Running head: Global patterns of cichlid diversification

Ecological opportunity alters the timing and shape of adaptive radiation

Edward D. Burress[1]* & Milton Tan[2]

[1]Department of Biological Sciences and Auburn University Museum of Natural History, Auburn University, Auburn, Alabama, U.S.A.

[2]Division of Infectious Diseases, School of Medicine, Emory University, Georgia, USA

*Corresponding author: edwarddburress@gmail.com

Keywords: BAMM; colonization; diversification; evolution; macroevolution; speciation

Acknowledgements

We thank Bradley Morgan and the HPC oversight committee for access and assistance for the CASIC supercomputer at Auburn University, and Robert A. Petit III and Timothy D. Read for access and assistance with computing at Emory University. Jamie Oaks kindly provided feedback on an earlier draft of this manuscript.

Abstract

The uneven distribution of diversity is a conspicuous phenomenon across the tree of life. Ecological opportunity is a prominent catalyst of adaptive radiation and therefore may alter patterns of diversification. We evaluated the distribution of shifts in diversification rates across the cichlid phylogeny and the distribution of major clades across phylogenetic space. We also tested if ecological opportunity influenced these patterns. Colonization-associated ecological opportunity altered the tempo and mode of diversification during the adaptive radiation of cichlid fishes. Clades that arose following colonization events diversified faster than other clades. Speciation rate shifts were non-randomly distributed across the phylogeny such that they were disproportionally concentrated around nodes that corresponded with colonization events (i.e., of continents, river basins, or lakes). Young clades tend to expand faster than older clades; however, colonization-associated ecological opportunity accentuated this pattern. There was an interaction between clade age and ecological opportunity that explained the trajectory of clades through phylogenetic space over time. Our results indicate that ecological opportunities afforded by continental- and ecosystem-scale colonization events explain the dramatic speciation rate heterogeneity and phylogenetic imbalance that arose during the evolutionary history of cichlid fishes.

Introduction

A central hypothesis about adaptive radiations is that rates of diversification will be fastest early in clade history as lineages rapidly disperse across novel regions of the adaptive landscape and subsequently slow as the niches associated with these regions saturate (Simpson 1953; Schluter 2000). Despite many examples from the fossil record (Foote 1994; Ruta et al. 2006; Hughes et al. 2013), early bursts of phenotypic evolution are rarely observed in comparative phylogenetic data (Harmon et al. 2010). Likewise, signals of diversity-dependent diversification in molecular phylogenies are often inconsistent and suggest that ecological opportunities rather than clade age regulates diversification rates (Rabosky and Lovette 2008; Rabosky 2009a,b; Rabosky et al. 2012; Rabosky and Hurlbert 2015). Regardless, the uneven distribution of diversity among extant clades is a conspicuous phenomenon across the tree of life. This discrepancy may be explained by diversification rate heterogeneity among clades (Kirkpatrick and Slatkin 1993; Chan and Moore 2002; Wiens 2011) or variable clade ages (McPeek and Brown 2007; Shi and Rabosky 2015).

Ecological opportunity has been characterized as relaxed competition in conjunction with ample resource availability (Schluter 2000) and is widely recognized a major catalyst of adaptive radiation (Rabosky & Lovette 2008; Rabosky 2009a; Losos 2010; Yoder et al. 2010; Mahler et al 2010). There are myriad sources of ecological opportunity that may alter a lineage’s evolutionary trajectory, including the evolution of a key innovation, extinction of a competitor, or the colonization of a novel environment that may provide access to new resources (Losos and de Queiroz 1997; Schluter 2000; Yoder et al. 2010; Wainwright et al. 2012). The occurrence of these events should elicit a burst in diversification rates in response to ecological opportunity and subsequent slowing as those opportunities saturate (Simpson 1953; Schluter 2000; Burress et al. 2016). Lineages proceed through these stages differently. For example, adaptive landscapes are variable across different organisms and environments and the magnitude of ecological opportunity, such as that afforded by the colonization of a novel ecosystem, will manifest in different ways and to different degrees. For example, among freshwater fishes, there is a large disparity in the ecological opportunities provided by lakes and rivers (Seehausen 2015). Additionally, ecosystem size constrains ecological opportunity and thereby *in situ* adaptive radiation (Wagner et al. 2014). Furthermore, ecological opportunity may interact with other mechanisms such as sexual selection (Wagner et al. 2012) or hybridization (Meier et al. 2017) that may accentuate the influence of ecological opportunity on adaptive radiation.

Cichlids are the third-most species-rich family of fishes and exhibit diversity that is not equally distributed across clades and considerable diversification rate heterogeneity among clades (McMahan et al. 2013). Some clades have diversified rapidly, particularly those of Lakes Malawi and Victoria. For example, the ~700 species endemic to Lake Victoria arose over the last 150,000 years (Meier et al. 2017), whereas other much older lineages have relatively few extant species (McMahan et al. 2013). Cichlids also have a complex biogeographic history. For example, cichlids colonized South America via trans-Atlantic dispersal from Africa (Friedman et al. 2013; Matschiner et al. 2017) and later colonized Middle America also via trans-Atlantic dispersal (Říčan et al. 2013). In addition to these major continental-scale colonization events, cichlids have also independently colonized and diversified within myriad lakes throughout East Africa (Wagner et al. 2014) and Middle America (Elmer et al. 2014). In some cases, lakes have been repeatedly colonized after multiple transitions between river and lake environments (Salzburger et al. 2005). These features make cichlid fishes an ideal model to assess the role of colonization-associated ecological opportunity on the tempo and mode of diversification during their adaptive radiation.

Our objectives were twofold. First, we estimated diversification rates across a family-wide phylogeny of 903 species of cichlid fishes that spans all major lineages. We focused on the relative rates among major clades, rates through time, and associations between rate shifts and colonization-associated ecological opportunity. We hypothesized that rate shifts would be non-randomly distributed across the phylogeny due to ecological opportunities afforded by colonization events during the evolutionary history of cichlids. Therefore, we predicted that rate shifts would be disproportionally concentrated around nodes that correspond with known colonization events (i.e., colonization of a continent, river basin, or lake). Secondly, we quantified the relative shapes of different regions of the phylogeny (i.e., the distribution of major clades in phylogenetic space). Lineages explore phylogenetic space as they proceed through the stages of adaptive radiation (Lewitus & Morlon 2016) such that lineages should exhibit differently shaped phylogenies because of different modes of diversification that are apparent depending upon clade age and temporal proximity to changes in ecological opportunity. We hypothesized that ecological opportunity also alters the shape of phylogenies by accentuating phylogenetic characteristics such as phylogenetic expansion, phylogenetic imbalance, and the distribution of branching events through time. Therefore, we predicted that the relationship between phylogenetic shape and clade age would be different between colonization-associated clades and clades that did not arise in association with colonization-associated ecological opportunity.

