

Retroposon Insertion Patterns of Neoavian Birds: Strong Evidence for an Extensive Incomplete Lineage Sorting Era

Andreas Matzke,¹ Gennady Churakov,¹ Petra Berkes,¹ Erin M. Arms,¹ Denise Kelsey,¹ Jürgen Brosius,¹ Jan Ole Kriegs,^{†1,2} and Jürgen Schmitz^{†*,1}

¹Institute of Experimental Pathology (ZMBE), University of Münster, Münster, Germany

²Landschaftsverband Westfalen-Lippe (LWL) Museum für Naturkunde, Westfälisches Landesmuseum mit Planetarium, Münster, Germany

[†]Both senior authors contributed equally.

*Corresponding author: E-mail: jueschm@uni-muenster.de.

Associate editor: Norihiro Okada

Abstract

More than 150 Ma, the avian lineage separated from that of other dinosaurs and later diversified into the more than 10,000 species extant today. The early neoavian bird radiations most likely occurred in the late Cretaceous (more than 65 Ma) but left behind few if any molecular signals of their archaic evolutionary past. Retroposed elements, once established in an ancestral population, are highly valuable, virtually homoplasy-free markers of species evolution; after applying stringent orthology criteria, their phylogenetically informative presence/absence patterns are free of random noise and independent of evolutionary rate or nucleotide composition effects. We screened for early neoavian orthologous retroposon insertions and identified six markers with conflicting presence/absence patterns, whereas six additional retroposons established before or after the presumed major neoavian radiation show consistent phylogenetic patterns. The exceptionally frequent conflicting retroposon presence/absence patterns of neoavian orders are strong indicators of an extensive incomplete lineage sorting era, potentially induced by an early rapid successive speciation of ancestral Neoaves.

Key words: bird orders, Neoaves, phylogeny, retroposons, cumulative retroposon fixation probability, incomplete lineage sorting.

The early diversification of neoavian orders most likely occurred in rapid succession, and attempts to resolve the higher level phylogenetic relationships have yielded some concordance but much controversy (summarized in Mayr 2011).

Phylogenetic reconstructions based on sequence data are vulnerable to systematic errors induced by phenomena such as long branch attractions and biased nucleotide compositional effects. In contrast, retroposon presence/absence data are free of such uncertainties owing to their unlimited character states corresponding to an infinite potential to insert nearly anywhere in the genome (for mathematical background, see Steel and Penny 2000). Although retroposon presence/absence data are not immune to conflicting phylogenetic patterns when they are extracted under strict criteria (unambiguous orthology by accurately defined insertion sites, clear identity of the retroposed element with diagnostic truncations and mutations), revealed discords are virtually restricted to the rare cases of incomplete lineage sorting or gene flow via hybridization (Shedlock et al. 2004).

CR1 retroposons were successfully used as clade markers in birds (Watanabe et al. 2006; Kaiser et al. 2007; Kriegs et al. 2007; Treplin and Tiedemann 2007; Suh et al. 2011), and other retroposed elements successfully identified phylogenetic conflict zones in placental mammals (Churakov et al. 2009; Nishihara et al. 2009; Churakov, Sadasivuni, et al. 2010) and cichlid fishes (Takahashi et al. 2001). Recently,

a data set of bird retroposon markers was published that included four retroposons with inconsistent presence/absence patterns that inserted during the rapid radiation of Neoaves (Suh et al. 2011).

In the present study, we more thoroughly investigated conflicting retroposon presence/absence patterns originating during or close to the assumed rapid radiation period of neoavian birds (see [supplementary methods, Supplementary Material](#) online). We hypothesized that the presence of substantially more inconsistent presence/absence patterns would confirm the rapid ancestral radiation as characterized by insertion polymorphism and incomplete lineage sorting in birds (see also Shedlock et al. 2004; Poe and Chubb 2004; Churakov et al. 2009).

