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: 3804

Number of references: 77

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 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 (polymerase chain reaction) 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 using touch-down PCR (TD-PCR). The sixty-two 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 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 could be consistently amplified in the Kentish plover and 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.

Keywords: primer, multiple gene, exon, polymorphism, phylogeny, coalescent theory

Introduction

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 and to reconstruct the demographic history of populations (Kingman, 1982; Tajima, 1983). Datasets that contain numbers from several to more than 100 nuclear genes (e.g. Bapteste et al., 2002; McCormack et al., 2013) can support accurate and reliable estimates of population parameters (Edwards & Bensch, 2009; Sunnucks, 2000), and have substantial power in phylogenetic and population genetic analysis (Beaumont et al., 2002; Brito & Edwards, 2009; Delsuc et al., 2005; Szöllősi et al., 2015). Birds are the most common and species-rich group of terrestrial vertebrates and exhibit tremendous diversity in their phenotypes, ecology, habitats and behaviors (Jetz et al., 2012). An immense effort has therefore been put into studies of their genetics, ecology and physiology to understand bird species' dispersion and adaptation (Burleigh et al., 2015). So far, a considerable effort has been devoted to resolve the phylogenetic relationships from higher taxonomic categories (e.g. Berv & Prum, 2014; Hackett et al., 2008; Jønsson et al., 2016;) to sister species (Backström et al., 2013; Chu et al., 2013; Dong et al., 2014; Wang et al., 2016; Yeung et al., 2011). 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 (e.g. Hung et al., 2016; Lim & Sheldon, 2011; Shaner et al., 2015; Wang et al., 2017). 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 (Burleigh et al., 2015) , which allows the application of sophisticated modeling and thus hypothesis testing.

By comparing molecular phylogenies based on different dataset sizes, Rokas et al. (2003) proposed that at least 20 loci are required to resolve incongruent branches of the Tree of Life (TOL). Thus, sufficient numbers of nuclear markers are required for effective evolutionary studies. Several sets of universal gene markers had been developed specially for fish (Li et al., 2007), reptiles (Townsend et al., 2008), beetles (Che et al., 2017) and even vertebrates (Shen et al., 2013). Only a few studies have developed avian universal nuclear markers (e.g. microsatellite loci: Dawson et al., 2010), and there is still no set of easily amplifiable nuclear genes that can fulfill the needs of modern coalescent-based analysis for most of bird species. Researchers therefore often need to design new primers for a bird species on which genomic data are not yet available. For instance, Kerr et al. (2014) developed 100 exonic markers from five avian genomes, and finally tested a subset of 25 markers in 12 avian orders. Kimball et al. (2009) tested the utility of 36 published markers on 199 species representing all avian orders. Backström et al. (2007) 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 should therefore be an ideal alternative for phylogenetic reconstruction (Thomson et al., 2010).

Nuclear Protein-coding Loci (NPCLs) are exons without flanking introns (Townsend et al., 2008), and are widely used in interspecific phylogenetic studies of Aves (e.g. RAG-1: Groth & Barrowclough, 1999; *c-myc*: Ericson et al., 2000; both exons: Johansson et al., 2002). NPCL markers possess favorable properties including homogeneous base composition, varied evolutionary rates and easy alignment across species or populations (Boekhorst & Snel, 2007; Zhan et al., 2015). Moreover, orthologous genes can be identified accurately using their annotations (Remm et al., 2001; Thomson et al., 2010).

However, some obstacles have hindered the development of universal NPCL markers. Firstly, widespread flanking introns make identification of the exon boundaries of a specific NPCL marker difficult (Townsend et al., Alegre et al., 2008). 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. However, using a small number of universal NPCL markers increases the probability of error when estimating species relationships due to conflict between loci, and researchers have advocated the use of more genes with longer sequences (Brito & Edwards, 2009). The development of a universal set of NPCL markers for birds would therefore significantly reduce the time required for future research as well as its cost.

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. We further estimated mutation rates by extracting universal NPCL markers from 48 published avian genomes (Jarvis et al., 2014). Samples from four populations of the Kentish Plover (*Charadrius alexandrinus*) were also amplified to estimated the intra-species polymorphic level of these universal NPCL markers.

Materials and 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 (Nam et al., 2010), the Red Junglefowl (GCA_000002315) and the Zebra

Finch (GCA_000151805). First we identified long (>600bp) single-copy exons within the genome of the Red Junglefowl and took these exons as templates. Then we aligned them with the genome of the Zebra Finch using BLAST (Basic Local Alignment Search Tool). We assumed that query sequences of the Zebra Finch with identity more than 80% and length more than 50% of templates length were orthologous exons and employed them as NPCL candidates.

We used Primer3 (Rozen & Skaletsky, 2000) to design the primers for NPCL candidates and used Touchdown PCR (TD-PCR), an improved standard PCR, by decreasing the annealing temperature 1℃/cycle from Tm+10℃ to Tm (Melting Temperature, Tm) (Don et al., 1991). For each primer pair, the oligomer ranged from 18bp to 25bp and GC content ranged from 20% to 80%.

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 (Table S2). We used a set of 10000 trees with 9993 operational taxonomic unites (OTUs) downloaded from http://birdtree.org/ (Jetz et al., 2012) to demonstrate the phylogenetic relationships among selected species.

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). 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

We downloaded 48 avian genomes covering all orders of Neoaves (Jarvis et al. 2014). Sixty-five gene sequences (the 63 universal NPCLs being tested, one frequently-used nuclear gene, RAG-1, and one mitochondrial gene, cytochrome *b* (*cyt b*) were aligned in these genomes using BLAST, and the results were filtered and adjusted by a Perl script in the Tianhe-2 server (School of Advanced Computing, Sun Yat-sen University). We BLASTed each NPCL of the Zebra Finch against 45 available genomes, (Three genomes, *Anas platyrhynchos*, *Meleagris gallopavo* and *Melopsittacus undulates* were excluded as their file type is megascaffold, which is not suitable for our script, so that the sequences could not be used in subsequent analysis.) To estimate the mutation rate of each gene, we firstly computed the overall mean genetic distance at each NPCL marker in MEGA v6.0 (Tamura *et al.* 2013) with 1000 bootstrap replicates. Then we calculated the ratio between the genetic distance at each NPCL and that in the *cyt b* gene. Finally, we multiplied the ratio by the average mutation rate in the *cyt b* gene (0.01035 mutations per site per million years) to get the average mutation rate of each NPCL (Li et al., 2010; Weir & Schluter, 2008).

