Molecular Phylogenetics and Evolution 57 (2010) 466–470

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Short Communication

Obtaining mtDNA genomes from next-generation transcriptome sequencing: A case study on the basal Passerida (Aves: Passeriformes) phylogeny

Benoit Nabholz [a,*], Erich D. Jarvis [b], Hans Ellegren [a]

[a] Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden
[b] Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, USA

ARTICLE INFO

Article history:
Received 15 March 2010
Revised 8 June 2010
Accepted 15 June 2010
Available online 25 June 2010

Keywords:
Songbirds
Passeriformes
Passerida
Mitochondrial DNA
Paridae
Passeroidea
Sylvioidea
Muscicapoidae
Molecular phylogeny
Next-generation sequencing

ABSTRACT

Classically, the mitochondrial genome is sequenced by a series of amplicons using conserved PCR primers. Here we show how shot-gun transcriptome sequencing can be used to obtain the complete set of protein-coding genes from the mtDNA of four passerine bird species. With these sequences, we address the still unresolved basal Passerida relationships (Aves: Passeriformes). Our analysis suggests a new hypothesis for the basal relationships of Passerida, namely a clade grouping Sylvioidea and Passeroidea, with Paridae and Muscicapidae as successive sister groups to this clade. This study demonstrates the usefulness of next-generation sequencing transcriptome sequencing for obtaining new mtDNA genomes.

1. Introduction

Passeriformes is by far the most speciose avian order, comprising nearly 60% of all living bird species (Clements, 2007). Since the pioneering DNA–DNA hybridization work of Sibley and Ahlquist (1990), significant progress has have been made in understanding the relationships among passerine birds (Ericson et al., 2002a,b; Barker et al., 2002, 2004; Ericson and Johansson, 2003; Johansson et al., 2008). For example, a consensus has emerged concerning the deepest relationships including the basal position of the enigmatic and species-poor New-Zealand wren family (Acanthisittidae) and the early dichotomy between Oscine (or songbirds) and Suboscine passerines (Barker et al., 2002; Ericson et al., 2002b). Within songbirds, several groups of Australo-papuan affinities such as lyrebirds (Menuridae) or bowerbirds (Ptilonorhynchidae), previously placed within the Corvoidea (i.e., crows and allied, *sensu* Sibley and Ahlquist, 1990), have a basal position within the Oscine clade (Barker et al., 2002, 2004). Finally, the huge clade of Passerida appears to be monophyletic with rockfowls (Picathartidae) and Australasian robins (Petroicidae) being sister groups to this clade (Ericson et al., 2002a,b; Ericson and Johansson, 2003; Barker et al., 2004).

Contrasting with these robust basal relationships, several more recent relationships are still ambiguous. Amongst the major uncertainties, the basal relationships within Passerida seem to be some of the most difficult. The use of several combinations of nuclear markers has lead to weakly supported alternative topologies (summarized in Fig. 1A–C; Barker et al., 2004; Treplin et al., 2008; Johansson et al., 2008). Johansson et al. (2008) reviewed the current knowledge of the early relationships among Passerida and concluded that they are best depicted by a polytomy of nine clades: Bombycillidae, African Hyliotas, Muscicapoidae, Certhioidae, Sylvioidea, Paridae (including Parinae and Remizinae), Regulidae, Stenostiridae, and Passeroidea.

Recently, complete mitochondrial genomes (mtDNA) have been used to infer basal Neoavian relationships with various degree of success and in some disagreement with comprehensive studies using nuclear markers (e.g., Hackett et al. (2008) versus Pratt et al. (2009)). For the shallower divergences, the mitochondrial genome has been, and remains to be, a valuable marker. The standard method of retrieving new mtDNA genomes has been to use conserved PCR primers to amplify a suite of overlapping fragments that correspond to the whole mtDNA molecule, which are then sequenced and assembled (e.g., Slack et al., 2006). Commonly, long-

