Structure versus time in the evolutionary diversification of avian carotenoid metabolic networks

ERIN S. MORRISON & ALEXANDER V. BADYAEV
Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA

Keywords:
functional module;
metabolic network evolution;
modularity;
network structure.

Abstract
Historical associations of genes and proteins are thought to delineate pathways available to subsequent evolution; however, the effects of past functional involvements on contemporary evolution are rarely quantified. Here, we examined the extent to which the structure of a carotenoid enzymatic network persists in avian evolution. Specifically, we tested whether the evolution of carotenoid networks was most concordant with phylogenetically structured expansion from core reactions of common ancestors or with subsampling of biochemical pathway modules from an ancestral network. We compared structural and historical associations in 467 carotenoid networks of extant and ancestral species and uncovered the overwhelming effect of pre-existing metabolic network structure on carotenoid diversification over the last 50 million years of avian evolution. Over evolutionary time, birds repeatedly subsampled and recombined conserved biochemical modules, which likely maintained the overall structure of the carotenoid metabolic network during avian evolution. These findings explain the recurrent convergence of evolutionary distant species in carotenoid metabolism and weak phylogenetic signal in avian carotenoid evolution. Remarkable retention of an ancient metabolic structure throughout extensive and prolonged ecological diversification in avian carotenoid metabolism illustrates a fundamental requirement of organismal evolution – historical continuity of a deterministic network that links past and present functional associations of its components.

Introduction
Over the past decade, two related findings have greatly influenced our understanding of evolutionary process. First is the realization of remarkable discordance of divergence between phenotypic and genomic levels of organization (Tatusov et al., 2003; Koonin, 2007; Povolotskaya & Kondrashov, 2010), and second, that only a small proportion of theoretically possible paths are traversed during organismal evolution (Raff, 1996; Carroll, 2005; Bershtein et al., 2006; Poelwijk et al., 2007; Harms & Thornton, 2014). These observations corroborate the visionary prediction (Maynard Smith, 1970) that dual requirements must be fulfilled for evolution to occur – historical continuity of a deterministic network that links past and present functional associations of adaptation components and survivability of intermediate steps (Schmalhausen, 1938; Gould, 2002; Gavrilets, 2004; Wagner, A., 2011; Wagner, G.P., 2014; Sarkisyan et al., 2016). This perspective and empirical observations contribute to the conceptual unification of the ‘structuralist’ and ‘selectionist’ views of evolution, because the patterns of connectivity among elements of a deterministic landscape (e.g. gene or metabolic networks) that underlie trait variation are formed by past functional and physical associations between these elements. It follows that these historical associations necessarily delineate the pathways available for subsequent evolution.
The dependency of evolutionary trajectories on the prior architecture of traits is particularly striking in the evolution of metabolic networks, where divergence of homologous networks preserves and expands on the structure of the preceding network. Structural properties of a metabolic network commonly scale with its expansion: compounds associated with the most enzymatic reactions preferentially gain new reactions (Barabási & Albert, 1999; Eisenberg & Levanon, 2003; Light et al., 2005; Wang et al., 2009), compound connectivity (the number of reactions per compound) is nearly universally associated with its evolutionary retention (Jeong et al., 2000; Fraser et al., 2002; Kunin et al., 2004; Liu et al., 2007; Bernhardsson et al., 2011), and many ancient biochemical pathways form modules comprised of interconnected reactions and compounds that persist intact through extensive diversifications (Wuchty et al., 2003; Campillos et al., 2006; Yamada et al., 2006; Peregrín-Alvarez et al., 2009; Wagner, 2009).