Of 161 experimentally investigated retroposon presence/absence candidate loci, 12 met the necessary stringent criteria for informative retroposon markers (see above). These were then experimentally investigated in 26 representative bird species from most of the major bird clades. Six of the 12 new retroposon loci exhibited inconsistent mosaic presence/absence insertion patterns that indicate ancestral polymorphism ([fig. 1](#)). The mosaics vary from single outliers (markers ZF19, Gy14) to more complex patterns (ZF42, ZF09, ZF10, ZF89). Interestingly, the ancestral polymorphisms also extended into apparently well-supported clades, such as the positioning of *Picus* and *Trogon* within higher landbirds ([fig. 2](#)). On the other hand, the retroposons that inserted before the earliest part of the neoavian radiation show

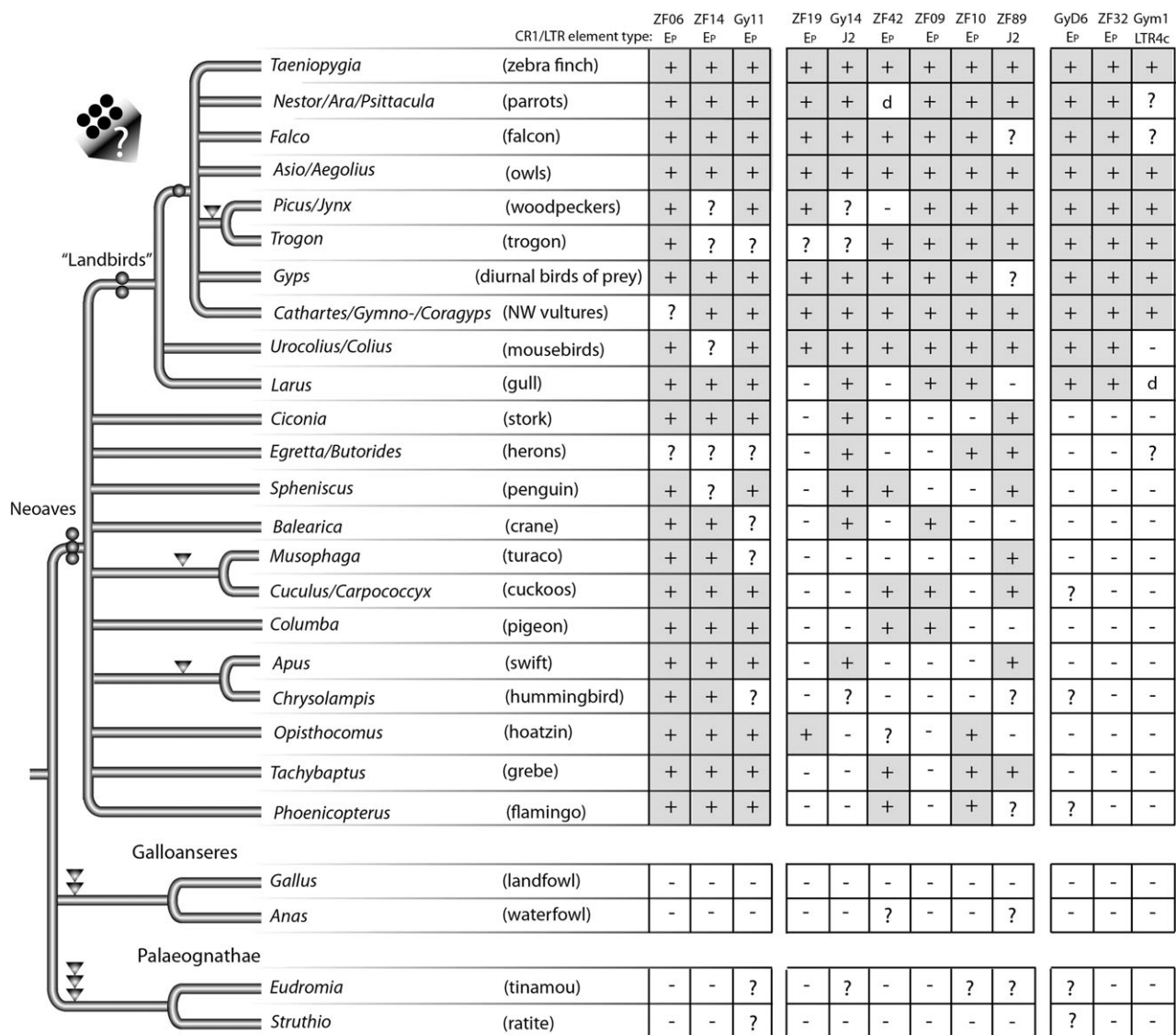


FIG. 1. Phylogenetic retroposon tree and presence/absence data matrix of birds. Phylogenetic reconstruction is based predominantly on the consistent insertion patterns of retroposed CR1 elements (gray circles) and random indels (insertions/deletions larger than 5 nt; triangles) found in this study. The table shows phylogenetically informative markers predating (left) and subsequent to the main neoavian radiation (right) and mosaic insertions (middle and black circles). Marker labels and element types are given above the data matrix. (+, shaded) Presence, (−) absence, (?) unknown state (d) large deletion including the element locus. For sequence data of diagnostic loci, see [Supplementary Material](#) online.

consistent and phylogenetically informative presence/absence patterns (ZF06, ZF14, Gy11) conforming to the well-accepted part of avian phylogeny (Hackett et al. 2008).

If the time of retroposon fixation overlaps succeeding speciation events, inconsistent patterns of insertions may appear and continue as relatively rare cases of ancestral polymorphism in the populations of the new species (Shedlock et al. 2004). Depending on the prevailing effective population size, several million years might be necessary to fix an inserted element (Kimura and Ohta 1969; Schmitz and Zischler 2002). Thus, polymorphic markers are of questionable phylogenetic value and should be considered carefully because they do not necessarily display the otherwise high reliability of retroposon-based phylogenetic reconstructions (for example, the parsimony reconstruction based on the

underlying data resulting in low bootstrap support and weak consistency indices; see [supplementary fig. S1, Supplementary Material](#) online). However, they can provide valuable information about the historical dynamics of populations during speciation and offer reliable evidence for incomplete lineage sorting or ancestral hybridization effects. The critical task is to distinguish such indicators of lineage sorting from misinterpretations of retroposon data that are sometimes introduced by applying relaxed stringency criteria for orthology and/or prejudiced definitions of species trees (e.g., Bashir et al. 2005; Han et al. 2011). We carefully inspected orthologous insertions for identical insertion sites, identical element types (in which small variations in subtype affiliation can occur caused by the deep divergences and occasionally short CR1 element fractions), and identical truncation

