General palaeontology, systematics, evolution (Evolutionary patterns)

Birds perching on bushes: Networks to visualize conflicting phylogenetic signals during early avian radiation

Des oiseaux perchés sur des arbustes : réseaux pour visualiser les conflits phylogénétiques au cours de la radiation évolutive des oiseaux

Antonio Hernandez-Lopez, Didier Raoult, Pierre Pontarotti

Aix-Marseille université, LATP UMR CNRS 7352, « Évolution biologique et modélisation », case 19, 3, place Victor-Hugo, 13331 Marseille cedex 3, France

Unité des rickettsies, faculté de médecine, UMR CNRS 6020, 13385 Marseille, France

ARTICLE INFO

Article history:
Received 23 October 2012
Accepted after revision 20 February 2013
Available online 19 April 2013

Presented by Philippe Taquet

Keywords:
Adaptive radiation
Reticulate evolution
Phylogenetic networks
Transposable elements

ABSTRACT

Hybridization is increasingly seen as an important source of adaptive genetic variation and biotic diversity. Recent phylogenetic studies on the early evolution of birds suggest that the early diversification of neoavian orders perhaps involved a period of extensive hybridization or incomplete lineage sorting. Phylogenetic error, saturation, long-branch attraction, and convergence make it difficult to detect ancient hybridization events and differentiate them from incomplete lineage sorting using sequence data. We used recently published retroposon marker data to visualize the early radiation of Neoaves within a phylogenetic network approach, and found that the most basal neoavian taxa indeed show a complex pattern of reticulated relationships. Moreover, the reticulation levels of different parts of the network are consistent with the insertion pattern of the retroposon elements. The use of network-based analyses on homoplasy-free data shows true conflicting signals and the taxa involved that are not represented in trees.


RÉSUMÉ

L’hybridation est de plus en plus perçue comme une importante source de variations génétiques adaptatives et de diversité biologique. Des études phylogénétiques récentes sur l’évolution des oiseaux suggèrent que la diversification précoce des ordres de Neoaves ait peut-être compris une période d’hybridation extensive ou de tri incomplet des lignées. En raison des erreurs phylogénétiques, de la saturation, de l’attraction des branches longues et de la convergence, il est très difficile de détecter des événements anciens d’hybridation et de les différencier du tri incomplet des lignées avec des données de séquence. Nous avons utilisé des marqueurs retroposons récemment publiés pour visualiser le rayonnement ancien des Neoaves, en utilisant une approche par réseau phylogénétique. On a constaté que les taxons les plus basaux des Neoaves montrent, en effet, un ensemble complexe de relations reticulées. De plus, les niveaux de réticulation des différentes parties du réseau sont compatibles avec le modèle d’insertion des éléments retroposons. L’utilisation des analyses de réseaux plus sophistiqués et rigoureux, basés sur des données sans homoplasie, va certainement aider à détecter des événements d’hybridation anciens et à les différencier du tri incomplet des lignées.


* Corresponding author.
E-mail address: antonio.hernandezlopez@univ-provence.fr (A. Hernandez-Lopez).


http://dx.doi.org/10.1016/j.crpv.2013.02.003
1. Trees vs. networks

The use of bifurcating trees as evolutionary models has greatly facilitated the discussion and testing of phylogenetic and evolutionary hypotheses. However, it is well known that such models poorly describe more complex evolutionary scenarios, particularly when reticulate events such as horizontal gene transfer, gene duplication, or recombination are involved (Huson and Bryant, 2006). Hybridization is of particular importance because gene flow between species can bring together adaptive variation from divergent genetic lineages for selection to act upon. Although the extent of its evolutionary significance remains controversial (Seehausen, 2004), it is increasingly seen as an important process for generating biotic diversity (Arnold, 1997; Ferguson and Sang, 2001; Rieseberg et al., 2003) and as a comparable, or even greater, source of adaptive genetic variation than mutation (Arnold and Martin, 2009; Grant and Grant, 1996; Kunte et al., 2011). Because hybridization is more common between closely related species, species-rich groups that are undergoing rapid adaptive radiations are those that hybridize most (Gourbière and Mallet, 2010; Price and Bouvier, 2002; Seehausen, 2004). If hybrid speciation takes place commonly in natural populations, it might also have played a major role in the origin of divergence of both major lineages and individual species. Hybridization is often inferred from incongruence among gene trees from independent loci (Fig. 1B; Linder and Rieseberg, 2004) or from trees in which species are not monophyletic (Funk and Omland, 2003). However, it is extremely difficult to detect the signal of a basal hybrid gene flow deep in a phylogeny using DNA sequence data, because phylogenetic reconstruction errors and artefacts based on sequence data induced by phenomena such as long-branch attraction, biased nucleotide compositional effects, saturation and perhaps even convergence due to selection, make distinguishing any weak remaining signal very difficult.