A prime example of these principles is carotenoid metabolism – an integral part of numerous biological functions from photosynthesis, to immunity, pigmentation and vision (Goodwin, 1980, 1984). First, core carotenoid metabolic pathways are highly conserved across prokaryotes and eukaryotes: homologous amino acid sequences and enzymes remain virtually unchanged between bacteria, where most carotenoid reactions originate (Vershinin, 1999), and plants and animals (Sandmann, 2002; Young et al., 2012; Badyaev et al., 2015). Second, carotenoid metabolism is hierarchical – species-specific branching pathways expand from highly conserved basal reactions involved in the initial assembly of carotenoid molecules from precursors (Umeno et al., 2005). Further, animals lack enzymes that synthesize carotenoids from noncarotenoids (Goodwin, 1984, 1986), necessitating that the basal pathways that link organismal carotenoid metabolism to environmentally available carotenoids have to be preserved in evolution in order to retain the ability to diversify and modify these carotenoids (Davies, 1985). Yet, despite this hierarchical structure and dependency on conserved basal pathways, the evolution of metabolic networks does not follow phylogenetic history in any of the animal taxa where it has been studied [e.g. fish: Seif et al. (2014), birds: Badyaev et al. (2015), insects: Kayser (1982)]. The primary reason for this is the small size and dense connectivity of the biochemical network on which animal carotenoid diversification takes place, leading to frequent convergence in biochemical pathways across taxa (Badyaev, 2007; Morrison & Badyaev, 2016a). For example, in birds, even large ecological divergence in dietary carotenoids – the starting point of carotenoid metabolism – does not preclude convergence in derived carotenoids because many of them can be reached from several dietary starting points (Badyaev et al., 2015; Fig. S1).

An explicit assumption for this explanation is that species retain access to carotenoid pathways they do not express, such that a lineage essentially subsamples a larger, pre-existing carotenoid network during its evolution (Maslov et al., 2009). Although there is growing empirical evidence for this assumption – for example, birds retain far more carotenoid compounds and reactions that they express in their integument (e.g. Walsh et al., 2012) – there are no formal tests. Such tests are proposed here.

We envision two alternative scenarios of pathway expansion in carotenoid network evolution (Fig. 1). First, diversification of carotenoid metabolism may proceed by the evolutionary gain or loss of enzymatic reactions that build on the existing structure of an ancestral network (Tanaka et al., 2006; Klassen, 2010), by the addition of novel reactions to the end of pathways (Horowitz, 1945; Szappanos et al., 2016), or by the branching of enzymatic pathways as the result of gene duplication, enzyme promiscuity, and subfunctionalization within pre-existing, functionally independent pathway modules (Jensen, 1976; Hartwell et al., 1999; Ravasz et al., 2002; Khersonsky et al., 2006). Over evolutionary time, this process will amount to either a phylogenetically structured expansion of the network from a set of ancestral core reactions or a phylogenetically structured exploration of the ancestral network. Under this scenario, the effect of past biochemical network structure should decrease with a lineage’s ecological and evolutionary diversifications, with reciprocally increasing effects of the species’ history of ecological specialization (Fig. 1b).

Alternatively, species may diverge by expressing different combinations of pre-existing biochemical modules of compounds and reactions (Hartwell et al., 1999; Peregrín-Alvarez et al., 2003, 2009; Zhang et al., 2006; Maslov et al., 2009), such that the network structure within each module would be conserved, whereas the combination of expressed modules would vary across species as a result of either differences in the regulation of enzyme activities among individual pathways (e.g. Lavington et al., 2014) or the acquisition of different dietary compounds (e.g. Badyaev et al., 2015). Under this scenario, the structural properties of a metabolic network should persist throughout ecological and evolutionary diversifications, because the connectivity between enzymes and compounds within each module remains intact (Fig. 1c). This scenario also accounts for large-scale shifts in carotenoid metabolism across the network as opposed to gradual, phylogenetically structured changes that would occur with the evolutionary expansion of functional modules under the first scenario.

To test these scenarios, we examined the extent to which the evolutionary diversification of reactions in carotenoid-producing networks of ornamental plumage of birds is consistent with the evolution of pathways...
versus subsampling of functional modules from a conserved metabolic network. Using carotenoid networks of 250 species from nine avian orders and networks of 217 ancestral taxa spanning 50 MYR of avian evolution, we tested whether the frequency of compound co-occurrence in species changed across evolutionary time periods in relation to the pathway lengths (number of reactions) between compounds (Fig. 1a). Modules in metabolic networks are comprised of groups of closely connected compounds separated by only a few reactions that are always co-expressed (Ravasz et al., 2002; Resendis-Antonio et al., 2012). If metabolic networks are derived from subsampling structurally conserved modules, we would expect pairs of compounds connected by short pathways to co-occur more frequently than pairs of compounds separated by longer pathways in both ancestral networks and networks of extant species (see below, Fig. 1c). The expansion of enzymatic pathways between ancestral and descendant species (e.g. Ebenhöh et al., 2004) would erase the influence of a common network structure, because as new reactions are added, compounds separated by longer pathways would co-occur as frequently as compounds separated by shorter pathways in descendant species (Fig. 1b).

