Mitogenomics clarifies the position of the Nearctic magpies (*Pica hudsonia* and *Pica nuttalli*) within the Holarctic magpie radiation

Alexey P. Kryukov[a,*], Kirill A. Kryukov[b], Kathleen Collier[c], Bohao Fang[d], Scott Edwards[d]

[a]Laboratory of Evolutionary Zoology and Genetics, Federal Scientific Center of the East Asia Terrestrial Biodiversity, Russian Academy of Sciences, Vladivostok 690022, Russia

[b]Center for Genome Informatics, Bioinformation and DDBJ Center, National Institute of Genetics, Mishima 411-8540, Japan

[c]Department of Biology and Wildlife, University of Alaska, Fairbanks 99709, USA

[d]Department of Organismic and Evolutionary Biology and the Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA

*Address correspondence to Alexey Kryukov. E-mail: alex.p.kryukov@gmail.com

Handling editor: Fumin Lei

Abstract

Partial separation of a peripheral population may lead to its divergence and, potentially, speciation due to genetic drift followed by selection and geographic isolation. This process may cause taxonomic uncertainty because reproductive isolation in allopatry cannot be verified directly. The two Nearctic allopatric species of magpies (Aves, Corvidae: *Pica*) serve as a good example of these problems. The Black-billed magpie *Pica hudsonia* is widely distributed in North America, whereas the Yellow-billed Magpie *Pica nuttalli* is endemic to restricted range in California. Their relationships with Palearctic species have been little studied. We obtained complete mitochondrial genomes of both Nearctic magpie species, along with the Eurasian Magpie (*Pica pica*) and the Oriental Magpie (*Pica serica*), 20 mitogenomes in total. Phylogenetic analysis reveals a basal position of *P. serica*, and *P. pica* as a sister clade to the two Nearctic species. *P. hudsonia* and *P. nuttalli* form reciprocal monophyletic subclades, showing recent divergence between and within them. Our data show that the Nearctic magpie lineage diverged from the common ancestor with *P. pica*, with a single migration wave via the Beringia. Within the Nearctic, we hypothesize a peripatric mode of speciation among *Pica* taxa due to divergence and separation of the small marginal population in California below the Sierra-Nevada mountains. Diversifying amino acid substitutions in *ND4-ND5-ND6* genes along the branch leading to the New World clade may indicate selection for heat-tolerance. Considering the clear phenotypic differences between *P. hudsonia* and *P. nuttalli*, our data, showing their reciprocal monophylies and genetic distinctness, is consistent with the two-species taxonomy.

Key words: mitochondrial genome, Corvidae, reciprocal monophyly, Sierra Nevada biogeographic barrier, peripatric speciation

The concept of peripheral isolation as a source of divergence and speciation is an important issue and mechanism in microevolution. In marginal, semi-isolated populations prone to fluctuations in size and bottlenecks, genetic drift, which operates more effectively in small population sizes, may lead to rapid fixation of neutral or even mildly deleterious mutations (Ohta 1992). When this process, followed by divergence in geographic isolation, produces features that serve as prezygotic or postzygotic barriers, reproductive isolation can be retained in the case of subsequent secondary contact with a parental population. This is known as peripatric speciation (Mayr 1963; Coyne and Orr 2004; Edwards et al. 2005). Such newly appeared smaller marginal population may be similar phenotypically and genetically to the parental source, which can cause difficulty in determining taxonomic ranks. Choosing between species and subspecies status can be challenging, as it is generally not feasible to directly check the genetic compatibility of two allopatric populations.

The diversification of the two Nearctic magpie species, *Pica hudsonia* (Sabine, 1823) and *Pica nuttalli* (Audubon, 1837) represents a possible case of peripatric speciation. The former is widely distributed in North America, while the latter occupies a very restricted range in California (Figure 1). No contact zone or hybrids have been reported between them. *P. nuttalli* is morphologically distinct, having a yellow bill (unique among magpies), and smaller body size than *P. hudsonia*, as well as chatter calls and colonial breeding habits different from the *P. hudsonia*. Thus, *P. nuttalli* has been usually considered a distinct species (Goodwin 1986; Birkhead 1991), although sometimes it has been treated as a subspecies of *P. pica* (Zink et al. 1995; Lee et al. 2003). *P. hudsonia* has been considered a subspecies of *P. pica* until relatively recently (American Ornithologists' Union 1957, Vaurie 1959; Howard and Moore 1984; Birkhead 1991; Zink et al. 1995). Most modern authors (Dickinson and Christidis 2014; del Hoyo and Collar 2016) and the latest checklists, including Birds of the World (Trost 2020; Koenig et al. 2022) and the IOC World Bird List (Gill et al. 2022) treat them both as full species. The widely-distributed Holarctic magpie species group also includes 5 Palearctic species: *Pica pica* (Linnaeus, 1758) with 6 subspecies, *Pica serica* Gould, 1845 with 3 subspecies, and the others monotypic: *Pica bottanensis* Delessert, 1840, *Pica asirensis* Bates, 1936 and *Pica mauritanica* Malherbe, 1845 (Madge et al. 2020a, b, c, d; del Hoyo et al. 2020). This latest arrangement is the taxonomy we use throughout this study.

Genetic studies provide important background knowledge to answer such taxonomic questions, but are currently both insufficient and controversial in the case of magpies. In early genetic studies, where DNA hybridization showed paraphyly in the genus *Pica* and close taxonomic linkage between *P. hudsonia* and *P. nuttalli* (Sibley and Ahlquist 1990), their status caused confusion. A high level of differentiation between the Palearctic *P. p. camtschatica* and the Nearctic *P. p. hudsonia* was found by restriction fragment length polymorphisms (RFLP) of mtDNA (*p*-distance of 0.039; Zink et al. 1995). Considered alongside their phenotypic differences, implied species rank, those authors, however, refrained from taxonomic changes, leaving them as subspecies. Although further studies using the same dataset came to the opposite conclusion, meaning *P. pica* and *P. hudsonia* deserve full species rank (Banks et al. 2000). Lee et al. (2003) tried to clarify the relationships between five taxa of *Pica,* including *P. hudsonia* and *P. nuttalli,* using an 813 bp of mitochondrial fragment concatenated from the *16s rRNA*, *tRNA-Leu* and *ND1* genes. They showed a close relationship between *P. p. camtschatica* and *P. p. pica*, which make together a sister group to the closely-related *P. p. hudsonia* and *P. p. nuttalli*, with *P. p. serica* as a more basal lineage. As a result, they preferred to retain both American forms as subspecies. However, using only a small part of the mitogenome can produce unreliable results when comparing closely related lineages (Duchene et al. 2011). Additionally, small sample sizes and limited taxonomic scope compromised this

study – they used only 3 *hudsonia* and 2 *nuttalli* samples, and only five taxa of *Pica* in total. Thus, their analysis lacked broad coverage of the genus *Pica*. Further, there was no correlation found between divergence in the mitochondrial *NADH dehydrogenase 2* (*ND2*) gene and nuclear RFLPs, on the one hand, and phenotypic differences between *P. hudsonia* and *P. pica,* on the other hand (Humphries and Winker 2011). Recently, ultraconserved elements (UCEs) were used for estimating divergence and trans-Beringian gene flow in these species, along with some others (McLaughlin et al. 2020). However, in both analyses, they pooled samples of highly divergent *P. p. camtschatica* and *P. serica* under *P. pica,* which makes their results inconclusive.

More geographically comprehensive work on intraspecific variation and phylogeography in *P. pica sensu lato* was conducted using mitochondrial *cytochrome b* (*CytB*) gene and control region (CR) by Kryukov et al. (2004, 2017, 2022) and Haring et al. (2007, 2012). They revealed a group of haplotypes from the western subspecies (Clade West), which were highly diverged from the East Asian group (Clade East), with a mean *p*-distance of 0.049. *P. p. camtschatica* appeared to be the sister group of Clade West, and *P. hudsonia* was the sister lineage of two main clades, while *P. mauritanica* was found as a basal node in the tree. Song et al. (2018) found sister position and low differentiation between *P. hudsonia* and *P. nuttalli* in their study of two mitochondrial genes, *CytB* and *ND2*, and two nuclear markers, *β-fibrinogen intron 5* (*FIB5*) and *transforming growth factor beta 2 intron 5* (*TGFB2*), but still treated them as full species. Additionally, Song et al. (2018) found basal position of the North African *P. mauritanica* and clear differentiation between *P. asirensis* from Arabia and *P. bottanensis* from Tibet. However, using small sets of sampled species and different non-congruent markers in different previous studies still limits our understanding of the phylogeny of *Pica*.

More recently, the advent of next-generation sequencing (NGS) techniques and the resulting large datasets offers new perspectives for comparative genomics (Prum et al. 2015; Smith et al. 2020; Feng et al. 2020). Whole-genome sequencing and large-scale efforts to sequence non-model organisms provide insights not only in the field of taxonomy, but also phylogeny, phylogeography, speciation and other related fields of evolutionary biology (Edwards et al. 2015; 2021). Specifically, complete mitochondrial genomes have become a low-cost technique for constructing phylogenies with better resolution and for searching for signs of divergence and selection at shallow taxonomic levels (Meiklejohn et al. 2007; Nosil et al. 2009; Duchene et al. 2011; Pacheco et al. 2011; DeSalle et al. 2017). However, complete mitochondrial genomes are still rarely sequenced for birds, compared to a bulk of papers exploring distinct mitochondrial markers. The most recent and the only mitogenomic survey of magpies was conducted by Kryukov et al. (2020). In that paper, the complete mitogenomes of five subspecies were obtained for the first time: *P. p. fennorum*, *P. p. leucoptera*, *P. p. melanotos*, *P. p. camtschatica*, and *P. s. jankowskii*. That study confirmed the distinctness of *P. pica* (which includes *P. p. fennorum, P. p. melanotos, P. p. leucoptera,* and *P. p. camtschatica*) from *P. serica* (consisting of *P. s. serica* and *P. s. jankowskii*).