* Corresponding author.
E-mail address: benoit.nabholz@ebc.uu.se (B. Nabholz).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.06.009

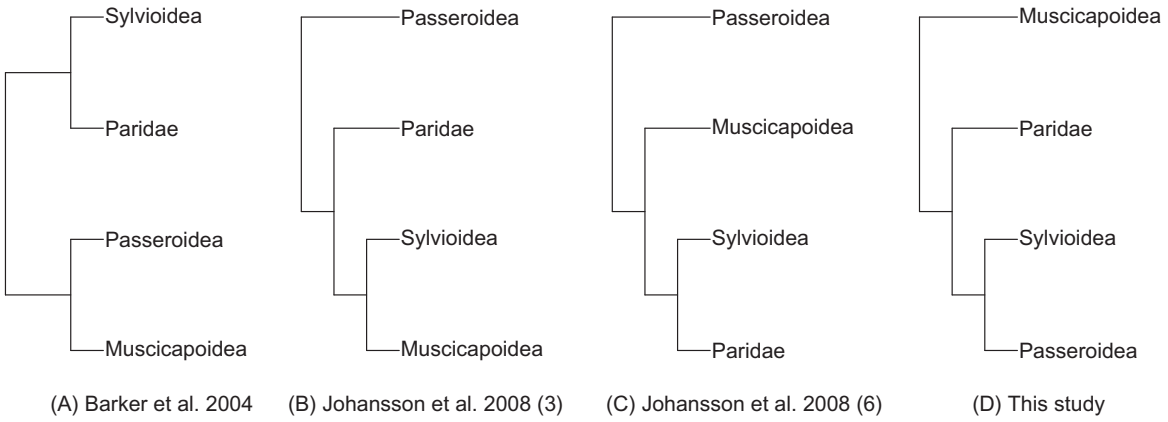

Fig. 1. Alternative Passerida topologies previously published (A–C) and proposed in this study (D). (A) is from Barker et al. (2004) using two nuclear exons corresponding to 4126 bp (similar to the Treplin et al. (2008) topology obtain with two additional markers for 6179 bp). (B) is the topology published by Johansson et al. (2008) using three nuclear markers corresponding to 2315 bp. (C) is the topology published by Johansson et al. (2008) using six nuclear markers corresponding to 7288 bp including the three using (B). (D) is the topology generated in this report. Note that as in the present study, relationships previously published received various statistical support, including very low (see original publications for details).

range PCR is first employed to generate templates for subsequent amplification of shorter fragments suitably sized for DNA sequencing. This two-step amplification procedure is thought to select for amplification products of true mitochondrial origin, as nuclear copies of mtDNA (numts) that have been transposed to the genomic DNA are usually short.

DNA sequencing is now gradually moving from Sanger-based technology, coupled with capillary electrophoresis, to next-generation sequencing (NGS) approaches based on sequencing-by-synthesis, or other concepts, coupled with massively parallel processing of templates (Meyer et al., 2008). While Sanger technology may yield up to 100 kb of sequence data per run on standard sequencing instruments, the new high-throughput platforms yield from several hundreds of Mb up to several tens of Gb of DNA sequence per run (although run times are typically much longer in the latter than in the former approach). Recently, NGS approaches for retrieving mtDNA genomes based on overlapping amplicons – obtained through long-PCR from genomic DNA – have been reported (Jex et al., 2008).

The exceptional amount of DNA sequence data generated by NGS methodology should imply realistic possibilities to obtain more or less complete mtDNA genomes from shot-gun sequencing without prior amplification of specific regions of the mtDNA molecule. In this study we evaluated the feasibility of this approach by analyzing brain transcriptome sequences from four passerine species obtained through Roche 454 sequencing as recently reported by Künstner et al. (2010). We then used these data together with 11 previously published avian mtDNA genomes to study Passerida relationships.