Materials and methods

Carotenoid metabolic networks of 250 species of birds (Appendix S1) were constructed based on an established universal metabolic network of carotenoid enzymatic reactions that connect the compounds identified in a species’ plumage (Badyaev et al., 2015; Morrison & Badyaev, 2016a). First, we collected an exhaustive list of all known compounds found in the plumage, plasma, and integument of a species, and these were mapped on the ‘avian space’ of the global carotenoid metabolic network. For each species, we recorded all of the biochemically possible reactions that linked starting dietary compounds to plumage compounds (Appendix S1). For species in which some of the dietary or intermediate compounds were not known, missing compounds and reactions between plumage compounds were assigned based on the location of the plumage compounds in the global network (e.g. an intermediate compound added between a known
dietary and a known expressed compound on a linear pathway; see Badyaev et al., 2015 for more details). Metabolic networks were only built for species with identified plumage carotenoids.

An ultrametric 50% majority-rule consensus tree of 243 species from 1,000 trees randomly sampled from the pseudo-posterior distribution of the Hackett All Species supertree (Jetz et al., 2014) was constructed using SumTrees (version 4.1.0) (Sukumaran & Holder, 2016) in DendroPy (version 4.1.0) (Sukumaran & Holder, 2010) (Fig. 2, Fig. S2, Appendix S2). Following a similar approach to ancestral network reconstruction (Ebenhöh et al., 2006), the ultrametric consensus tree was used in individual joint maximum-likelihood estimations for ancestral state reconstructions of each of the 55 compounds and 91 reactions that occurred in species’ networks. The compounds and reactions were considered discrete (present or absent) and unordered traits, and we tested two Markov models of binary trait evolution in the program r8s (version 1.8) (Sanderson, 2003; Marazzi et al., 2012) for each compound and reaction: the Binary-1 model assumed equal rates and the Binary-2 model assumed different rates of gain and loss across the phylogeny. In each Markov model, the reconstructed ancestral state (present or absent) for a compound or a reaction with the highest likelihood (Pupko et al., 2000) was assigned to each of the 217 internal nodes in the phylogeny by r8s, and the ancestral states in the model with the lower Akaike information criterion (AIC) value (Akaike, 1974) were retained. The ancestral network for each internal node in the phylogeny was comprised of all of the compounds and reactions that were present in the ancestral reconstructions (Fig. S2, Appendix S1).

Fig. 2 Persistence of the same metabolic network structure throughout the evolutionary history of birds. Compounds that are separated by shorter structural distances (number of reactions) in a metabolic network co-occur in the same species (phylogenetic profile similarity) more often than compounds separated by longer pathways. The effect of structural distance on the co-occurrence of compounds is equally strong in (a) networks of extant species and (b) in ancestral networks across all time periods. Asterisks denote significance of $p_{ST}$ (standardized regression coefficient, in standard deviation of the mean) at $P < 0.01$. Shown is the majority-rule consensus tree of 243 species of birds (see Fig. S1 for a more detailed image of the phylogeny). The width of each shaded column in the phylogeny denotes the time span (Myr) corresponding to each group of ancestral networks, located below the column.
To establish functional and evolutionary associations between compounds, we assigned a phylogenetic profile to each compound \( (\text{sensu Pellegrini et al., 1999}) \). The profile represents the subset of species and ancestral networks that contain the compound (ibid.) and comprises a binary string that encodes the presence (1) or absence (0) of a compound in each of the 250 species and 217 ancestral networks (Fig. 1a). The similarity between two compounds’ phylogenetic profiles was used to measure how often pairs of compounds co-occurred in the same species and ancestral carotenoid networks. This was calculated with the Jaccard coefficient \( (J_{ij}) \) (Jaccard, 1912) using the package vegan (version 2.4-1) (Oksanen et al., 2016) in R (version 3.2.2) (R Development Core Team, 2016):

\[
J_{ij} = \frac{n_{ij}}{n_i + n_j - n_{ij}}
\]

