Diversification Rate is Associated with Rate of Molecular Evolution in Ray-Finned Fish (Actinopterygii)

Andrew M. Ritchie1,3 · Xia Hua1,2 · Lindell Bromham1

Received: 8 June 2021 / Accepted: 24 February 2022 / Published online: 9 March 2022
© The Author(s) 2022

Abstract
Understanding the factors that drive diversification of taxa across the tree of life is a key focus of macroevolutionary research. While the effects of life history, ecology, climate and geography on diversity have been studied for many taxa, the relationship between molecular evolution and diversification has received less attention. However, correlations between rates of molecular evolution and diversification rate have been detected in a range of taxa, including reptiles, plants and birds. A correlation between rates of molecular evolution and diversification rate is a prediction of several evolutionary theories, including the evolutionary speed hypothesis which links variation in mutation rates to differences in speciation rates. If it is widespread, such correlations could also have significant practical impacts, if they are not adequately accounted for in phylogenetic inference of evolutionary rates and timescales. Ray-finned fish (Actinopterygii) offer a prime target to test for this relationship due to their extreme variation in clade size suggesting a wide range of diversification rates. We employ both a sister-pairs approach and a whole-tree approach to test for correlations between substitution rate and net diversification. We also collect life history and ecological trait data and account for potential confounding factors including body size, latitude, max depth and reef association. We find evidence to support a relationship between diversification and synonymous rates of nuclear evolution across two published backbone phylogenies, as well as weak evidence for a relationship between mitochondrial nonsynonymous rates and diversification at the genus level.

Keywords  Sister pairs · Molecular rates · Diversification · Life history · Comparative methods · Phylogeny · Synonymous substitution rates

Introduction
Variation in the tempo and mode of evolution is an important feature of the tree of life, which can potentially reveal evolutionary patterns and processes. At the largest and most obvious scales, the wide variation in diversity observed among higher-order taxa has been a target of evolutionary study for decades (Fischer 1960; May 1988; Currie 1991; Rohde 1992; Farrell 1998; Fine 2015). Although some of this pattern likely reflects differences in the amount of time lineages have to accumulate species (McPeek and Brown 2007; Hedges et al. 2015; Marin et al. 2018), it is clear that the net rate of species diversification varies quite significantly among taxa even after accounting for time (Cardillo 1999; Alfaro et al. 2009; Rabosky et al. 2012; Kozak and Wiens 2016; Scholl and Wiens 2016). At the other end of the scale, the microevolutionary process governing the rate of genomic evolution, driven by spontaneous mutation, segregation and fixation of molecular substitutions, is itself subject to variation on evolutionary timescales. Empirical patterns of molecular rate variation may be significantly correlated to variation in life history characters such as body size (Barrera-Redondo et al. 2018; Berv and Field 2018), lifespan (Galtier et al. 2018) or the rate of reproduction (Welch et al. 2008). These findings of consistent patterns of rate variation between lineages open the door to viewing the
rate of molecular evolution as a trait that can itself evolve and diversify (Bromham 2009, 2011), subject both to selective pressures and random variation (Dobzhansky 1970; Lynch et al. 2016).

An increasing number of empirical studies explore the connection between microevolutionary change at the molecular level and macroevolutionary change at the level of biodiversity. In particular, over the past two decades, a pattern of correlation between the rate of molecular evolution and net diversification rate, as reflected in species richness of lineages, has been observed in several groups of organisms, including plants (Lancaster 2010; Bromham et al. 2015), birds (Lanfear et al. 2010a; Iglesias-Carrasco et al. 2019) and reptiles (Eo and DeWoody 2010). Similar relationships have been reported across a large range of phylogenies (Pagel et al. 2006). Several possible explanations have been given for the observation of a connection between substitution rates and species richness. The process of speciation may itself give rise to a period of accelerated molecular evolution, whether due to population structure, rapid adaptation, or rapid genomic divergence following from reproductive isolation (Barton and Charlesworth 1984; Venditti and Pagel 2010). Higher rates of molecular evolution may cause separated populations to achieve reproductive isolation more rapidly, increasing the chance that temporary separation leads to long-term incompatibility and speciation (Hua and Bromham 2017). Alternatively, diversification rates and rates of molecular evolution may be indirectly related if both are correlated with other variables. For example, climate and environment may be underlying drivers of variation in both molecular evolution and diversification (Barraclough and Savolainen 2001; Gillman and Wright 2014), or both diversification rate and mutation rate might be impacted by species traits such as sexual selection (Iglesias-Carrasco et al. 2019).

A relationship between molecular evolutionary rates and diversification rates is not only of theoretical interest but has direct implications for the relationship between observed molecular variation and the shape of molecular phylogenies. This could affect the inference of phylogenies, divergence times and macroevolutionary processes. Molecular rate variation can complicate the inference of diversification rates from phylogenies (Shafir et al. 2020), especially when there is a link between diversification and molecular rates (Duchêne et al. 2017). In Bayesian divergence time estimation, diversification (“tree”) models and rate variation (“clock”) models are typically specified independently (Bromham et al. 2018), and interactions between them are rarely considered (Sarver et al. 2019). While the impact of variation in rates of molecular evolution with diversification has been studied by simulations, the impact on empirical molecular dating and macroevolutionary studies is currently unknown because the extent of these relationships has not yet been characterised for most groups of organisms. It is, therefore, desirable to investigate how widespread the relationship between diversification rates and rates of molecular evolution is across the tree of life and evaluate the strength of the relationship.

Ray-finned fish (Actinopterygii) are the largest and most diverse major group of vertebrates, with currently over 35,000 described species (Fricke et al. 2022). The degree of diversity varies to a stunning extent among taxonomic groups, from the astounding richness of true gobies (Gobiidae, > 1900 species; McCrane et al. 2020) to the sparseness of the bowfin (Amiidae, 1 species) or sunfish lineages (Molidae, 5 species; Bass et al. 2005; Thompson et al. 2021; Fricke et al. 2022). This variability in diversity extends to biological characteristics: some groups show extreme morphological and ecological variation, such as the cichlids (Henning and Meyer 2014; Salzburger et al. 2014), while others show extreme morphological stasis, such as the paddlefish and bichir (Near et al. 2014; Cheng et al. 2021). It is particularly fascinating to see such diversity in marine or lake environments, because many of the mechanisms of terrestrial speciation, involving geographical barriers, are largely absent (Wiens 2015). Ecological gradients have been suggested to play a prominent role in driving speciation in fish (Schluter et al. 1996; Puebla 2009), and there are many accounts of such processes in specific species pairs, genera or families (e.g. Schliewen et al. 1994; Nagel and Schluter 1998; Rocha et al. 2008; Ingram 2011; May-McNally et al. 2015; Burress et al. 2017).

Other key features of fish diversity include the influence of coral reefs and coastal habitats as hotbeds of species richness and endemism (Cowman and Bellwood 2011; Parravicini et al. 2013; Cowman 2014) and the role of depth gradients in promoting both speciation and ecological divergence (Ingram 2011; Jennings et al. 2013; Ohlberger et al. 2013). An association between depth changes and speciation has been demonstrated in the speciose rockfish genus (Ingram 2011). Body size, usually measured as maximum specimen length, has been found to be negatively related to species richness (Fisher et al. 2010), but size has also been observed to be positively correlated with diversification rates in some reef fish, leading to uncertainty about the nature of the influence of size on speciation and extinction (Siqueira et al. 2020). The rate of body size evolution in fish is also positively correlated with speciation rates (Rabosky and Adams 2012). A latitudinal diversity gradient exists for both marine and freshwater fish species and may reflect factors such as temperature, primary productivity and ecological opportunity (Macpherson 2002; Tittensor et al. 2010; Chaudhary et al. 2016; Hanly et al. 2017), as well as time and historical glaciation cycles (Bernatchez and Wilson 1998; April et al. 2013; Miller and Román-Palacios 2021). However, despite the trend for reduced extant diversity with distance from the equator, rates of diversification have been shown to associate...
positively with latitude across the entire fish tree of life, potentially reflecting fish lineages that have radiated to fill new niches during their recolonisation of the higher latitudes following the last ice age (Rabosky et al. 2018).

