiformes (tropicbirds + sunbitterns) are identical with previous studies<sup>41,45,46</sup>. However, we also found discordances between this phylogenetic tree and previous results<sup>41,44–46</sup>, specifically in some branches with conflict placements with low support. For example, Columbiformes (doves), Pteroclidiformes (sandgrouses) and Mesitornithiformes (mesites) are not clustered into Columbiformes. The placement of Charadriiformes (plovers), Gruiformes (cranes) and Opisthocomiformes (hoatzins) are incongruence with Jarvis *et al.*<sup>41</sup>, respectively.

**Intraspecific polymorphism of 17 randomly selected NPCL markers.** A total of 12,420 bp DNA sequences, including 11,196 bp of 17 NPCLs and 1,224 bp of two 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 *KBTD8* (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 haplotype and nucleotide diversity (mean  $Hd = 0.84$  and  $0.92$ , mean  $\pi = 0.0047$  and  $0.0040$ , respectively). In contrast, a mitochondrial gene *ND3* had only one polymorphic site, yielding a low haplotype diversity (mean  $Hd = 0.36$ ) and nucleotide diversity (mean  $\pi = 0.0009$ ). When we compared the results of interspecific mutation rates and intraspecific polymorphism, we found that the inter- and 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 polymorphism at nine NPCL markers was higher than that of two mitochondrial genes.

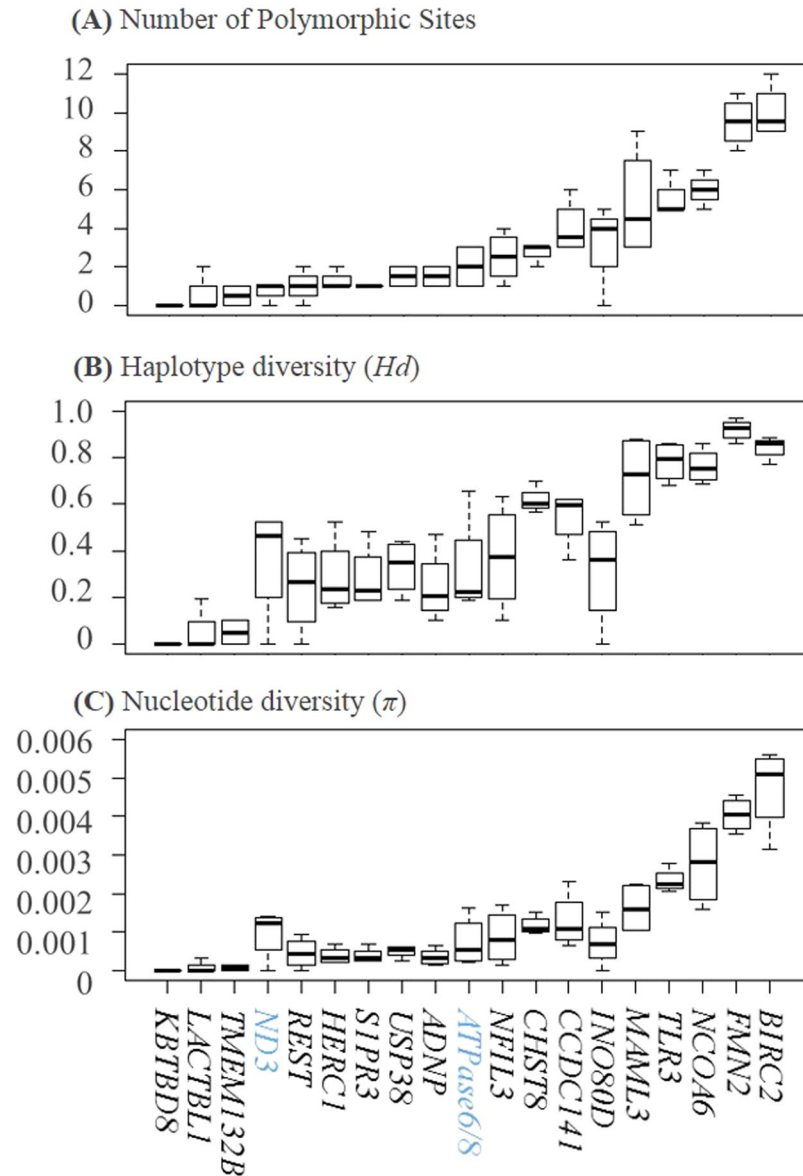
The genetic polymorphism parameters varied greatly, not only among genes, but among populations as well. For example, the  $Hd$  value of the *MAML3* gene was lowest in the Taiwan population ( $Hd = 0.51$ ) and highest in the Qinghai population ( $Hd = 0.88$ ). The measure for nucleotide diversity,  $\pi$  of the *NCOA6* gene was lowest in the Taiwan population ( $\pi = 1.59$ ) and highest in the Guangxi population ( $\pi = 3.81$ ). 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  $\geq 80\%$ ), and denoted different levels of inter- and intraspecific polymorphism. Therefore, our NPCLs 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 contributes multiple independent estimates to species trees<sup>47,48</sup>, 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<sup>49,50</sup>. 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<sup>41,44</sup>, introns<sup>20</sup>, ultraconserved elements (UCEs)<sup>46,51</sup> and retroposon presence/absence matrix<sup>45</sup>. The congruent parts of topology reveal multiple cluster clades, such as Telluraves (core landbirds), Aequornithia (core waterbirds) and Phoenicopteriformes, Otidiformes, Caprimulgiformes and Phaethontiformes. We also find some unresolved placements comparing with Jarvis *et al.*<sup>41</sup>. These include Columbiformes (doves), Pteroclidiformes (sandgrouses), Mesitornithiformes (mesites), Charadriiformes (plovers) and Gruiformes (cranes), which exhibits hard polytomies in the avian tree of life. Though recent efforts in avian phylogenomic studies using whole-genome<sup>41</sup> or genome-level data<sup>44</sup>, irresolvable relationships have been found in some clades<sup>41,44–46</sup>. Suh *et al.*<sup>45</sup> investigated the causes of phylogenetic irresolvabilities and concluded that such phylogenetic discordances were originated from prevalent ancestral polymorphism denoted by incomplete lineage sorting (ILS)<sup>52</sup>, which is probably associated with an initial near-K-Pg super-radiation<sup>41</sup> 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 Neoaves, containing several unresolved lineages, leads extreme ILS and associated network-like phylogenetic relationships<sup>45,46</sup>. On one hand, again, the topology reconstructed by the present set of 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 sufficient polymorphism to resolve phylogenetic relationships among lineages with less ILS in Neoaves.