where \( n_{ij} \) is the number of networks with both compounds \( i \) and \( j \), \( n_i \) is the number of networks with compound \( i \), and \( n_j \) is the number of networks with compound \( j \). Pairs of compounds with similar phylogenetic profiles \( (J_{ij} = 1) \) are assumed to be functionally correlated and coevolve as part of evolutionary conserved modules, whereas pairs of phylogenetically distinct compounds \( (J_{ij} = 0) \) evolve independently and occur in different modules (Yamada et al., 2006; Zhao et al., 2007). The distance between any two compounds in the avian carotenoid metabolic network (Figs 1a and S1) was represented by the fewest number of reactions between the compounds. Pairs of compounds that were not connected by reactions were excluded from this study. To assess the evolutionary expansion of networks, the relationships between phylogenetic profile similarity and structural distance between pairs of compounds over evolutionary time were assessed separately for all 250 extant species and for groups of ancestral networks representing six evolutionary time periods (Fig. 2: 0-5 MYR, 5-10 MYR, 10-15 MYR, 15-20 MYR, 20-30 MYR, 30-50 MYR). All statistical analyses were implemented in SAS v. 9.4.

Results and discussion

In this study, we examined whether evolutionary diversification in avian carotenoid networks was consistent with the phylogenetically structured expansion of enzymatic pathway modules or utilization of a pre-existing network with selective expression of conserved biochemical modules (Fig. 1). We found that the structural associations between compounds and reactions involved in carotenoid metabolism persisted largely intact throughout avian ecological and evolutionary diversification (Fig. 1c): compounds connected by fewer reactions consistently co-occurred more often than compounds separated by several reactions and the strength of this relationship was equally as strong in both extant species networks (Fig. 2a; \( n = 250 \) species, \( b_{ST} = -0.37, P < 0.01 \), and in ancestral networks from six different time periods across 50 MYR of avian evolution (Fig. 2b): 0-5 MYR \( (n = 46 \) networks, \( b_{ST} = -0.44, P < 0.01 \), 5-10 MYR \( (n = 57 \), \( b_{ST} = -0.52, P < 0.01 \), 10-15 MYR \( (n = 49 \), \( b_{ST} = -0.57, P < 0.01 \), 15-20 MYR \( (n = 37 \), \( b_{ST} = -0.49, P < 0.01 \), 20-30 MYR \( (n = 13 \), \( b_{ST} = -0.53, P < 0.01 \), 30-50 MYR \( (n = 18 \), \( b_{ST} = -0.45, P < 0.01 \). Notwithstanding the occasional addition of lineage-specific biochemical pathways (Fig. S2, Appendix S3), subsampling and selective expression of different combinations of highly conserved modules thus appear to be the dominant mode of avian carotenoid evolution.

These findings emphasize two points. First, much of avian carotenoid evolution consists of expressing and recombining entire biochemical modules, with biochemical structure within these modules remaining virtually intact over long periods of time (Figs 1c and 2). Second, despite extensive diversification, birds, evidently, retain access to a far larger carotenoid metabolic network than is expressed in plumage-bound carotenoid pathways at any particular time. What are the mechanisms that produce these patterns?