2. Materials and methods

2.1. Passerine mitochondrial data set

We built a mitochondrial DNA data set composed of all 12 heavy-strand protein-coding genes: COX1–3, ND1–5, ND4L, ATP6, ATP8, and CYTB. As in prior studies (Paton et al., 2002; Slack et al., 2006), the NADH dehydrogenase 6 (ND6) gene was excluded because it is encoded on the opposite strand and has a very different base composition compared to the other mitochondrial protein-coding genes. We focused on protein-coding genes because the RNA isolation method used in Künstner et al. (2010) was designed to select for transcripts with terminal poly-As (see below). We first downloaded all complete passerine mitochondrial genomes available (May 2010) in GENBANK from NCBI (http://www.ncbi.nlm.nih.gov/), including gray-headed broadbill (*Smithornis sharpei*, NC_000879), fuscous flycatcher (*Cnemotriccus fuscatus*, NC_007975), superb lyrebird (*Menura novaehollandiae*, NC_000880), rook (*Corvus frugilegus*, NC_002069), eastern-orphean warbler (*Sylvia crassirostris*, NC_010229), village indigobird (*Vidua chalybeata*, NC_007883), zebra finch (*Taeniopygia guttata*, NC_007897), Taiwan (Styan's) Bulbul (*Pycnonotus taivanus*, NC_013483), blackcap (*Sylvia atricapilla*, NC_010228), and Eurasian reed warbler (*Acrocephalus scirpaceus*, NC_010227). We also download the near complete mitochondrial genome of the rifleman (*Acanthisitta chloris*, AY325307), considered to belong to the sister group of all other passerines (Acanthisittidae, New-Zealand wrens) (Barker et al., 2002, 2004; Ericson et al., 2002b). We also perform an analysis using the mitochondrial genomes of two non-passerine species, the great roadrunner (*Geococcyx californianus*, NC_011711) and the kelp gull (*Larus dominicanus*, NC_007006).

Four passerine species were subject to brain transcriptome sequencing: golden-collared manakin (*Manacus vitellinus*, Pipridae, SRR029477–78, available in Sequence Read Archive, http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi), American crow (*Corvus brachyrhynchos*, Corvidae, SRR029463–64), blue tit (*Cyanistes* [*Parus*] *caeruleus*, Paridae, SRR029162), and pied flycatcher (*Ficedula hypoleuca*, Muscicapidae, SRR029159–61). The source of brain samples, methods for RNA isolation, cDNA preparation and next-generation sequencing are given in Künstner et al. (2010). For each species, the reads, which averaged 250 bp in length, were *de novo* assembled into contigs using the NEWBLER software (version 2.0) distributed with the ROCHE 454-instrument. Then we chose a closely related species for which the complete mitochondrial genome was available and performed a reciprocal best-hit BLAST analysis between contigs from the transcriptome data and the reference species, using BLATZ (Schwartz et al., 2003). We used the rook as a reference for the American crow, the eastern-orphean warbler for the pied flycatcher and the blue tit, and the fuscous flycatcher for the golden-collared manakin. We removed the last 10 bp of ATP8 and the last 7 bp of ND4L that overlap with ATP6 and ND4, respectively. Amino acid sequence alignments were performed using the MAFFT software with high accuracy parameters (Katoh et al., 2002) and verified by visual inspection.

2.2. Phylogenetic analyses

We used both maximum likelihood (ML) and a Bayesian method to estimate the topology of the passerine tree. We performed the ML analyses with PHYML (version 3.0, Guindon and Gascuel,

2003) and GARLI (version 0.951, Zwickl, 2006). The best model of sequence evolution was selected by AIC criteria using JMODELTEST, with a neighbor-joining tree as fixed topology (Posada, 2008). We constructed trees from the full nucleotide data and from a dataset where third-codon positions were coded as RY (purines coded as R and pyrimidines as Y). RY coding of the third codon position limits the influence of base composition heterogeneity among species and among-site rate variation on phylogenetic reconstruction (e.g., Delsuc et al., 2003), and has proven valuable for more accurate phylogenetic inference in previous studies of avian mtDNA genomes (Harrison et al., 2004; Gibb et al., 2007). For the Bayesian analyses we used the MRBAYES software (version 3.1, Ronquist and Huelsenbeck, 2003) applying the substitution model selected by JMODELTEST with a partition of the dataset according to the genes or to the codons positions. We assessed the robustness of the ML analyses using 1000 bootstrap replicates and using posterior probabilities for the Bayesian analyses (summarized in the Table 2).