Complete mitochondrial genomes can also be used for detecting selection. While occupying and settling in America, magpies should have adapted to new local environments, i.e., cold arctic in the North and hot semiarid in the South. The metabolic functions of mitochondria and respiratory activity of a cell are sensitive to temperature. Heat tolerance is a common adaptation known to be controlled by sites in the mitochondrial genome (Lamb et al. 2018; Sokolova 2018;

Baker et al. 2019). In several species structural and regulatory variation in mitochondrial genome were found to be responsible for fitness and adaptation of the host organism to local conditions (Rand 2001; Bazin et al. 2006; Kristensen et al. 2019). A bulk of published data on this matter regarding variety of creatures, including humans (Mishmar et al. 2003; Balloux et al. 2009), led to the hypothesis of "mitochondrial climatic adaptation" (Camus et al. 2017). Therefore, we were interested in searching for such selected sites in magpie mitochondrial genomes.

The aim of this study was to reconstruct the mitochondrial phylogeny of the genus *Pica* with regards to *P. hudsonia* and *P. nuttalli*, a question that has not yet been explored extensively. Thus, we sequenced and compared complete mitochondrial genomes of 20 magpie samples. We sought to shed light on the origin of these New World lineages, presumably derived from an East Asian ancestral species, and estimate their divergence times and time to the most recent common ancestor (tMRCA). We also conducted Bayesian skyline modeling to trace the population demography and searched for any evidence of differential, positive selection within the mitogenomes that might shape the distribution and formation of each lineage. The integration of new mitogenomic data with prior taxonomic studies should help to clarify the phylogenetic position of both New Word lineages and evolutionary history of the entire genus.

Materials and Methods

Sampling

Tissue samples kept in ethanol or high-quality frozen museum tissues were obtained from four museums in USA (Supplementary Table 1). This sampling regime spans nearly the longitudinal extent of *P. hudsonia*'s range, from the central Aleutians at its westernmost to central Colorado at its easternmost (n = 9), as well as the entire range of *P. nuttalli* (n = 5; Supplementary Figure 1). In addition, three samples of *P. p. camtschatica* from the Kamchatka Peninsula and three *P. s. jankowskii* from South Russian Far East were newly studied (Supplementary Table 1). All specimens were associated with vouchered study skins, and Figure 1 shows sampling localities of all publicly available *Pica* mitogenomes used in this study, including those from GenBank. Accession numbers of all 20 de novo assembled sequences are the following: OQ076672–OQ076691 (Supplementary Table 1a). We additionally used all the relevant available data from GenBank (four *P. pica* and two *P. serica*). Notably, one of those *P. serica* samples was listed in GenBank as *P. pica* under ID HQ915867.1, but should be treated as *P. serica serica* according to its known sampling coordinates (Prof. Gang Song, Institute of Zoology, Chinese Academy of Sciences, personal comm.). Also, we used four previously-published corvid mitogenomes (*Garrulus glandarius*, *Nucifraga caryocatactes*, *N. columbiana*, and *Podoces hendersoni*) as outgroup (Supplementary Table 1b).

DNA extraction and sequencing

DNA was extracted from the tissue samples with a DNeasy Blood and Tissue kit (Qiagen, Valencia, California USA), following the manufacturer's instructions. For samples Phud-1 to Phud-4 and Pnut-1 to Pnut-5 (Supplementary Table 1) the procedure was used as explained below. The transposase-mediated fragmentation-based library protocol was applied to mitochondrial DNA (mtDNA), because the fundamental principles of DNA fragmentation, adapter ligation, and sequencing are the same for both nuclear and mitochondrial genomes. We assumed a mitochondrial genome size of 16.6 kb (human mt-genome size) and 0.13% of reads mapping to the mitochondrial genome (data from previous sequencing users in Harvard Bauer Core Facility). Using an Illumina NovaSeq S4 2 x 150 single lane at the same facility, we could achieve 0.5X for a 1.1 GB sized genome and approximately 44X of the mitochondrial genome. Accordingly, we lowered the DNA input to obtain a mitogenome sequencing depth ranging from 10X to 15X per sample.

The remaining samples (Phud-5 to Phud-9 from Alaska, Ppic-1 to Ppic-3 and Pser-1 to Pser-3 from Russia, Supplementary Table 1) were processed as dual-indexed DNA libraries using the methods of Glenn et al. (2017) and calibrated using a Qubit fluorimeter (Invitrogen, Inc. Carlsbad CA, USA). Pooled libraries were sequenced using PE150 sequencing on an Illumina HiSeq 2500 at Facility of University of Alaska Museum (Winker et al. 2019). The quality of sequencing data from both experimental series was assessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Mitogenome assembly

All reads were compared with the reference mitochondrial genome of *P. serica* (GenBank ID HQ915867.1) using a blastn search with BLAST 2.13.0 (Altschul et al. 1997) with parameters "-task blastn -evalue 1e-3 -dbsize 3200000000 -num_alignments 500000". Only pairs where at least one read had a match were used for assembly. Assembly was done using GetOrganelle 1.7.5.3 (Jin et al. 2020), bowtie 2 2.4.5 (Langmead and Salzberg 2012), and SPAdes 3.15.4 (Prjibelski et al. 2020) within the GetOrganelle software package). The following command was used for assembling the mitochondrial genomes: "get_organelle_from_reads.py -F animal_mt --target-genome-size 17000 --overwrite -R 15 -w 35 -k 21,35,45,55 --which-spades SPAdes-3.15.4-Linux/bin -s 0-ref/Pica-pica-HQ915867.1.fasta -1 1-reads/$1-R1.fastq -2 1-reads/$1-R2.fastq -o 2-assembly/$1".

Mitogenome annotation and analysis

Annotation was performed using an original simple script in order to get gene coordinates based on a multiple alignment with several previously annotated *Pica* mitochondrial genomes in GenBank, with accession numbers of HQ915867.1, MT792352.1, MT792353.1, MT792354.1, MT792355.1, and MT792356.1. The script is available at: https://kirill-kryukov.com/study/tools/mitochondrial-annotator/. Sequence alignment was performed using MAFFT 7.505 (Katoh and Standley 2013) with default parameters. AliView 1.28 (Larsson 2014) was used for editing the alignment. GC content and nucleotide composition for protein coding genes (PCGs) were estimated in MEGA-X v.10.1.7 (Kumar et al. 2018).

Population genetics parameters, neutrality tests and phylogenetic analysis

We used DnaSP (Librado and Rozas 2009) for estimating nucleotide and haplotype diversity, tests of neutrality: *D* (Tajima 1989), F_s (Fu 1997) and R_2 (Ramos-Onsins and Rozas 2002) as well as the raggedness index *r* (Harpending 1994). Genetic uncorrected *p*-distances between lineages were estimated for complete mitogenomes and PCG alignments with MEGA-X. The same two kinds of alignments were used for constructing several phylogenetic trees. The maximum likelihood (ML) tree was made with MEGA-X; the general time reversible (GTR)+G+I substitution model was found to be as optimal under the BIC scores, with the methods implemented in MEGA-X. Significance of nodes was supported with 1000 bootstrap replicates.

Time-calibrated Bayesian inference (BI) trees were constructed with Beast v.2.6.7 package which includes BEAUti2 (Bouckaert et al. 2014), using the GTR best-fit substitution model and Yule model as a prior, with Markov chain Monte Carlo (MCMC) length as 10^7, stored every 1,000 generations, and discard the first 10^6 trees form each run as burn-in. For this analysis, the molecular clock model was set as strict, and the evolutionary rate was estimated as 0.01 substitutions/site/million years, as the most used value for *cytochrome b* and other protein-coding mitochondrial genes in birds (Shields and Wilson 1987, Randi 1996; Garcia-Moreno 2004; Lovette 2004; Weir and Schluter 2008). To check if each parameter in the log file reached an effective sample size (ESS) of 200 in Bayesian analysis, we used Tracer v.1.7.1 (Rambaut et al. 2018). The maximum clade credibility tree was obtained with TreeAnnotator v.2.6.2 (Drummond et al. 2012) and a target tree visualized with FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree). tMRCA estimates from PCGs for each lineage were computed in two ways: from BI trees, and with the Coalescent Bayesian Skyline model in Beast and Tracer programs.

We remain cautious about the times of divergence in our study, due to uncertainties in estimations of mitogenomic substitution rates. Applying commonly used estimates of substitution rates to *Pica* is problematic due to the variability of evolutionary rates among sites and lineages (Arbogast et al. 2002), the known poor calibration by bird fossils and the time-dependence of evolutionary rates (Ho and Larson 2006; Harmon et al. 2021). Lerner et al. (2011) estimated substitution rates for each of 18 mitochondrial genes and partitions in honeycreepers (Drepanididae) and found high variability in rates among genes (from 0.014 subst/site/My for *CytB* to 0.029 for *ND2*), making any reasonable estimate for the whole mitogenome or even concatenated protein coding genes problematic. As for calibration by fossils, only a few records in the America are known for *Pica*. One originates from the early Pleistocene of Texas and was attributed to *P. hudsonia* (Miller and Bowman 1956), but later re-identified as the jay *Protocitta ajax* sp. nov. (Brodkorb 1972); the other was from the late Pleistocene of California, identified as *P. nuttalli* (Miller 1929). Because neither clearly represent a species that could be ancestral to one or more *Pica* species, neither were useful for calibration. Additionally, our methods assume that the coalescence times of mitogenomes are synonymous with the speciation times, which is not correct (Edwards and Beerli 2000; Jennings and Edwards 2005; McCormack et al. 2011). The divergence times of course represent coalescence times and not speciation times, which are likely closer to the present (Arbogast et al. 2002).

To trace past population dynamics through time, we conducted Bayesian skyline plot analysis (BSP; Drummond et al. 2012) for all main lineages under study. We analyzed alignments of protein coding genes, totaling 11,403 bp, using

the Beast package, with the same substitution and clock models, substitution rate and MCMC parameters as above. The dimension of population and group sizes was set to 3. For reconstructing the Bayesian skyline plots, we used Tracer. After checking if the values of all statistic parameters had ESS > 200, we obtained plots of population dynamics (i.e., effective female population size per time), which were visualized in Excel 2019.