Groups of co-expressed compounds can occur when multiple sequential reactions are activated together (Krebs, 1957; Fell, 1992), or as a result of a single enzyme complex that catalyses more than one reaction (Srere, 1987). Many biochemical pathway modules that are expressed and recombined in avian evolution include links to one of two specific dietary carotenoids (lutein and zeaxanthin), such that these modules are expressed only when a particular dietary compound is encountered in the evolution of an ecologically specialized lineage (Figs S1 and S2). Other modules are comprised entirely of metabolically derived carotenoids and are typically located at the intersection of several biochemical pathways that are interchangeably linked to several dietary carotenoids (Fig. S1, Appendix S3; Badyaev et al., 2015) so that they are expressed when any of the dietary precursors are encountered (e.g. Edelman & Gally, 2001). Such configuration accounts for evolutionary persistence of modules consisting of derived carotenoids despite frequent shifts in dietary precursors (Badyaev et al., 2015). This arrangement also explains why the two most common dietary carotenoids in birds – lutein and zeaxanthin (Appendix S1) – are not a part of any biochemical module (Fig. S1, Appendix S3). Overall, the highly modular pathway structure of the ancestral carotenoid metabolic network likely facilitates ecological diversification and specialization of organisms that depend on it (e.g. Edelman & Gally, 2001; Wagner & Fell, 2001; Kreimer et al., 2008).
reversible switches between modules could be achieved when enzymes in the ancestral network are nonspecific, such as when a single enzyme catalyses multiple reactions with distinct substrate or product compounds (Khosla & Harbury, 2001; Morrison & Badyaev, 2016b). When such an enzyme is affiliated with several pathway modules in the ancestral network, changes in acquired dietary compounds lead to the expression of different modules without the loss of modules associated with absent dietary compounds. Persistence of modules associated with unique enzymes could be facilitated by rapid and reversible changes in the regulation of the enzymes that control metabolic flux through the network (e.g. Campagna et al., 2017; Morrison & Badyaev, 2017). On a microevolutionary level, such modulation is readily accomplished by a variety of genetic and epigenetic regulators of metabolic flux that activate or deactivate enzymes (Cohen, 1988; Ketterson & Nolan, 1999; Hau, 2007; Varum et al., 2011; Gordon & Ruvinsky, 2012). Further, a majority of carotenoid compounds and reactions are utilized in biochemical functions unrelated to plumage coloration including vision, transmembrane transport and the immune system (e.g. Krensly & Yeum, 2003; Toomey et al., 2015), assuring their evolutionary retention and likelihood of recruitment into production of plumage-bound carotenoids, such as through the upregulation of metabolic flux (e.g. Lu & Li, 2008). Frequent and reversible utilization of distinct biochemical modules in avian evolution (Morrison & Badyaev, 2016a; Badyaev & Morrison, 2018), along with the evolutionary conservation of modules comprising of derived carotenoids reachable from a diverse array of dietary precursors (Badyaev et al., 2015), could further maintain access to the same biochemical network during evolution.

The finding that recombination of biochemical modules underlies the diversification of carotenoid networks across species allows for qualitative predictions of carotenoid network diversification in birds. For example, compounds linked to multiple modules should be more conserved over evolutionary history and across species’ networks than compounds that only have within-module connections (Guimerà & Amaral, 2005). Indeed, carotenoids with the most enzymatic reactions were the hotspots of avian carotenoid diversification, despite high evolutionary lability in expression of the modules of which they are a part (Morrison & Badyaev, 2016a), allowing even closely related species to express structurally distant modules (Fig. S2). Furthermore, the periodic convergence of distantly related and ecologically distinct taxa in carotenoid networks (Badyaev et al., 2015) suggests that an ancestral network structure delineates avian carotenoid diversification.

The extent to which these patterns translate into diversification in plumage colour is not explored in this study. At least three factors affect the correspondence between the carotenoid metabolic network and colour diversity: the relationship between metabolic flux and structure in the network, resulting in differing efficiency of metabolic pathways and variable mixtures of compounds deposited in feathers, incongruence of colour and metabolic changes across the avian carotenoid network where some metabolic neighbourhoods generate greater colour diversity than others, and variation in integration of carotenoid compounds and feather keratins that ultimately produce plumage coloration. Whether these factors affect, mask or exaggerate the patterns of combinatorial metabolic evolution found in this study requires further research.

The striking realization that the biochemical structure of an ancient carotenoid network persists through avian diversification and likely plays an important role in the evolution of contemporary adaptations in avian carotenoids, focuses the search for the mechanism by which selective expression and recombination of historically conserved biochemical modules are accomplished in evolution.

**Acknowledgments**

We thank V. Belloni, M. Coope, V. Farrar and J. Andrews for help with the data collection and R. Raff, M. J. Sanderson, S. Gavrilets, R.A. Duckworth, C. Schmidt-Dannert and L. Orr for comments and discussion. Funding was provided by grants from the National Science Foundation (DEB-0077804 and IBN-0218313) and the David and Lucille Packard Foundation to A.V.B., and Amherst College Graduate Fellowships and the University of Arizona Galileo Circle Scholarship to E.S.M.

**References**


**Supporting information**

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1** Visualization of the avian carotenoid network.

**Figure S2** Labeled phylogeny of all taxa included in this study and the distribution of biochemical modules across taxa.

**Appendix S1** Data set for extant species and ancestral reconstructed networks.

**Appendix S2** Nexus file of phylogenies used in this study.

**Appendix S3** List of biochemical modules found in the avian carotenoid metabolic network.

Received 20 September 2017; revised 14 February 2018; accepted 20 February 2018