3. Results and discussion

3.1. mtDNA genome assembly based on shot-gun RNA sequencing

We evaluated the usefulness of deep transcriptome sequencing for retrieving mitochondrial genomes by focusing on the four passerine species included in Künstner et al. (2010), namely the American crow, the pied flycatcher, the blue tit and the golden-collared manakin. We used Roche 454 reads to construct contigs of each gene and reciprocal BLAST analyses to map these onto several reference mtDNA genomes (see Section 2). It was clear from this analysis that all 12 mtDNA protein-coding genes encoded from the heavy strand were expressed in adult avian brain. Most genes were fully assembled with <10% of missing data in the total set of genes, except for in blue tit where 33% was missing (Table 1). Although the cDNA was normalized in order to avoid repeated sequencing of abundant transcripts, we explored the level of expression of the mitochondrial genes by mapping the reads onto the assembled gene-specific contigs in each species and quantified the mean number of reads per base per gene (i.e., coverage). We found that mtDNA gene expression was on average 4–8 times higher than the mean levels of nuclear gene expression (cf. Table 1 of Künstner et al., 2010), suggesting that normalization did not completely compensate for the much higher copy number of mtDNA compared to nuclear DNA (Hale et al., 2009).

We found that the expression level of individual mitochondrial genes varied within an order of magnitude; for the American crow they ranged from 12 to 104 reads per base per gene (excluding ND4L which was not detected), for the pied flycatcher from 19 to 64 (excluding ATP8) and for the golden-collared manakin from 9 to 132. There was no correlation in levels of gene expression among these species (American crow-pied flycatcher, $r = 0.254$, $p = 0.42$; American crow-golden-collared manakin, $r = 0.057$, $p = 0.86$; pied flycatcher-golden-collared manakin, $r = 0.55$, $p = 0.064$). The blue tit was an outlier relative to the other species; three genes were not detected at all (ATP8, ND1, and ND4L) and coverage for the other genes varied extensively, from 9 reads per base (ND3) to 3080 (ATP6). Such variation indicates that although all 12 genes are thought to be transcribed from one common heavy strand promoter (Scarpulla, 2008), their expression levels differ, perhaps due to either technical, species-specific, or brain state differences in the expression of mitochondrial genes controlled by other DNA regulatory elements.

3.2. Phylogenetic analyses

In addition to the four new species for which we retrieved mtDNA genome sequence data by transcriptome sequencing, complete mtDNA data was obtained from GENBANK for another 11 passerines (see Section 2). After alignment, this resulted in a matrix of 15 species and a maximum of 10,830 bp across the coding sequences of the 12 genes. We computed the GC content at the third codon position (GC3) for all the species. GC3 varies significantly across species, from 38% in the fuscous flycatcher to 55% in the

Table 2
Statistical supports obtain for the interesting nodes (as labeled in Fig. 2).

Bootstrap support (ML analyses)					Posterior probability (Bayesian analyses)	
Node No.	PHYML	GARLI	PHYML with RY-coding	PHYML without blue tit	Genes partitioning	Codons position partitioning
1	85.6	74.1	76.4	87.5	1.00	0.77
2	97.8	99.8	100	100	1.00	1.00
3	60.6	60.6	73.9	90.0	1.00	0.99
4	42.8	38.5	36.1	–	0.26	0.99
5	100	99.6	100	100	1.00	1.00
6	100	100	100	100	1.00	1.00
7	100	99.0	91.3	100	1.00	1.00

Table 1
Size and coverage of the sequences obtain through 454 transcriptome sequencing. Coverage is computed by mapping the reads onto the assembled gene-specific contigs in each species and quantifying the mean number of reads per base per gene.