Signatures of natural selection in protein-coding genes

To reveal any signs of selection on particular branches or sites, we conducted several commonly used analyses based on the ratio of nonsynonymous (*dN*) to synonymous (*dS*) substitutions, or *dN*/*dS* values. In all analyses of selection, we used protein coding genes only with all stop-codons deleted. McDonald-Kreitman (MK) tests were conducted in DnaSP. The test implies neutrality if the ratio of replacement to synonymous changes fixed between species is the same as ratio of nonsynonymous and synonymous polymorphisms within species (McDonald and Kreitman 1991). A prevailing portion of fixed differences between species over the expected within species value would suggest positive selection. Statistical significance was estimated with Fisher's exact test and G-test.

We also used a set of methods implemented in HyPhy (Kosakovsky Pond et al. 2020), accessed via the Datamonkey 2.0 platform (datamonkey.org; Weaver et al. 2018) and choosed only the fixed effects likelihood (FEL) analysis to identify site-specific *dN*/*dS* ratios along the known phylogeny (Kosakovsky Pond and Frost 2005). This test assumes constant selection pressure for each site along the phylogeny and uses a likelihood method to determine the significance of *dN*/*dS* ratios. Another important approach to detecting natural selection are genome scans, which may reveal loci exhibiting increased divergence than entirely neutral loci (Storz 2005; Rogers and Bernatchez 2005; Stinchcombe and Hoekstra 2008; Nosil et al. 2009). To identify genome regions with excessive divergence between lineages, and to estimate pairwise number of nucleotide substitutions per site (D_{xy}, Nei 1978), we used the sliding window method, implemented in DnaSP. Divergence was estimated with Jukes-Cantor method, with 100 bp windows and 25 bp as step size. Patterns of substitutions in each codon position were checked manually.

Results

Mitogenome size and gene content

We obtained complete circular assemblies for our 20 samples, and the assembled genomes ranged from 16,928 to 16,946 bp in length, with coverage ranging between 18X and 217X (Supplementary Table 2). All new mitogenomes contained the complete set of genes corresponding to the "standard" gene order common to avian mitogenomes (Gibb et al. 2007). We recovered 13 protein coding genes, two rRNA genes (*12S* and *16S*), 22 tRNA genes and the control region. The total length of 13 concatenated PCGs was 11,403 bp (~67.3% of the total genomes), and the majority of the PCGs and tRNAs were found on the H-strand, with only *ND6* and eight tRNAs encoded on the L-strand. For the PCGs, the most common start codon was ATG, while GTG used for *COX1*. As a

stop-codons, TAA, AGG or AGA used. GC content in PCGs ranged from 43.9% – 44.1%, and average A, T, C and G content was ~29.3, 26.7, 30.3 and 13.7%, respectively, with no difference between lineages (Supplementary Table 3).

Genetic variation and phylogenetic trees

Nucleotide diversity (π) was estimated as lowest in *P. p. camtschatica*, intermediate in both *P. hudsonia* and *P. nuttalli* and highest in *P. serica*, where standard deviation was 3–4 times higher than in *P. nuttalli* and *P. hudsonia*. Haplotype diversity is similar in all lineages (Table 1). All the trees obtained, regardless of the method of estimation (BI, ML or MP) or dataset (complete mitogenome or PCG), had the exact same topology (Figures 2, 3 and Supplementary Figure 2). The clades of each species and subspecies were consistently reciprocally monophyletic and well-supported, with high Bayesian inference and bootstrap support values. The eastern Palearctic *P. serica* resides at a basal position after a set of outgroup species. It is represented by five samples: four *P. serica jankowskii* from the Russian Far East and one *P. serica serica* from China. The European-Siberian group of subspecies, presented by *P. p. melanotos, P. p. leucoptera, P. p. fennorum* and *P. p. camtschatica*, together form the other well-diverged clade, with *P. p. melanotos* in a basal position. The four samples of *P. p. camtschatica* are very closely related and form a distinct clade within *P. pica*. The sister clade for *P. pica* is represented by the well-supported branch of *P. hudsonia* and *P. nuttalli*. In turn, it is subdivided into two distinct subclades corresponding to these forms. Genetic *p*-distances between *P. hudsonia* and *P. nuttalli* are the lowest among all other comparisons, whereas *P. serica* is the most divergent when compared to other lineages (Supplementary Table 4).

Tests of neutrality, divergence times and population demography

Each neutrality test – Tajima's *D*, Fu's F_s and R_2 – were insignificant in nearly all cases (with one exception), so null-hypothesis of neutrality cannot be rejected. However, the non-insignificance of the raggedness *r*-index may support closeness of observed and expected mismatch distribution for all taxa, and thus suggest population expansion (Table 1).

Estimations of divergence times and tMRCA should be conducted with caution (explained in the Methods), so we can only make preliminary calculations. The time calibrated tree shows that the clade of closely related *P. hudsonia* and *P. nuttalli* diverged from *P. pica* ~2.2 million years ago (Mya), in the late Pliocene, while *P. serica* diverged ~2.8 Mya (Figure 3). Separation between *P. hudsonia* and *P. nuttalli* occurred ~195 kya (Figure 3). tMRCA for the clades of *P. hudsonia* and *P. nuttalli* was estimated by PCGs as 48–66 kya, what is somewhat earlier than for *P. p. camtschatica*, which started ca. 22–37 kya (Table 1). tMRCA within *P. serica* was the longest, ca. 226–250 kya. The same difference in tMRCA is revealed by the skyline plots on Figure 4, where *P. hudsonia* and *P. nuttalli* show similar patterns. Similar signals were found in the *r* indexes, which suggest population growth; *P. nuttalli* seems have experienced N_{ef} growing much faster than in *P. hudsonia* (Figure 4). Almost complete identity of four studied *P. p. camtschatica* mitogenomes did not permit us to make any BSP plot for it.

Evidence of natural selection

In the MK test, all pairwise comparisons of several taxa showed that *dN*/*dS* were always higher within species than between species, as has been found in many mammalian species (Supplementary Table 5; Nachman 1996). We found significant deviation from neutrality only in the case of *P. p. camtschatica* / *P. serica* (neutrality index NI > 1, *p*-value of Fisher's independence test 0.002 and G test *p*-value 0.002), supporting purifying rather than positive selection: the *dN*/*dS* of fixed substitutions was 0.09 as compared to 0.25 *pN*/*pS* for polymorphisms. In four others pairwise comparisons, tests were insignificant and thus could not reject the hypothesis of neutrality. Between *P. nuttalli* and *P. hudsonia*, only one fixed nonsynonymous substitution was revealed, despite four nonsynonymous polymorphic sites within both species. Additionally, one nonsynonymous polymorphic site found within *P. nuttalli* (in *CytB*), and 3 such sites within *P. hudsonia*: in *ATP8*, *COX3* and *ND5*.

In the site-by-site based FEL analysis, in all major internal branches of the trees, a small number of sites were found to be under positive selection, less than the number of sites under purifying selection. We found signs of positive selection at 11 sites along the branch leading to the common ancestor of *P. nuttalli* and *P. hudsonia*, if consider *p*-value threshold ≤ 0.1. Among 11 sites under diversifying selection with different replacements of amino acids, 8 were located in the *ND4*, *ND5* and *ND6* genes, with Isoleucine/Valine being the most common. Eight sites experiencing positive selection were found on the branch leading to *P. serica*, 4 such sites on the branch leading to *P. pica*, and 19 on the branch leading to the outgroup (Figure 3). The number of sites under purifying selection was as 19, 26, 14 and 240, respectively, for those same branches. Between *P. nuttalli* and *P. hudsonia*, only one site of diversifying selection was revealed. The other tests from HyPhy set, namely aBSREL, MEME and BUSTED, did not show any clear result.

In addition, we studied variation and divergence between lineages across their mitogenomes. The sliding window method revealed seven regions of extended divergence between *P. nuttalli* and *P. hudsonia*, corresponding to the genes for *ND1*, *COX1*, *ATP6*, *ND4*, *ND5*, *CytB* and *ND6* (Figure 5). The highest levels of divergence were found in the genes for *ND1*, *ND4*, *ND5* and *ND6*, being maximal at the *ND6*. The only nonsynonymous fixed substitution between *nuttalli* and *hudsonia* was revealed within the *ND4* gene, with 27 fixed synonymous substitutions across all PCGs (almost all transitions at the 3rd position of codons). The triplet ATC at site #2208 was found in all lineages except *hudsonia*, where it changed to GTC, replacing Isoleucine in *P. nuttalli* and others to Valine in *P. hudsonia*. At each of the four *ND* genes mentioned, from 3 to 7 mutations were found to be fixed. Among 13 ML trees, constructed for each protein coding genes, only four trees, those made by the same four *ND* genes, upheld the *hudsonia* and *nuttalli* clades at a significant level (with 97–100% bootstrap support). Comparing PCG alignments of *hudsonia* and *camtschatica*, we found no such peaks of divergence, where mutations were distributed rather uniformly, though with the highest peak also for *ND6* (Figure 5). Almost the same pattern was obtained for the comparison of *P. hudsonia* and *P. p. leucoptera* (not shown). Pairwise analysis of nucleotide divergence per site between lineages (D_{xy}), conducted by the same method

(Supplementary Table 6), confirms the results obtained by *p*-distances (Supplementary Table 4), with close affinity of *P. hudsonia* and *P. nuttalli*, as well as between *P. p. camtschatica* and *P. p. leucoptera*.

Discussion

Phylogenetic reconstructions

In this paper, we revealed the genetic relationships of the two New World magpie species by means of constructing mitochondrial phylogeny of the genus *Pica*. Using complete mitochondrial genomes of most Holarctic magpie lineages, our analysis shows a clear and unambiguous pattern, clarifying their position in the genus (Figures 2 and 3). Our phylogeny suggests early separation of the East Palearctic species *P. serica* from all others, later divergence for ancestor of *P. pica* and common ancestor of the two American species which, in turn, separated very recently. Notably, the high divergence of both Nearctic species from *P. pica*, evident from our phylogenetic reconstructions, indicates that they certainly do not belong to the Eurasian species *P. pica* as was suggested before (Zink et al. 1995; Lee et al. 2003). Despite the geographic proximity, *P. p. camtschatica* appears as not closely related to the Nearctic magpies, but instead belongs to the clade of the widely distributed Eurasian *P. pica*. We confirm the overall phylogenetic structure of Lee et al. (2003) both generally and with specific regard to the Nearctic *Pica*, and extend its scope. Our study presents the most comprehensive molecular phylogeny of the *Pica* genus available to date.