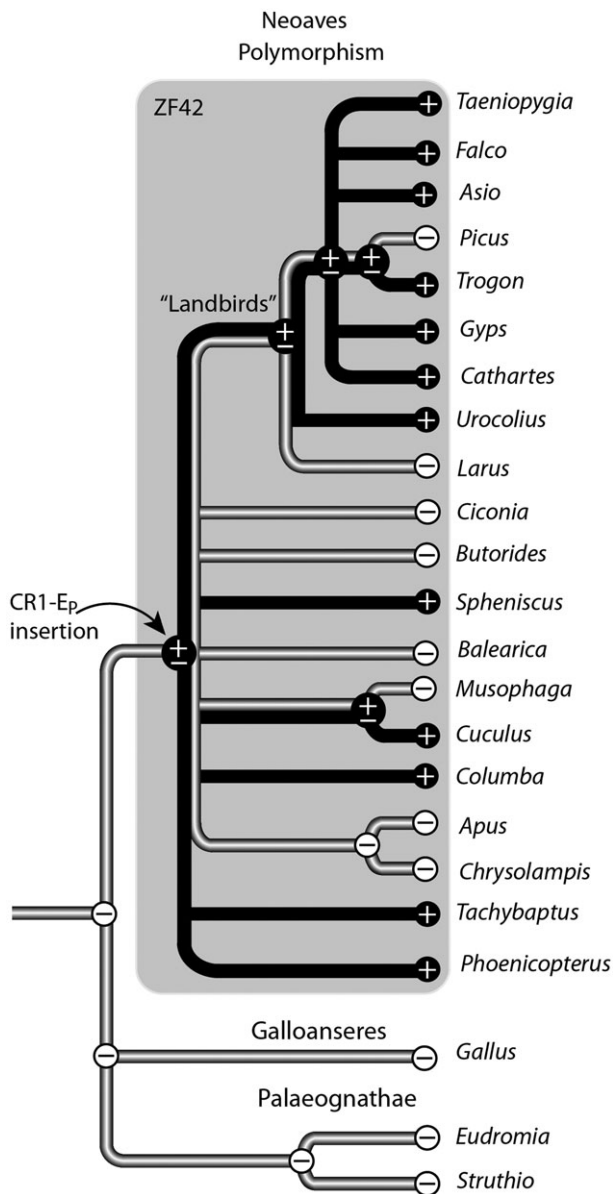


FIG. 2. Neoavian polymorphism of ZF42. The CR1-E_{Pass} element probably inserted shortly before speciation in the common ancestor of Neoaves. The polymorphic state (when a given element is both present and absent in the population of an ancestral lineage) is retained in ancestral Neoaves and in the relatively recent ancestral lineage of *Picus* and *Trogon* or *Musophaga* and *Cuculus*. (±) denotes the polymorphic situation that is further indicated with parallel gray and black branches. (+) Presence, (−) absence of the orthologous retroposon.

points (LINE-derived elements like CR1s usually truncate randomly during the process of insertion), as well as unoccupied sequences for their corresponding flanking sequence similarities.

Favored by incremental changes leading to small effective population sizes (and thereby rapid fixation during population bottlenecks) or/and long periods between speciation, many phylogenetically informative retroposon insertions were established in the early phase of neoavian evolution. This is indicated by a period of increased cumulative retroposon fixation probability. A graphical cumulative activity

pattern can be derived from Transposition in Transposition (TinT) data representing the chronological order of retroposon family activity (fig. 3, left peak; see also [supplementary figs. S2 and S3, Supplementary Material](#) online; for general information about TinT, see [Kriegs et al. 2007](#); [Churakov, Grundmann, et al. 2010](#)). Later, the survivors may have spread into available environments across huge parts of the globe, leading, for a relatively short time, to larger population sizes. Assuming that the individual retroposons distributed relatively constant during this time frame, such hypothetical conditions would have yielded elongated fixation times and the occurrence of polymorphic retroposon insertion patterns during fast speciation succession. The reduced probability of cumulative fixation is indicated in the zebra finch fixation pattern and contrasts with the processes in domestic fowl (fig. 3, saddle between peaks). Incomplete lineage sorting effects resulted in a variety of mutually contradictory retroposon mosaics, involving several active families of CR1 elements during ancient neoavian radiations. The unlikelihood of parallel insertions or exact deletions causing random phylogenetic noise, make the bird retroposon markers particularly suitable for identifying such phylogenetic lineage sorting areas ([Suh et al. 2011](#)). Another possible bottleneck or elongated internal branch in Passeriformes may lead to another significant cumulative retroposon fixation peak (fig. 3, right peak).

Indeed, the high percentage of discordant retroposons found in this study, together with the four described previously ([Suh et al. 2011](#)), provide evidence to support the rapid ancestral radiation of Neoaves as characterized by insertion polymorphism and incomplete lineage sorting. Historical variations in the effective population sizes might have played one key role in the observed phenomenon. Although there is little information available about these historic population sizes, the variable cumulative TinT patterns indicate an inverse relationship to the assumed historical fluctuations currently thought to have occurred (supposing a relatively constant generation time and constant activity of involved retroposons). It should be noted that, alternatively to the expected drastic change in population size and rapid radiation in ancient Neoaves, an extreme burst of retrotransposon activity might also explain the differences in the cumulative TinT pattern of the zebra finch. The appearance of many new retroposon subfamilies in both peaks of the zebra finch cumulative TinT could be interpreted in this direction. However, because of the virtually homoplasy-free character of retroposon presence/absence markers, they are valuable means of differentiating the random phylogenetic noise of sequence-based data from incomplete lineage sorting effects. They also help to identify phylogenetic issues that have a high chance of being reliably resolved by sequence-based data. However, due to the somewhat questionable nature of the polymorphic retroposon markers, the phylogeny of early neoavian birds remains partially unresolved, and only some internal branches (such as that of the landbird clade) stand out from the nebulous history of this group.