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 (Küpper et al., 2009) and compare the degree of polymorphism between nuclear and mitochondrial genes, we added two mitochondrial genes, 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 Zhuang (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 (Swindell & Plasterer, 1997). Some parameters of the DNA polymorphism, the number of polymorphic sites (*S*) and haplotypes (*H*), haplotype diversity (*Hd*), and nucleotide diversity (π) were calculated using DnaSP v5.0 (Librado & Rozas, 2009) . The neutrality of each locus was tested using the Hudson-Kreitman-Aguade (HKA) test (Hudson, Kreitman, & Aguadé, 1987) and Tajima's D (Tajima, 1989) implemented in DnaSP v5.0.

Results

Pan-avian order amplifications of the novel NPCL markers

The genome alignment and BLAST procedures described above resulted in 136 NPCL marker candidates, which were broadly distributed across 24 autosomal chromosomes and the Z chromosome of the Red Junglefowl genome. Their fragment length ranged from 656bp to 1190bp (Table S1).

In total, 5,146 PCRs were performed to amplify the 136 NPCL markers in 41 species representing 23 avian orders. Among them, 2,875 (55.9%) of PCR performances produced a target band (Table S3). For the 136 candidates, we successfully amplified 12 NPCL markers 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 ≥ 80%) (Figure 2A); all of them were successfully amplified in *Caprimulgiformes* and *Gruiformes*, and the PSR ranged from 65% to 97% in other orders (Figure 1B, Table S3).This set of 63 avian nuclear markers was evenly distributed across 17 autosomal chromosomes and the Z chromosome (Figure 3).

Interspecific mutation rate of the 63 universal NPCL markers

The genome-based BLAST showed that the widely used NPCL marker RAG-1 was recorded in all 45 genomes, whilst 42 genomes recorded the mitochondrial gene *cyt b*. Of the 63 novel NPCL markers, 43 genes (68%) were recorded in all genomes, 18 genes (29%) were recorded in 44 genomes, and one gene (BCR) was recorded in 43 genomes (Table S4). The locus 'RNF' was recorded in just eight genomes and was excluded from subsequent analysis. These results confirmed the universality of the 62 remaining NPCL markers, which were orthologous among the 45 genomes considered in this study.

The range of the estimated mutation rates for the universal avian NPCL markers is broad; it ranged from 0.0864 to 0.58×10^{-8} per site per million years (Figure 2B). Among these 62 NPCL markers, the mutation rates of 22 NPCL were slower than the mutation rate of RAG-1, an extensively used nuclear gene (Cibois & Cracraft, 2004; Paton et al., 2003), whilst the other 40 NPCL markers were faster. All NPCL markers showed a slower mutation rate than that of the mitochondrial *cyt b* gene. The wide range of mutation rates of these universal nuclear markers, approximately 7-fold,

indicates its potential to resolve phylogenetic and population genetics questions at multiple levels of the hierarchy.

Intraspecific polymorphism of 17 randomly selected NPCL markers

A total of 12,420bp DNA sequences, including 11,196bp of 17 NPCL and 1,224bp of two mtDNA genes 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 '12_52' (Figure4). There were 10 polymorphic sites in loci 'LOC10022531' and '3_119', while there were only 1-6 polymorphic sites in other loci. Correspondingly, 'LOC10022531' and '3_119' possessed the highest values of 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 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 intraspecific genetic diversity of our gene markers were incongruent; although the estimated mutation rates at our NPCL markers 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 4_108 gene was lowest in the TW population (*Hd*= 0.51) and highest in the QH population (*Hd*= 0.88). The measure for nucleotide diversity, π, of the 20_7 gene was lowest in the TW population (π= 1.59) and highest in the GX population (π= 3.81). Detailed information on measures from each population is available in 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 (Table S5).

Discussion

We developed a set of 62 avian universal NPCL markers with diverse mutation rates and levels of intraspecific polymorphism. Our results showed that the 62 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 marker 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.

Applicability of the novel NPCL markers

It is commonly assumed that NPCLs are conservative loci, highly suitable to address questions concerning high-level systematics (Brito & Edwards, 2009). However, the information available from individual genes depends on their mutation rates and evolutionary history (Nosenko et al., 2013). Inherited from different chromosomes, nuclear markers contribute multiple independent estimates to estimates of species trees (Salas-Leiva et al., 2014; Waters et al., 2010), in order to alleviate the node conflicts of phylogenetic trees caused by incomplete lineage sorting, horizontal gene transfer, inconsistent evolutionary rates, gene duplication and/or gene loss and so on (Edwards, 2009; Szöllősi et

al., 2015). Our novel set of NPCL markers has a wide range of mutation rates, providing a rich source to constitute diverse multilocus datasets for different taxonomic levels in avian evolutionary studies, which have substantially more resolving power than the use of mtDNA alone (see also Zink & Barrowclough, 2008). The gene RAG-1 is one of the most widely-used markers in avian research, including phylogenetic studies among and within avian orders. Therefore, we recommend that, when studying high-level avian systematics, it is advisable to select NPCL markers with slow mutation rates, i.e. slower than that of RAG-1 to resolve phylogenies of deep nodes. Moreover, a dataset containing markers with different mutation rates is beneficial to estimate and calibrate the divergence time between branch nodes (Smith & Klicka, 2013; Thorne & Kishino, 2002). The PSR of each gene marker, the location on the chromosome, annotations and their mutation rates are useful information when we select at least 20 NPCLs, as Rokas et al (2003) proposed, to resolve the phylogenetic relationships among all taxa within Aves.