Our results show that both *P. nuttalli* and *P. hudsonia* form reciprocally monophyletic sister clades, with low divergence both between and within their mitochondrial lineages and with similar tMRCA within their respective mitoclades (Figure 3 and Table 1). Surprisingly, the mitochondrial genomes of *P. hudsonia* and *P. nuttalli* show a fairly low degree of divergence, even lower than between some subspecies of *P. pica* (Figures 2 and 3), high nucleotide diversity in which is likely a consequence of multiple diverged lineages. Our ML and BI trees do not separate the Northern (Alaska) and Southern (California and New Mexico) *P. hudsonia* populations (Figures 2 and 3), however a slight inter-population differentiation can be seen in the MP tree (Supplementary Figure 2). Our results provide an important clarification of the relationship of Nearctic *Pica* species to Palearctic ones, and to each other.

Divergence times and colonization of North America

Our data allow quantification of the degree of divergence between magpie species and subspecies (Figures 3 and Supplementary Figure 2, Supplementary Tables 4 and 6). We estimated that the divergence of Nearctic magpies from the Eurasian common ancestor occurred in the late Pliocene (~2.2 Mya in our data, Figure 3). According to branching off at the mitochondrial trees (Figures 2 and 3), the ancestor of Nearctic magpies most probably migrated to North America from Eurasia via the Beringia land bridge. This bridge appeared somewhere from nine to twenty or more times between northeastern Asia and North America between 2.6 Mya and 10 kya (Hopkins 1967; McLaughlin et al. 2020). The

American clade is not clustered together with the Kamtchatka clade in any of our phylogenetic trees, therefore the Kamchatka clade could not have given rise to the American magpies. The current Kamchatka population experienced a bottleneck, inferred from a low nucleotide variation and a short tMRCA (Table 1), and an excess of nonsynonymous polymorphisms (Supplementary Table 5). The short and similar tMRCA of both *P. nuttalli* and *P. hudsonia* (Figure 4 and Table 1) may suggest that they also experienced bottlenecks during climatic oscillations in Quaternary. The more rapid increase of effective population size in *P. nuttalli* than in *P. hudsonia* in the Bayesian skyline plot (Figure 4) might be explained by fluctuations in lineages with smaller effective population size and occupying smaller range.

Voous (1960) was the first to hypothesize that magpies colonized North America in two waves. According to him, the newcomer population became initially widely distributed, but later restricted by ice sheet to the south to produce the *P. nuttalli* lineage, and after the ice sheet retreated, the second wave entered from Asia and gave rise to the *P. hudsonia* lineage. However, our phylogenetic trees show the two Nearctic magpie species in a single monophyletic clade (Figures 2 and 3 and Supplementary Figure 2), also shown in MP and NJ trees in Lee et al. (2003). Furthermore, each of these two species forms its own monophyletic clade, and the two lineages appear relatively recently diverged from a common ancestor. This evidence supports only one wave of colonization as the most likely scenario, also claimed in Lee et al. (2003).

Diversification and speciation of magpies in America

The establishment of the magpie ancestor in America conforms to the common scenario of allopatric speciation. The dispersion of America by magpies might have been linked with grazing ungulates (caribou, bisons and others, see, e.g., Londei 2018 for analogous association in Europe) in grasslands with trees and forested river valleys. The pattern of further magpie diversification in America is most easily explained by a model of peripatric speciation via isolated peripheral population (Mayr 1963; Frey 1993; Coyne and Orr 2004; Edwards et al. 2005). Peripatric speciation implies asymmetry in range sizes between both clades involved (Barraclough and Vogler 2000) and dispersal but not vicariance (Heads 2004). This model has been suggested for a variety of avian groups: Sulidae (Friesen and Anderson 1997), Alcidae (Jones 1999), *Empidonax* flycatchers (Johnson and Cicero 2002), Tetraoninae (Drovetski 2003), for mammals: the Prairie dog *Cynomys mexicanus* (Castellanos-Morales et al. 2016), and other groups and species. An alternative centrifugal speciation model implies periods of range expansion and contraction, but predicts an apomorphic state and potentially speciation at a central population instead of marginal one (Brown 1957; Frey 1993), which is not our case. Other scenario might imply speciation with gene flow; however, this seems less viable because magpies are sedentary birds without seasonal migrations and avoid long-distance dispersal.

Examining the two American magpie ranges, one may suppose some role of the Sierra Nevada Mountain range in their isolation. The California Floristic Province, including the Sierra Nevada, is an important biodiversity hotspot which contains biogeographic boundaries, glacial refugia and multiple endemic species (Macey et al. 2001; Calsbeek et al. 2003; Rubinoff et al. 2021). The Sierra Nevada range has served as a powerful biogeographic barrier for a number of different species in the region, which is obviously applicable to a peripatric endemic and its more widespread counterpart at the adjacent regions. Distinct lineages, isolated by the Sierra Nevada, have been documented for multiple taxa,

including birds (*Baeolophus* titmice, Cicero 2004; owl *Strix nebulosa yosemitensis*, Hull et al. 2010), amphibians (California tiger salamander *Ambystoma californiense*, Shaffer et al. 2004; California newt *Taricha torosa sierrae*, Kuchta and Tan 2006; salamander *Hydromantes brunus*, Rovito 2010), and others. Analogously, the Sierra Nevada would have served as isolation barrier for ancestral magpie populations.

Ancestral magpie population might have been distributed from Alaska until coastal California, alongside the Sierra Nevada, and occupied both sides of the mountain range. This range already existed at that time, because our inferred timing of divergence of *P. hudsonia* and *P. nuttalli* (~0.2 Mya, Figure 3) is much later than the Late Cenozoic uplift of the Sierra Nevadas (3–5 Mya, Christensen 1966; Unruh 1991). During the Pleistocene glacial oscillations, this initial magpie population would have experienced range expansions and retractions, as did many other American bird species (Klicka and Zink 1999; Weir and Shluter 2004). Magpie gene flow along the sea shore might have slowed down, which resulted in semi-isolation of the southern marginal population, and finally in complete isolation from the parental northern population, in addition to the existing isolation by the Sierra Nevada from the east. That diverging marginal population would have experienced genetic drift, resulting in random fixation of numerous alleles and development of the phenotypic differentiation currently observed in *P. nuttalli* (e.g., yellow bill and skin below the eye), which are the most apparent differences between the species. Both species are also distinguishable by whine-call and slightly higher-pitched chatter call (which are however more similar between these species than with *P. pica*) (Enggist-Dublin and Birkhead 1992; Kryukov et al. 2017). These differences, particularly those influencing conspecific recognition or breeding behavior, likely contributed as a premating barrier to reproductive isolation, and possibly facilitated peripatric speciation.

Inferring adaptation in North American magpies

In this study, we found an increased number of sites under positive selection along the common branch leading to *P. hudsonia* and *P. nuttalli*: there are 11 sites, what is much more than number of diversifying sites along the branches to *P. pica* and *P. serica* in the FEL analysis (Figure 3). The corresponding replacements were located mainly in the *ND4-ND5-ND6* genes, and the most common was Isoleucine/Valine. The single non-synonymous substitution separating *P. hudsonia* and *P. nuttalli* is in the *ND4* gene and involves the same two amino acids, the finding supported by the MK test (Supplementary Table 5). This gene, alongside synonymous mutations in three *ND* genes and three other mitochondrial coding genes, show corresponding clear peaks of variation between *P. hudsonia* and *P. nuttalli* in our sliding-window analysis (Figure 5). Most of these selected sites are located in mitochondrial genes for the complex I of the oxidative phosphorylation (OXPHOS) pathway.

Natural selection in the course of local adaptation to semi-arid environments might have occurred during the establishment of the southern marginal population. Genetic drift was likely responsible for the origin of several apomorphic phenotypic traits in it, but even without drift, selection might explain high divergence of a marginal population when gene flow was interrupted (Garcia-Ramos and Kirkpatrick 1997; Räsänen and Hendry 2008). We

assume the two species experienced divergent selection in terms of energy usage and oxygen metabolism. Such traits as decreased body size in *P. nuttalli* (in accordance with Bergman's rule), its other morphological features, colonial breeding etc. are most likely controlled by nuclear genes and thus are beyond the scope of this paper. However, other traits, such as heat tolerance, might be influenced by mitochondrial genome variation in the complex I of OXPHOS metabolic pathway, discovered in this study.

The OXPHOS pathway contains 5 metabolic complexes of mitochondrially-encoded polypeptides (Blier et al. 2001). Genes responsible for selection for heat/cold tolerance belong mainly to ND1-ND6 peptides form the first, NADH–ubiquinone oxidoreductase complex. Here are a few examples out of many such cases: Two amino-acid replacements in the OXPHOS genes, i.e., Isoleucine/Valine and Histidine/Aspartic Acid, were discovered to relate to environmental gradients in South African arid birds Karoo scrub-robin (Ribeiro et al. 2011). Clinal variation in the frequency of *ND3* gene haplotypes was correlated with elevational gradient in the Rufous-collared sparrow *Zonotrichia capensis* (Cheviron and Brumfield 2009). Several examples of involving *ND* genes in adaptations for hypoxia at high altitudes were reported for mammals of Tibet (Luo et al. 2013). While searching for positive selection at mitogenomes among 237 species and higher animal taxa, Gavrin et al. (2015) found complex I of the OXPHOS (mostly subunit *ND5*) as a target for proteins involving in a positive selection in at least 17 cases. These reports support the possibility that diversifying selection in the complex I of OXPHOS metabolic pathway might have taken place in genetic adaptations to cold/heat-tolerance in the evolution of New World *Pica* lineages.