Methods

Phylogeny construction

To provide a phylogenetic hypothesis for comparative analyses while accommodating phylogenetic uncertainty, we sought to infer a distribution of densely-sampled, time-calibrated phylogenetic trees. Synthetic time trees were inferred using several steps, including 1) compiling sequence data from GenBank, 2) bootstrapping maximum-likelihood phylogenetic analysis on these sequence data to infer a distribution of bootstrap trees representing all cichlids, 3) performing time-calibrated phylogenetic analysis to infer a posterior distribution of reference time trees representing a subset of cichlids, and finally, 4) integrating these results to infer divergence times on the bootstrap trees based on the reference time trees.

First, to compile phylogenetic data, we used the phlawd pipeline (Smith et al. 2009) to query GenBank for several commonly used genes in cichlid phylogenetics. Phlawd was designed to rapidly compile phylogenetic datasets for densely-sampled species-level phylogenetic analysis. A GenBank database was downloaded and set up using phlawd (accessed 2016 Dec 16). For each locus, phlawd queries the GenBank database and generates alignments where a single sequence represents each taxon (generally species, but also subspecies if present) (Smith et al. 2009). We queried the mitochondrial genes 12S, 16S, COI, cyt b, ND2, ND4 and the mitochondrial control region; and the nuclear genes 4c4, enc1, RAG1, RAG2, S7 intron 1, SH3PX3, glyt, myh6, plagl2, ptr, sreb2, and tbr1. Outgroup taxa were selected from other ovalentarians used in a time-calibrated phylogenetic analysis of cichlids (Friedman et al. 2013). Sequences were excluded if they represented unidentified taxa marked with sp., aff., cf., and hybrids by NCBI taxonomy. We also manually included several sequences as user sequences (Table S1). For some species, subspecies were also present in the alignment generated by the phlawd pipeline. For these species, we selected the longest sequence from among the subspecies present for each locus to represent the species.

Sequences were concatenated and alignment formats converted using AMAS (Borowiec 2016). PartitionFinder 2.1.1 was used to determine the best model of evolution and partitioning scheme available for RAxML (Lanfear et al. 2016). A preliminary maximum likelihood tree was inferred, followed by manual inspection to remove incorrectly placed taxa. Finally, a maximum likelihood tree was inferred, and 100 bootstrap replicates were analyzed with branch lengths estimated. A distribution of reference time trees for subsequent congruification were inferred using the xml file made available by Friedman et al. (2013) for their time-calibrated phylogenetic analysis of percomorphs (doi:10.5061/dryad.48f62). Friedman et al. (2013) used 14 fossil calibrations for the ages of 13 outgroup nodes distributed across Percomorpha and the root node using a relaxed-clock analysis in BEAST to calibrate their phylogeny, which included 156 percomorph species, including

91 cichlids. Friedman et al. (2013) provides a relatively young estimate of cichlid age that infers trans-Atlantic dispersal from Africa to South America, which has been corroborated by subsequent analyses (Matschiner et al. 2017). We performed two runs using BEAST 1.8.2 (Drummond et al. 2012). Using burntrees.pl (https://github.com/nylander/Burntrees), burn-in was excluded and 50 evenly-spaced time trees were sampled from the remaining posterior distribution of each run, resulting in 100 time trees. We then determined the maximum clade credibility (MCC) tree from these samples using TreeAnnotator.

We timed the maximum-likelihood tree using the MCC tree as a reference by the congruification method (Eastman et al. 2013) implemented in geiger v2 (Pennell et al. 2014). The congruification software automatically finds matching nodes between a reference time tree to a target tree based on shared taxa. The node ages of shared nodes from the time tree are then used as secondary calibrations to calibrate the target tree, and divergence times were estimated using treePL (Smith and O'Meara 2012). We used the congruify and write.treePL functions to perform the congruification and generate the treePL input files for analysis. The treePL analysis was primed to determine optimal settings, and the tree was then time-calibrated with the thorough setting. This was also repeated for the 100 bootstrap trees and 100 sampled time trees, to allow for comparative analyses accommodating uncertainty in tree topology and divergence time estimates. To maximize the number of shared nodes between each bootstrap tree and time tree pair, trees were paired to minimize the Robinson-Foulds distance between them, computed using phangorn (Schliep 2011). This was done iteratively without replacement; hence, for each bootstrap tree, the nearest time tree was selected, and this time tree was excluded from the set paired to the remaining bootstrap replicates. As with the ML and MCC tree congruification, each treePL analysis was primed to determine optimal settings, and divergence times were estimated with the thorough setting.

Delimitation of focal clades

We delimited 27 non-overlapping clades across the cichlid phylogeny that exhibit variable species diversities and clade ages (Table 1). Ten of the 27 focal clades are defined by nodes that correspond to known colonization events that were represented in our phylogeny (i.e., of continents, river basins, or lakes). The focal nodes (i.e., colonization nodes hereafter) for these clades were defined as the node where all descending tips occur in the same region, except for South America and Lake Tanganyika, which included subsequent colonization events into other regions (i.e., Middle America and the Malagarasi River, respectively). The remaining 17 clades are not associated with colonization events. To evaluate the role of colonization-associated ecological opportunity, we used these two sets of clades – colonization-associated clades that arose after a colonization event and non-colonization-associated clades that did not arise following such events. We defined all clades that met these criteria: (i) clades must be non-overlapping and (ii) clades must consist of at least eight taxa to permit estimation of diversification rate and phylogeny shape. We first defined colonization associated clades based on the literature and our phylogeny, then defined all possible non-colonization-associated clades given the two aforementioned criteria. A third criteria was used for the non-colonization associated clades, which were (iii) defined as large as possible to maximize the number of taxa represented in our focal clades. Our global-scale has some drawbacks because of low resolution in some parts of the phylogeny. We exclude some small radiations that arose following colonization events due to poor representation in our phylogeny that prevented robust estimation of diversification rates and phylogeny shape; for example, some of the small African lake radiations such as the *Alcolapia* flock of Lake Natron (Ford et al. 2016; Kavembe et al. 2016) and the *Coptodon* flock of Lake Bermin (Martin et al. 2015). Additionally, some radiations are simply too small (i.e., species pairs) to robustly assess speciation rates and phylogeny shape (e.g., *Amphilophus* of the Nicaraguan crater lakes; Elmer et al. 2014). We also excluded one potential clade of *Crenicichla* that includes several such small radiations and therefore could not be unambiguously coded as a non-colonization-associated clade. Similarly, some taxa had long branches that were not allied with a clade and therefore were not encompassed in a focal clade. The species composition of each clade is delimited in Figure S1.