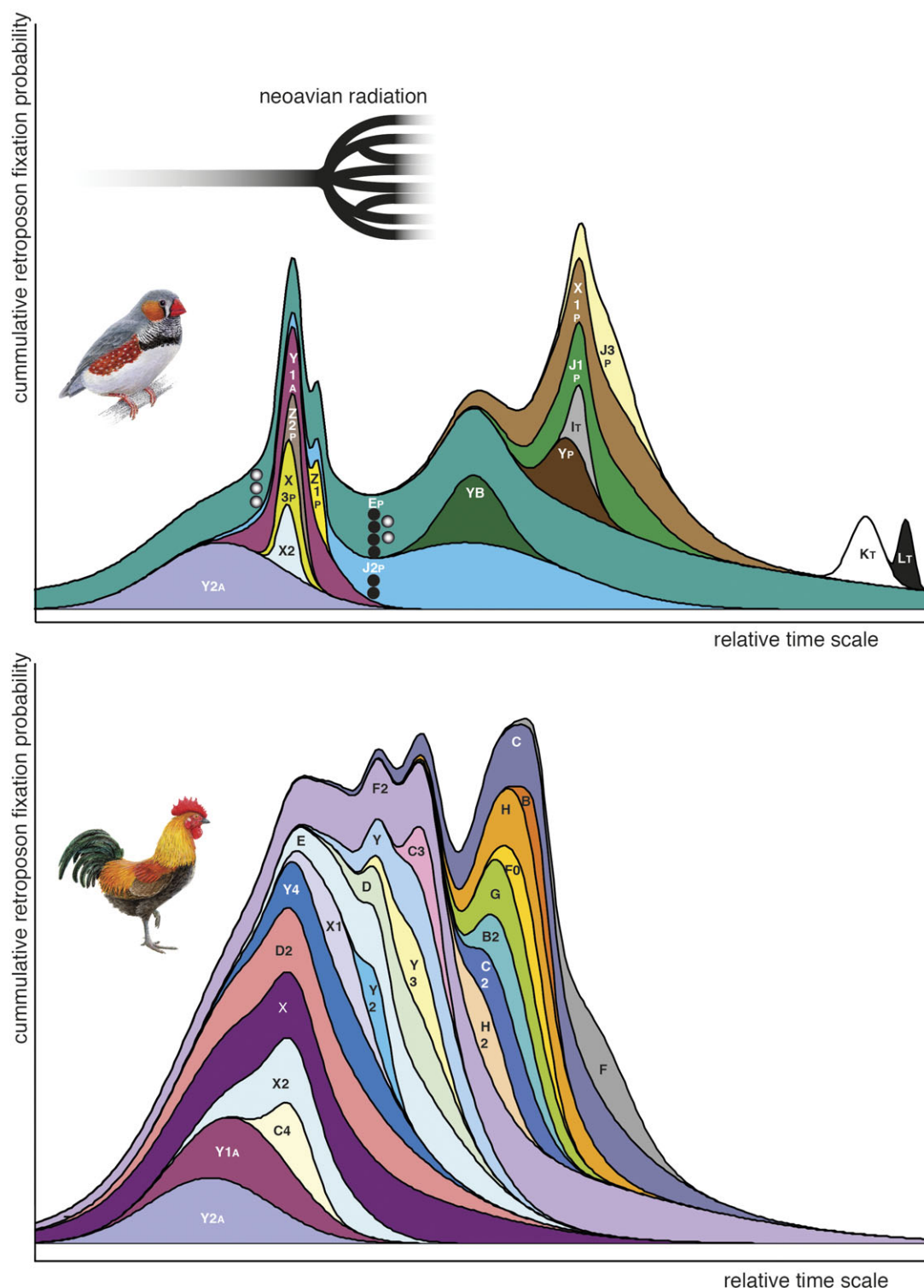


Fig. 3. Cumulative probability of CR1 retroposon fixation in zebra finch and domestic fowl. The classification of CR1 element families in zebra finch (top) and domestic fowl (bottom) are adopted from the RepeatMasker library (A = Aves, P = Passeriformes, T = Taeniopygia). CR1-Y2 and CR1-Y1 were active before the divergence of Neoaves and Galloanseres (see Suh et al. 2011). CR1-Y2 activity was used to calibrate the two cumulative curves, and a neognath-specific CR1-Y1 insertion from Suh et al. 2011 to set the left boundary for the preneoavian radiation. The domestic fowl cumulative element fixation profile indicates a homogenous accumulation, whereas the pattern for the zebra finch shows a fissure indicating a period of rapid radiation and reduced probability of retroposon fixation, flanked by periods of high probabilities of retroposon fixation. It is important to mention that these curves are optimized for CR1 elements (see [supplementary figs. S2 and S3, Supplementary Material](#) online). Gray circles show the probable location of phylogenetically informative retroposon markers; black circles indicate conflicting markers.