Taxonomic implications

The evaluation of the taxonomic ranks of both American magpie lineages depends on the choice of the species concept. The two taxa would be considered full biological species under the phylogenetic and evolutionary species concepts, because they are readily distinguishable by phenotypes and underwent long enough independent evolution. However, the key criterion of reproductive isolation under the biological species concept, the failure to interbreed, cannot be checked directly due to absence of any physical contact reported between them. When taking genetic distances into account, it appears that divergence between *P. nuttalli* and *P. hudsonia* (*p*-distance ~0.3% in our estimates) is much smaller compared to divergence level between *P. serica* and *P. pica*, and coalescence times are less than that of subspecies within *P. pica* (Figures 2, 3 and Supplementary Figure 2, Tables 1 and Supplementary Tables 4 and 6). By these criteria, *nuttalli* and *hudsonia* would retain subspecies status within *P. hudsonia* (according to the priority rule of the Codex). On the other hand, *nuttalli* and *hudsonia* lineages show 1) geographic unities; 2) reciprocal monophyly and coalescent-based species delimitation (see, e.g., Rannala and Yang 2003; Fujita et al. 2012 for discussion of these criteria for discrimination a species); and 3) clear phenotype differences – three criteria of a species rank, suggested by Hosner et al. (2018). In species delimitation practice, genetic and genomic data should be used together with phenotypic and any other traits in context of integrative taxonomy (Dayrat 2005; Padial et al. 2010; Sangster 2018). Normally we would try to resolve this uncertainty (species vs. subspecies) by assessing the status of allopatric forms, comparing their divergence level with that evaluated for related species coexisting in sympatry, by using phenotypes, genetics, vocalization, or other

traits (Helbig et al. 2002, Donegan 2018). However, magpies do not provide any cases of sympatry, they only provide situations of allopatry, leaving several taxonomic decisions uncertain (for a general review of this problem, see, e.g., Mayr 1969; Winker 2021). Therefore, based on all available evidence, we suggest to continue treating *P. hudsonia* and *P. nuttalli* as two distinct species.

In conclusion, our in-depth mitochondrial phylogeny reveals the Nearctic magpie clade as the sister group of the Palearctic nominate species *P. pica*, with divergence time estimated as late Pliocene. *P. hudsonia* and *P. nuttalli* form reciprocal monophyletic clades in all our phylogenetic trees, showing low divergence between and within these lineages. This finding, together with asymmetry in allopatric distribution of both American lineages, and their distinct phenotypic and ecological traits, support the hypothesis of peripatric speciation. The current study is the most comprehensive investigation of the phylogeny of the *Pica* genus done to date. In the future, more thorough analysis can be done using the complete nuclear genome data. Such analysis will help to further clarify the relationship between northern and southern populations of *P. hudsonia*, and assess of a gene flow between them, as well as such possibility between *P. hudsonia* and *P. nuttalli*. Also, future studies may incorporate the remaining *Pica* species: *P. mauritanica*, *P. asirensis* and *P. bottanensis*. However, strong statistical support of branching patterns in our trees suggests that such addition will not change our reconstructed relationship between species studied in this work. The North American magpies present an interesting model for more detailed investigation of peripatric speciation, selection, and phylogeography.

Acknowledgments

We are grateful to several museums and individuals who supplied tissues for this project: The Museum of Vertebrate Zoology, University of California, Berkeley (Carla Cicero and Rauri Bowie), Denver Museum of Natural Science (Garth Spellman and Andrew Doll), University of Alaska Museum (Kevin Winker and Jack Withrow), and University of Washington Burke Museum. We thank Yaroslav Red'kin for providing the corrected map of magpie ranges. We are much thankful for anonymous reviewers for their valuable comments and suggestions.

Funding

This work was supported by Harvard University and a postdoctoral fellowship to BF from the Finnish Foundation. BF was also supported by a grant from the Harvard Global Institute for research on biodiversity of China. The research was carried out within the state assignment of Ministry of Science and Higher Education of the Russian Federation (theme No. 121031500274-4).

Authors Contribution

SE conceived the experiment; AK organized the project, made calculations and wrote the manuscript; KC and BF conducted sequencing; KK assembled and annotated the mitogenomes, and constructed multiple alignments; AK and KK constructed phylogenetic trees; SE, KK, KC and BF edited the manuscript; all authors have read and approved the final manuscript.

Competing Interests Statement

There is no conflict of interests to declare.

Data depository

The mitochondrial DNA sequences have been deposited in the National Centre for Biotechnology Information (NCBI); the accession numbers are OQ076672–OQ076691. The WGS reads are available on request.

ORCID

Alexey P. Kryukov https://orcid.org/0000-0001-7010-7338

Kirill A. Kryukov https://orcid.org/0000-0002-0286-0288

Kathleen Collier https://orcid.org/0000-0002-4057-4282

Bohao Fang https://orcid.org/0000-0001-5283-067X

Scott Edwards https://orcid.org/0000-0003-2535-6217

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z et al., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* 25:3389–3402.

Arbogast BS, Edwards SV, Wakeley J, Beerli P, Slowinski JB, 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annu Rev Ecol Syst* 33:707–740.

Baker EP, Peris D, Moriarty RV, Li XC, Fay JC et al., 2019. Mitochondrial DNA and temperature tolerance in lager yeasts. *Sci Adv* 5:eaav1869.

Balloux F, Handley LJL, Jombart T, Liu H, Manica A, 2009. Climate shaped the worldwide distribution of human mitochondrial DNA sequence variation. *Proc R Soc Biol Sci B* 276:3447–3455.

Banks RC, Cicero C, Dunn JL, Kratter AW, Ouellet H et al., 2000. Forty-second supplement to the American Ornithologists' Union check-list of North American birds. *Auk* 117:847–858.

Barraclough TG, Vogler AP, 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155:419–434.

Bazin E, Glémin S, Galtier N, 2006. Population size does not influence mitochondrial genetic diversity in animals. *Science* 312(5773):570–572.

Birkhead TR, 1991. *The Magpies: The Ecology and Behaviour of the Black-billed and Yellow-billed Magpies.* London: T. & A.D. Poyser.

Blier PU, Dufresne F, Burton RS, 2001. Natural selection and the evolution of mtDNA-encoded peptides: Evidence for intergenomic co-adaptation. *Trends Genet* 17: 400–406.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu C-H et al., 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537.

Brodkorb P, 1972. Neogene fossil jays from the Great Plains. *Condor* 74:347–349.

Brown JWL, 1957. Centrifugal speciation. *Quart Rev Biol* 32(3):247–277.

Calsbeek R, Thompson JN, Richardson JE, 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: The California Floristic Province. *Mol Ecol* 12:1021–1029.

Camus MF, Wolff JN, Sgrò CM, Dowling DK, 2017. Experimental support that natural selection has shaped the latitudinal distribution of mitochondrial haplotypes in Australian *Drosophila melanogaster*. *Mol Biol Evol* 34:2600–2612.

Castellanos-Morales G, Gámez N, Castillo-Gámez RA, Eguiarte LE, 2016. Peripatric speciation of an endemic species driven by Pleistocene climate change: The case of the Mexican prairie dog *Cynomys mexicanus*. *Mol Phylogenet Evol* 94:171–181.

American Ornithologists' Union, 1957. *Check-list of North American Birds*, 5th edn. Washington, D.C.: American Ornithologists' Union.

Cheviron ZA, Brumfield RT, 2009. Migration - selection balance and local adaptation of mitochondrial haplotypes in rufous-collared sparrows *Zonotrichia capensis* along an elevational gradient. *Evolution* 63(6):1593–1605.

Christensen MN, 1966. Late Cenozoic crustal movements in Sierra Nevada of California. *Geol Soc Amer Bull* 77:163–181.

Cicero C, 2004. Barriers to sympatry between avian sibling species (Paridae: *Baeolophus*) in local secondary contact. *Evolution* 58(7):1573–1587.

Coyne JA, Orr HA, 2004. *Speciation*. Massachusetts: Sinauer Associates Inc.

Dayrat B, 2005. Towards integrative taxonomy. *Biol. J. Linnean Soc. 85*(3):407-417.

del Hoyo J, Collar N, Christie DA, 2020. Maghreb Magpie (*Pica mauritanica*), version 1.0. In: Hoyo J del, Elliott A, Sargatal J, Christie DA, Juana E de eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

del Hoyo J, Collar NJ, 2016. *HBW and BirdLife International Illustrated Checklist of the Birds of the World.* Vol. 2: Passerines. Barcelona: Lynx.

DeSalle R, Schierwater B, Hadrys H, 2017. MtDNA: The small workhorse of evolutionary studies. *Front Bioscience-Landmark* 22(5):873–887.

Dickinson EC, Christidis L, 2014. *The Howard & Moore Complete Checklist of the Birds of the World*. 4th edn, Vol. 2. Eastbourne: Aves Press.

Donegan TM, 2018. What is a species? A new universal method to measure differentiation and assess the taxonomic rank of allopatric populations, using continuous variables. *ZooKeys* 757:1–67.

Drovetski SI, 2003. Plio-Pleistocene climatic oscillations, Holarctic biogeography and speciation in an avian subfamily. *J Biogeogr* 30:1173–1181.

Drummond AJ, Suchard MA, Xie D, Rambaut A, 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973.

Duchene S, Archer FI, Vilstrup J, Caballero S, Morin PA, 2011. Mitogenome phylogenetics: The impact of using single regions and partitioning schemes on topology, substitution rate and divergence time estimation. *PLoS ONE* 6(11): e27138.

Edwards S, Kingan SB, Calkins JD, Balakrishnan CN, Jennings WB et al., 2005. Speciation in birds: Genes, geography, and sexual selection. *Proc Natl Acad Sci* 102:6550–6557.

Edwards S, Schultz A, Campbell-Staton S, 2015. Next-generation sequencing and the expanding domain of phylogeography. *Folia Zool* 64:187–206.

Edwards SV, Beerli P, 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.

Edwards SV, Robin VV, Ferrand N, Moritz C, 2021. The evolution of comparative phylogeography: Putting the geography (and more) into comparative population genomics. *Genome Biol Evol* 14:evab176.

Enggist-Düblin P, Birkhead TR, 1992. Differences in the calls of European and North American blackbilled magpies and the yellow-billed magpie. *Bioacoustics* 4:185–194.

Feng S, Stiller J, Deng Y et al., 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature* 587:252–257.

Frey JK, 1993. Modes of peripheral isolate formation and speciation. *Syst Biol* 42:373–381.

Friesen VL, Anderson DJ, 1997. Phylogeny and evolution of the Sulidae (Aves: Pelecaniformes): A test of alternative modes of speciation. *Mol Phylogenet Evol* 7:252–260.