Rates of speciation

To quantify time-varying diversification rates, we implemented BAMM v2.5.0 (Rabosky 2014; Rabosky et al. 2014; Shi and Rabosky 2015). Priors were computed using setBAMMpriors in BAMMtools. Clade-specific sampling probabilities were specified for major clades of cichlids based on the proportion of species in the tree out of the total number of species described according to www.cichlidae.com (accessed May 2016). Four chains were run with 100M generations. MCMC parameters and event data were sampled every 20,000 generations. We discarded the first 10% of the run as burn-in. We used the BAMMtools package to compute and visualize the highest *a posteriori* credible shift configurations, marginal odds ratios of a shift occurring for all branches, rates through time across cichlids and for major clades, compute Bayes Factors (BF) for the number of shifts, and to compare the prior and posterior distributions of shifts. The posterior distribution of the number of rate shifts can depend upon the prior (Moore et al. 2016); therefore, we emphasize the marginal odds ratios BF for each shift and BF for the number of shifts, which are statistics that are computed by explicitly comparing the prior and posterior to evaluate evidence for the presence of each shift and number of shifts (Rabosky et al. 2017).

To assess the influence of colonization-associated ecological opportunity on the distribution of rate shifts across the cichlid phylogeny, we compared the marginal odds ratios of shifts at 11 nodes that correspond with colonization events and nodes that do not correspond to colonization events. Colonization nodes were defined to capture major biogeographic shifts during the evolutionary history of cichlids. These 11 nodes include the crown of each of the 10 aforementioned colonization-associated clades (Table 1) and the node corresponding with the initial colonization of Lake Tanganyika. A shift to a new diversification rate regime may not be coincident with a colonization event, but rather occur after some degree of lag following the colonization event (Burress et al. 2016); therefore, we also included the two subsequent descendant nodes for each colonization node, as well as each of their descendant nodes (up to four nodes, if they did not end in tips). This scheme can capture both immediate and delayed shifts to new rate regimes following each colonization event.

Reconstruction of phylogenetic space

We compared the shapes of the phylogenies of 27 focal clades (Table 1) using the RPANDA package, which allows for model-free assessment of macroevolutionary patterns by summarizing information contained in trees (Morlon et al. 2016). In this procedure, spectral density profiles are calculated using information about tree shape from the evolutionary distances between nodes (Lewitus & Morlon 2016). These profiles include three summary statistics (λ^*, ψ, and η) that describe different aspects of tree shape. The ln-transformed principal eigenvalue (λ^*) describes the distance between nodes such that speciation-poor and speciation-rich regions of the tree correspond with longer and shorter distances, respectively. Therefore, smaller eigenvalues depict regions of the tree undergoing phylogenetic expansion. Skewness (ψ) indicates the stem-to-tip distribution of branching events. In other words, if the eigenvalues are disproportionately small or large, the phylogeny is comprised of primarily stemmy or tippy trees, respectively. Lastly, peak height (η) depicts the heterogeneity of branch lengths in the phylogeny such that low and high values indicate heterogeneous and homogeneous branch lengths, respectively.

We evaluated the structure of the 27 focal clades across phylogenetic space using Ward's minimum variance clustering. Clusters were evaluated using confidence intervals derived from 10,000 multiscale bootstrap resampling. Cluster analysis and *P*-values were calculated using pvclust package (Suzuki & Shimodaira 2011). Approximately unbiased *P*-values are between 0 and 1 and depict how strongly the cluster is supported by the data (Suzuki & Shimodaira 2011). Clusters were evaluated using λ^*, ψ, and η.

Lineages often explore phylogenetic space over time (Lewitus & Morlon 2016); therefore, we tested relationships between clade age and λ^*, ψ, and η. Because these parameters describe the shape of regions of the cichlid phylogeny (i.e., the focal clades) rather than characteristics of the tips themselves and do not assume a model of evolution, we evaluated these relationships using a simple linear model. We determined if models were statistically significant by calculating effect sizes (correlation coefficient) and determining if 0 fell outside of the 95% confidence interval. For any significant models, we then evaluated the influence of ecological opportunity by testing for an interaction between ecological opportunity and clade age using ANCOVA. If present, we evaluated if interactions significantly improved the model using a drop test via the anova function in R.

Results

The phylogeny of cichlids

We inferred the most densely-sampled hypothesis to date of the cichlid phylogeny based on seven mitochondrial and 12 nuclear loci (Figure S1). Consistent with previous family-wide analyses, we found poor node support along the backbone of the African clade as well as among species within the East African Great Lakes (Genner et al. 2007; McMahan et al. 2013; Friedman et al. 2013; Figure S1). The major tribes were recovered as monophyletic; however, most of the East African tribes had poor support (Figure S1). The relationships among tribes were generally consistent with previous analyses of extant cichlids (Genner et al. 2007; McMahan et al. 2013; Friedman et al. 2013).

The distribution of rate shifts across the phylogeny

The maximum shift credibility tree included nine core shifts (Figure 1A). Shift configurations sampled in the posterior ranged from six to 18 shifts, with nine being the most sampled number of shifts (Figure 1B). Bayes factors indicate that seven is the most probable number of shifts, with a BF of 32.67 over six shifts (Table S2). Although the distribution of the number of shifts in the posterior was sensitive to the prior; the most probable number of shifts supported by Bayes Factors was robust to the number specified in the prior (Figure S2). Seven rate shifts were still recovered as most probable when a prior of five shifts was used, whereas six shifts were recovered as most probable using a prior of 10 shifts based on Bayes Factors (Figure S2). Due to the size of the tree and thus many potential shift configurations, no shift configurations were frequently sampled (Figure S3). For example, the nine most frequently sampled shift configurations from the 95% credible shift set included configurations that had seven to twelve shifts and represent only 0.018% of the sampled shift configurations; however, the approximate location of most shifts were conserved such that they merely alternated around sequential nodes (Figure S3). Thus, despite the low frequency of any particular shift configuration, the shift locations are consistent among configurations. Diversification rates through time began gradually increasing approximately 12 Mya and then exponentially increased over the last one My (Figure 1C). The distribution of diversification rates across the phylogeny was highly skewed such that most rates were relatively slow and fast rates occurred mostly in isolated bouts (Figure 1D).

The relative rates of diversification of the focal clades

The distribution of core rate shifts across the phylogeny corresponds with colonization events (Figure 1E) and the rate shifts with the highest marginal odds ratios are conspicuously associated with a handful of lineages, namely the Geophagini, Heroini, Oreochromini, and the Malawian and Victorian super flocks (Figure 1F). By far, the fastest and most exponentially increasing diversification rates occur within the lineages that arose with the African Lakes Malawi, Victoria, and Barombi Mbo (Figure S4). Three river-dwelling clades also exhibit diversification rates that markedly increase over time, albeit at distinctly slower rates than their lake-dwelling counterparts. These include the heroine clade distributed throughout Middle America, the *Crenicichla lacustris* clade distributed throughout subtropical South America, and the *Serranochromis* clade distributed throughout Southern Africa (Figure S4). In contrast, all Lake Tanganyikan clades (i.e., Bathybatini and allies, Ectodini, Lamprologini, and Tropheini) exhibit diversification rates that distinctly slow through time (Figure S4).