The link between molecular rates and diversification rates has not received as much attention as the effects of geography or ecological traits. In addition to the practical interest, this correlation could have important implications for hypotheses of fish diversification. For instance, such a correlation would be consistent with a prediction of the evolutionary speed hypothesis, which suggests that temperature drives faster mutation rates and/or more rapid life history cycles (Rohde 1992; Gillman and Wright 2014). Several correlates of rate of molecular evolution have been identified in fish. Intraspecific genetic diversity has been related to surface temperature and region-specific factors (Manel et al. 2020) and surface temperature is also associated with differences in molecular rates between closely related species (Wright et al. 2011). Molecular rates have been shown to vary with ecological and life history characters, including body size, depth range and association with coral reefs, accounting for phylogenetic covariance across the fish tree of life (May et al. 2020). Body size in particular is frequently identified as a negative correlate of molecular rates (April et al. 2013; Hua et al. 2015; May et al. 2020). Since some of these correlates are also related to species richness or diversification rates, it is possible that molecular rates are related to diversification through life history or habitat, but this has not been thoroughly tested yet.

Rabosky et al. (2018) examined the relationship of tipspecific speciation rates, measured by the DR statistic (Jetz et al. 2012), and root-to-tip substitution path lengths, but the focus was on molecular rate as a confounding factor for the inverse latitudinal speciation gradient, and the relationship was not tested formally. Therefore, while some correlates of both diversification rates and rates of molecular evolution have been identified, it is not yet clear whether a relationship between molecular and diversification rates exists in ray-finned fish, either as an indirect effect of the relationship between life history and habitat with diversification and molecular rates, or as a direct connection between diversification and molecular evolution. This question is timely, as last decade has seen the coverage of the fish tree of life improve in leaps and bounds, providing more and more raw data to address this question (Near et al. 2012; Hughes et al. 2018; Rabosky et al. 2018). This provides the opportunity to determine whether ray-finned fish may be added to the growing number of taxa in which these scales of macro- and micro-evolution are related.

Common approaches for investigating relationships between molecular rates, diversification and life history or ecological traits can be divided into sister-pair analyses and tree-based analyses (Lanfear et al. 2010b). The method of sister-pair comparison is a robust way of exploring interactions between molecular evolution, life history and diversification (Garland et al. 1992; Freckleton 2000; Welch and Waxman 2008; Lanfear et al. 2010b; Bromham et al. 2015). In a sister-pair analysis, taxa are chosen that form pairs of non-overlapping sister clades, with each pair of sister clades sharing a common ancestor that is not included in any other sister comparison (Garland et al. 1992). Tree-based analyses instead infer substitutions for all branches in the phylogeny and correlate these with node numbers on root-to-tip paths in the tree (Webster et al. 2003).

Analysing correlates of diversification rates using sister clades has several important advantages. First, it satisfies the requirement of statistical independence because the differences in substitution rates and diversification rates between sister pairs are independent of any other such comparison, because they represent separate evolutionary histories. Second, because the two sister clades share an initial divergence date, the contrast in net diversification rates between the two clades can be estimated from the difference in number of extant species in these clades (clade sizes), and the contrast in molecular rates can be derived from the average number of substitutions occurring in each clade since divergence. The contrast in clade sizes can then be related to the contrast in inferred substitution rates using ordinary least squares regression. Calculating these contrasts does not require an knowledge of the absolute ages of these clades to serve as the denominator for calculating the rate, although an estimate of relative pair age, which may be based on molecular dates or the sum of branch lengths in the clade, can be used to correct for heteroscedasticity (Garland et al. 1992; Hua et al. 2015). Beyond requiring that the sister pairs be phylogenetically independent so that they are mutually monophyletic and do not overlap on a phylogeny, the method does not require knowledge of the deeper topology of the tree.

The method of sister pairs allows an uncomplicated means of estimating and comparing relative diversification rates and rates of molecular evolution, while satisfying the requirements for statistical independence needed for the analyses and minimises the impact of the many assumptions required to construct time-resolved molecular phylogenies. For example, phylogenetic estimates of substitution rates can be strongly affected by choice and placement of fossil calibrations or the choice of phylogenetic model (Duchêne et al. 2014; Bromham et al. 2018). Sister pairs also offer relatively simple ways of controlling confounding artefacts such as the node-density effect. The node-density effect in phylogenetics refers to a phenomenon in which genetic distances tend to be underestimated along branches in a phylogeny where there are fewer intervening nodes (Venditti et al. 2006; Hugall and Lee 2007). Longer, unbroken branches are more likely to have ‘hidden’ substitutions that are not inferred by the maximum likelihood algorithm unless evidence for their
existence arises from sequences that from intervening lineages that inherited that substitution. Because more substitutions are inferred along paths when there are more intervening lineages, substitution rate estimates could be inflated in clades where there are more species, leading to a spurious positive correlation between species richness and rates of molecular evolution. In a sister-pairs analysis, the effect can be ameliorated by selecting an even number of sequences for each of the two sister clades when estimating substitution rates (Lanfear et al. 2010b).

Despite these many advantages, a perceived disadvantage of sister-pair analyses is that they use only a subset of the available data, unlike tree-based analyses where all available tips in the phylogeny are included, as well as all internal branches. However, tree-based approaches are highly dependent on model-based inference of both speciation and substitution rates which are not only dependent on the assumptions of the model, but are also subject to sampling and interference biases such as incomplete sampling or the node-density effect. To provide a contrast to our sister-pair analyses, we also perform an analysis using the tree-based method of Webster et al. (2003), in which the root-to-tip molecular path length (sum of branch lengths) is regressed against the number of inferred nodes on the root-to-tip path accounting for covariation due to shared history via a linear mixed model (Webster et al. 2003; Pagel et al. 2006; Venditti et al. 2006; Venditti and Pagel 2010). This method is designed to detect the effect of punctuational bursts of substitutions or morphological traits related to speciation events, but could also detect correlated evolution between traits and speciation rates (Rabosky 2012).

Here, we examine pairs of sister clades at different taxonomic levels to investigate the relationship between synonymous and nonsynonymous substitution rates and diversification rates in ray-finned fishes (Actinopterygii). We also perform tree-based analyses for mitochondrial and nuclear gene sequences and for substitutions inferred from a published supermatrix (Rabosky et al. 2018). We investigate potential confounding effects of life history, depth and latitude and examine whether molecular rate variation has a role in mediating their relationships with fish diversity.

Materials and Methods

Molecular Phylogenies

We used two alternative phylogenies to select species pairs for analysis. First, we used a recent ray-finned fish phylogeny incorporating almost all presently described species (Rabosky et al. 2018). The backbone for this phylogeny has been constructed via maximum likelihood estimation using a concatenated matrix of 27 mitochondrial and nuclear coding genes for 11,638 species. The timescale was derived from a penalised likelihood rate smoothing procedure (treePL; Smith and O’Meara 2012). The molecular backbone contains representatives of every extant ray-finned fish family, with an accompanying taxonomy that places nearly all named actinopterygian species on the tree. The authors extended this backbone to make a series of complete trees by inserting species without molecular data using a birth–death-sampling interpolation procedure. This produces a distribution of 100 complete ‘all-taxon-assembled’ time trees with 32,100 tips total (Chang et al. 2020), which we obtained from https://fishtreeoflife.org/.

