Molecular Phylogenetics and Evolution 116 (2017) 69–77

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

The development of scientific consensus: Analyzing conflict and concordance among avian phylogenies

Joseph W. Brown*, Ning Wang, Stephen A. Smith*

Department of Ecology & Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

ARTICLE INFO

Keywords:
Phylogenetic conflict
Concordance
Aves
Supertrees

ABSTRACT

Recent developments in phylogenetic methods and data acquisition have allowed for the construction of large and comprehensive phylogenetic relationships. Published phylogenies represent an enormous resource that not only facilitates the resolution of questions related to comparative biology, but also provides a resource on which to gauge the development of concordance across the tree of life. From the Open Tree of Life, we gathered 290 avian phylogenies representing all major groups that have been published over the last few decades and analyzed how concordance and conflict develop among these trees through time. Nine large scale phylogenetic hypotheses (including a new synthetic tree from this study) were used for comparisons. We found that conflicts were over-represented both along the backbone (higher-level neoavian relationships) and within the oscine Passeriformes. Importantly, although we have made major strides in the resolution of major clades, recent published comprehensive trees, as well as trees of individual clades, continue to contribute significantly to the resolution of relationships throughout the avian phylogeny. Our analyses highlight the need for continued research into the resolution of avian relationships.

1. Introduction

Large and comprehensive phylogenies (i.e., including hundreds of taxa and based on genome-scale datasets) have become more common as inference methods and sequencing techniques capable of constructing enormous datasets have been developed (e.g., Smith and Donoghue, 2008; Rabosky et al., 2013; Zanne et al., 2014; Prum et al., 2015; Simion et al., 2017). These phylogenies have, in many cases, given fresh views to macroevolution and transformed our ability to address diverse sets of comparative biological questions ranging from lineage diversification to morphological evolution to rate heterogeneity (Brockington et al., 2015; Lin et al., 2016; Scholl and Wiens, 2016). Comprehensive phylogenies that include all or nearly all taxa constructed from supertree techniques also provide a means of determining where data collection efforts should be focused (Davis and Page, 2014; Jetz et al., 2012; Hinchliff et al., 2015). While these trees may facilitate interesting biological inquiries, they also provide a resource by which we can better assess the development of congruence among evolutionary hypotheses (e.g., Davis and Page, 2014; Suh, 2016; Reddy et al., 2017). Recent efforts to better understand the development of conflict and concordance among trees have been conducted primarily with molecular data (e.g., Hinchliff and Smith, 2014; Smith and Stamatakis, 2013; Smith et al., 2015). Nevertheless, phylogenetic resources, including TreeBASE (Sanderson et al., 1994) and more recently the Open Tree of Life (Hinchliff et al., 2015; McTavish et al., 2015), are now available to better analyze both the novelty and congruence of inferred relationships across the tree of life across studies.

The Open Tree of Life is an NSF funded project whose aim is to construct a comprehensive tree of life using published phylogenetic trees along with taxonomic data (Hinchliff et al., 2015). To facilitate this research, the Open Tree of Life has developed and provided the community with several important resources. The Open Tree Taxonomy (hereafter OTT; Rees and Cranston, 2017), unlike many other synthetic taxonomies available, attempts to include only phylogenetically appropriate taxa (i.e., through exclusion of names of dubious taxonomic status). It is also more comprehensive than other more commonly used taxonomies (e.g., NCBI) as it includes taxa regardless of whether they have molecular data associated. The Open Tree of Life also constructs and serves a draft synthetic tree of all described species (Hinchliff et al., 2015), through the grafting of OTT together with published trees identified, uploaded, and curated by the community. This resource, while continually improving, provides significant opportunities to address broad evolutionary questions that previously would have been impossible. Finally, the project also openly provides the database of published phylogenies that have been curated by the community (McTavish et al., 2015). Importantly, the taxa included in each

---

* Corresponding authors.
E-mail addresses: josephwb@umich.edu (J.W. Brown), eebsmith@umich.edu (S.A. Smith).

http://dx.doi.org/10.1016/j.ympev.2017.08.002
Received 31 March 2017; Received in revised form 3 August 2017; Accepted 6 August 2017

phylogeny have been mapped to a common taxonomy (i.e., OTT), which allows for comparisons to be performed across datasets without an additional tedious and error prone step of name reconciliation. Instead, this reconciliation has already been performed by those who uploaded the tree, often researchers with close knowledge of the focal organisms.

Here, by utilizing the database of curated phylogenies from the Open Tree of Life, we assess the concordance and conflict among the growing number avian phylogenies that have been published during the last few decades. Methods that are used in this study can also be applied to other living groups on Earth based on the Open Tree of Life resources. As the most diverse extant tetrapod lineage with ~10,800 recognized extant species (Gill and Donsker, 2016) [and potentially more than twice as many cryptic lineages; Barrowclough et al. (2016)], birds have experienced a rapid inter-ordinal radiation where extremely short internodes exist (Hackett et al., 2008; McCormack et al., 2013; Burleigh et al., 2015; Suh, 2016; Reddy et al., 2017). Although substantial progress has been made on reconstruction of the Aves phylogeny, discovering successive divergence of three monophyletic groups [i.e., Palaeognathae (the tinamous and flightless ratites), Galloanserae (game birds and waterfowl), and Neoaves (all other living birds), Groth and Barrowclough, 1999; Cracraft et al., 2004], resolving the avian phylogeny (especially within Neoaves) has continued to prove a difficult task for the avian systematics community since the pioneering efforts of Sibley and Ahlquist (1990).