Fu Y-X, 1997. Statistical test of neutrality against population growth, hitchhiking and background selection. *Genetics* 147:915–925.

Fujita MK, Leachen AD, Burbrink FT, McGuire JA, Moritz C, 2012. Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol Evol* 27:480–488.

Garcia-Moreno J, 2004. Is there a universal mtDNA clock for birds? *J Avian Biol* 35:465–468.

Garcia-Ramos G, Kirkpatrick M, 1997. Genetic models of adaptation and gene flow in peripheral populations. *Evolution* 51(1):21–28.

Gavrin MR, Bielawski JP, Sazanov LA, Gharrett AJ, 2015. Review and meta-analysis of natural selection in mitochondrial complex I in metazoans. *J Zool Syst Evol Res* 53(1):1–17.

Gibb GC, Krdailsky O, Kimball RT, Braun EL, Penny D, 2007. Mitochondrial genomes and avian phylogeny: Complex characters and resolvability without explosive radiations. *Mol Biol Evol* 24:269–280.

Gill F, Donsker D, Rasmussen P, 2022. *IOC World Bird List* (v 12.1). Available at https://www.worldbirdnames.org/

Glenn TC, Bayona-Vásquez NJ, Kieran TJ, Pierson TW, Hoffberg SL et al., 2017. Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1 USD per Sample (3RAD). *BioRxiv* 20579.

Goodwin D, 1986. *Crows of the world.* Seattle: University of Washington Press.

Haring E, Däubl B, Pinsker W, Kryukov A, Gamauf A, 2012. Genetic divergences and intraspecific variation in corvids of the genus *Corvus* (Aves: Passeriformes: Corvidae): A first survey based on museum specimens. *J Zool Syst Evol Res* 50:230–246.

Haring E, Gamauf A, Kryukov A, 2007. Phylogeographic patterns in widespread corvid birds. *Mol Phylogenet Evol* 45:840–862.

Harmon LJ, Pennell MW, Henao-Diaz LF, Rolland J, Sipley BN et al., 2021. Causes and consequences of apparent timescaling across all estimated evolutionary rates. *Ann Rev Ecol Evol Syst* 52:587–609.

Harpending HC, 1994. Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biol* 66:591–600.

Heads M, 2004. What is a node? *J Biogeogr* 31:1883–1891.

Helbig AJ, Knox AG, Parkin DT, Sangster G, Collinson M, 2002. Guidelines for assigning species rank. *Ibis* 144:518–525.

Ho SYW, Larson G, 2006. Molecular clocks: When times are a-changin. *Trends Genet* 22:79–83.

Hopkins DM, 1967. *The Bering Land Bridge.* Palo Alto: Stanford Univ Press.

Hosner PA, Campillo LC, Andersen MJ, Sanchez-Gonzalez LA, Oliveros CH et al., 2018. An integrative species delimitation approach reveals fine-scale endemism and substantial unrecognized avian diversity in the Philippine Archipelago. *Conserv Genet* 19:1153–1168.

Howard R, Moore A, 1984. *A complete Checklist of the Birds of the World.* London: Macmillan.

Hull JM, Keane JJ, Savage WK, Godwin SA, Shafer JA et al., 2010. Range-wide genetic differentiation among North American great gray owls *Strix nebulosa* reveals a distinct lineage restricted to the Sierra Nevada, California. *Mol Phylogenet Evol* 56:212–221.

Humphries EM, Winker K, 2011. Discord reigns among nuclear, mitochondrial, and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. *Mol Ecol* 20:573–583.

Jennings WB, Edwards SV, 2005. Speciational history of Australian grass finches (*Poephila*) inferred from 30 gene trees. *Evolution* 59:2033–2047.

Jin JJ, Yu WB, Yang JB, Song Y, dePamphilis CW et al., 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 21:241.

Johnson NK, Cicero C, 2002. The role of ecologic diversification in sibling speciation of *Empidonax* flycatchers (Tyrannidae): Multigene evidence from mtDNA. *Mol Ecol* 11:2065–2081.

Jones IL, 1999. Assessing the role of sexual selection in adaptive radiation of the auklets (Alcidae, Aethiini). In: Adams NJ, Slotow RH eds. *Proc. 22 Int. Ornithol. Congr., Durban.* Johannesburg: BirdLife South Africa, 1115-1125.

Katoh K, Standley DM, 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772–780.

Klicka RM, Zink RM, 1999. Pleistocene effects on North American songbird evolution. *Proc R Soc Lond B* 266:695–700.

Koenig WD, Reynolds MD, Airola DA, 2022. Yellow-billed Magpie (*Pica nuttalli*), version 2.0. In: Kirwan GM, Keeney BK eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Kosakovsky Pond SL, Frost SDW, 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22:1208–1222.

Kosakovsky Pond SL, Poon AFY, Velazquez R, Weaver S, Hepler NL et al., 2020. HyPhy 2.5: A customizable platform for evolutionary hypothesis testing using phylogenies. *Mol Biol Evol* 37:295–299.

Kristensen TN, Loeschcke V, Tan Q, Pertoldi C, Mengel-From J, 2019. Sex and age specific reduction in stress resistance and mitochondrial DNA copy number in *Drosophila melanogaster*. *Sci Rep* 9:12305.

Kryukov A, Iwasa MA, Kakizawa R, Suzuki H, Pinsker W et al., 2004. Synchronic east-west divergence in azure-winged magpies *Cyanopica cyanus* and magpies *Pica pica*. *J Zool Syst Evol Res* 42:342–351.

Kryukov AP, Goroshko OA, Arkhipov VY, Red'kin YA, Lee S et al., 2022. Introgression at the emerging secondary contact zone of magpie *Pica pica* subspecies (Aves: Corvidae): Integrating data on nuclear and mitochondrial markers, vocalizations and field observations. *Org Divers Evol* 22:1037–1064.

Kryukov AP, Spiridonova LN, Mori S, Arkhipov VYu, Red'kin YA et al., 2017. Deep phylogeographic breaks in magpie *Pica pica* across the Holarctic: Concordance with bioacoustics and phenotypes. *Zool Sci* 34:185–200.

Kryukov AP, Spiridonova LN, Tyunin AP, Kryukov KA, Dorda BA, 2020. Complete mitochondrial genomes of five subspecies of the Eurasian magpie *Pica pica*, obtained with Oxford Nanopore MinION, and their interpretation regarding intraspecific taxonomy. *Mitochondrial DNA B* 5:3792–3793.

Kuchta SR, Tan A-M, 2006. Lineage diversification on an evolving landscape: Phylogeography of the California newt *Taricha torosa* (Caudata: Salamandridae). *Biol J Linn Soc* 89:213–239.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K, 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549.

Langmead B, Salzberg SL, 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9:357–359.

Larsson A, 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278.

Lee SI, Parr CS, Hwang, Mindell DP, Choe JC, 2003. Phylogeny of magpies (genus *Pica*) inferred from mtDNA data. *Mol Phylogenet Evol* 29:250–257.

Lerner HRL, Meyer M, James HF, Hofreiter M, Fleischer RC, 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. *Current Biol* 21:1838–1844.

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

Londei T, 2018. Association of *Pica* magpies with grazing ungulates: A clue to the genus' origins. *Rivista Ital Ornitol* 87:39–42.

Lovette IJ, 2004. Mitochondrial dating and mixed support for the "2% rule" in birds. *The Auk* 121:1–6.

Luo Y, Yang X, Gao Y, 2013. Mitochondrial DNA response to high altitude: A new perspective on high-altitude adaptation. *Mitochondrial DNA* 24:313–319.

Macey JR, Strasburg JL, Brisson JA, Vredenburg VT, Jennings M et al., 2001. Molecular phylogenetics of western North American frogs of the *Rana boylii* species group. *Mol Phylogenet Evol* 19:131–143.

Madge S, Christie D, Kirwan GM, 2020a. Eurasian magpie *Pica pica*. version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Madge S, Christie DA, Kirwan GM, 2020b. Black-rumped magpie *Pica bottanensis*. version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Madge S, Christie DA, Kirwan GM, 2020c. Oriental magpie *Pica serica*. version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Madge S, Kirwan GM, 2020d. Asir magpie *Pica asirensis*. version 1.0. In: Billerman SM, Keeney BK, Rodewald PG, Schulenberg TS eds. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Mayr E, 1963. *Animal Species and Evolution*. Cambridge: Belknap Press.

Mayr E, 1969. *Principles of Systematic Zoology*. New York: McGraw-Hill.

McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL, 2011. Calibrating divergence times on species trees versus gene trees: Implications for speciation history of *Aphelocoma* Jays. *Evolution* 65:184–202.

McDonald JH, Kreitman M, 1991. Adaptive evolution at the *Adh* locus in *Drosophila*. *Nature* 351:652–654.

McLaughlin JF, Faircloth BC, Glenn TC, Winker K, 2020. Divergence, gene flow and speciation in eight lineages of trans-Beringian birds. *Mol Ecol* 29:3526–3542.

Meiklejohn CD, Montooth KL, Rand DM, 2007. Positive and negative selection on the mitochondrial genome. *Trends Genet* 23(6):259-263.

Miller AH, 1929. Additions to the Rancho La Brea Avifauna. *Univ Calif Publ Bull Dept Geol Sci* 19:223–224.

Miller AH, Bowman RI, 1956. A fossil magpie from the Pleistocene of Texas. *Condor* 58:164–165.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG et al., 2003. Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 100:171–176.

Nachman MW, 1996. Non-neutral mitochondrial DNA variation in humans and chimpanzees. *Genetics* 142:953–963.

Lamb AM, Gan HM, Greening C, Joseph L, Lee YP et al., 2018. Climate-driven mitochondrial selection: A test in Australian songbirds. *Mol Ecol* 27:898–918.

Nei M, 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89:583–590.

Nosil P, Funk DJ, Ortiz-Barrientos D, 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol* 18:375–402.

Ohta T, 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Systematics* 23:263–286.

Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S et al., 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. *Mol Biol Evol* 28(6):1927–1942.

Padial JM, Miralles A, De la Riva I, Vences M, 2010. The integrative future of taxonomy. *Front Zool* 7:16.

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A, 2020. Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* 70:e102.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP et al., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:569–573.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA, 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biol* 67:901–904.