All phylogenies have some limitations in terms of taxon sampling, analysis method or assumptions made in the analysis. Since Rabosky et al. (2018) take a supermatrix approach, the molecular data are sparse and some genes are represented in many more taxa than others, particularly the mitochondrial barcoding gene cytochrome oxidase I (COI). This means that the data are not equally informative at all resolutions and backbone relationships may be uncertain. To test whether the choice of underlying phylogeny influenced the results, we also ran the analysis on the Euteleost Tree of Life dataset (Betancur-R et al. 2013). The Euteleost Tree of Life backbone phylogeny consists of 1416 taxa and was reconstructed via maximum likelihood. While this phylogeny has fewer species than the phylogeny of Rabosky et al. (2018), the data matrix is more complete, meaning that most species have most of the sequences available.

DNA Sequence Selection

A sister pair consists of two sister clades that share an immediate common ancestor. We use the word ‘sister pair’ for the containing group and ‘sister clades’ for each of the two daughter groups throughout, that is each sister clade is one half of a sister pair (Fig. 1). To maximise the amount of available data for the sister-pairs analysis, we chose genes to provide broad coverage across many sister pairs.

For the analysis based on the phylogeny of Rabosky et al. (2018), the genes chosen were mitochondrial 12S and 16S ribosomal RNA (12s, 16s), cytochrome oxidase I (coI) and cytochrome B (cyTB), and nuclear recombination activating gene 1 (rag1). These sequences were obtained from GenBank using the accessions provided by Rabosky et al. (2018). All four mitochondrial genes (Mitochondrial Total sequences, labelled Mito.All) were concatenated for the tree-based analysis and to estimate total substitution rates. For determining relative nonsynonymous and synonymous substitution rates (dS and dN), only protein-coding sequences can be used, and so coI and cyTB coding sequences (Mitochondrial Coding sequences, labelled Mito.Coding) were concatenated without the rRNA sequences. The nuclear data consisted only of protein-coding sequences and so were used...
for inferring Total substitutions, dS and dN (Nuclear Coding with rag1, labelled Nuc.RAG1).

For the analysis based on the published phylogeny and dataset of Betancur-R et al. (2013), we selected all 19 protein-coding nuclear exonic markers taken from this dataset (labelled Nuc.ETOL). Non-protein-coding and intron data were not used. This dataset allowed for a smaller but more data-rich set of sister pairs, which could provide more power for substitution rate estimation. We used 19 loci from this study, each of which includes exon sequence from nuclear protein-coding genes (ENC1, FICD, GlyT, KIAA1239, Myh6, Panx2, Plagl2, Ptdch1, Rag1, Rag2, rh, Ripk4, SH3PX3, SIDKEY, SREB2, SVEP1, Tbr1, VCPiP, Zic1).

**Sister-Pair Selection**

The sister-pair selection process is illustrated in Fig. 2. The same procedure applies to both phylogenies. We first assigned all species on the backbone phylogeny to families according to the molecular taxonomy provided with the tree on https://fishtreeoflife.org. We then identified families in which at least one tip possessed the target sequences for the dataset—rRNA, col and cytB for Mito.All, col and cytB only for Mito.Coding, and rag1 for Nuc.RAG1. We produced a tree in which each of these families was represented by its own tip, the first in the order in which the phylogeny was presented in the tree file, and all other families were absent. From this tree, we chose all pairs of families that descended from the same node, with no intervening tips – equivalent to all remaining clades of size 2. The pairs obtained by this method constituted our full set of “family-level” sister pairs. To examine the effect of pair age and taxonomic level, we repeated this procedure using a tree in which tips represented genera with available sequence rather than families. We refer to these as our “genus-level” pairs. At both family and genus levels, this was repeated for each of the Mito.All, Mito.Coding and Nuc.RAG1 target sequences for a total of six sets of sister pairs (see Table 1).

For the phylogeny of Rabosky et al. (2018) only, because of the sparse nature of the sequence data in the backbone phylogeny, genus-level relationships were often poorly resolved, which renders true clade sizes uncertain and may affect the comparative analysis. To examine this, we also produced a ‘Filtered’ set by performing an additional filter for mutual monophyly of pairs to the ‘full’ sets of sister pairs. If members of any genus were found in both sister clades of a given sister pair in a given tree of the all-taxon-assembled pseudoposterior (generated by stochastic placement of tips without molecular data), we considered the sister clades not to be mutually monophyletic in that tree. For our optional filtration step, we removed all pairs which were not mutually monophyletic in 80% or more of the 100 all-taxon-assembled trees. We performed this step for each of the six datasets described above (Table 1), producing separate ‘Full’ and ‘Filtered’ analyses for each. Comprehensive code for reproducing our initial sister-pair selections is available online (https://github.com/amritchie/fish-diversity-molecular-evolution).

As a check for possible confounding or mediation relationships between either clade size or substitution rates and life history, we also attempted to construct a set of datasets filtered for the availability of life history data, including body size and depth range. (Supplementary Methods S2 available as online Supplementary Information). However, we found that this extra filtering step reduced the power sufficiently that results found on the Full and Reduced datasets could not be confirmed. We provide some comments on this issue in the Discussion.

**Species Richness (Clade Sizes)**

Each of our comparisons consist of two closely related sister clades, defined by a unique shared common ancestor that is shared with no other such pairs in the dataset. Because of this, each sister clade has had the same amount of time to accumulate species. Therefore, the number of extant lineages in each sister clade can be used to infer differences in net diversification rate between each sister lineage: because they are defined by the same ancestor, they are the same age, and any difference in species richness must indicate a difference in the speciation rate or extinction rate or both. Therefore, we want to know how many extant species can be assigned to each sister clade of each comparison.
If the phylogeny contains all extant species, then we can use clade size to estimate net diversification rate. The Rabosky et al. (2018) tree includes nearly all named actinopterygian species; however, the placement on these species varies between alternative phylogenetic solutions. We estimate clade sizes for in our sister pairs by averaging the number of tips descending from the most recent common ancestor (MRCA) of that clade in each of the 100 all-taxon-assembled sister pairs. In detail, for each all-taxon-assembled tree, we counted how many tips were contained in each sister clade, including species with no molecular data. We then averaged these clade sizes over all 100 all-taxon-assembled trees and took these values as our approximate clade sizes. This procedure accounts for uncertainty in the placement of species without molecular data. We checked whether these clade sizes were a good match to the sizes obtained from formal taxonomies, as done for example in (Lanfear et al. 2010a).

To check whether the average clade sizes produced in this way are a good reflection of species richness, we also generated species richness estimates from formal taxonomy. We used Family- and Genus-level classifications from FishBase and produced clade sizes by counting all valid species names for all families or genera within each sister clade of each sister pair. Because the correlation between clade sizes produced from all-taxon-assembled trees and from formal taxonomies was strong (Supplementary Figs. S1, S2), we used the averaged clade sizes only.