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 (Dawson et al., 2010; Galbusera et al., 2000; Wang et al., 2017). However, microsatellites have high levels of length homoplasy (Sunnucks, 2000). Studying 17 NPCL markers in the Kentish Plover, we found that 16 had low to moderate levels of intraspecific polymorphism and nine of them showed higher genetic diversity than mitochondrial genes. This dataset provides sufficient information to study population genetics in the Kentish Plover in East Asia (Wang et al., 2018). Although a previous study showed a low level of genetic differentiation across Eurasian populations (Küpper et al., 2012), this species exhibits variability in morphology and

behavior among and within populations in East Asia (Rheindt et al., 2011), warranting further coalescent-based analysis of their evolutionary history. Combined, our set of NPCL markers provides polymorphic data which can be widely used to estimate species trees using methods such as MP-EST (Liu et al., 2010), *BEAST (Drummond et al., 2012) and BP&P (Yang & Rannala, 2010).They could also be selected for population genetic analyses using coalescent-based approaches such as the Isolation with Migration (IM) model (Hey, 2004) and Approximate Bayesian Computation (ABC) simulations (Beaumont et al., 2002) to infer divergence models and estimate the direction and magnitude of gene flow, effective population size (N_e) and timing of divergence (reviewed in Hey 2007). Indeed some studies have used these novel NPCL to apply the aforementioned analyses (Hung et al., 2014, Wang et al., 2017).

The challenge of developing NPCL markers

Compared with traditional Sanger sequencing, the fast development of next generation sequencing (NGS) techniques has enabled researchers to obtain genetic polymorphisms easily (McCormack et al., 2013). For example, Jarvis et al.(2014) performed a highly resolved phylogenetic tree of 48 species using phylogenomic methods, and Prum et al (2015) 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 future avian molecular studies. Sanger sequencing technique of NPCL markers is less sensitive to the quality of template DNA than genomic approaches. Moreover, CPU time and computational burden are linearly correlated with data size (Roure et al. 2013) and, thus, genomic data places high demands on computing power. High-quality samples for

sequencing and high-performance servers are not available to all researchers. It is still useful and necessary to align orthologous sequences across multiple hierarchical levels using NPCL markers, especially for a pilot study.

However, there are some limitations when using this set of universal NPCL markers. First, PCR performance was simultaneously tested under a unified protocol (e.g. Tm=50℃), so that the PSR of each NPCL marker could 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 (Lande 1976). 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 (Felsenstein, 1978). Kimball et al. (2013) proposed that adopting various analytical methods might overcome these adverse effects.

In conclusion, we have developed 62 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 is 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.

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. 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.

References

Backström, N., Fagerberg, S., & Ellegren, H. (2007). Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*, *17*(4),

964–980.

Backström, N., Sætre, G.-P., & Ellegren, H. (2013). Inferring the demographic history of european ficedula, flycatcher populations. *BMC Evolutionary Biology*, *13*(1), 2.

Bapteste, E., Brinkmann, H., Lee, J. A., Moore, D. V., Sensen, C. W., Gordon, P., … Philippe, H. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1414–1419.

Beaumont, M. A., Nielsen, R., Robert, C., Hey, J., Gaggiotti, O., Knowles, L., … Excoffier, L. (2010). In defence of model-based inference in phylogeography. *Molecular Ecology*, *19*(3), 436–446.

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

Berv, J. S., & Prum, R. O. (2014). 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. *Molecular Phylogenetics and Evolution*, *81*, 120–136.

Boekhorst, J., & Snel, B. (2007). Identification of homologs in insignificant blast hits by exploiting extrinsic gene properties. *BMC Bioinformatics*, *8*(1), 356.

Brito, P. H., & Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, *135*(3), 439–455.

Burleigh, J. G., Kimball, R. T., & Braun, E. L. (2015). Building the avian tree of life using a large-scale, sparse supermatrix. *Molecular Phylogenetics and Evolution*, *84*, 53–63.

Che, L.-H., Zhang, S.-Q., Li, Y., Liang, D., Pang, H., Ślipiński, A., & Zhang, P. (2017). Genome-wide

survey of nuclear protein-coding markers for beetle phylogenetics and their application in resolving both deep and shallow-level divergences. *Molecular Ecology Resources*. *17*(6), 1342-1358.

Chu, J.-H., Wegmann, D., Yeh, C.-F., Lin, R.-C., Yang, X.-J., Lei, F.-M., … Li, S.-H. (2013). Inferring the geographic mode of speciation by contrasting autosomal and sex-linked genetic diversity. *Molecular Biology and Evolution*, *30*(11), 2519–2530.

Cibois, A., & Cracraft, J. (2004). Assessing the passerine "Tapestry": phylogenetic relationships of the Muscicapoidea inferred from nuclear DNA sequences. *Molecular Phylogenetics and Evolution*, *32*(1), 264–273.

Curto, M. A., Puppo, P., Ferreira, D., Nogueira, M., & Meimberg, H. (2012). Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). *Molecular Phylogenetics and Evolution*, *63*(3), 758–767.

Dawson, D. A., Horsburgh, G. J., Küper, C., Stewart, I. R. K., Ball, A. D., Durrant, K. L., … Burke, T. (2010). New methods to identify conserved microsatellite loci and develop primer sets of high cross-species utility - as demonstrated for birds. *Molecular Ecology Resources*, *10*(3), 475–494.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, *6*(5), 361–375.

Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K., & Mattick, J. S. (1991). 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Research*, *19*(14), 4008–4008.

Dong, F., Li, S.-H., Zou, F.-S., Lei, F.-M., Liang, W., Yang, J.-X., & Yang, X.-J. (2014). Molecular systematics and plumage coloration evolution of an enigmatic babbler (Pomatorhinus ruficollis)

in East Asia. *Molecular Phylogenetics and Evolution*, *70*, 76–83.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology & Evolution*, *29*(8), 1969-1973.

Edwards, S., & Bensch, S. (2009). Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, *18*(14), 2930–2933.

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

Ericson, P. G. P., Johansson, U. S., & Parsons, T. J. (2000). Major divisions in oscines revealed by insertions in the nuclear gene c-myc: a novel gene in avian phylogenetics. *The Auk*, *117*(4), 1069-1178.