	Total size gene	Blue tit		Pied flycatcher		Golden-collared manakin		American crow	
		Size gene	Mean coverage	Size gene	Mean coverage	Size gene	Mean coverage	Size gene	Mean coverage
ATP6	681	270	3080	642	39	642	40	672	81
ATP8	156	0		0		156	26	144	12
COX1	1548	1448	1123	1539	25	1464	55	1464	93
COX2	681	672	478	675	31	681	74	681	104
COX3	783	783	2272	756	24	783	99	762	98
CYTB	1140	930	1152	1107	29	1119	43	1134	92
ND1	975	0		927	19	909	132	480	93
ND2	1038	1017	509	1026	21	1038	68	1035	72
ND3	348	348	9	345	64	348	108	348	44
ND4	1377	851	1358	1377	27	1362	53	1368	39
ND4L	288	0		273	33	234	9	0	
ND5	1815	889	192	1794	23	1692	43	1800	45
Total sequence size (% missing data)	10,830 (0)	7208 (34)		10,461 (3)		10,428 (4)		9888 (9)	
Mean coverage per gene (SD)			1130 (1007)		30 (13)		63 (36)		71 (30)

blue tit (52% for the European reed warbler). This variation could be problematic for phylogenetic reconstruction if a substitution model assuming constant GC content is used. RY coding provides an empirical test to the influence of GC content on phylogenetic reconstruction (Delsuc et al., 2003).

All the phylogenetic analyses performed (ML and Bayesian analyses with and without RY-recoding of the third codon position) provided the same topology with most of the nodes obtaining systematically good statistical support (Fig. 2 and Table 2). The mitochondrial dataset recovered without ambiguity the basal Oscine–Suboscine split. Within Suboscines, the golden-collared manakin was grouped with the fuscous flycatcher (*Cnemotriccus*), reflecting the well-known Old World–New World Suboscine split (Barker et al., 2002, 2004; Ericson et al., 2002a,b). Within Oscines, we obtained strong statistical support for the basal position of the lyrebird (*Menura*) (Ericson et al., 2002a).

The extremely short internodes at the base of the Passerida group (Fig. 2) indicate a rapid radiation and probably explain why previous work have yielded conflicting results (e.g., Ericson and Johansson, 2003; Barker et al., 2004; Treplin et al., 2008; Johansson et al., 2008; summarized in Fig. 1A–C). The fast evolving mitochondrial genome could potentially provide some phylogenetic signal for these nodes. In this regard, our analysis suggests a new hypothesis for the Passerida relationship (summarized in Fig. 1D). Specifically, we obtain moderate to strong support for grouping Sylvioidea (represented by Acrocephalidae, the Eurasian reed warbler, two Sylviidae, the blackcap and eastern-orphean warbler and one Pycnonotidae, the Taiwan bubul) and Passeroidea (represented by an Estrildidae, the zebra finch, and Viduidae, the village indigobird) (Fig. 2 and Table 2, node No. 3). Within the Sylvioidea, we obtain support for grouping the Pycnonotidae and the Acrocephalidae (Fig. 2 and Table 2, node No. 1). Finally, the Paridae, represented by the blue tit, appeared to be an immediate sister group of this clade, followed by the Muscicapidae, represented by the pied flycatcher. However, the branching pattern of Paridae and Muscicapidae obtained a low bootstrap support (Fig. 2 and Table 2, node No. 4) and thus remains unresolved.