Fig. 2 Schematic of the process used to select sister pairs and analyse the relationship between molecular evolution and diversification. (1) Tips are grouped into the selected taxonomic level, such as families. (2) A family-level tree is produced by pruning each family to a single tip. (3) Families with no sequence data available in the downloaded data set, if any, are removed. (4) From the remaining tips, we select all immediately adjacent pairs of tips (once without sequence data are removed) as our sister pairs. (5) To form our Filtered dataset, we perform checking for mutual monophyly of the clades within each sister pair. Pairs where one or more genera are present in both clades in > 80% of all-taxon-assembled trees in Rabosky et al. (2018) are removed. This was only done for the Rabosky et al. phylogeny, as most families in the Euteleost Tree of Life tree are monophyletic and no all-taxon-assembled trees are available. (6) To reduce the node-density effect, we randomly delete tips from the larger clade of each sister pair until both clades have the same number of tips. (7) To create outgroups for analysis, each sister pair is attached to its nearest sister pair to form a quartet. (8) Molecular sequence alignments are produced for each quartet. (9) Our final dataset consists of a set of contrasts, each consisting of a pair of sister clades, for which we have data on (a) clade sizes for each pair, and (b) molecular substitution rates (branch lengths) inferred from the sequence alignment. (10) A regression is conducted for contrasts in log clade sizes against contrasts in log substitution rates to test for a relationship between substitution and diversification rates.
Molecular Rate Estimation

We used the molecular data to estimate average synonymous (dS) and nonsynonymous (dN) substitutions per site across root-to-tip paths in each sister clade within each sister pair. To do this, for each sister pair, we collected all tips that had target genes available for each sister pair (Fig. 2). We then evened out the number of tips in the two sister clades within each pair by sampling the appropriate number of tips uniformly at random from all tips with sequence data in the larger sister clade. This step helps reduce the impact of the node-density artefact, which can result in clades with fewer tips having lower estimates of substitution rate (Hugall and Lee 2007).

Sister pairs comparisons of rates of molecular evolution require an outgroup to allow the position of the root to be defined. The outgroup should ideally be the nearest relative to the sister pair that is certain not to fall within the sister comparison, yet has the minimum length of branch joining the two pairs in order to minimise saturation. For the sparse dataset of Rabosky et al. (2018), to ensure an outgroup with sufficient data coverage, we selected the most closely related sister pair to form a quartet. Each sister pair in the quartet acts as the outgroup to the other sister pair, rooting the tree and allowing substitutions to be inferred for each sister pair. For the more complete dataset of Betancur-R et al. (2013), we simply selected the nearest single tip with the most complete data as outgroup.

For both phylogenies, we produced preliminary alignments of all mitochondrial and nuclear genes using MAFFT (Katoh and Standley 2013) with the L-INS-I algorithm. We then generate an alignment for each comparison by extracting just the tips included in each quartet. These alignments were then pruned to remove any codons that did not have data for at least one tip from each sister in each pair and from the corresponding outgroup. The monophyly filter removes pairs in which members of one or more genera are found on both sides of the pair in 80% or more of all-taxon-assembled trees incorporating taxa without molecular data. The set of pairs filtered by both Monophyly and Welch filters formed our Filtered dataset, whereas the set of pairs filtered only by the Welch filter constituted our Full dataset.

| Rank   | Alignment | Genes          | Phylogeny | Substitutions | Number of pairs |
|--------|-----------|----------------|-----------|---------------|-----------------|
|        |           |                |           | All           | Mono only (not tested) | Filtered (Mono + Welch) | Full (Welch) |
| Families | Mito.All | rRNA, col, cytB | Rabosky   | Total         | 54              | 41               | 41           | 50          |
| Families | Mito.Coding | col, cytB     | Rabosky   | dS            | 88              | 59               | 58           | 77          |
| Families | Nuc.RAG1 | ragl           | Rabosky   | Total         | 119             | 82               | 82           | 113         |
| Families | Nuc.RAG1 | ragl           | Rabosky   | dS            | 475             | 193              | 144          | 235         |
| Families | Nuc.Coding | 19 nuclear protein-coding exons | Betancur-R | Total         | 118             | 98               | 78           | –           |
| Genera | Mito.All | rRNA, col, cytB | Rabosky   | dN            | 475             | 204              | 83           | 138         |
| Genera | Mito.Coding | col, cytB     | Rabosky   | dN            | 547             | 249              | 121          | 256         |
| Genera | Nuc.RAG1 | ragl           | Rabosky   | dN            | 547             | 239              | 131          | 247         |
| Genera | Nuc.ETOL | 19 nuclear protein-coding exons | Betancur-R | Total         | 118             | 98               | 75           | –           |

Rank is the taxonomic rank at which pairs are selected. Three marker sets are used for the main analysis, with the Mito. All set including non-coding sequences for additional power in estimating total substitution rates. An additional analysis uses the euteleost phylogeny of Betancur-R et al. (2013) and an associated dataset of 19 nuclear exonic sequences (Table S2 available as Supplementary Information). Datasets are further divided by the type of substitution rate estimated. Different substitution types may have different numbers of usable pairs because some PAML analyses fail due to short branch lengths or because pairs are deleted by the Welch filter, which removes shallow pairs with high variance in branch length estimates. The number of pairs remaining is shown after filtering for mutual monophyly (Mono) and after conducting the Welch filter (Welch). The monophyly filter removes pairs in which members of one or more genera are found on both sides of the pair in 80% or more of all-taxon-assembled trees incorporating taxa without molecular data. The set of pairs filtered by both Monophyly and Welch filters formed our Filtered dataset, whereas the set of pairs filtered only by the Welch filter constituted our Full dataset.
maximum likelihood tree search using RAxML v8 (Stamatakis 2014): topology was constrained to ensure each sister pair and its dependent sister clades were monophyletic. We estimated average substitutions along all branches of these quartet phylogenies separately for each dataset (see Table 1) using baseml within the PAML 4 software suite (Yang 2007) using the RAxML topology. Similarly, for each clade within each sister pair, we estimated dS (average synonymous substitutions per site) and dN (average nonsynonymous substitutions per site) with codeml in PAML 4. For the underlying mutation model in codeml and the substitution model in baseml, we used the general time-reversible (GTR) model with gamma-distributed rate variation and 4 gamma categories. A separate codon model was assigned to each sister clade within each sister pair in the quartet. The molecular clock option was not used, meaning that we inferred average substitutions per site rather than branch-specific molecular rates. For each sister clade in each sister pair, a phylogenetic average substitution rate was estimated by successively averaging the lengths of sister branches from the tip down and adding the result to the length of the parent branch. This provides an estimate of the average substitutions per site from the root of a clade to each tip of the clade while ensuring that the estimate is not dominated by speciose subclades, so the contrast in this estimate between sister clades reflects difference in the substitution rate between the sister clades. We removed any pairs for which the estimated average number of synonymous substitutions per site was greater than 2, as this would likely indicate saturation of the available sequence.

As suggested by Welch and Waxman (2008), we also tested for a negative relationship between the absolute value of the normalised molecular rate contrasts and the square root of the inferred age of the pair (derived from the published phylogeny: see Supplementary information). This relationship arises because inferred substitution rates in shallow pairs have higher variance due to the uncertainty of rate estimation with few differences between sequences. The presence of these pairs can reduce the power of the test and lead to false negative results. As suggested by the study, we successively remove the shallowest pair from each dataset until no negative relationship remains, as tested by a simple linear model (Figs. S21, S22, S23 available as Supplementary Information). We repeated this test using Total, dS and dN substitution rate estimates, removing different numbers of pairs for each dataset. Following this procedure, we also observed a distinct grouping of pairs with very small mean substitution rates that nevertheless had extremely large contrasts in substitution rates, which we attributed to the unreliability of the maximum likelihood optimization in PAML for very short branches. We, therefore, removed these points from the analysis (summarised in Table 1). Following this “Welch Filter”, we tested for a relationship between molecular rate and diversification rate contrasts by ordinary least squares regression through the origin using R v3.13 (R Core Team 2019). The regression is forced through the origin because we do not have any expectation regarding the direction of the substitution rate contrast for two clades of identical size, so the expected substitution rate contrast must be zero.