The distribution of clades across phylogenetic space

The focal clades were widely distributed across phylogenetic space defined by λ^*, ψ, and η (Figure 2A) and varied in their degree of phylogenetic expansion, the relative stem-to-tip distribution of branching events, and heterogeneity of branch lengths. Most notably, the Lake Malawi super flock occupied a distinct region of phylogenetic space defined by high η values (Figure 2A). Clades varied much more continuously along the λ^* and ψ axes. Clades associated with smaller λ^* values depicting rapid phylogenetic expansion include the clades of Lakes Victoria and Barombi Mbo, whereas clades associated with higher λ^* values depicting slow phylogenetic expansion include the Tanganyikan Lamprologini and the Neotropical Cichlasomatini, Geophagini, and Heroini (Figure 2A). The frequency distribution of spectral curves based on λ^* values indicates high degrees of overlap among most clades, except for the handful of clades at each extreme (Figure 2B). Clades associated with smaller ψ values depicting branching events skewed towards the stem include the Lamprologini, Tropheini, and Chromidotilapiini, whereas clades associated with larger ψ values depicting branching events skewed towards the tips includes the Ptychochrominae, *Bathybates* and allies, and *Orthochromis* clades (Figure 2A).

Of the parameters that describe phylogenetic space, only λ^* values were significantly correlated with clade age (Figure 3C) such that younger clades tend to exhibit more phylogenetic expansion (Figure 2D). The focal clades were well distributed across λ^* and clade age values (Figure 2D). Cluster analysis revealed well-supported spatial structure in the distribution of the focal clades across phylogenetic space; however, there were no intuitive cases where clades from similar regions or ecosystems were clustered together (Figure 2E).

The influence of ecological opportunity on the rates and shape of adaptive radiation

The probability of a shift occurring at one of the 11 colonization nodes was significantly higher than the probability of a shift at another node (Figure 3A). Likewise, the marginal odds ratios of rate shifts were 45.9 times higher, on average, for colonization nodes than for non-colonization nodes (Figure

3B; Table S3). The shift with the highest marginal odds ratio, by far, was associated with the colonization of Lake Malawi (Figure 1E,F; Table S4). In fact, there were highly probable shifts on the colonization node and descendant nodes (Table S4). Rate shifts associated with the colonization of Lakes Victoria, Tanganyika, Barombi Mbo and the La Plata Basin also had high marginal odds ratios (Table S4). The highest marginal odds ratios were coincident with the colonization node for Lake Barombi and the La Plata Basin; however, in Lakes Victoria and Tanganyika, the highest marginal odds ratios occurred on a descendant node (Table S4). The rate shift coincident with the colonization of Middle America also had a high marginal odds ratio; however, the rate shift with the highest marginal odds ratio occurred on a descendant node following the colonization of South America and colonization of the Congo River by *Nanochromis* and *Steatocranus* (Table S4). In contrast, the secondary colonization of Lake Tanganyika by the Tropheini and colonization of the Malagarasi River by *Orthochromis* were not associated with a well-supported shift at the colonization nodes or their immediately descendant nodes (Table S4).

There was a significant positive relationship between ln-transformed eigenvalues and clade age for both colonization-associated and non-colonization-associated clades (Figure 3C) such that younger clades exhibit more phylogenetic expansion than older clades (Lewitus & Morlon 2016); however, we found that there was a significant interaction between ecological opportunity (i.e., colonization- vs. non-colonization-associated clades) and their λ^* values (F=7.629; P=0.014; Figure 4C; Table S5) such that clades that arose following colonization events explored phylogenetic space faster than other clades. A drop test indicated that including the interaction significantly improved the model (F=7.629; P=0.011; Tables S6 and S7).

Discussion

Ecological opportunity is considered one of the prominent features of adaptive radiation that may alter adaptive landscapes and therefore, evolutionary trajectories (Rabosky & Lovette 2008; Rabosky 2009a; Losos 2010; Yoder et al. 2010; Mahler et al. 2010). Ecological opportunity has been invoked to explain the rates and modes of diversification among adaptive radiations ranging from fishes, lizards, and birds, among others (Rabosky & Lovette 2008; Mahler et al 2010; Burress et al. 2016); however, the extent to which ecological opportunity influences broader adaptive radiations remains unclear (Slater 2015). We demonstrate that colonization-associated ecological opportunity has influenced the rates and modes of diversification and has led to the uneven distribution of extant taxa across the cichlid phylogeny.

The colonization of lakes has long been considered a major source of ecological opportunity for fishes by providing access to novel resources and physical dimensions relative to the marine or riverine conditions from which the colonists dispersed (Seehausen 2015; Burress 2015). Ecological diversification following lake colonization results in a few prodigious patterns across disparately related groups. For example, repeated diversification along the benthic-to-pelagic habitat axis by stickleback and by cichlids is a major theme among glacial, crater, and rift lakes (Rundle et al. 2000; Hulsey et al. 2013; Elmer et al. 2014; Machado-Schiaffino et al. 2014). We found that bursts in diversification occurred following the colonization of Lakes Tanganyika, Malawi, and Victoria. Although Lake Tanganyika may have served as the cradle from which the entire East African radiations arose, their geographic dispersal involved a series of complex colonization events. For example, there were multiple invasions of Lake Tanganyika from riverine ancestors. The first gave rise to the Tanganyikan species flock and a subsequent colonization gave rise to the Tropheini (Salzburger et al. 2002). In contrast, Lake Malawi may have been colonized directly via dispersal from Lake Tanganyika (Salzburger et al. 2002; Joyce et al. 2011). Lastly, the Lake Victoria species flock may have arisen via hybridization between divergent riverine lineages (Meier et al. 2017). We also found evidence of a burst in diversification following the colonization of Lake Barombi Mbo, which is located near the Cameroonian coast. Similar to the resident assemblages of the East African Great Lakes, the Lake Barombi Mbo species flock also arose from riverine ancestors and may have a complex history of hybridization (Martin et al. 2015). There are clear themes associated with the colonization of lakes by cichlids. Namely, colonization-associated ecological opportunity results in bursts in diversification, ecological differentiation along the benthic-to-pelagic habitat axis, and complex instances of repeated gene flow (Joyce et al. 2011; Hulsey et al. 2013; Martin et al. 2015; Meier et al. 2017). Ecosystem size may ultimately constrain the degree to which *in situ* diversification unfolds such that larger lakes may provide more ecological opportunity (Wagner et al. 2014).