Excoffier, L., Estoup, A., & Cornuet, J.-M. (2005). Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*, *169*(3), 1727–1738.

Felsenstein, J. (1978). Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Biology*, *27*(4), 401–410.

Felsenstein, J. (2005). Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, *23*(3), 691–700.

Galbusera, P., Dongen, S. van, & Matthysen, E. (2000). Cross-species amplification of microsatellite primers in passerine birds. *Conservation Genetics*, (1), 163–168.

Groth, J. G., & Barrowclough, G. F. (1999). Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Molecular Phylogenetics and Evolution*, *12*, 115–123.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., … Yuri, T. (2008). A phylogenomic study of birds reveals their evolutionary history. *Science*, *320*(5884), 1763–

1768.

Hey, J. (2004). Multilocus Methods for Estimating Population Sizes, Migration Rates and Divergence Time, With Applications to the Divergence of *Drosophila pseudoobscura* and *D. persimilis. Genetics*, *167*(2), 747–760.

Hudson, R. R., Kreitman, M., & Aguadé, M. (1987). A test of neutral molecular evolution based on nucleotide data. *Genetics*, *116*(1), 153–159.

Hung, C.-M., Drovetski, S. V., & Zink, R. M. (2016). Matching loci surveyed to questions asked in phylogeography. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 20152340.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., … Nabholz, B. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, *346*, 1320–1331.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448.

Johansson, U. S., Irestedt, M., Parsons, T. J., & Ericson, P. G. P. (2002). Basal phylogeny of the tyrannoidea based on comparisons of cytochrome b and exons of nuclear c-myc and RAG-1 genes. *The Auk*, *119*(4), 984-995.

Jønsson, K. A., Fabre, P.-H., Kennedy, J. D., Holt, B. G., Borregaard, M. K., Rahbek, C., & Fjeldså, J. (2016). A supermatrix phylogeny of corvoid passerine birds (Aves: Corvides). *Molecular Phylogenetics and Evolution*, *94*, 87–94.

Kerr, K. C. R., Cloutier, A., & Baker, A. J. (2014). One hundred new universal exonic markers for birds developed from a genomic pipeline. *Journal of Ornithology*, *155*(2), 561–569.

Kimball, R. T., Braun, E. L., Barker, F. K., Bowie, R. C. K., Braun, M. J., Chojnowski, J. L., … Heimer-Torres, V. (2009). A well-tested set of primers to amplify regions spread across the avian genome.

Molecular Phylogenetics and Evolution, *50*(3), 654–660.

Kimball, R. T., Wang, N., Heimer-McGinn, V., Ferguson, C., & Braun, E. L. (2013). Identifying localized biases in large datasets: A case study using the avian tree of life. *Molecular Phylogenetics and Evolution*, *69*(3), 1021–1032.

Kingman, J. F. C. (1982). On the genealogy of large populations. *Journal of Applied Probability*, *19*(A), 27–43.

Küpper, C., Augustin, J., Kosztolányi, A., Burke, T., Figuerola, J., & Székely, T. (2009). Kentish versus Snowy Plover: Phenotypic and Genetic Analyses of *Charadrius alexandrinus* Reveal Divergence of Eurasian and American Subspecies. *The Auk*, *126*(4), 839–852.

Küpper, C., Edwards, S. V., Kosztolányi, A., Alrashidi, M., Burke, T., Herrmann, P., … Székely, T. (2012). High gene flow on a continental scale in the polyandrous Kentish plover *Charadrius alexandrinus*. *Molecular Ecology*, *21*(23), 5864–5879.

Li, C., Ortí, G., Zhang, G., & Lu, G. (2007). A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evolutionary Biology*, *7*(1), 44.

Li, J. W., Yeung, C. K. L., Tsai, P. W., Lin, R. C., Yeh, C. F., Yao, C. T., … Li, S. H. (2010). Rejecting strictly allopatric speciation on a continental island: prolonged postdivergence gene flow between Taiwan (*Leucodioptron taewanus* , Passeriformes Timaliidae) and Chinese (*L. canorum canorum*) hwameis. *Molecular Ecology*, *19*(3), 494–507.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451–1452.

Lim, H. C., & Sheldon, F. H. (2011). Multilocus analysis of the evolutionary dynamics of rainforest bird populations in Southeast Asia: population history of Sundaland birds. *Molecular Ecology*,

20(16), 3414–3438.

Liu, L., Yu, L., & Edwards, S. V. (2010). A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology*, *10*(1), 302.

McCormack, J. E., Harvey, M. G., Faircloth, B. C., Crawford, N. G., Glenn, T. C., & Brumfield, R. T. (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE*, *8*(1), e54848.

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526–538.

Nam, K., Mugal, C., Nabholz, B., Schielzeth, H., Wolf, J. B., BackströM, N., … Ellegren, H. (2010). Molecular evolution of genes in avian genomes. *Genome Biology*, *11*, R68.

Nosenko, T., Schreiber, F., Adamska, M., Adamski, M., Eitel, M., Hammel, J., … Wörheide, G. (2013). Deep metazoan phylogeny: When different genes tell different stories. *Molecular Phylogenetics and Evolution*, *67*(1), 223–233.

Paton, T. A., Baker, A. J., Groth, J. G., & Barrowclough, G. F. (2003). RAG-1 sequences resolve phylogenetic relationships within Charadriiform birds. *Molecular Phylogenetics and Evolution*, *29*(2), 268–278.

Prum, R. O., Berv, J. S., Dornburg, A., Field, D. J., Townsend, J. P., Lemmon, E. M., & Lemmon, A. R. (2015). A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*, *526*(7574), 569–573.

Remm, Maido, Christian, E. S., & Erik, L. S. (2001). Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. *Journal of Molecular Biology*, *314*(5), 1041–1052.

Rheindt, F. E., Székely, T., Edwards, S. V., Lee, P. L. M., Burke, T., Kennerley, P. R., … Küpper, C. (2011). Conflict between genetic and phenotypic differentiation: the evolutionary history of a 'lost and rediscovered' shorebird. *PLoS ONE*, *6*(11), e26995.