Regression Analysis of Sister Pair Contrasts in Molecular Rates and Diversification

To explore the relationship between molecular rates and diversification rates, we calculated differences between the logarithms of the phylogenetic average dS and dN values for the two clades within each sister pair. Likewise, to estimate contrasts in net diversification rate, we used the difference between the logarithm of the clade size (as calculated from the average of the number of tips in the 100 all-taxon-assembled phylogenies) between the two sister clades of each sister pair. We performed additional tests and data transformations to check whether the size of contrasts was affected by the depth of the pair, as represented by the inferred age of the MRCA of each sister pair, or by the ancestral value of the diversification or substitution rates (using values of time and rates derived from the published phylogeny: see Table S1 and Supplementary Methods S1 available as online Supplementary Information), but these analyses did not materially alter the results, so the analysis presented here uses comparisons without any adjustment for age or ancestral values.

Tree-based Analysis

For the tree-based analysis, we analysed three subsets of the molecular backbone phylogeny inferred by Rabosky et al. (2018). The first was the complete molecular backbone. This is a large tree of 11,638 tips with substitutions inferred from a supermatrix of 27 genes. Although being a supermatrix with significant incomplete data means that inferred substitutions may be less reliable in some parts of the tree and means that we cannot determine separate effects for nuclear and mitochondrial genes, this is by far the most comprehensive dataset presently available and allows the most powerful test for whether a correlation between molecular evolution and diversification is likely to exist. To investigate results for mitochondrial and nuclear sequences separately and enable comparison with our sister-pair analysis, we construct a second set of substitution rate inferences directly from a dataset of all species in the molecular backbone tree having sequences for the full set of four mitochondrial genes (Mito.All) used in the sister-pair analysis (12s, 16s, col and cytB), and a third dataset of all species with nuclear rag1 sequences is available (Nuc.RAG1). We removed 50%
nodes on path = β × path length + intercept + ε,

where β is the regression coefficient, residual vector ε has a multivariate normal distribution with covariance matrix derived from the phylogeny, and δ is a parameter controlling the degree of non-linearity in the response. An optimised value of δ > 1 in combination with a significant likelihood ratio test for β > 0 is taken as evidence for the presence of the node-density effect (Venditti et al. 2006). Simulations indicate that the test for this artefact is quite robust, with a false positive rate of only 1.7% (Venditti et al. 2006). We ran this test with both node numbers and path lengths as the response model and inspected the resulting plots in cases of serious discrepancy, as the three-parameter model can be affected by outliers or groups of tips with the same node number.

Results

Data Selection

The number of pairs selected by the procedure above is shown in Tables 1 and 2, including the Full Dataset (all sister pairs) and the Filtered Dataset (after additional filtering steps). Imposing a requirement that all genera be monophyletic resulted in the removal of up to three quarters of the initial genus-level pairs (Table 1). Additionally, the procedure of Welch and Waxman (2008) for removing shallow sister pairs resulted in greater reductions for dN than dS datasets because of the approximately tenfold smaller number of nonsynonymous substitutions.

Sister-Pair Analysis

For each of our datasets (listed in Table 1), we performed linear regressions of transformed clade size contrasts against log dS or dN substitution rate contrasts (Table 2). For Full Dataset comparisons chosen from the Rabosky et al. tree, clade size had a significantly positive relationship with Total rate of molecular evolution for the nuclear sequences at both family and genus level. Clade size was also significantly associated with dN at family level. For the Filtered Dataset, there was a significant relationship between total nuclear substitution rate at the family level and for nonsynonymous mitochondrial substitution rate (dN) at the genus level. For the comparisons chosen from the Euteleost Tree of Life, both total substitution rate and dS were significantly positively associated with clade size.

Tree-Based Analysis

The tree-based analysis related the number of nodes on each root-to-tip path on the full backbone, mitochondrial and nuclear trees to the total number of substitutions estimated along the path (the path length) (Table 3, Figs. S13, S14 available as Supplementary Information). The values with path length as response variable are shown because inspection of the associated plots indicated that the fitted curves for the mitochondrial tree were strongly affected by outlying values when node number was the response variable. Significant positive correlations were discovered for the backbone tree (P < 10⁻⁶) and the mitochondrial tree (P < 10⁻⁴). The nuclear tree did not show a significant correlation between path lengths and number of nodes per path. For
the mitochondrial tree, $\delta$ is much greater than 1, indicating that some of the trend may be a node-density effect caused by underestimation of long branches. For the Rabosky et al. backbone tree, $\delta$ is close to one and becomes slightly less than one if the node path is used as response variable (0.95), indicating a small or no contribution to the significant positive trend from the node-density effect.

### Power to Detect Rate Differences

If molecular rates are too slow in the loci analysed to generate consistent difference in branch length, or if the pairs used diverged too short a time ago, there may be too few sequence differences to correctly reconstruct the difference in molecular rates. We consider the expected effect of the number of pairs, relative age of pairs and the length of the alignment on the ability to correctly infer substitution rate differences. For the tree-based analyses, the alignments were about 1.38 kb for the nuclear analysis and 1.6 kb for the mitochondrial

---

**Table 2** Regressions through the origin relating contrasts in log substitution rates to contrasts in transformed clade sizes across sister pairs

| Phylogeny       | Mono. Filter | Rank            | Sequence set | Subst. Rate | No. pairs | Coeff | Std. Err | $t$ | P value (>|t|) |
|-----------------|--------------|-----------------|--------------|-------------|-----------|-------|----------|-----|----------|
| Rabosky et al. (2018) | Full Families | Mito.All Total | 50        | 0.82        | 0.68      | 1.20  | 0.24     |     |          |
|                 |              | Mito.Coding dS  | 77        | −0.68       | 1.11      | −0.61 | 0.54     |     |          |
|                 |              | dN              | 78        | −0.07       | 0.35      | −0.21 | 0.83     |     |          |
|                 |              | Nuc.RAG1 Total  | 113       | 0.92        | 0.25      | 3.63  | <0.01    |     |          |
|                 |              | dS              | 65        | 0.72        | 0.28      | 2.60  | 0.01     |     |          |
|                 |              | dN              | 62        | 0.47        | 0.33      | 1.43  | 0.16     |     |          |
| Genera          |              | Mito.All Total  | 64        | 0.07        | 0.55      | 0.12  | 0.90     |     |          |
|                 |              | Mito.Coding dS  | 235       | 0.08        | 0.14      | 0.62  | 0.54     |     |          |
|                 |              | dN              | 138       | −0.05       | 0.10      | −0.53 | 0.60     |     |          |
|                 |              | Nuc.RAG1 Total  | 256       | 0.22        | 0.10      | 2.14  | 0.03     |     |          |
|                 |              | dS              | 355       | 0.02        | 0.08      | 0.29  | 0.77     |     |          |
|                 |              | dN              | 247       | 0.18        | 0.10      | 1.90  | 0.06     |     |          |
| Filtered Families | Mito.All Total | 41              | 0.97       | 0.79        | 1.24    | 0.22  |          |     |          |
|                 |              | Mito.Coding dS  | 58        | 0.24        | 0.57      | 0.42  | 0.68     |     |          |
|                 |              | dN              | 63        | 0.03        | 0.37      | 0.07  | 0.94     |     |          |
|                 |              | Nuc.RAG1 Total  | 82        | 0.74        | 0.31      | 2.38  | 0.02     |     |          |
|                 |              | dS              | 44        | 0.43        | 0.37      | 1.17  | 0.25     |     |          |
|                 |              | dN              | 43        | 0.50        | 0.42      | 1.20  | 0.24     |     |          |
| Genera          |              | Mito.All Total  | 48        | −0.19       | 0.55      | −0.35 | 0.73     |     |          |
|                 |              | Mito.Coding dS  | 144       | −0.05       | 0.13      | −0.40 | 0.69     |     |          |
|                 |              | dN              | 83        | 0.21        | 0.10      | 2.13  | 0.04     |     |          |
|                 |              | Nuc.RAG1 Total  | 121       | 0.32        | 0.19      | 1.71  | 0.09     |     |          |
|                 |              | dS              | 199       | −0.02       | 0.11      | −0.19 | 0.85     |     |          |
|                 |              | dN              | 131       | 0.07        | 0.14      | 0.49  | 0.63     |     |          |
| Betancur-R et al. (2013) | Yes Families | Nuc.ETOL Total | 78        | 0.64        | 0.30      | 2.13  | 0.04     |     |          |
|                 |              | dS              | 75        | 0.81        | 0.36      | 2.26  | 0.03     |     |          |
|                 |              | dN              | 78        | 0.08        | 0.21      | 0.36  | 0.72     |     |          |