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. For example, microsatellites were developed for specific species or orders to detect differences in genotypes and further to quantify intraspecific genetic diversity<sup>35,53,54</sup>. But introns like microsatellites have high levels of length homoplasy<sup>55</sup>. It is commonly assumed that NPCLs are conservative loci, highly suitable to address questions concerning high-level systematics<sup>40</sup>. However, some population genetic studies highlight the importance of using functional exonic SNPs in population genetic studies<sup>11,56</sup>, comparing to neutral markers (such as microsatellite and mitochondrial DNA). Datasets that contain numbers of several to more than 100 exon genes<sup>51,57</sup> can support accurate and



**Figure 5.** Polymorphism at 17 avian universal NPCL markers and two mitochondrial loci in four 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 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 ranging from 0 to 5.59.

reliable estimates of population genetic parameters<sup>55,58</sup>, and have a substantial power in population genetic analysis<sup>5,40,50,59</sup>. 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<sup>60</sup>, this species exhibits variability in morphology and behavior among and within populations in East Asia<sup>61</sup>, 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<sup>62</sup>. 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*<sup>7</sup> and some loci at mtDNA<sup>63</sup> provide a reference for marker choice (Figs 2B and 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<sup>64,65</sup> and effective population sizes ( $N_e$ )<sup>59</sup>. The associated analyses, such as species-tree estimation, e.g. MP-EST<sup>66</sup>, \*BEAST<sup>67</sup>, BP&P<sup>68</sup>, and demographic analysis, such as Isolation with Migration (IM) model<sup>69,70</sup> and Approximate Bayesian Computation (ABC) simulations<sup>59</sup> require multiple

independent loci with different demographic histories and mutation rates. In this regard, markers from different genomic segments, such as introns (developed previously<sup>35–37</sup>) and exons we developed are preferred to combine to be used. It is no doubt that the present marker set is a useful resource to generate 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<sup>25,30</sup>.