Processes other than hybridization, such as gene duplication (Small et al., 2004) and incomplete lineage sorting (Buckley et al., 2006; Holder et al., 2001; Huson et al., 2005; Joly et al., 2009; Machado et al., 2002; Sang and Zhong, 2000; Wang et al., 1997), can also produce incongruent gene–species phylogenies (Fig. 1A). Gene duplication is of more concern for recent divergences, because duplicated genes tend to diverge rapidly, and end up becoming sub-functional, acquiring new functions, or losing function and being lost. Incomplete lineage sorting refers to the retention and incomplete random sorting of alleles at many loci independently because the intervals between divergence events were brief (Fig. 1C). This is more likely if the effective population sizes are large relative to the divergence time after a speciation event. In this case, genetic drift is unlikely to have time to bring genes to fixation before subsequent divergences (Pamilo and Nei, 1988).

Retroposons (RTs) are highly repetitive mobile sequences that transpose into eukaryotic DNA through RNA intermediates. In many species, RTs are a ubiquitously dispersed feature of the whole genome, often constituting a significant (~ 10%) mass fraction of total DNA. RTs cause mutations both by their retro-transposition within genes and by unequal recombination, and are widely considered examples of ‘selfish’ or ‘parasitic’ DNA. RTs have proven to be excellent markers for phylogenetic analysis: their mode of evolution is predominantly homoplasy-free, because they do not insert in the same locus of two unrelated lineages, and unidirectional, because they are not precisely excised from a locus with the flanking sequences preserved (Shedlock and Okada, 2000). The use of RTs for diagnosing common ancestry is being expanded to examine the differential sorting of lineages through the course of speciation events. They have been used to show that hippopotamuses are the closest living relatives of whales (Nikaido et al., 1999; Shimamura et al., 1997), to determine phylogenetic relationships among cichlid fishes (Takahashi et al., 2001; Terai et al., 2003), and to elucidate the phylogeny of eight Primate species, providing the strongest evidence yet that chimps are the closest living relatives of humans (Salem et al., 2003).

Incongruent insertion patterns revealed by ancient fixed RT loci contain information that can be useful in detecting cases of incomplete lineage sorting/hybridization that are otherwise obscured by accumulated mutations, homoplasy and phylogenetic uncertainty in sequence data (Churakov et al., 2009; Shedlock et al., 2004). Among vertebrates, birds are particularly prone to hybridization (reviewed and summarized in Mallet, 2005). Phylogenetic evidence suggests that the early diversification of neovian orders most likely occurred in a rapid adaptive radiation more than 65 million years ago (Hackett et al., 2008; Haddrath and Baker, 2012; Matzke et al., 2012; Poe and Chubb, 2004; Suh et al., 2011), and that rapid adaptive radiations have been abundant during both early and ongoing bird radiations (Jetz et al., 2012). Attempts to resolve the higher-level phylogenetic relationships have resulted in contradictory or unresolved relationships at deep nodes (Mayr, 2011). RTs have been useful to clarify relationships between game birds (Kriegs et al., 2007), and in a recent study, Matzke et al. (2012) used chicken repeat 1 (CR1) RT markers to elucidate the early radiation of Neoaves. Because there is a high percentage of discordant retroposons found in neovian orders, the authors suggest a rapid ancestral radiation of Neoaves characterized by insertion polymorphism and incomplete lineage sorting (and perhaps hybridization). Phylogenetic networks have an important role to play in the representation and eventual reconstruction of evolutionary history when applied to RT-type data. Implicit models such as split networks are very useful for exploring and visualizing the different signals in a data set, whereas explicit models such as hybridization and recombination networks can be used to model reticulate evolution.