Ramos-Onsins SE, Rozas J, 2002. Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* 19:2092–2100.

Rand DM, 2001. The units of selection on mitochondrial DNA. *Annu Rev Ecol Syst* 32:415–448.

Randi E, 1996. A mitochondrial cytochrome *B* phylogeny of the *Alectoris* partridges. *Mol Phylogenet Evol* 6:214–227.

Rannala B, Yang Z, 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164:1645–1656.

Räsänen K, Hendry AP, 2008. Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecol letters* 11(6):624–636.

Ribeiro ÂM, Lloyd P, Bowie RC, 2011. A tight balance between natural selection and gene flow in a southern African arid-zone endemic bird. *Evolution* 65(12):3499–3514.

Rogers SM, Bernatchez L, 2005. Integrating QTL mapping and genomic scans towards the characterization of candidate loci under parallel directional selection in these lake whitefish *Coregonus clupeaformis*. *Mol Ecol* 14:351–361.

Rovito SM, 2010. Lineage divergence and speciation in the web-toed salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada, California. *Mol Ecol* 19:4554–4571.

Rubinoff D, Doorenweerd C, McElfresh JS, Millar JG, 2021. Phylogeography of an endemic California silkmoth genus suggests the importance of an unheralded central California province in generating regional endemic biodiversity. *Mol Phylogenet Evol* 164:107256.

Sangster G, 2018. Integrative taxonomy of birds: The nature and delimitation of species. In: Tietze DT editor. *Bird Species: How They Arise, Modify and Vanish*. Switzerland: Springer, 9-37.

Shaffer HB, Pauly GB, Oliver J, Trenham PC, 2004. The molecular phylogenetics of endangerment: Cryptic variation and historical phylogeography of the California tiger salamander *Ambystoma californiense*. *Mol Ecol* 13:3033–3049.

Shields GF, Wilson AC, 1987. Calibration of mitochondrial DNA evolution in geese. *J Mol Evol* 24**:**212–217.

Sibley GC, Ahlquist JE, 1990. *Phylogeny and Classification of Birds: A Study in Molecular Evolution.* New Haven: Yale Univ Press.

Smith SD, Pennell MW, Dunn CW, Edwards SV, 2020. Phylogenetics is the new genetics (for most of biodiversity). *Trends Ecol Evol* 35:415–425.

Sokolova I, 2018. Mitochondrial adaptations to variable environments and their role in animals' stress tolerance. *Integrat Compar Biol* 58(3):519–531.

Song G, Zhang R, Alström P, Irestedt M, Cai T et al., 2018. Complete taxon sampling of the avian genus *Pica* (magpies) reveals ancient relictual populations and synchronous Late-Pleistocene demographic expansion across the Northern Hemisphere. *J Avian Biol* 49:e01612.

Stinchcombe JR, Hoekstra HE, 2008. Combining population genomics and quantitative genetics: Finding the genes underlying ecologically important traits. *Heredity* 100(2):158–170.

Storz JF, 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Mol Ecol* 14(3):671–688.

Tajima F, 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.

Trost CH, 2020. Black-billed magpie *Pica hudsonia*. version 1.0. In: Billerman S editor. *Birds of the World*. Ithaca: Cornell Lab of Ornithology.

Unruh JR, 1991. The uplift of the Sierra Nevada and implications for late Cenozoic epeirogeny in the western Cordillera. *GSA Bull* 103:1395–1404.

Vaurie Ch, 1959. *The Birds of the Palearctic Fauna*: *Order Passeriformes.* London: Witherby.

Voous KH, 1960. *Atlas of European Bird.* London: Nelson.

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV et al., 2018. Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 35:773–777.

Weir JT, Schluter D, 2004. Ice sheets promote speciation in boreal birds. *Proc R Soc Lond B* 271:1881–1887.

Weir JT, Schluter D, 2008. Calibrating the avian molecular clock. *Mol Ecol* 17:2321–2328.

Winker K, 2021. An overview of speciation and species limits in birds. *Ortithology* 138:1–27.

Winker K, Glenn TC, Withrow J, Spencer G, Sealy SG et al., 2019. Speciation despite gene flow in two owls (*Aegolius* ssp.): Evidence from 2,517 ultraconserved element loci. *Auk* 136:1–12.

Zink RM, Rohwer S, Andreev AV, Dittmann DL, 1995. Trans-Beringia comparisons of mitochondrial DNA differentiation in birds. *Condor* 97:639–649.

Figure Captions

Figure 1. Distribution of *Pica pica sensu lato* subspecies (From Kryukov et al. 2022, with changes) and geographic origin of samples used for mitogenomic analysis. Red dots – this work, black dots – data from GenBank.

Figure 2. ML tree based on complete mitogenomes of *Pica* species and outgroup. Numbers above the branches indicate bootstrap support. Short indices correspond to the newly obtained data, others are GenBank accession numbers (Supplementary Table 1). The images of *Pica hudsonia*, *P. nuttalli*, *P. pica* and *P. serica*, by artist Brian Small, used at the Figure 2, are copyrighted by Lynx Edicions.

Figure 3. Time-calibrated Bayesian inference tree based on concatenated mitochondrial protein-coding genes of *Pica* species and outgroup. Numbers above the branches indicate Bayesian posterior probability values. Blue bars next to nodes indicate 95% credibility intervals for their age estimates. The figures in bold at the nodes are divergence times in million years before present (Mya). Short indices correspond to the new obtained mitogenomes, others are GenBank accession numbers (Supplementary Table 1). Selection parameters are shown in bold: number of sites under positive selection along the associated branches and number of sites under purified selection after slash, revealed by fixed effects likelihood (FEL) analysis.

Figure 4. Bayesian skyline plots for population dynamics over time for three taxa of *Pica*. X-axis is times in thousand years before present. Y-axis represents median values of effective female population sizes in millions on a log scale.

Figure 5. Sliding window plots of pairwise comparisons of nucleotide variability along the alignments of *Pica hudsonia* and *Pica nuttalli* (a), *Pica hudsonia* and *Pica pica camtschatica* (b). X-axes: nucleotide position at alignment of protein coding genes, Y-axes: net nucleotide variation per site centered at the midpoints of the windows of 100 sites. Note the differently scaled Y-axes.

Table Legends

Table 1. Parameters of variation, neutrality tests and tMRCA estimations for lineages based on concatenated protein coding genes alignments.

Table 1. Parameters of variation, neutrality tests and tMRCA estimations for lineages based on concatenated protein coding genes alignments

Lineage	N	S	k	π ± SD (%)	h	Hd ± SD	*D*	*p (D)*	F_s	*p* (F_s)	R_2	*p* (R_2)	*r*	*p (r)*	tMRCA from BI tree	tMRCA from Skyline plots
P. hudsonia	9	24	7.67	0.067±0.013	6	0.833±0.127	-0.654	0.26	0.949	0.69	0.117	0.05*	0.14	0.69	62 (40-100)	48.0 (28.2-69.7)
P. nuttalli	5	16	7.00	0.061±0.022	5	1.00±0.126	-1.044	0.16	-0.832	0.19	0.231	0.51	0.24	0.51	66 (40-100)	56.5 (29.5-85.5)
P. pica camtschatica	4	6	3.00	0.026±0.011	3	0.833±0.222	-0.808	0.2	0.731	0.60	0.344	0.71	0.3	0.47	37 (20-60)	22.3 (2.3-40.8)
P. serica	5	68	35.8	0.314±0.064	5	1.00±0.126	0.733	0.75	1.178	0.47	0.199	0.42	0.16	0.32	250 (190-310)	226.0 (170.9-285.8)

N: sample size; S: number of polymorphic sites; k: average pairwise nucleotide differences; π: nucleotide diversity with standard deviation; h: number of haplotypes; Hd: haplotype diversity with standard deviation; *D*: Tajima's, F_s: Fu's, and R_2: Ramos-Onsins & Rozas's tests; their *p*-values: * $p \leq 0.05$. r: Harpending's raggedness index and its p-value, all obtained in DnaSP. tMRCA was estimated from BI trees and Coalescent Bayesian Skyline model in Beast, with 95% HPD range, in kyr.