Compared with traditional Sanger sequencing, the fast development of next generation sequencing (NGS) techniques has enabled researchers to obtain genetic polymorphisms easily<sup>71</sup>. For example, Jarvis *et al.*<sup>41</sup> performed a highly resolved phylogenetic tree of 48 species using phylogenomic methods, and Prum *et al.*<sup>44</sup> conducted a comprehensive phylogeny of 198 species within 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 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 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 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 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 analysis. Processing genomic data always places high demands on bioinformatics and computational power<sup>72</sup>. High-quality samples for NGS, project budget, bioinformatic facilities are not available to all 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.  $T_m = 50^\circ\text{C}$ ), so that the PSR of each NPCL marker might be underestimated. Reducing the annealing temperature by  $1\sim 2^\circ\text{C}$  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<sup>73</sup>. 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<sup>74</sup>. Kimball *et al.* proposed that adopting various analytical methods might overcome these adverse effects<sup>75</sup>.

In conclusion, we have developed 63 avian universal NPCL markers, evenly distributed across 17 autosomal chromosomes and the Z chromosome. This set of universal NPCL markers had high PCR success rates (PSR  $\geq 80\%$ ) 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 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 have widespread application in phylogenetic and population genetics studies.

## Methods

**Development of NPCL markers and primer design.** To screen NPCL marker candidates, we aligned parts of the genome of two species with a distant phylogenetic relationship<sup>76</sup>, the Red Junglefowl (GCA\_000002315) and the Zebra Finch (GCA\_000151805). Firstly we identified long ( $>600$  bp) 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 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 orthologous exons and employed them as NPCL marker candidates.

We used the program Primer3<sup>77</sup> to design the primers for NPCL marker candidates. We selected exon sequences of Zebra Finch as templates, focusing on High-scoring Segment Pairs region (HSP, 700–1200 bp). For each primer pair, the oligomer ranged from 18 bp to 25 bp and GC content ranged from 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 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.

**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<sup>18</sup>.

Total genomic DNA was extracted from ethanol-preserved muscle tissue or blood using a TIANamp Genomic DNA kit (TIANGEN, China) and stored at  $4^\circ\text{C}$ . DNA concentration and purity were estimated by NanoDrop 2000 (Thermo Scientific, USA). We used Touchdown PCR (TD-PCR)<sup>78</sup>, an improved standard PCR to test the utility of primers sensitively, by decreasing the annealing temperature  $1^\circ\text{C}/\text{cycle}$  from  $T_m + 10^\circ\text{C}$  to  $T_m$  (Melting



Temperature). The Touchdown PCR was performed in a Veriti96 PCR thermal cycler system (ABI, USA) using a 10  $\mu$ l reaction containing 2  $\mu$ l template DNA (10–70 ng totally), with mixed concentrations of 10  $\times$  PCR buffer, 20  $\mu$ M dNTP, 10 mM of each forward and reverse primer, and 5U Taq polymerase (Takara, China). The initial temperature profile was 2 min at 94 °C, 10 cycles at 94 °C for 30 s, 60–50 °C (decreasing the annealing temperature by 1 °C per cycle) for 30 s and 72 °C for 90 s followed by 30 similar cycles but with a constant annealing temperature of 50 °C. This process was concluded with an extra elongation step at 72 °C 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 120 V for 30 min.

**Estimation of inter-species mutation rates and construction of Neoavian phylogeny.** We downloaded 48 avian genomes covering all orders of Neoaves<sup>41</sup>. 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 Finch (abbreviation as *Tgu1*) using BLAST. These sequences were extracted and filtered in batches by own-developed Perl script (shared in DRYAD <https://doi.org/10.5061/dryad.ht3823d>) in the Tianhe-2 server (School of Advanced Computing, Sun Yat-sen 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 *Melospittacus undulatus*<sup>41</sup>. 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 orthologous sequences using MEGA v6.0<sup>79</sup>.

To estimate the mutation rate of each gene, we firstly computed the overall mean genetic distance 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 rate in the *cyt b* (0.01035 mutations per site per million years)<sup>80</sup> to get the average mutation rate of each NPCL<sup>81</sup>.

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<sup>82</sup>, with GTRCAT model and 1,000 bootstrap runs. Maximum-likelihood-bootstrap proportions  $\geq 70\%$  were considered strong support<sup>83</sup>.

**Intra-species polymorphism measurements.** 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 European populations of the Kentish Plover<sup>84</sup> and compare the degree of polymorphism between nuclear and mtDNA, we added two mtDNA loci, ATPase subunit six concatenated with partial ATPase 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) (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 Institute (BGI, China).

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

## Data Availability

BLAST alignment of 63 universal NPCL markers and Perl script: DRYAD <http://dx.doi.org/XXXX>.