The evolutionary implications of colonizing rivers are less clear than lakes. In general, riverine species flocks are far less common (Seehausen 2015). This pattern suggests that diversification tends to occur via vicariance among rivers rather than *in situ*. We assessed four colonization-associated riverine clades that represent all known riverine species flocks among cichlids (Salzburger et al. 2002; Schwarzer et al. 2011; Piálek et al. 2012); although, it is noteworthy that the *Orthochromis* clade may have arose in an ancient paleo-lake that subsequently dried (Joyce et al. 2005). We found support for rate shifts associated with three of these four clades. Also in contrast with lake-dwelling clades, there are fewer themes across the riverine clades. For example, *Crenicichla* exhibit a burst in diversification rates following the colonization of the La Plata Basin, particularly within the Paraná and Uruguay Rivers, and also exhibit dramatic trophic diversity, including the rise of parallel ecomorphs (Piálek et al. 2012; Burress et al. 2013; Piálek et al. 2015). *Steatocranus* and *Nanochromis* exhibit some ecomorpholocial diversity associated with specialization among different microhabitats in the lower Congo River (Schwarzer et al. 2011), whereas *Orthochromis* have not ecomorphologically diversified much within the Malagarasi River (Seegers and De Vos 1988). Therefore, the nature of adaptive radiation following the colonization of rivers is variable and may be strongly constrained by the history of each lineage. For example, *Crenicichla* already colonized a novel functional region of the adaptive landscape characterized by high velocity ram feeding (López-Fernández et al. 2013), and thus may have been predisposed to exploiting any subsequent trophic-associated ecological opportunities. Alternatively, rivers themselves are heterogeneous and factors such as depth and habitat heterogeneity may constrain ecological opportunities for resident lineages. Disentangling the relative importance of intrinsic characteristics of the lineages and characteristics of the ecosystems in moderating or constraining adaptive radiations within rivers may be a fruitful topic for future research.

Many organisms have successfully colonized the Americas via trans-Atlantic dispersal from Africa, including freshwater fishes, reptiles, mammals, and plants (Lundberg 1993; de Quiroz 2005; Oaks 2011; Beaulieu et al. 2013). For cichlids, the active nature of this colonization event is a relatively recent paradigm shift from the long-held hypothesis of Gondwanan vicariance (Genner et al. 2007; Friedman et al. 2013; Matschiner et al. 2017). Cichlids likely used a combination of island hopping, dilute surface water from the outflow of a large river system such as the present-day Congo River, as well as some salinity tolerance to successfully colonize freshwaters separated by marine environments (Friedman et al. 2013 and references therein). Likewise, the subsequent colonization of Middle America from South America likely occurred via a combination of land bridges and island hopping through the Greater Antilles (Říčan et al. 2013; Tagliacollo et al. 2017). Both of these continental-scale colonization events corresponded with bursts in diversification rates, albeit generally less probable shifts than those associated with ecosystem-scale colonization events.

There were few unifying themes across the cichlid phylogeny such that diversification rates were heterogeneous and clade-specific phylogenetic shape was highly variable, even among clades that have classically been associated (i.e., the East African Great Lakes; Kocher 2004; Seehausen 2015). One apparent unifying theme is the role that colonization-associated ecological opportunity has played during the evolutionary history of cichlid fishes. Colonization events, ranging from continental- to ecosystem-scale, elicited bursts in diversification rates and altered the movement of clades across phylogenetic space. There were several other emergent patterns: (1) lake- and river-associated clades do not occupy distinct regions of phylogenetic space and (2) the most probable rate shifts are associated with ecosystem-scale colonization events rather than continental-scale events. Colonization-associated ecological opportunities may be responsible for much of the dramatic diversification rate heterogeneity and phylogenetic imbalance that arose during the evolutionary history of cichlid fishes and has served as a catalyst for numerous adaptive radiations of cichlid fishes.

Cichlids have colonized and flourished within incredibly diverse and sometimes harsh environments, including hot, alkaline, and hypersaline conditions (Trewavas 1983; Ford et al. 2015; 2016; Kavembe et al. 2016), torrential streams (Alter et al. 2017), and abyssal conditions (Schobert et al. 2013; Hahn et al. 2017). Ecological opportunities that lineages capitalize upon (i.e., successful colonization events) are likely only a subset of those that arise. The strong correspondence between colonization events and bursts in diversification not only emphasizes the critical role that colonization-associated ecological opportunity played during the evolutionary history of cichlids, but also emphasizes the importance of any intrinsic traits that permitted cichlids to successfully colonize given the opportunity. Most notably, colonizing alkaline lakes, often from comparatively acidic rivers, or colonizing freshwater ecosystems after dispersing through marine environments, points towards a clear demand upon cichlid physiological systems. Hence, we pose the hypothesis that physiological plasticity and adaptability predisposed cichlids to capitalizing upon colonization opportunities that arose throughout their evolutionary history may be the principal factor in driving their bewildering diversity.

Literature Cited

Beaulieu, J. M., D. C. Tank, and Donoghue, M. J. 2013. A Southern Hemisphere origin for campanulid angiosperms, with traces of the break-up of Gondwana. BMC Evolutionary Biology 13:80.

Borowiec, M. L. 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4:e1660.

Burress, E. D., J. M. Holcomb, M. Tan, and J. W. Armbruster. 2016. Ecological diversification associated with the benthic-to-pelagic transition by North American minnows. Journal of Evolutionary Biology 30:549-560.

Burress, E. D. 2015. Cichlid fishes as models of ecological diversification: patterns, mechanisms, and consequences. Hydrobiologia 748:7-27.

Chan, K. M., and B. R. Moore. 2002. Whole-tree methods for detecting differential diversification rates. Systematic Biology 51:855-865.

Burress, E. D., A. Duarte, W. S. Serra, M. Loueiro, M. M. Gangloff, and L. Siefferman. 2013. Functional diversification within a predatory species flock. PloS one 8:e80929.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution 29:1969-1973.

Eastman, J. M., L. J. Harmon, and D. C. Tank. 2013. Congruification: support for time scaling large phylogenetic trees. Methods in Ecology and Evolution 4:688-691.

Elmer, K. R., S. Fan, H. Kusche, M. L. Spreitzer, A. F. Kautt, P. Franchini, and A. Meyer. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nature communications 5.

Foote, M. 1994. Morphological disparity in Ordovician-Devonian crinoids and the early saturation of morphological space. Paleobiology 20:320-344.

Ford, A. G., L. Rüber, J. Newton, K. K. Dasmahapatra, J. D. Balarin, K. Bruun, and J. J. Day. 2016. Niche divergence facilitated by fine-scale ecological partitioning in a recent cichlid fish adaptive radiation. Evolution 70:2718-2735.

Friedman, M., B. P. Keck, A. Dornburg, R. I. Eytan, C. H. Martin, C. D. Hulsey, ... and T. J. Near. 2013. Molecular and fossil evidence place the origin of cichlid fishes long after Gondwanan rifting. Proceedings of the Royal Society B. 280:20131733.

Hahn, C., M. J. Genner, G. T. Turner, and D. A. Joyce. 2017. The genomic basis of adaptation to the deep water twilight zone in Lake Malawi cichlid fishes. bioRxiv 102830.

Harmon, L. J., J. B. Losos, T. Jonathan Davies, R. G. Gillespie, J. L. Gittleman, W. B. Jennings, ... and A. Purvis. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution 64:2385-2396.