Supplementary Material

Supplementary methods, material, and figures S1, S2, and S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Gerald Mayr for his important comments on a previous version of the manuscript and Jón Baldur Hlíberg for providing the bird paintings. This work was supported by the Deutsche Forschungsgemeinschaft (grant numbers KR3639/1-1, SCHM1469/4-1). Additional support came from Conservation, Research and Education Opportunities (CREO), Seattle, USA, and the Universitätsklinikum Münster. We thank Marsha Bundman for editorial assistance and Martin Kieffmann, Rüdiger Dierkes, Lars Gaedicke, and Sven Niehues for assistance with polymerase chain reactions. We are very grateful for bird samples from: Gerald Mayr (Senckenberg Museum, Frankfurt, Germany), Christian Dietzen, Werner Beckmann, Ommo Hüppop, Sharon Birks (Burke Museum, Seattle), Sandra Silinski and Andrew Fidler (Allwetterzoo Münster, Germany), Renate van den Elzen (Museum König, Bonn, Germany), Michael Wink, Landschaftsverband Westfalen-Lippe-DNA- und Gewebearchiv (Naturkundemuseum Münster, Germany), Jos Custers, Holger Schielzeth, Herbert Grimm (Naturkundemuseum Erfurt), Franz Müller, Tang Thean Hock, and Gino Ooi Hong Chin (Penang Bird Park, Malaysia).

References

- Bashir A, Ye C, Price AL, Bafna V. 2005. Orthologous repeats and mammalian phylogenetic inference. *Genome Res.* 15:998–1006.
- Churakov G, Grundmann N, Kuritzin A, Brosius J, Makalowski W, Schmitz J. 2010. A novel web-based TinT application and the chronology of the Primate *Alu* retroposon activity. *BMC Evol Biol.* 10:376.
- Churakov G, Kriegs JO, Baertsch R, Zemmann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res.* 19:868–875.
- Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J. 2010. Rodent evolution: back to the root. *Mol Biol Evol.* 27:1315–1326.
- Hackett SJ, Kimball RT, Reddy S, et al. (18 co-authors). 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.
- Han KL, Braun EL, Kimball RT, et al. (17 co-authors). 2011. Are transposable element insertions homoplasy free?: an examination using the avian tree of life. *Syst Biol.* 60:375–386.
- Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol.* 24:338–347.
- Kimura M, Ohta T. 1969. The average number of generations until fixation of a mutant gene in a finite population. *Genetics* 61:763–771.
- Kriegs JO, Matzke A, Churakov G, Kuritzin A, Mayr G, Brosius J, Schmitz J. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol.* 7:190.
- Mayr G. 2011. Metaves, Mirandornithes, Strisores and other novelties—a critical review of the higher-level phylogeny of neornithine birds. *J Zool Syst Evol Res.* 49:58–76.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A.* 106:5235–5240.
- Poe S, Chubb AL. 2004. Birds in a bush: five genes indicate explosive evolution of avian orders. *Evolution* 58:404–415.
- Schmitz J, Zischler H. 2002. Molecular cladistic markers and the infraordinal phylogenetic relationships of primates. In: Kay RF, Ross C, editors. In anthropoid origins: new visions. New York: Kluwer Academic Press.
- Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol.* 19:545–553.
- Steel M, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol Biol Evol.* 17:839–850.
- Suh A, Paus M, Kieffmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.
- Takahashi K, Terai Y, Nishida M, Okada N. 2001. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. *Mol Biol Evol.* 18:2057–2066.
- Treplin S, Tiedemann R. 2007. Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picathartes*) to crows and ravens (*Corvidae*). *Mol Phylogenet Evol.* 43:328–337.
- Watanabe M, Nikaido M, Tsuda TT, Inoko H, Mindell DP, Murata K, Okada N. 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. *Gene* 365:57–66.