We performed two control analyses to check the robustness of our topology. Firstly, we assessed the influence of the outgroup on topology reconstruction. We added two non-passerine species to the data set, namely the great roadrunner (*Geococcyx californianus*) and the kelp gull (*Larus dominicanus*). We replicated the ML analysis using PHYML with RY coding and obtained a similar topology, except for that the blue tit and pied flycatcher position was swapped. The Passeroidea–Sylvioidea relationship obtain a comparable support (68%, Supplementary Fig. S1). Secondly, given that the blue tit had the highest proportion of missing data, we removed it from the alignment, and replicated the ML analysis using PHYML. The Sylvoidea + Passeroidea clade was confirmed with a higher bootstrap support than before (90% compared to 60–70% in the analysis including blue tit) (Table 2).

3.3. Nuclear copies of mtDNA (numts)

Numts can create problems for phylogenetic reconstruction due to their mistaken identity as mitochondrial sequences (Sorenson and Quinn, 1998). However, a previous study suggested that numts are rare in the avian genome (Pereira and Baker, 2004). Numts are generally considered pseudogenes, although some are transcribed and mutations in them can cause human disease (Hazkani-Covo et al., 2010). In theory, if numts are transcribed in birds, our shot-gun cDNA sequencing should generate numerous numt sequences. To test for the presence of numts in the shot-gun transcriptome data, we were in a few cases able to compare our assembled sequences with available mitochondrial gene sequences of the same species obtained by PCR amplification from genomic DNA: for blue tit COX1 (GQ481677) and CYTB (DQ474043), and for pied flycatcher COX1 (GQ481901) and ND2 (DQ146345). For the blue tit, this revealed in each case four mismatches out of 618 and 809 sites compared, respectively. For the pied flycatcher, we found no mismatch out of 692 sites for COX1 and two out of 1011 sites for ND2. For all comparisons, the amount of observed divergence (<1%) is within the expected level of polymorphism of bird species which is around 3% mean pairwise synonymous divergence (Berlin et al., 2007; Nabholz et al., 2009). These findings indicate that numts may have not significantly influenced our phylogenetic analyses.

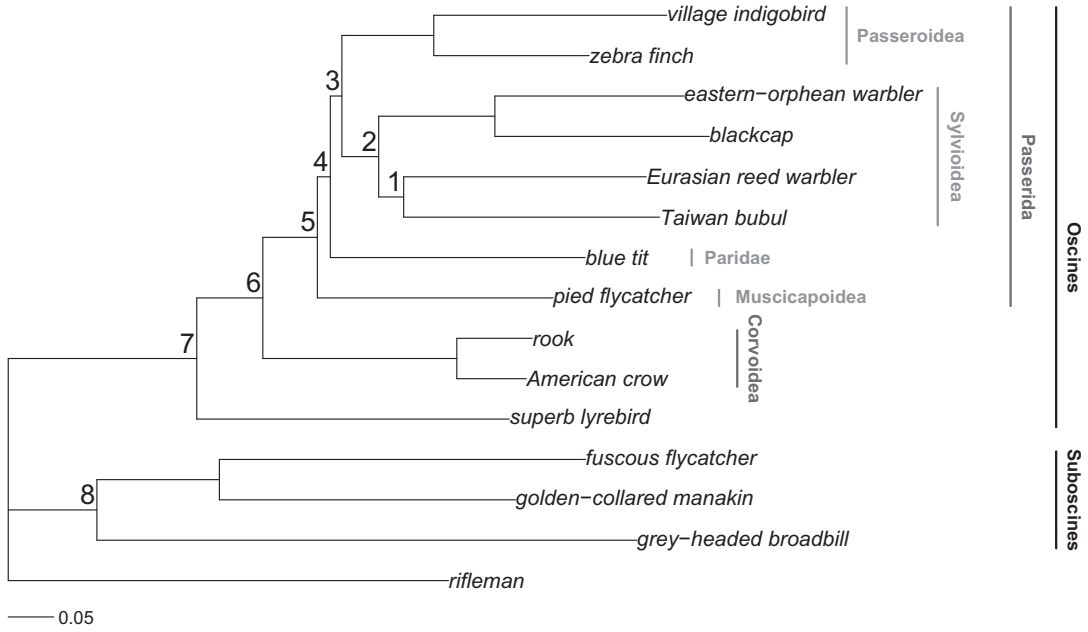

Fig. 2. Maximum likelihood (ML) tree with a GTR + I + Gamma6 model. Numbers close to the nodes indicate statistical support as recorded in Table 2. All the other nodes are supported by bootstrap or posterior probabilities equal to 100%. Scale indicates substitution/site.