Hughes, M., S. Gerber, and M. A. Wills 2013. Clades reach highest morphological disparity early in their evolution. Proceedings of the National Academy of Sciences 110:13875-13879.

Hulsey, C. D., R. J. Roberts, Y. H. Loh, M. F. Rupp, and J. T. Streelman. 2013. Lake Malawi cichlid evolution along a benthic/limnetic axis. Ecology and evolution 3:2262-2272.

Joyce, D. A., D. H. Lunt, R. Bills, and G. F. Turner. 2005. An extant cichlid fish radiation emerged in an extinct Pleistocene lake. Nature 435:90.

Kavembe, G. D., A. F. Kautt, G. Machado-Schiaffino, and A. Meyer. 2016. Eco-morphological differentiation in Lake Magadi tilapia, an extremophile cichlid fish living in hot, alkaline and hypersaline lakes in East Africa. Molecular ecology 25:1610-1625.

Kirkpatrick, M., and M. Slatkin. 1993. Searching for evolutionary patterns in the shape of a phylogenetic tree. Evolution 47:1171-1181.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution 34:772-773.

Lewitus, E., and H. Morlon. 2016. Natural constraints to species diversification. PLoS Biology 14:e1002532.

López-Fernández, H., J. H. Arbour, K. Winemiller, and R. L. Honeycutt. 2013. Testing for ancient adaptive radiations in Neotropical cichlid fishes. Evolution 67:1321-1337.

Losos, J. B. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism: American Society of Naturalists EO Wilson Award address. The American Naturalist 175:623-639.

Losos, J. B., and K. D. Queiroz. 1997. Evolutionary consequences of ecological release in Caribbean Anolis lizards. Biological Journal of the Linnean Society 61:459-483.

Lundberg, J. G. 1993. African-South American freshwater fish clades and continental drift: problems with a paradigm. In Biological relationships between Africa and South America (Goldblatt, P., ed.). New Haven, CT. Yale University Press, pp. 156-199.

Machado-Schiaffino, G., A. F. Kautt, H. Kusche, and A. Meyer. 2015. Parallel evolution in Ugandan crater lakes: repeated evolution of limnetic body shapes in haplochromine cichlid fish. BMC evolutionary biology 15:9.

Mahler, D. L., L. J. Revell, R. E. Glor, and J. B. Losos. 2010. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. Evolution 64:2731-2745.

Martin, C. H., J. S. Cutler, J. P. Friel, C. Dening Touokong, G. Coop, and P. C. Wainwright. 2015. Complex histories of repeated gene flow in Cameroon crater lake cichlids cast doubt on one of the clearest examples of sympatric speciation. Evolution 69:1406-1422.

Matschiner, M., Z. Musilová, J. M. Barth, Z. Starostová, W. Salzburger, M. Steel, and R. Bouckaert. 2017. Bayesian phylogenetic estimation of clade ages supports trans-Atlantic dispersal of cichlid fishes. Systematic Biology 66:3-22.

McMahan, C. D., P. Chakrabarty, J. S. Sparks, W. L. Smith, and M. P. Davis. 2013. Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). PloS one 8:e71162.

McPeek, M. A., and J. M. Brown. 2007. Clade age and not diversification rate explains species richness among animal taxa. The American Naturalist 169:E97-E106.

Meier, J. I., D. A. Marques, S. Mwaiko, C. E. Wagner, L. Excoffier, and O. Seehausen. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature Communications 8.

Moore, B. R., S. Höhna, M. R. May, B. Rannala, and J. P. Huelsenbeck. 2016. Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proceedings of the National Academy of Sciences 113:9569-9574.

Morlon, H., E. Lewitus, F. L. Condamine, M. Manceau, J. Clavel, and J. Drury. 2016. RPANDA: an R package for macroevolutionary analyses on phylogenetic trees. Methods in Ecology and Evolution 7:589-597.

Oaks, J. R. 2011. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. Evolution 65:3285-3297.

Pennell, M. W., J. M. Eastman, G. J. Slater, J. W. Brown, J. C. Uyeda, R. G. FitzJohn, ... and L. J. Harmon. 2014. geiger v2. 0: an expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. Bioinformatics, 30:2216-2218.

Rabosky, D. L., and I. J. Lovette. 2008. Density-dependent diversification in North American wood warblers. Proceedings of the Royal Society of London B: Biological Sciences 275:2363-2371.

Rabosky, D. L. 2009a. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. Ecology letters 12:735-743.

Rabosky, D. L. 2009b. Ecological limits on clade diversification in higher taxa. The American Naturalist 173:662-674.

Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PloS one 9:e89543.

Rabosky, D. L., M. Grundler, C. Anderson, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. Methods in Ecology and Evolution 5:701-707.

Rabosky, D. L., and A. H. Hurlbert. 2015. Species richness at continental scales is dominated by ecological limits. The American Naturalist 185:572-583.

Rabosky, D. L., J. S. Mitchell, and J. Chang. 2017. Is BAMM flawed? Theoretical and practical concerns in the analysis of multi-rate diversification models. Systematic Biology 66:477-498.

Říčan, O., L. Piálek, R. Zardoya, I. Doadrio, and J. Zrzavý. 2013. Biogeography of the Mesoamerican Cichlidae (Teleostei: Heroini): colonization through the GAARlandia land bridge and early diversification. Journal of Biogeography 40:579-593.

Rundle, H. D., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287:306-308.

Ruta, M., P. J. Wagner, and M. I. Coates. 2006. Evolutionary patterns in early tetrapods. I. Rapid initial diversification followed by decrease in rates of character change. Proceedings of the Royal Society of London B: Biological Sciences 273:2107-2111.

Salzburger, W., T. Mack, E. Verheyen, and A. Meyer. 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evolutionary Biology 5:17.

Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592-593.

Schluter, D. 2000. *The ecology of adaptive radiation*. OUP Oxford.

Schobert, C. S., M. L. Stiassny, I. R. Schwab, C. Zeiss, R. C. Schelly, and R. R. Dubielzig. 2013. Comparative ocular anatomy in a blind African cichlid fish, Lamprologus lethops. Veterinary ophthalmology 16:359-364.

Seegers, L., and L. De Vos. 1998. Seven new *Orthochromis* species (Teleostei: Cichlidae) from the Malagarasi, Luiche and Rugufu basins (Lake Tanganyika drainage), with notes on their reproductive biology. Ichthyological Exploration of Freshwaters:371-420.

Seehausen, O. 2015. Process and pattern in cichlid radiations–inferences for understanding unusually high rates of evolutionary diversification. New Phytologist 207:304-312.

Shi, J. J., & D. L. Rabosky. 2015. Speciation dynamics during the global radiation of extant bats. Evolution 69:1528-1545.

Slater, G. J. 2015. Iterative adaptive radiations of fossil canids show no evidence for diversity-dependent trait evolution. Proceedings of the National Academy of Sciences 112:4897-4902.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC evolutionary biology 9:37.