Regressions were conducted on the phylogenies of Rabosky et al. (2018) and Betancur-R et al. (2013), and with or without filtering pairs for mutual monophyly of formal taxa (Mono. Filter). Data is shown for mitochondrial sequences with rRNA (Mito.All) and without (Mito.Coding). Nuclear data consists of RAG1 only for the Rabosky tree and of 19 nuclear marker from the Euteleost tree of life (ETOL) for the Betancur tree. Total, synonymous (dS) and nonsynonymous (dN) substitution rates are tested as predictors of clade size. $t$ statistics and P values are for Wald tests. Results with P < 0.05 are in italics

---

**Table 3** Results of the tree-based analyses performed on the full molecular backbone of Rabosky et al. (2018) and two sampled trees with inferred mitochondrial and nuclear substitutions

| Tree             | Slope     | Likelihood ratio | P value  | $\delta$ |
|------------------|-----------|------------------|----------|----------|
| Backbone         | $2.4 \times 10^{-4}$ | 27.05          | $1.98 \times 10^{-7}$ | 1.03     |
| Mitochondrial    | $2.4 \times 10^{-3}$ | 25.80          | $3.79 \times 10^{-5}$ | 1.98     |
| Nuclear          | $1.2 \times 10^{-5}$ | 0.54           | 0.46      | 1.00     |

The results are shown with root-to-tip path length as the response variable. A significant likelihood ratio test (bold type) indicates a likely correlation between substitutions and node numbers. The combination of a significant likelihood ratio test and a $\delta$ curvilinearity value greater than one indicates the likely presence of the node-density artefact in the Actinopterygian tree. Trends are plotted in Figs. S13, S14 available as Supplementary Information
genes. Our sister-pair alignments averaged 1.2 kb for nuclear and mitochondrial coding alignments (Nuc.RAG1, Mito. Coding) and 2.5 kb for the full mitochondrial alignment (Mito.All). Sister-pair analyses of similar sized datasets have yielded significant results in previous studies (Bromham et al. 2015). The mean age of sister pairs in this study was about 74 million years for families and 48 million years for genera, and molecular rates were on the order of 10^{-3} substitutions per site per million years. Without resorting to extensive simulations, we consider the informativeness of the dataset by approximating the number of substitutions for our average alignment as a Poisson distribution, with mean 74 Myr x 10^{-3} subs/site/Myr x 2.5 x 10^{3} sites = 185 substitutions. We can then consider how frequently the contrast in underlying rates of molecular evolution would show up as positive contrasts in the number of substitutions in each clade within each sister pair. Assuming a conservative average 10% difference in underlying substitution rates between sister pairs, about 82% of underlyingly positive rate contrasts would show up as positive substitution rate contrasts for the family-level mitochondrial dataset, whereas number would only be about 67% for the nuclear genus-level dataset with average 1.2 kb alignments. By this calculation, adding an extra marker with 1.5 kb of alignment would improve the power to reconstruct substitution rate contrasts by about the same amount as adding 6 additional sister pairs. In the case of the fish data, having more stringent pair selection, requiring taxa with more available sequences to increase the alignment length, would have reduced the number of possible pairs by considerably more than half a dozen pairs. Therefore, we consider our sister-pair selection to be as good as practically possible given the available data, and that given the available data, we could not have increased our power to detect consistent rate differences by only choosing pairs with longer alignments.

**Discussion**

Our results provide evidence for a relationship between molecular evolution and diversification in ray-finned fish, with significant positive associations between clade size and rates of molecular evolution in nuclear sequences. This association was found for nuclear total substitution rates and synonymous substitution rates, as well as for mitochondrial nonsynonymous substitution rates (in the Filtered dataset only). The result is consistent with the finding of a relationship between root-to-tip path lengths and number of nodes on the full backbone phylogeny of Rabosky et al. (2018).

Our results are necessarily dependent on the accuracy of the phylogenies and taxonomies used, which may affect estimates of both diversification rates and molecular rates, in particular if the relationships between sister families or genera are reconstructed incorrectly. However, the findings for nuclear nonsynonymous substitution rates are replicated on two separate phylogenies with different molecular datasets and are, therefore, unlikely to be artefacts of particular markers or of the details of phylogenetic reconstruction. A tree-based analysis of the complete supermatrix from Rabosky et al. (2018) also found an association between molecular path lengths to node counts along root-to-tip paths, as has been found in many other taxa across the tree of life, and this is probably not primarily a result of the node-density effect based on a value of δ close to 1. This relationship is also found in a mitochondrial tree-based dataset, though here the node-density effect cannot be ruled out. We, therefore, consider that these results are consistent with a general association between nuclear synonymous substitution rates and diversification rates in bony fish.

These results from both sister-pairs analyses and whole-tree analyses suggest that ray-finned fish conform to a widespread relationship between speciation and molecular divergence (Webster et al. 2003; Pagel et al. 2006; Venditti et al. 2006), consistent with previous studies in plants, birds and other reptiles (Lanfear et al. 2010a; Eo and DeWoody 2010; Bromham et al. 2015). Detecting an association between substitution rates and clade size does not distinguish between alternative explanations for the relationship between diversification rate and rate of molecular evolution. However, the positive relationship between nuclear synonymous substitutions and diversification rates from our sister-pairs analysis indicates an association between mutation rate and diversification. This observation suggests a causal direction for the association. A correlation between mutation rate and diversification rate could arise if generation of variation during molecular evolution drives reproductive isolation or ecological divergence (Hua and Bromham 2017). Higher mutation rates cause more rapid accumulation of Bateson–Dobzhansky–Muller incompatibilities, speeding the acquisition of genetic barriers between species. It is more challenging to think of a plausible way that a higher rate of diversification could consistently raise the mutation rate. If speciation involves a strong population bottleneck it could cause the fixation of slightly deleterious changes to DNA repair proteins, but presumably these would be subject to negative selection after the speciation event to reduce the mutation rate. Further, in this instance we might expect to see strong relationships between dN rates and diversification rates, whereas this is not found for nuclear data in our analysis and is only found in one Filtered dataset for mitochondrial data. An association between net diversification and synonymous substitution rate could potentially be generated if lower mutation rate was linked to higher rates of extinction, for example through reduced capacity to respond to environmental change, though this would be challenging to
evaluate given the lack of independent evidence of extinction rates in bony fish lineages across the phylogeny.

These results add to a growing picture linking change at the genomic level to the generation of biodiversity. If lineages that produce more diversity tend to have higher rates of molecular evolution, and particularly if this relationship is observed for synonymous substitution rates which are considered to predominantly reflect differences in the underlying mutation rate, this suggests that increased supply of changes to DNA sequences is somehow connected to increased speciation rate (or decreased extinction rate or both). Since the genes we have examined here are predominantly “housekeeping genes” associated with basic biochemical, genetic or metabolic functions, it seems unlikely that the association between rate of molecular change and diversification is primarily driven by selective processes such as competition, adaptation to new niches, mate recognition or reinforcement. Instead, the results are consistent with higher mutation rate driving a higher substitution rate which drives faster genome divergence and more rapid completion of speciation through genomic incompatibility.