Finally, we performed phylogenetic reconstruction of all species for each of the 12 genes independently using PHYML with 200 bootstrap replications (Supplementary Fig. S2). Comparisons between the genes trees and the topology presented in Fig. 2 did not reveal any strongly supported incongruence (bootstraps >70%). The only exception was for CYTB, where the monophyly of Suboscines is not recovered. To further explore the congruence between the single genes and full matrix analyses, we applied the supertree method PHYSIC_IST (Scornavacca et al., 2008), considering only the node with the bootstrap support higher than 70% (PHYSIC_IST option -b 70). The supertree topology was identical to the one obtained with the full matrix analyses with Paridae/Muscicapidae node being the only node unresolved.

4. Conclusions

Our finding that more or less the complete protein-coding part of the mitochondrial genome can be obtained as a byproduct in NGS transcriptome sequencing is encouraging for further phylogenetic work because the amount of sequence data generated by NGS methodologies is expected to increase significantly in the near future. Here we found support for a novel hypothesis concerning basal relationships within Passerida in the form of a Passeroidea + Sylvioidea clade. Increased taxonomical sampling to include Bombycillidae, *Hyliota*, Certhioidae, *Regulus* or Stenostiridae would be necessary to further support or refute this hypothesis and would provide additional information on the relationships within the highly diversified Passerida group (>2000 species).

Acknowledgments

Computational work was performed at the Uppsala University UPPMAX platform for high-performance computing. H.E. acknowledges funding from the Swedish Research Council, The Knut and Alice Wallenberg Foundation and a European Research Council Advanced Investigator Grant. E.D.J. acknowledges funding from NIH and HHMI. The authors thank the editor and the three anonymous reviewers for their valuable suggestions.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.06.009.

References

Barker, F.K., Barrowclough, G.F., Groth, J.G., 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proc. Biol. Sci. 269, 295–308.

Barker, F.K., Cibois, A., Schikler, P., Feinstein, J., Cracraft, J., 2004. Phylogeny and diversification of the largest avian radiation. Proc. Natl. Acad. Sci. USA 101, 11040–11045.

Berlin, S., Tomaras, D., Charlesworth, B., 2007. Low mitochondrial variability in birds may indicate Hill–Robertson effects on the W chromosome. Heredity 99, 389–396.

Clements, J.F., 2007. The Clements Checklist of Birds of the World, sixth ed. Cornell University Press.

Delsuc, F., Phillips, M.J., Penny, D., 2003. Comment on "Hexapod origins: monophyletic or paraphyletic?". Science 301, 1482.

Ericson, P.G.P., Johansson, U.S., 2003. Phylogeny of Passerida (Aves: Passeriformes) based on nuclear and mitochondrial sequence data. Mol. Phylogenet. Evol. 29, 126–138.

Ericson, P.G.P., Christidis, L., Irestedt, M., Norman, J.A., 2002a. Systematic affinities of the lyrebirds (Passeriformes: *Menura*), with a novel classification of the major groups of passerine birds. Mol. Phylogenet. Evol. 25, 53–62.

Ericson, P.G.P., Christidis, L., Cooper, A., Irestedt, M., Jackson, J., Johansson, U.S., Norman, J.A., 2002b. A Gondwanan origin of passerine birds supported by DNA sequences of the endemic New Zealand wrens. Proc. Biol. Sci. B 269, 235–241.