Rokas, A., Williams, B. L., King, N., & Carroll, S. B. (2003). Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, *425*(6960), 798.

Rozen, S., & Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. In *Bioinfrormatics methods and protocols* (Vol. 132, pp. 365–386). Totowa, NJ: Humana Press.

Salas-Leiva, D. E., Meerow, A. W., Francisco-Ortega, J., Calonje, M., Griffith, M. P., Stevenson, D. W., & Nakamura, K. (2014). Conserved genetic regions across angiosperms as tools to develop single-copy nuclear markers in gymnosperms: an example using cycads. *Molecular Ecology Resources*, *14*(4), 831–845.

Shaner, P.-J. L., Tsao, T.-H., Lin, R.-C., Liang, W., Yeh, C.-F., Yang, X.-J., … Li, S.-H. (2015). Climate niche differentiation between two passerines despite ongoing gene flow. *Journal of Animal Ecology*, *84*(3), 829–839.

Shen, X. X., Liang, D., Feng, Y. J., Chen, M. Y., & Zhang, P. (2013). A versatile and highly efficient toolkit including 102 nuclear markers for vertebrate phylogenomics, tested by resolving the higher level relationships of the Caudata. *Molecular Biology and Evolution*, *30*(10), 2235–2248.

Smith, B. T., & Klicka, J. (2013). Examining the role of Effective population size on mitochondrial and multilocus divergence time discordance in a songbird. *PLoS ONE*, *8*(2), e55161.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, *15*(5), 199–203.

Swindell, S. R., & Plasterer, T. N. (1997). SEQMAN. *Sequence Data Analysis Guidebook*, 75–89.

Szöllősi, G. J., Tannier, E., Daubin, V., & Boussau, B. (2015). The Inference of Gene Trees with Species Trees. *Systematic Biology*, *64*(1), e42–e62.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*(2), 437–460.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585–595.

Thomson, R. C., Wang, I. J., & Johnson, J. R. (2010). Genome-enabled development of DNA markers for ecology, evolution and conservation. *Molecular Ecology*, *19*(11), 2184–2195.

Thorne, J. L., & Kishino, H. (2002). Divergence time and evolutionary rate estimation with multilocus data. *Systematic Biology*, *51*(5), 689–702.

Tonnabel, J., Olivieri, I., Mignot, A., Rebelo, A., Justy, F., Santoni, S., … Douzery, E. J. P. (2014). Developing nuclear DNA phylogenetic markers in the angiosperm genus Leucadendron (Proteaceae): A next-generation sequencing transcriptomic approach. *Molecular Phylogenetics and Evolution*, *70*, 37–46.

Townsend, T. M., Alegre, R. E., Kelley, S. T., Wiens, J. J., & Reeder, T. W. (2008). Rapid development of multiple nuclear loci for phylogenetic analysis using genomic resources: An example from squamate reptiles. *Molecular Phylogenetics and Evolution*, *47*(1), 129–142.

Wang, B., Xie, X., Liu, S., Wang, X., Pang, H., & Liu, Y. (2017). Development and characterization of novel microsatellite markers for the Common Pheasant (Phasianus colchicus) using RAD-seq. *Avian Research*, *8*, 4.

Wang, N., Liang, B., Wang, J., Yeh, C.-F., Liu, Y., Liu, Y., … Li, S.-H. (2016). Incipient speciation with gene flow on a continental island: Species delimitation of the Hainan Hwamei (Leucodioptron

canorum owstoni, Passeriformes, Aves). *Molecular Phylogenetics and Evolution*, *102*, 62–73.

Wang, P., Liu, Y., Liu, Y., Chang, Y., Wang, N., & Zhang, Z. (2017). The role of niche divergence and geographic arrangement in the speciation of Eared Pheasants (Crossoptilon, Hodgson 1938). *Molecular Phylogenetics and Evolution*, *113*, 1–8.

Waters, J. M., Rowe, D. L., Burridge, C. P., & Wallis, G. P. (2010). Gene Trees versus Species Trees: Reassessing life-History evolution in a freshwater fish radiation. *Systematic Biology*, *59*(5), 504–517.

Weir, J. T., & Schluter, D. (2008). Calibrating the avian molecular clock. *Molecular Ecology*, *17*(10), 2321–2328.

Yang, Z., & Rannala, B. (2010). Bayesian species delimitation using multilocus sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9264–9269.

Yeung, C. K. L., Lin, R.-C., Lei, F., Robson, C., Hung, L. M., Liang, W., … Yang, X. (2011). Beyond a morphological paradox: Complicated phylogenetic relationships of the parrotbills (Paradoxornithidae, Aves). *Molecular Phylogenetics and Evolution*, *61*(1), 192–202.

Zhan, X., Dixon, A., Batbayar, N., Bragin, E., Ayas, Z., Deutschova, L., … Bruford, M. W. (2015). Exonic versus intronic SNPs: contrasting roles in revealing the population genetic differentiation of a widespread bird species. *Heredity*, *114*(1), 1–9.

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, *17*(9), 2107–2121.

Data accessibility

BLAST alignment of 62 universal NPCL markers and Perl script: DRYAD http://dx.doi.org/XXXX.

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.

Supporting Information

Table S1 Summary information for the 136 NPCL candidates.

Table S2 List of all samples used in the universality and polymorphism tests.

Table S3 PCR performance of 136 NPCL markers in 41 species representing 23 avian orders.

Table S4 BLAST results for 64 universal NPCLs and one mitochondrial gene (cytb) in 48 avian genomes.

Table S5 Polymorphic information of 16 NPCLs and two mitochondrial loci (ATPase subunit six concatenated with partial ATPase subunit 8 (ATPase6/8) and NADH dehydrogenase subunit 3 (ND3))

Figure caption:

Figure 1. PCR performance for the 136 NPCL candidates in 23 avian orders. (A) Genetic relationships among our experimental samples. 41 species are highlighted in different colors representing 23 avian orders widely distributed in the avian phylogenetic tree. **(B)** PCR performance for 136 NPCL candidates. Each square represents a PCR result. Success is shown in black and 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 success rate of each avian order is indicated at the bottom of the matrix of 63 NPCL markers.