Life history variation is a potential confounding or mediating factor in an analysis of the association between molecular rates and diversification. Body mass in fish has previously been found to be negatively associated with molecular rates (May et al. 2020) and has been shown to associate with diversification rates for reef fish, although this association was positive, whereas the relationship in our data was consistently negative (Siqueira et al. 2020). Other traits known or believed to associate with fish diversity. Specifically, depth range has been to correlate with extant species richness (Smith and Brown 2002; Fisher et al. 2010), but this may be the result of accumulation of species over longer stable time periods rather than diversification rate variation. Unfortunately, it was not possible to test comprehensively for these effects because, for most of our datasets, adding an additional filtering step for the availability of life history variables reduced the power of the analysis so much that the relationship between molecular rates and diversification rates could no longer be detected. The one dataset where a weak relationship remained (Full data, Family level, Nuc. Rag1) indicated that a model including both a body size measure (maximum length) and synonymous substitution (dS) rates explained significantly more variation in clade size than a model with body size alone, suggesting that the observed relationship between dS rates and clade sizes is not wholly explained by an indirect relationship with body size. However, the utility of this test is low because of the reduced power of the dataset and the lack of confirmation from other substitution rate types. It will be valuable for future studies to revisit this question as more molecular data are obtained for resolving genus- and species-level relationships across a wider range of ray-finned fish taxa.

It has previously been suggested that higher rates of mutation at warmer temperatures lead to faster molecular evolution and more rapid diversification (Rohde 1992; Gillman and Wright 2014; Orton et al. 2019) and that fish diversification is associated with temperature (Wright et al. 2011), although the exact mechanism linking temperature and mutation is unclear. These studies identify interesting correlates of diversification rate that suggest a role for rate of genomic divergence in fish diversification; however, it is likely to be one contributing factor in a complex pattern of diversification influenced by many environmental and evolutionary factors. The importance of glaciation history (Bernatchez and Wilson 1998; April et al. 2013; Miller and Román-Palacios 2021), recent radiations at high latitudes (Rabosky et al. 2018) and coral reefs (Cowman and Bellwood 2011; Cowman 2014) makes it clear that there is much more to the story of fish diversification than general patterns in energy and molecular evolution, but these remain a plausible part of the tale.

Conclusion

We find evidence for a link between rate of molecular evolution and rates of diversification in ray-finned fish. Significant associations between clade size and synonymous, but not nonsynonymous, nuclear substitution rates suggest that the observed relationship may be driven by lineage-specific differences in mutation rate. We confirm this relationship on two separate phylogenies and molecular datasets and at family and genus taxonomic scales. Future investigations should be undertaken as more species-level molecular data accumulate, finer-scale taxonomic resolution is achieved, and more comprehensive data on life history and ecology are compiled.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00239-022-10052-6.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This work was supported by the Australian Research Council Grant DP190101916.

Data Availability All data are from freely available online sources. Data and code for this project are also available online at https://github.com/amritchie/fish-diversity-molecular-evolution.

Declarations

Conflict of interest There is no competing interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source.
provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Alfaro ME, Santini F, Brock C, Alamillo H, Dornburg A, Rabosky DL, Carnevale G, Harmon LJ (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Natl Acad Sci USA 106:13410

April J, Hanner RH, Mayden RL, Bernatchez L (2013) Metabolic rate and climatic fluctuations shape continental wide pattern of genetic divergence and biodiversity in fishes. PLoS ONE 8:e70296

Barrera-Redondo J, Ramirez-Barahona S, Eguiarte LE (2018) Rates of molecular evolution in tree ferns are associated with body size, environmental temperature, and biological productivity. Evolution 72:1050

Bass AL, Dewar H, Thys T, Streelman JT, Karl SA (2005) Evolu-
tionary divergence among lineages of the ocean sunfish family, Molidae (Tetraodontiformes). Mar Biol 148:405

Bromham L, Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol Phylogenet Evol 78:277

Duchêne DA, Hua X, Bromham L (2017) Phylogenetic estimates of diversification rate are affected by molecular rate variation. J Evol Biol 30:1884

Fo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proc R Soc B 277:3387

Farrell BD (1998) “Inordinate fondness” explained: why are there so many beetles? Science 281:555

Fine PVA (2015) Ecological and evolutionary drivers of geographic variation in species diversity. Annu Rev Ecol Ecol Syst 46:369

Fischer AG (1960) Latitudinal variations in organic diversity. Evolution 14:64

Fisher JAD, Schmieder WN, Van der Laan R (2022) Schmieder's catalog of fishes: genera, species, references

Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU (2009) Mitochondrial diversity plus high turnover explain species diversity in fishes. PLoS Curr. https://doi.org/10.1371/currents. tol.53ba26ed4db01ca675bb165c82c62288

Cheng P, Huang Y, Lv Y, Du H, Ruan Z, Li C, Ye H, Zhang H, Wu J, Wang C, Ruan R, Li Y, Bian C, You X, Shi C, Han K, Xu J, Shi Q, Wei Q (2021) The American paddlefish genome provides novel insights into chromosomal evolution and bone mineralization in early vertebrates. Mol Biol Evol 38:1595

Cowman PF (2014) Historical factors that have shaped the evolution of tropical reef fishes: a review of phylogenies, biogeography, and remaining questions. Front Genet. https://doi.org/10.3389/fgene.2014.00394

Cowman PF, Bellwood DR (2011) Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. J Evol Biol 24:2543

Currie DJ (1991) Energy and large-scale patterns of animal- and plant-species richness. Am Nat 137:27

Dobzhansky TG (1970) Genetics of the evolutionary process. Columbia University Press, New York

Duchêne S, Lanfear R, Ho SYW (2014) The impact of calibration and clock-model choice on molecular estimates of divergence times. Mol Phylogenet Evol 78:277

Duchêne DA, Hua X, Bromham L (2017) Phylogenetic estimates of diversification rate are affected by molecular rate variation. J Evol Biol 30:1884

Eo SH, DeWoody JA (2010) Evolutionary rates of mitochondrial genomes correspond to diversification rates and to contemporary species richness in birds and reptiles. Proc R Soc B 277:3387

Farrell BD (1998) “Inordinate fondness” explained: why are there so many beetles? Science 281:555

Fine PVA (2015) Ecological and evolutionary drivers of geographic variation in species diversity. Annu Rev Ecol Ecol Syst 46:369

Fischer AG (1960) Latitudinal variations in organic diversity. Evolution 14:64

Fisher JAD, Frank KT, Leggett WC (2010) Global variation in marine fish body size and its role in biodiversity–ecosystem functioning. Mar Ecol Prog Ser 405:1

Freckleton RP (2000) Phylogenetic tests of ecological and evolutionary hypotheses: checking for phylogenetic independence. Funct Ecol 14:129

Fricke R, Schmieder WN, Van der Laan R (2022) Schmieder's catalog of fishes: genera, species, references

Galtier N, Jobson RW, Nabholz B, Glémin S, Blier PU (2009) Mitochondrial diversity plus high turnover explain species diversity in fishes. PLoS Curr. https://doi.org/10.1371/currents. tol.53ba26ed4db01ca675bb165c82c62288
Hughes LC, Orti G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Arctica D, Betancur-R R, Li C, Becker L, Bellora N, Zhao X, Li X, Wang M, Fang C, Xie B, Zhou Z, Huang H, Chen S, Venkatess B, Shi Q (2018) Comprehensive phylogeny of ray-finned fishes (Actinopterygii) based on transcriptomic and genomic data. Proc Natl Acad Sci USA 115:6249

Iglesias-Carrasco M, Duchene DA, Head ML, Moller AP, Cain K, Hughes LC, Ortí G, Huang Y, Sun Y, Baldwin CC, Thompson AW, Kozak KH, Wiens JJ (2016) Testing the relationships between divergence rate and species richness in passerines. Biol Lett. https://doi.org/10.1098/rsbl.2019.0257