Gibb, G.C., Kardailsky, O., Kimball, R.T., Braun, E.L., Penny, D., 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24, 269–280.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K., Harshman, J., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Steadman, D.W., Witt, C.C., Yuri, T., 2008. A phylogenomic study of birds reveals their evolutionary history. Science 320, 1763–1768.

Hale, M.C., Mccormick, C.R., Jackson, J.R., Dewoody, J.A., 2009. Next-generation pyrosequencing of gonad transcriptomes in the polyploid lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. BMC Genomics 10, 203.

Harrison, G.L., Mclenachan, P.A., Phillips, M.J., Slack, K.E., Cooper, A., Penny, D., 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous. Mol. Biol. Evol. 21, 974–983.

Hazkani-Covo, E., Zeller, R.M., Martin, W., 2010. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. PLoS Genet. 6, e1000834.

Jex, A.R., Hu, M., Littlewood, D.T., Waeschenbach, A., Gasser, R.B., 2008. Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics 9, 11.

Johansson, U.S., Fjeldså, J., Bowie, R.C.K., 2008. Phylogenetic relationships within Passerida (Aves: Passeriformes): a review and a new molecular phylogeny based on three nuclear intron markers. Mol. Phylogenet. Evol. 48, 858–876.

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066.

Künstner, A., Wolf, J.B., Backström, N., Whitney, O., Balakrishnan, C.N., Day, L., Edwards, S.V., Janes, D.E., Schlinger, B.A., Wilson, R.K., Jarvis, E.D., Warrens, W.C., Ellegren, H., 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. Mol. Ecol. 19 (Suppl. 1), 266–276.

Meyer, M., Stenzel, U., Hofreiter, M., 2008. Parallel tagged sequencing on the 454 platform. Nat. Protoc. 3, 267–278.

Nabholz, B., Glémin, S., Galtier, N., 2009. The erratic mitochondrial clock: variations of mutation rate, not population size, affect mtDNA diversity across birds and mammals. BMC Evol. Biol. 9, 54.

Paton, T., Haddrath, O., Baker, A.J., 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proc. Biol. Sci. B 269, 839–846.

Pereira, S.L., Baker, A.J., 2004. Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. BMC Evol. Biol. 4, 17.

Posada, D., 2008. JModelTest: phylogenetic model averaging. Mol. Biol. Evol. 25, 1253–1256.

Pratt, R.C., Gibb, G.C., Morgan-Richards, M., Phillips, M.J., Hendy, M.D., Penny, D., 2009. Toward resolving deep neoaves phylogeny: data, signal enhancement, and priors. Mol. Biol. Evol. 26, 313–326.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Scarpulla, R.C., 2008. Transcriptional paradigms in mammalian mitochondrial biogenesis and function. Physiol. Rev. 88, 611–638.

Schwartz, S., Kent, W.J., Smit, A., Zhang, Z., Baertsch, R., Hardison, R.C., Haussler, D., Miller, W., 2003. Human–mouse alignments with BLASTZ. Genome Res. 13, 103–107.

Scornavacca, C., Berry, V., Lefort, V., Douzery, E.J., Ranwez, V., 2008. PhySIC_IST: cleaning source trees to infer more informative supertrees. BMC Bioinformatics 9, 413.

Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press.

Slack, K.E., Jones, C.M., Ando, T., Harrison, G.L., Fordyce, R.E., Arnason, U., Penny, D., 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol. 23, 1144–1155.

Sorenson, M.D., Quinn, T.W., 1998. Numts: a challenge for avian systematics and population biology. Auk 115, 214–221.

Treplin, S., Siegert, R., Bleidorn, C., Thompson, H.S., Fotso, R., Tiedemann, R., 2008. Molecular phylogeny of songbirds (Aves: Passeriformes) and the relative utility of common nuclear marker loci. Cladistics 24, 328–349.

Zwickl, D., 2006. Ph.D. Thesis. University of Texas.