## References

1. Ansorge, W. J. Next-generation DNA sequencing techniques. *New Biotechnol.* **25**, 195–203 (2009).
2. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving 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).
4. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
5. Delsuc, F., Brinkmann, H. & Philippe, H. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* **6**, 361–375 (2005).
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 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 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 insertions in the nuclear gene *c-myc*: a novel gene in avian phylogenetics. *The Auk* **117**, 1069–1178 (2000).
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 genes. *The Auk* **119**, 984 (2002).
10. Boekhorst, J. & Snel, B. Identification of homologs in insignificant blast hits by exploiting extrinsic gene properties. *BMC Bioinformatics* **8**, 356 (2007).
11. Zhan, X. *et al.* Exonic versus intronic SNPs: contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity* **114**, 1–9 (2015).
12. Remm, M., Christian, E. S. & Erik, L. S. Automatic clustering of orthologs and in-paralogs 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 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. Jönsson, K. A. *et al.* A supermatrix phylogeny of corvid 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 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 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-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 birds. *Science* **346**, 1320–1331 (2014).
42. Cibois, A. & Cracraft, J. Assessing the passerine “Tapestry”: phylogenetic relationships of the Muscipoidea inferred from nuclear DNA sequences. *Mol. Phylogenet. Evol.* **32**, 264–273 (2004).
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 (2003).
44. Prum, R. O. *et al.* A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526**, 569–573 (2015).
45. Suh, A., Smeds, L. & Ellegren, H. The dynamics of incomplete lineage sorting across the 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. *Zool. Scr.* **45**, 50–62 (2016).
47. Salas-Leiva, D. E. *et al.* Conserved genetic regions across angiosperms as tools to develop single-copy nuclear markers in gymnosperms: an example using cycads. *Mol. Ecol. Resour.* **14**, 831–845 (2014).
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).
49. Edwards, S. V. Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19 (2009).
50. Szöllösi, G. J., Tannier, E., Daubin, V. & Boussau, B. The Inference of Gene Trees with Species Trees. *Syst. Biol.* **64**, e42–e62 (2015).
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).
52. Maddison, W. P. & Knowles, L. L. Inferring Phylogeny Despite Incomplete Lineage Sorting. *Syst. Biol.* **55**, 21–30 (2006).
53. Galbusera, P., Dongen, S. van & Matthysen, E. Cross-species amplification of microsatellite primers in passerine birds. *Conserv. Genet.* **163–168** (2000).
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).
55. Sunnucks, P. Efficient genetic markers for population biology. *Trends Ecol. Evol.* **15**, 199–203 (2000).
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 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 amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 1414–1419 (2002).
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).
59. Beaumont, M. A., Zhang, W. & Balding, D. J. Approximate Bayesian computation in population genetics. *Genetics* **162**, 2025–2035 (2002).

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).
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).
62. Nosenko, T. *et al.* Deep metazoan phylogeny: when different genes tell different stories. *Mol. Phylogenet. Evol.* **67**, 223–233 (2013).
63. Zink, R. M. & Barrowclough, G. F. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* **17**, 2107–2121 (2008).
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).
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. USA* **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. U. S. A.* **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 *Bioinformatics 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., Filipiński, A. & Kumar, S. MEGA6: Molecular 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 (2008).
81. Li, J. W. *et al.* Rejecting strictly allopatric speciation on a continental island: prolonged postdivergence gene flow between Taiwan (*Leucodioptron taewanus*, Passeriformes Timaliidae) and Chinese (*L. canorum canorum*) hwameis. *Mol. Ecol.* **19**, 494–507 (2010).
82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
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).
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* **126**, 839–852 (2009).
85. Swindell, S. R. & Plasterer, T. N. SEQMAN. *Seq. Data Anal. Guideb.* 75–89 (1997).
86. Librado, P. & Rozas, J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
87. Hudson, R. R., Kreitman, M. & Aguadé, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
88. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).

## Acknowledgements

We are grateful to Per Alström, Fuming Lei, Fasheng Zou, Xiaojun Yang, Ulf Johansson, Chung-Yu 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 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 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 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

Y.L., S.H.L. and D.L. designed this study. C.F.Y. and N.Z. carried out the primer design and BLAST procedures. G.L.C. and P.J.Q. provided materials and technical support in the lab. S.M.L. completed wet lab experiments, analyzed the data and wrote the manuscript with Y.L.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-33646-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018