Smith, S. A., and B. C. O'Meara. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689-2690.

Suzuki, R., and H. Shimodaira. 2011. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap. R package version 1.2-2.

Tagliacollo, V. A., S. M. Duke-Sylvester, W. A. Matamoros, P. Chakrabarty, and J. S. Albert. 2015. Coordinated dispersal and pre-isthmian assembly of the Central American ichthyofauna. Systematic biology, 66:183-196.

Trewavas, E. 1983. Tilapiine fishes of the genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum (Natural History).

de Quiroz, A. 2005. The resurrection of oceanic dispersal in historical biogeography. Trends in Ecology and Evolution 20:68-73.

Wainwright, P. C., W. L. Smith, S. A. Price, K. L. Tang, J. S. Sparks, L. A. Ferry, ... and T. J. Near. 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Systematic Biology 61:1001-1027.

Wagner, C. E., L. J. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366-369.

Wagner, C. E., L. J. Harmon, and O. Seehausen. 2014. Cichlid species-area relationships are shaped by adaptive radiations that scale with area. Ecology Letters 17:583-592.

Wiens, J. J. 2011. The causes of species richness patterns across space, time, and clades and the role of "ecological limits". The Quarterly Review of Biology 86:75-96.

Yoder, J. B., E. Clancey, S. Des Roches, J. M. Eastman, L. Gentry, W. Godsoe, ... and B. A. J. Sarver. 2010. Ecological opportunity and the origin of adaptive radiations. Journal of evolutionary biology 23:1581-1596.

Table 1. Clades used to evaluate the influence of ecological opportunity on the shape of phylogenies. Asterisks denotes clades that were considered to have incurred ecological opportunities associated with colonization events. Abbreviations correspond to those used in Figure 3.

Clade	N	Abbreviation	Lineage	Region/Ecosystem	Age (My)	λ^*	ψ	η
Apistogramma+allies	16	Api	Geophagini	South America	23.524	7.082	-0.027	1.585
Bathybates+allies	13	Bat	Unnamed clade	Lake Tanganyika	4.751	5.725	1.686	1.064
Cichlasomatini	48	Cich	Cichlasomatini	South America	21.136	8.179	0.295	1.756
Cichlini	14	Cic	Cichlini	South America	6.383	5.646	-0.015	1.395
Chromidotilapia+allies	8	Chr	Chromidotilapiini	West Africa	17.965	5.710	-0.129	1.036
Coptodon+allies	26	Cop	Unamed clade	Africa/Middle East	11.370	6.752	0.874	1.733
Cyphotilapia+allies	12	Cyph	Unnamed clade	Lake Tanganyika	5.253	5.227	0.423	1.097
Cyprichromis+allies	19	Cyp	Cyprichromini	LakeTanganyika	5.534	5.581	0.088	1.661
Ectodini	30	Ect	Ectodini	Lake Tanganyika	5.941	6.248	0.069	1.797
Etroplinae	12	Etr	Etroplinae	India/Madagascar	29.075	6.768	1.070	1.310
Geophagus+allies	26	Geo	Geophagini	South America	23.056	7.715	0.359	1.706
Hemichromis+allies	12	Hem	Unnamed clade	West Africa	20.544	6.608	0.776	1.684
Lamprologini	84	Lam	Lamprologini	Lake Tanganyika	6.0358	7.353	-0.300	1.969
Mesonauta+allies	12	Mes	Heroini	South America	10.032	5.885	0.254	1.750
Oreochromis+allies	21	Ore	Oreochromini	Africa/Middle East	5.437	5.988	0.808	1.282
Ptychochrominae	10	Pty	Ptychochrominae	Madagascar	18.120	6.890	2.157	1.935
Serranochromis+allies	30	Ser	Haplochromini	South/Central Africa	4.943	6.149	0.719	1.583
Amphilophines*	45	Amp	Heroini	Middle America	10.759	7.248	0.824	1.873
Herichthyines*	65	Her	Heroini	Middle America	10.861	7.773	0.397	2.156
*Crenicichla lacustris**	19	Cre	Geophagini	La Plata Basin	10.025	6.329	0.618	1.144
Malawian super flock*	147	Mal	Haplochromini	Lake Malawi	1.833	6.717	1.045	4.986
*Nanochromis**	11	Nan	Chromidotilapiini	Congo River	11.645	5.917	1.281	1.563
*Orthochromis**	8	Ort	Haplochromini	Malagarasi River	4.552	4.790	1.439	1.060
*Steatochromis**	8	Ste	Tilapiini	Congo River	3.796	4.531	0.259	1.002
Stomatepia+allies*	11	Sto	Oreochromini	Lake Barombi Mbo	0.633	2.994	0.090	1.192
Tropheini*	21	Tro	Haplochromini	Lake Tanganyika	3.404	5.454	-0.119	1.313
Victorian super flock*	49	Vic	Haplochromini	Lake Victoria	1.561	3.782	0.247	1.151

Figure legends

Figure 1. Diversification rates across the cichlid phylogeny based on the maximum credibility shift set from BAMM analysis (A). Circles indicate the location of core rate shifts. Prior and posterior distributions of the number of rate shifts (B), diversification rate through time (C), and the distribution of rates (D) across the cichlid phylogeny. Major clades and colonization events are depicted in the adjacent phylogeny for reference (E). The marginal odds ratio tree with major clades color-coded (F).