Researchers have started to explicitly assess progress in avian phylogenetics. By constructing a consensus tree based on six genome-scale phylogenies from five independent studies (i.e., Hackett et al., 2008; McCormack et al., 2013; Jarvis et al., 2014; Suh et al., 2015; Prum et al., 2015), Suh (2016) assessed the reproducibility of various avian phylogenetic hypotheses. Due to the overwhelming conflict among the source trees used (i.e., no higher-level clade could be supported by at least two out of the six trees), Suh (2016) suggested that the very onset of the neoavian radiation produced an irresolvable nine-taxon hard polytomy. Reddy et al. (2017) constructed a nearly identical summary consensus tree to Suh (2016) using a smaller sample of three major hypotheses (i.e., Jarvis et al., 2014; Prum et al., 2015; Reddy et al., 2017), but were more optimistic that more realistic biological-modelling and, importantly, careful selection of data types, will enable further progress. We note that none of the trees considered by Suh (2016) or Reddy et al. (2017) had sufficient sampling of Passeriformes (songbirds; roughly 60% of extant avian species), so conflict could not be ascertained within that clade. To date, these and other studies have mainly focused on identifying causes of conflict, attributing tree differences to various factors including gene tree discordance due to incomplete lineage sorting (ILS; Jarvis et al., 2014; Suh et al., 2015), differences in phylogenetic signal content among data types (Jarvis et al., 2014; Reddy et al., 2017), and the influence of taxon sampling (Prum et al., 2015). However, while these issues of inference are important to keep in mind for future research, little effort has been made to summarize the development and growth of consensus when considering the entire corpus of published phylogenetic hypotheses.

In this study, eight large-scale avian trees (Table 1) published in different time intervals are used as exemplars to assess trends of concordance and conflict. Additionally, after filtering 290 avian source trees publicly available from the Open Tree of Life, we constructed a new comprehensive synthetic bird tree and use it for assessment as the largest avian tree to date. This synthetic tree also serves as a resource for other researchers, and as a summary point from which we can compare future comprehensive avian phylogenies.

2. Methods

2.1. Source trees

Avian phylogenetic hypotheses that have been published in the last few decades were curated through the Open Tree of Life online curator (https://tree.opentreeoflife.org/curator), following the protocol of Hinchliff et al. (2015). Generally, published trees (as newick, NEXUS, or NeXML format) were obtained by appealing to authors, or imported from TreeBASE (Sanderson et al., 1994) and Dryad. We attempted to incorporate the source trees from the Davis and Page (2014) supertree study. However, we found that many trees from this resource were some form of consensus hypothesis (e.g., between parsimony and maximum likelihood) and/or included unsampled taxa (both extinct and extant) from the Davis and Page (2014) taxonomy. In sum, these trees reflected neither a specific hypothesis nor the extent of sampling of the original publication, and so were not included here. The full species-level tree of Sibley and Ahlquist (1990) has, to our knowledge, never been available in electronic format. As part of this study, JWB constructed the tree with branch lengths from Figs. 357–368, 371–385 of Sibley and Ahlquist (1990); this is the UPGMA tree commonly known as the “Sibley-Ahlquist Tapestry”, and is now freely available from the Open Tree of Life curator (study id: ot_427, tree id: tree5). The taxon labels for each source tree sampled here were mapped to the Open Tree of Life taxonomy (i.e., OTT) and trees were rooted with outgroups identified from the original study.

In total, 290 avian phylogenetic hypotheses were gathered from the existing resources in the Open Tree of Life. These are all openly available in the git-based phylesystem repository (McTavish et al., 2015; https://github.com/OpenTreeOfLife/phylesystem). The distribution of trees sampled through time (Fig. 1) reflects data availability rather than research effort, as historically phylogenetic hypotheses have not been archived in machine-readable formats (Stoltzfus et al., 2012; Drew et al., 2013). Among the sampled trees, seven major hypotheses were used as focal trees for the assessment of concordance and conflict against the remaining tree set: Sibley and Ahlquist (1990), Livezey and Zusi (2007), Hackett et al. (2008), Jetz et al. (2012), Davis and Page (2014), Jarvis et al. (2014), and Prum et al. (2015). We note that we used the Jetz et al. (2012) tree limited to taxa with genetic data (6670 tips) and constrained based on the Hackett et al. (2008) backbone. In addition to the seven trees above, the Open Tree of Life synthetic tree version 7 (hereafter Opentree7, updated in Sep 2016; https://tree.opentreeoflife.org/about/synthesis-release/v7.0) was also included as one of the backbone resources (see Table 1 for a summary of the properties of these trees).

Source trees may not be independent of each other. For example, the datasets used to construct phylogenies may have partial overlap (e.g., Wang et al., 2011; Kimball et al., 2013) or datasets can share constraints (e.g., the Jetz et al. (2012) tree has a backbone constraint based on Hackett et al. (2008)). We attempted to minimize these non-independent comparison as they may cause overestimation of conflicts or concordance. To this end, we filtered the source trees by including only one tree from each study (to avoid largely overlapped or same trees from the same study). However, and despite these efforts, we note that overlap can hardly be avoided among studies (especially given the high frequency use of certain genes). So, while we try to avoid this as much as was possible, there are likely to be non-independent edges between trees that were included in these analyses.

2.2. Construction of a new synthetic tree of Aves

In addition to the individual phylogenetic trees that we collected from the Open Tree of Life, we also assembled a novel synthetic avian tree using the “propinquity” pipeline from Redelings and Holder (2017). This supertree method takes as input a taxonomy tree (i.e., OTT) and a set of ranked source trees. Of the 290 avian trees collected above, 183 were selected that reflect community consensus about phylogenetic hypotheses. In general, the propinquity method constructs a supertree that displays the largest number of input tree edges while avoiding the inclusion of edges in the final tree that are unsupported by any input phylogeny. Because synthesis relies upon supertree