Figure 2. PCR success rate distribution for the 136 NPCL candidates and mutation rates for the 62 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 candidates. **(B)** Mutation and success rates for the 62 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 RAG-1 (bold) and mitochondrial gene *cyt b* (in blue) were selected as references.

Figure 3. Chromosome mapping of the 63 avian NPCL markers in the genome of Red Junglefowl (*Gallus gallus*). 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. 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. **(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.

Figure 1

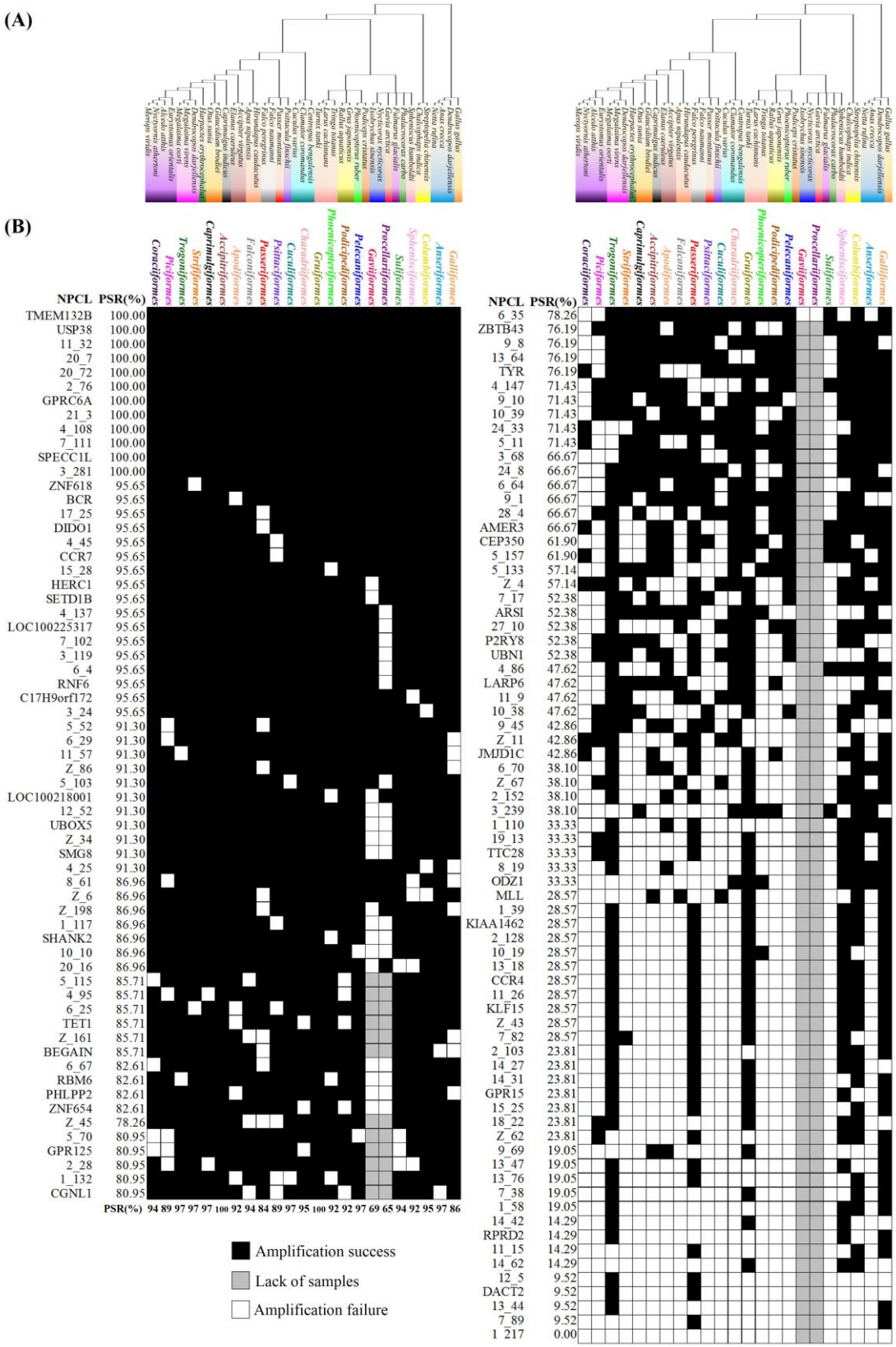
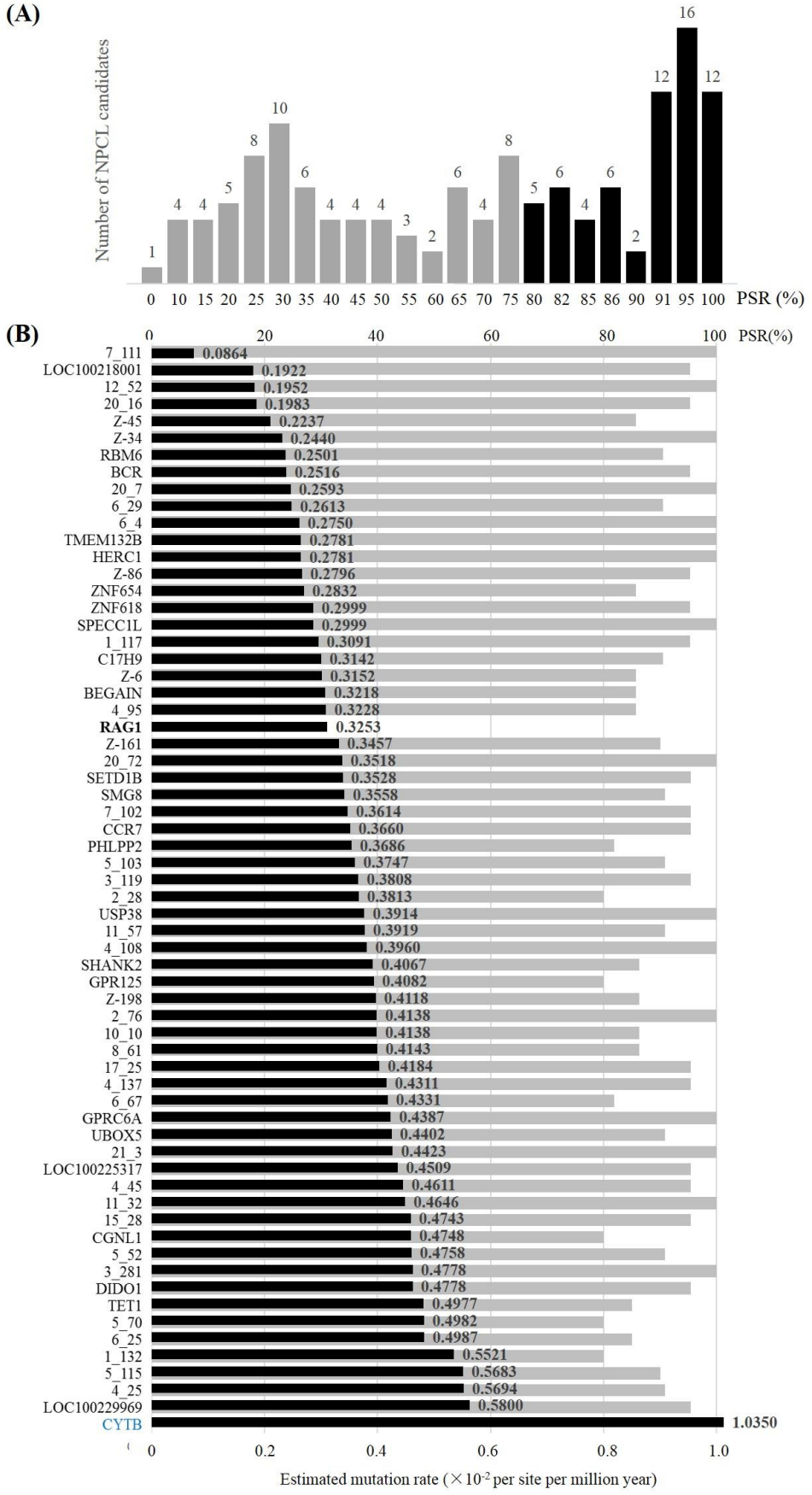