Figure 1

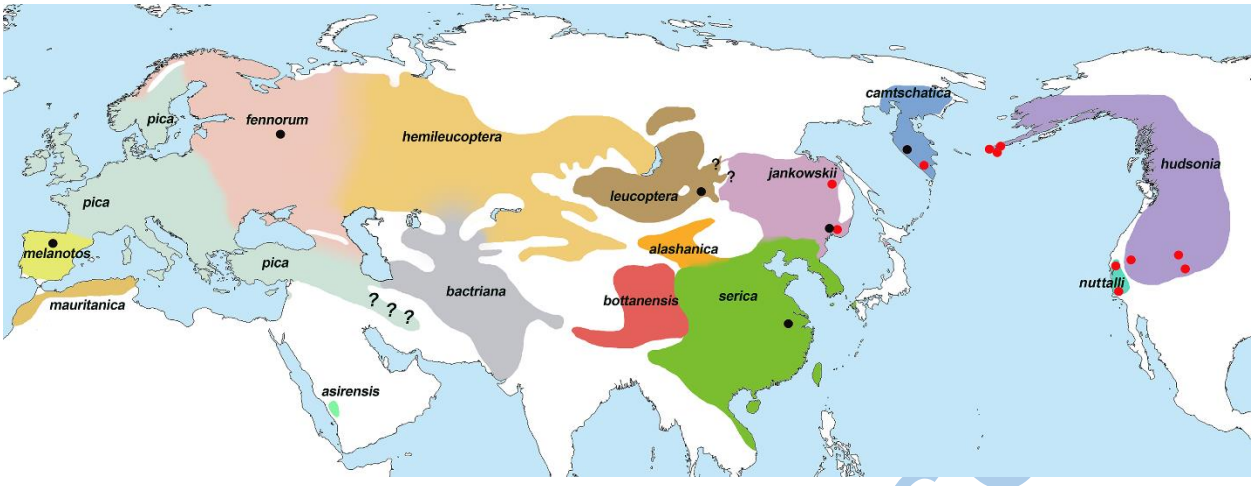

Figure 2

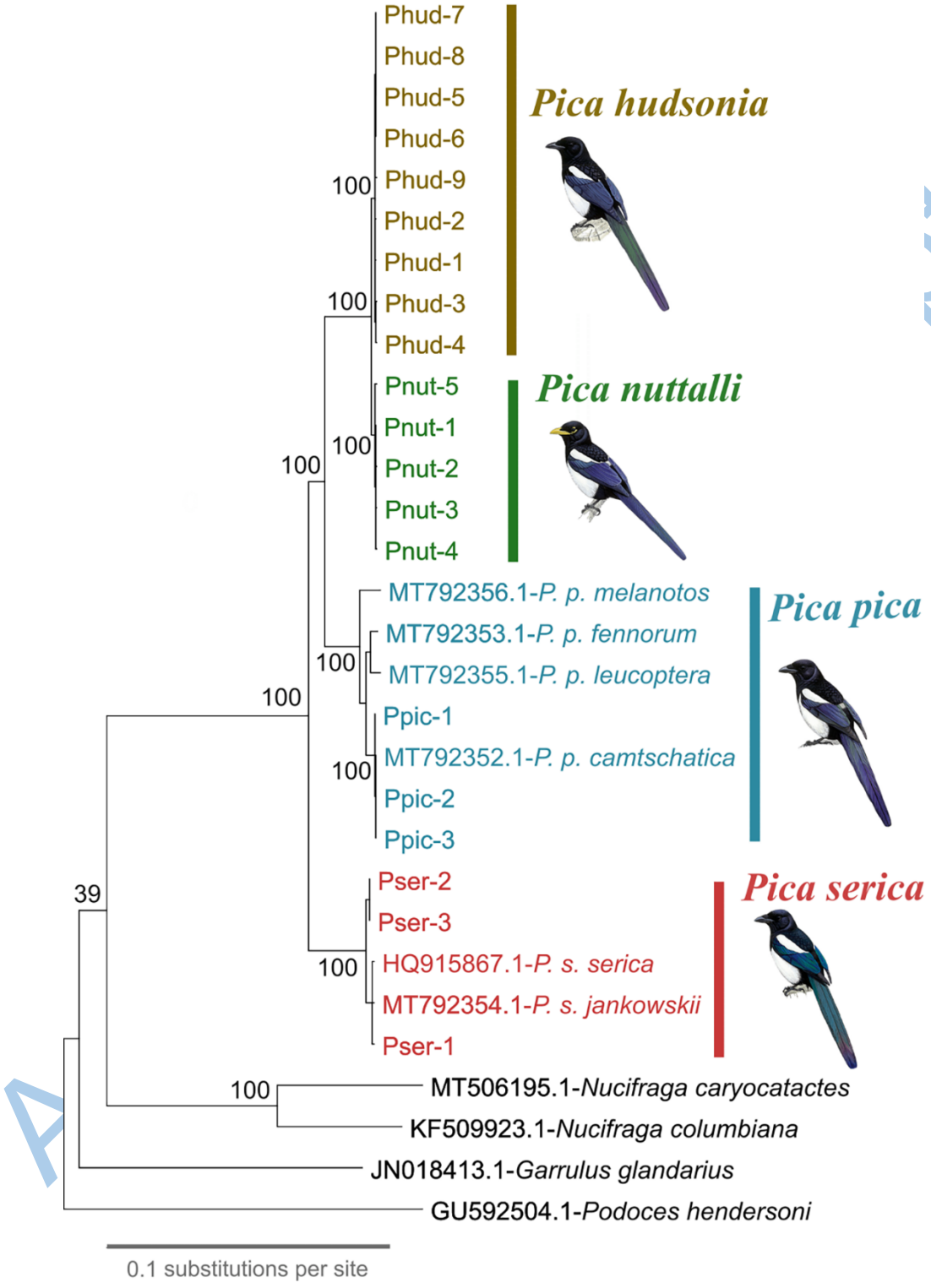

Figure 3

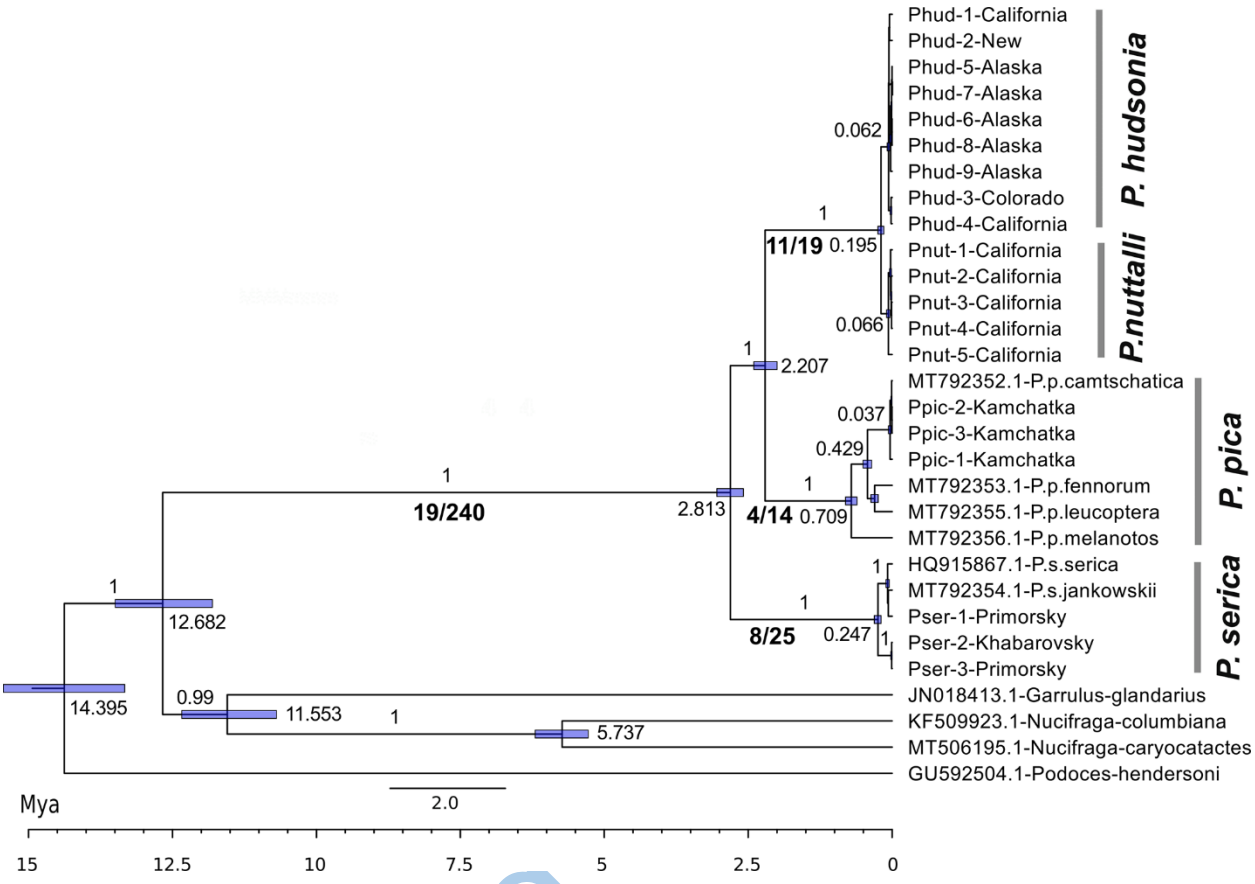

Figure 4

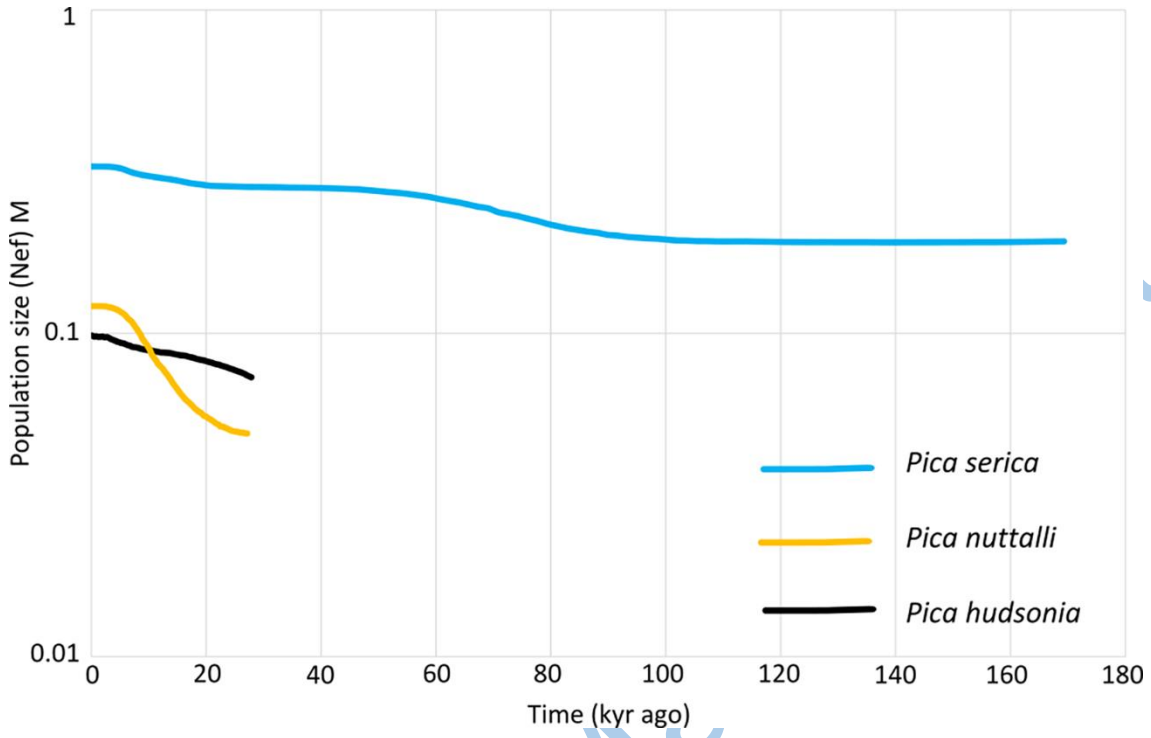

Figure 5