Ingram T (2011) Speciation along a depth gradient in a marine adaptive radiation. Proceedings of the Royal Society B-Biological Sciences 278:613

Jennings RM, Etter RJ, Ficarra L (2013) Population differentiation and species formation in the deep sea: the potential role of environmental gradients and depth. PLoS ONE 8:e77594

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. Nature 491:444

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772

Kozak KH, Wiens JJ (2016) Testing the relationships between diversification, species richness, and trait evolution. Syst Biol 65:975

Lancaster LT (2010) Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. BMC Evol Biol 10:162

Lanfear R, Ho SYW, Love D, Bromham L (2010a) Mutation rate influences diversification rate in birds. Proc Natl Acad Sci USA 107:20423

Lanfear R, Welch JI, Bromharn L (2010b) Watching the clock: studying variation in rates of molecular evolution between species. Trends Ecol Evol 25:495

Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, Foster PL (2016) Genetic drift, selection and the evolution of the mutation rate. Nat Rev Genet 17:704

Macpherson E (2002) Large-scale species-richness gradients in the Atlantic Ocean. Proc Biol Sci 269:1715

Manel S, Guérin P-E, Mouillot D, Blanchet S, Velez L, Albouy C, Pellissier L (2020) Global determinants of freshwater and marine fish genetic diversity. Nat Commun 11:692

Marin J, Rapacciuolo G, Costa GC, Graham CH, Brooks TM, Young BE, Radeloff VC, Behm JE, Helmus MR, Hedges SB (2018) Evolutionary time drives global tetrapod diversity. Proc R Soc B 285:20172378

May RM (1988) How many species are there on earth? Science 241:1441

May JA, Feng Z, Orton MG, Adamowicz SJ (2020) The effects of ecological traits on the rate of molecular evolution in ray-finned fishes: a multivariable approach. J Mol Evol 88:689

May-McNally SL, Quinn TP, Taylor EB (2015) Low levels of hybridization between sympatric Arctic char (Salvelinus alpinus) and Dolly Varden char (Salvelinus malma) highlight their genetic distinctiveness and ecological segregation. Ecol Evol 5:3031

McCraney WT, Thacker CE, Alfaro ME (2020) Supermatrix phylogeny resolves goby lineages and reveals unstable root of Gobiaria. Mol Phylogenet Evol 151:106862

McPeek Mark A, Brown Jonathan M (2007) Clade age and not diversification rate explains species richness among animal taxa. Am Nat 169:E97

Miller EC, Román-Palacios C (2021) Evolutionary time best explains the latitudinal diversity gradient of living freshwater fish diversity. Glob Ecol Biogeogr 30:749

Nagel L, Schluter D (1998) Body size, natural selection, and speciation in sticklebacks. Evolution 52:209

Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL (2012) Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci USA 109:13698

Near TJ, Dornburg A, Tokita M, Suzuki D, Brandley MC, Friedman M (2014) Boom and bust: ancient and recent diversification in bichirs (Polypteridae: Actinopterygii), a relicual lineage of ray-finned fishes. Evolution 68:1014

Ohlberger J, Brännstrom A, Dieckmann U (2013) Adaptive phenotypic diversification along a temperature-depth gradient. Am Nat 182:359

Orton MG, May JA, Ly W, Lee DJ, Adamowicz SJ (2019) Is molecular evolution faster in the tropics? Heredity 122:513

Pagel M, Venditti C, Meade A (2006) Large punctuational contribution of speciation to evolutionary divergence at the molecular level. Science 314:119

Parravicini V, Kulbicki M, Bellwood DR, Friedlander AM, Arias-Gonzalez JE, Chabanet P, Floeter SR, Myers R, Vigliola L, D’Agata S, Mouillot D (2013) Global patterns and predictors of tropical reef fish species richness. Ecosphere 36:1254

Puebla O (2009) Ecological speciation in marine v. freshwater fishes. J Fish Biol 75:960

R Core Team (2019) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rabosky DL (2012) Positive correlation between diversification rates and phenotypic evolvability can mimic punctuated equilibrium on molecular phylogenies. Evolution 66:2622

Rabosky DL, Adams DC (2012) Rates of morphological evolution are correlated with species richness in salamanders. Evolution 66:1807

Rabosky DL, Slater GI, Alfaro ME (2012) Clade age and richness are decoupled across the Eukaryotic tree of life. PLoS Biol 10:e1001381

Rohde K (1992) Latitudinal gradients in species-diversity—the search for the primary cause. Oikos 65:514

Salzburger W, Van Bocxlaer B, Cohen AS (2014) Ecology and evolution of the African great lakes and their faunas. Annu Rev Ecol Evol Syst 45:519

Salzburger W, Van Bocxlaer B, Cohen AS (2014) Ecology and evolution of the African great lakes and their faunas. Annu Rev Ecol Evol Syst 45:519

Sarver BAJ, Pennell MW, Brown JW, Keeble S, Hardwick KM, Sullivan J, Harmon LJ (2019) The choice of tree prior and molecular clock does not substantially affect phylogenetic inferences of diversification rates. PeerJ 7:e6334

Scheuwlen UK, Tautz D, Pääbo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368:629

Sclater D, Rambaut A, Clarke BC, Grant PR (1996) Ecological speciation in postglacial fishes. Philos Trans R Soc Lond Ser B 351:807

Schluter D, Wiens JJ (2016) Diversification rates and species richness across the tree of life. Proc Biol Sci 283:20161334

Shafrir A, Azouri D, Goldberg EE, Mayrose I (2020) Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates. Evolution 74:1620

Siqueira AC, Morais RA, Bellwood DR, Cowman PF (2020) Trophic innovations fuel reef fish diversification. Nat Commun 11:2669

Smith KE, Brown JH (2002) Patterns of diversity, depth range and body size among pelagic fishes along a gradient of depth. Glob Ecol Biogeogr 11:313

Smith SA, O’Meara BC (2012) treePL: divergence time estimation using penalized likelihood for large phylogenies. Bioinformatics 28:2689
Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312
Thompson A, Hawkins M, Parey E, Wcisel D, Ota T, Kawasaki K, Funk E, Losilla M, Fitch O, Pan Q, Feron R, Milhes M, Racicot B, Childs K, Fontenot Q, Ferrara A, David S, McCune A, Dornburg A, Yoder J, Guiguen Y, Crollius HR, Berthelot C, Harris M, Braasch I (2021) The bowfin genome illuminates the developmental evolution of ray-finned fishes. Nat Genet 53:1373
Tittensor DP, Mora C, Jetz W, Lotze HK, Ricard D, Berghe EV, Worm B (2010) Global patterns and predictors of marine biodiversity across taxa. Nature 466:1098
Venditti C, Pagel M (2010) Speciation as an active force in promoting genetic evolution. Trends Ecol Evol 25:14
Venditti C, Meade A, Pagel M (2006) Detecting the node-density artifact in phylogeny reconstruction. Syst Biol 55:637
Webster AJ, Payne RJH, Pagel M (2003) Molecular phylogenies link rates of evolution and speciation. Science 301:478
Welch JJ, Waxman D (2008) Calculating independent contrasts for the comparative study of substitution rates. J Theor Biol 251:667
Welch JJ, Bininda-Emonds ORP, Bromham L (2008) Correlates of substitution rate variation in mammalian protein-coding sequences. BMC Evol Biol 8:53
Wiens JJ (2015) Faster diversification on land than sea helps explain global biodiversity patterns among habitats and animal phyla. Ecol Lett 18:1234
Wright SD, Ross HA, Jeanette Keeling D, McBride P, Gillman LN (2011) Thermal energy and the rate of genetic evolution in marine fishes. Evol Ecol 25:525
Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.