(A)
(B)
NPCL
PSR(%)
Amplification success
Lack of samples
Amplification failure

Figure 2


(A)
Number of NPCL candidates
1
4
4
5
8
10
6
4
4
4
3
2
6
4
8
5
6
4
6
2
12
16
12
0 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 82 85 86 90 91 95 100
PSR (%)
(B)
0 20 40 60 80 100 PSR(%)
7_111 0.0864
LOC100218001 0.1922
12_52 0.1952
20_16 0.1983
Z-45 0.2237
Z-34 0.2440
RBM6 0.2501
BCR 0.2516
20_7 0.2593
6_29 0.2613
6_4 0.2750
TMEM132B 0.2781
HERC1 0.2781
Z-86 0.2796
ZNF654 0.2832
ZNF618 0.2999
SPECC1L 0.2999
1_117 0.3091
C17H9 0.3142
Z-6 0.3152
BEGAIN 0.3218
4_95 0.3228
RAG1 0.3253
Z-161 0.3457
20_72 0.3518
SETD1B 0.3528
SMG8 0.3558
7_102 0.3614
CCR7 0.3660
PHLPP2 0.3686
5_103 0.3747
3_119 0.3808
2_28 0.3813
USP38 0.3914
11_57 0.3919
4_108 0.3960
SHANK2 0.4067
GPR125 0.4082
Z-198 0.4118
2_76 0.4138
10_10 0.4138
8_61 0.4143
17_25 0.4184
4_137 0.4311
6_67 0.4331
GPRC6A 0.4387
UBOX5 0.4402
21_3 0.4423
LOC100225317 0.4509
4_45 0.4611
11_32 0.4646
15_28 0.4743
CGNL1 0.4748
5_52 0.4758
3_281 0.4778
DIDO1 0.4778
TET1 0.4977
5_70 0.4982
6_25 0.4987
1_132 0.5521
5_115 0.5683
4_25 0.5694
LOC100229969 0.5800
CYTB 1.0350
0 0.2 0.4 0.6 0.8 1.0
Estimated mutation rate (×10-2 per site per million year)