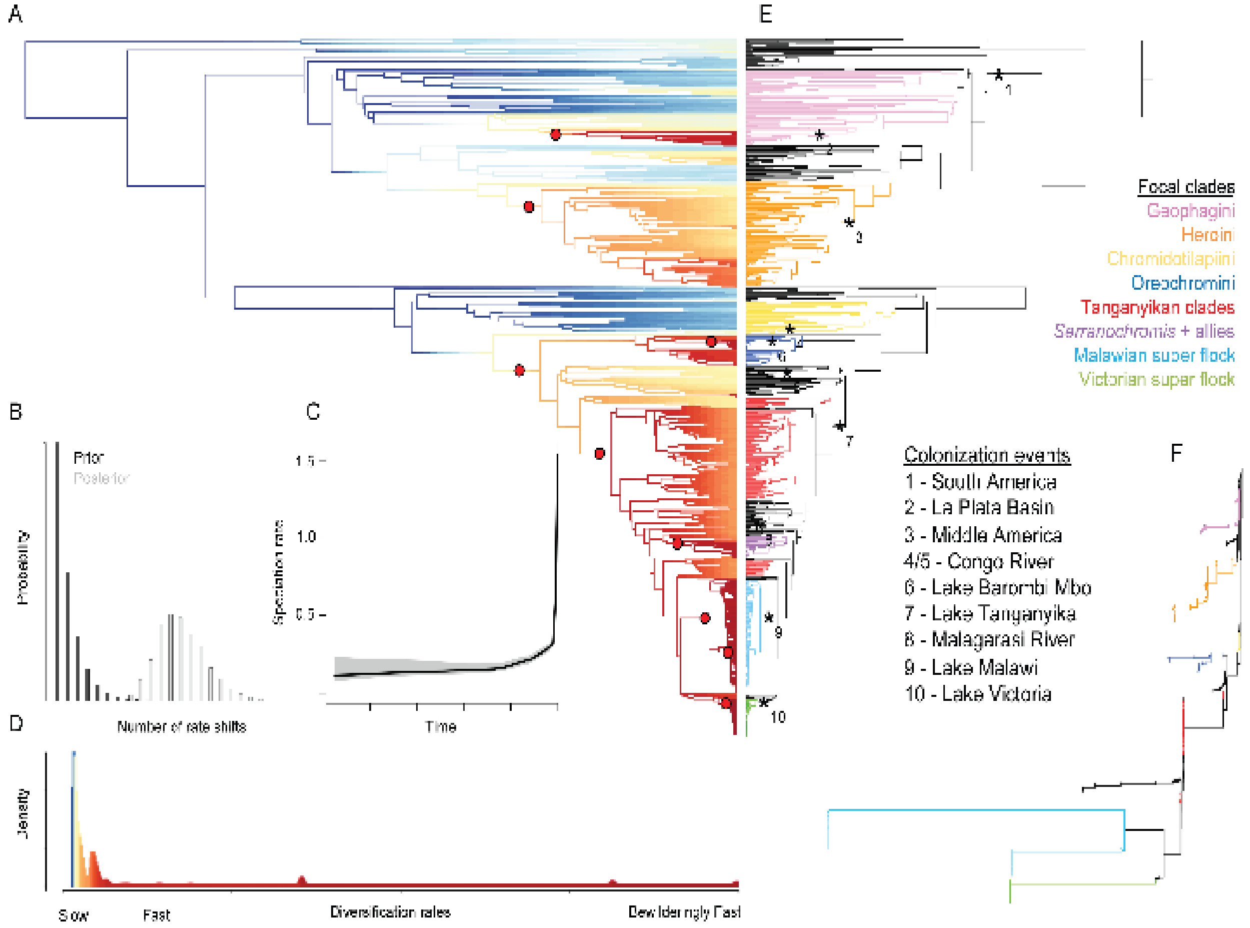

Figure 2. The distribution of the focal clades across phylogenetic space defined by λ^*, ψ, and η (A). The distributions of ln-transformed eigenvalues (λ^*) among the focal clades (B). Effect sizes (r±95%C.I.) calculated from the correlations between clade age and the parameters that describe phylogenetic space (C). The distribution of the focal clades across the relationship between clade age and ln-transformed eigenvalues (D). Hierarchical cluster analysis of the focal clades based parameters (λ^*, ψ, and η) that describe phylogenetic shape (; E). See Table 2 for the clades that correspond to the abbreviations used in panel E.

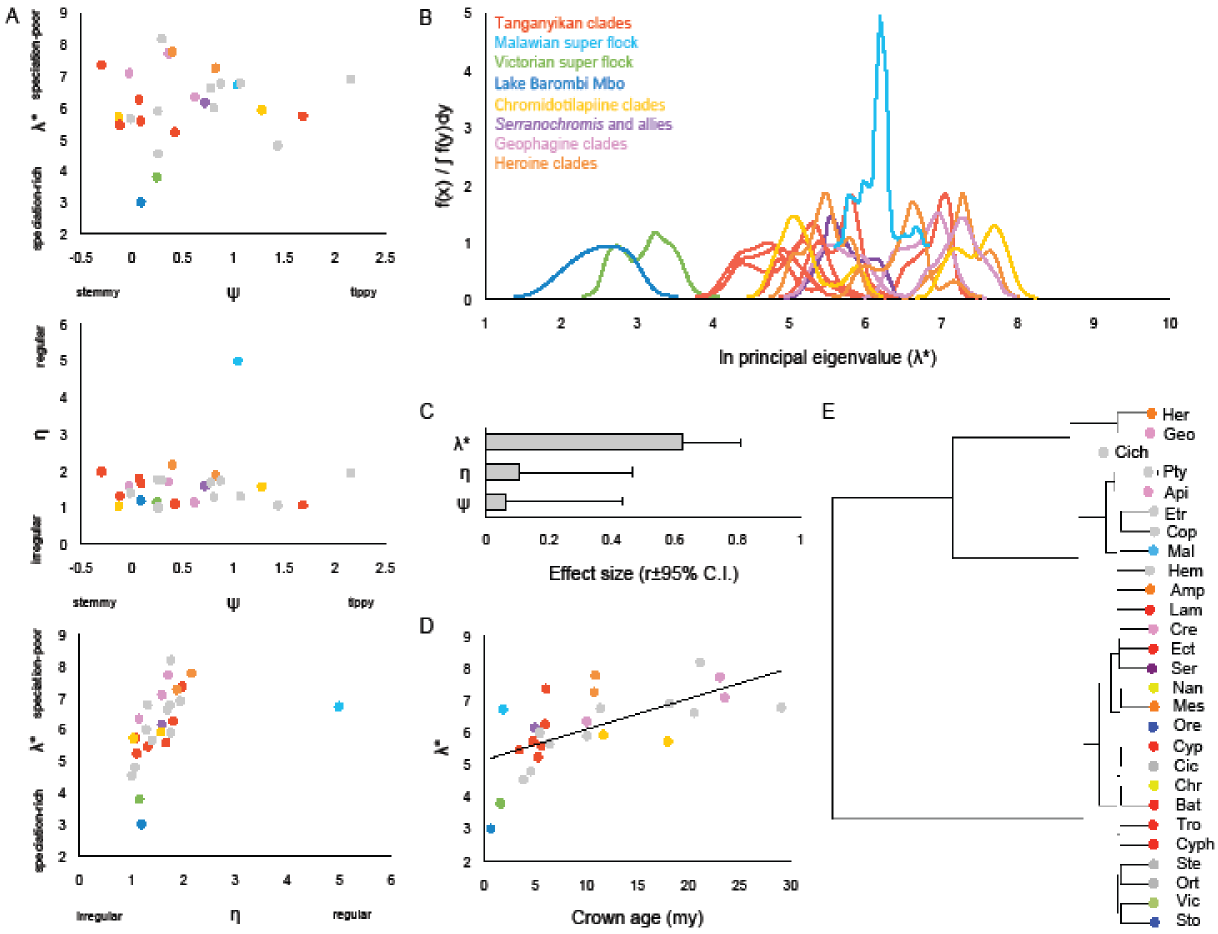

Figure 3. The role of ecological opportunity in the timing and shape of adaptive radiation. The probability of rate shifts at colonization-associated nodes and other nodes (A). The marginal odds ratios of rate shifts at colonization-associated nodes and other nodes (B). The relationship between clade age and ln-transformed eigenvalues (λ*) among colonization-associated clades and other clades (C). Effect sizes (r±95%C.I.) for these correlations are inset.