Median networks, a type of unrooted phylogenetic networks, are particularly suited to visualize conflicting data in a set of binary characters on a set of taxa, such as the presence/absence of CR1 insertion data. In a median network, every sequence of characters is represented by a node, and additional nodes represent unobserved characters (i.e. insertion patterns). Two nodes are connected by an edge only if they differ by one character (as when used to visualize haplotype relationships). We used the 12 retroposition markers used by Matzke et al. (2012) and the
four conflicting markers used by Suh et al. (2011) to construct a median-joining split network as implemented in SplitsTree4, with count-splits filter. In this case, the CR1 presence/absence data are analogous to a group of haplotypes, and the network represents the relationships among them (i.e. each “haplotype” is the insertion pattern of the 16 markers for each species).

2. Taxa with conflicting patterns of chicken repeat 1 insertions

The network has three regions with different levels of reticulation: the ratites–tinamou–Galloanseres branch and the “landbirds” branch with low reticulation, and the basal Neoaves species with high reticulation (Fig. 2). Six retroposon loci published by Matzke et al. (2012) exhibited inconsistent mosaic presence/absence insertion patterns that indicate ancestral polymorphisms. These polymorphic markers are of questionable phylogenetic value, but can provide valuable information about the historical population dynamics during speciation and offer compelling evidence for incomplete lineage sorting or ancestral hybridization effects (Matzke et al., 2012). The high reticulation (higher number of reticulate nodes, as opposed to bifurcating nodes) in the Neoaves part of the network (Fig. 2) corresponds to the polymorphic mosaic loci, and is consistent with a rapid radiation scenario. But, whereas in the phylogenetic tree of Matzke et al., the more basal species in the Neoaves group show comb-like unresolved relationships, they are grouped in the higher reticulation part of the network, and there are clearly different reticulation levels within Neoaves (retroposon insertions mark dividing nodes between reticulated/less-reticulated parts of the network, i.e. the “Landbirds” divide; Fig. 2 following Figure 1 in Matzke et al., 2012).

Because the probability of RT elements transposing twice independently at the same position is negligible, and the precise loss of an entire retroposon after its fixation in the genome is extremely rare and very unlikely to happen in multiple genomes, incomplete lineage sorting was the most parsimonious conclusion derived from the contradictory retroposon insertion patterns (Matzke et al., 2012). Conflicts in ancient population relationships can be better represented, in terms of taxa involved and levels of
reticulation (i.e. conflict), by the representation as a "haplotype" network, as shown on Fig. 2. This pattern can better represent the complicated radiation during the early radiation of Neoaves, and how it slowed down and eventually disappeared along the edge leading to Passerines (zebra finch), parrots and falcons. Although a simple analysis such as the one we used here cannot definitely distinguish between incomplete lineage sorting and hybridization, it does highlight the extent of taxa involved in reticulate evolution. The implementation of more accurate hybridization and phylogenetic network analyses (Holland et al., 2008; Yu et al., 2011, 2012) on presence/absence data (i.e. free from homoplasy) could help achieve a more precise differentiation of hybridization versus incomplete lineage sorting, and further exploration along these lines is certainly needed to shed light on ancient adaptive radiations underlying major groups. However, given that:

- hybridization is common in a wide range of contemporary species (Mallet, 2005);
- adaptive radiations have been, and still are, widespread in all bird groups (Jetz et al., 2012), hybridization could also be responsible for the observed reticulation patterns of RT insertions during an ancient adaptive radiation.

Fig. 2. Median-joining split network for 16 chicken repeat 1 retroposon presence/absence data. Neoaves and "Landbirds" clades are delimited by the horizontal and transversal lines, respectively, according to Figure 1 in Matzke et al. (2012). Grey circles represent the insertion patterns of retroposed chicken repeat 1 elements corresponding to Matzke et al. (2012) reconstruction.

Fig. 2. Median-joining split network pour les données de présence/absence des marqueurs rétroposons. Le groupe des Neoaves et le groupe des oiseaux terrestres, correspondant à ceux de la Figure 1 dans Matzke et al. (2012), sont délimités par les lignes horizontales et transversales, respectivement. Les cercles gris représentent les points d'insertion des éléments chicken repeat 1 correspondant à la reconstruction de Matzke et al. (2012).

References


Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mito-