582

Figure 3

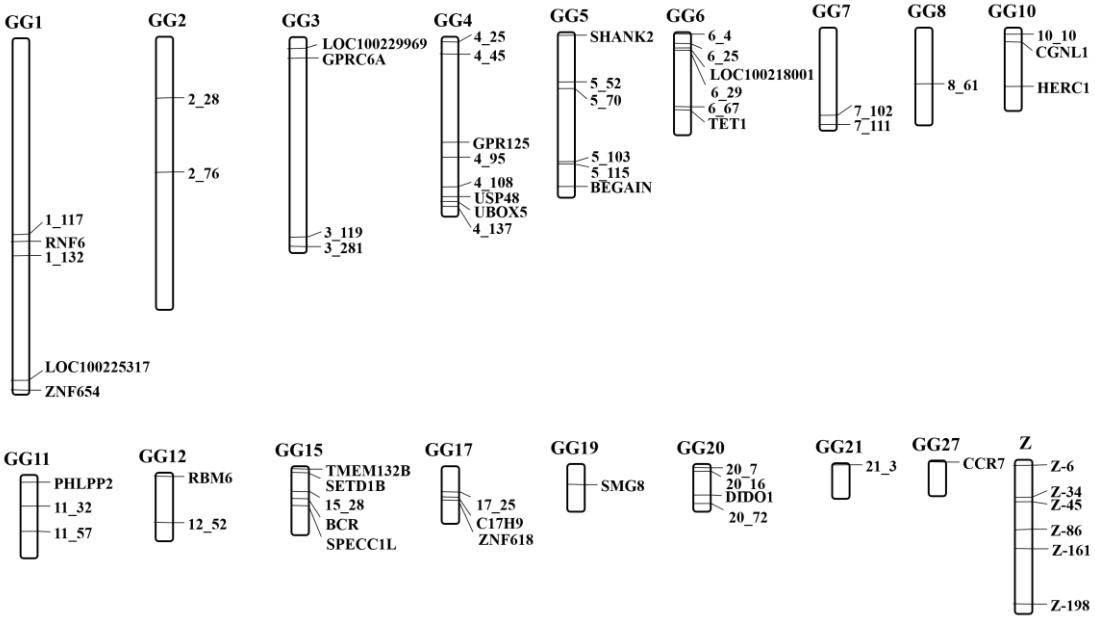
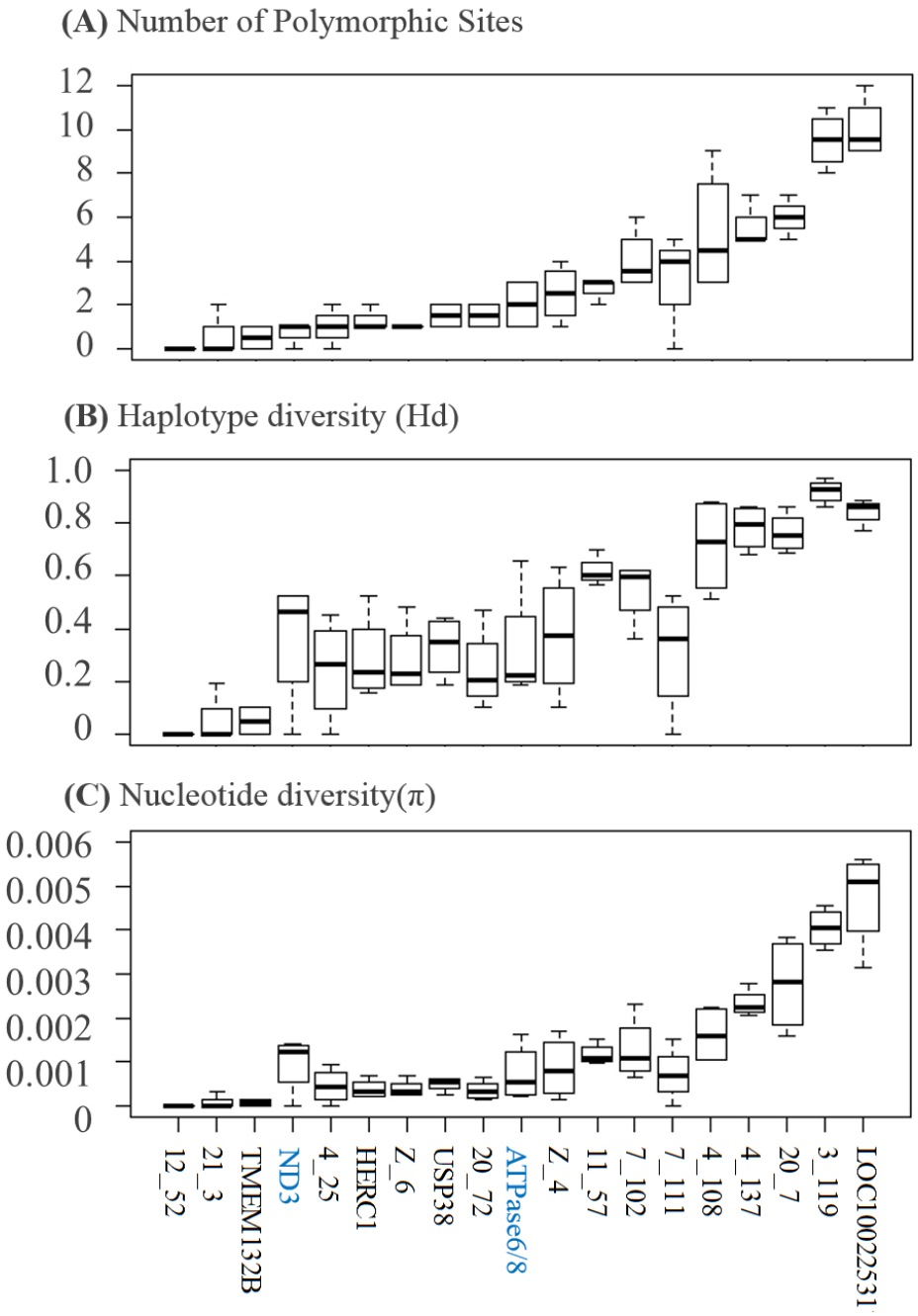

GG1
1_117
RNF6
1_132
LOC100225317
ZNF654
GG2
2_28
2_76
GG3
LOC100229969
GPRC6A
3_119
3_281
GG4
4_25
4_45
GPR125
4_95
4_108
USP48
UBOX5
4_137
GG5
SHANK2
5_52
5_70
5_103
5_115
BEGAIN
GG6
6_4
6_25
LOC100218001
6_29
6_67
TET1
GG7
7_102
7_111
GG8
8_61
GG10
10_10
CGNL1
HERC1
GG11
PHLPP2
11_32
11_57
GG12
RBM6
12_52
GG15
TMEM132B
SETD1B
15_28
BCR
SPECC1L
GG17
17_25
C17H9
ZNF618
GG19
SMG8
GG20
20_7
20_16
DIDO1
20_72
GG21
21_3
GG27
CCR7
Z
Z-6
Z-34
Z-45
Z-86
Z-161
Z-198

Figure 4


(A) Number of Polymorphic Sites
12
10
8
6
4
2
0
(B) Haplotype diversity (Hd)
1.0
0.8
0.6
0.4
0.2
0
(C) Nucleotide diversity(π)
0.006
0.005
0.004
0.003
0.002
0.001
0
12_52
21_3
TMEM132B
ND3
4_25
HERC1
Z_6
USP38
20_72
ATPase6/8
Z_4
11_57
7_102
7_111
4_108
4_137
20_7
3_119
LOC100225 31

588