COMPARATIVE PHYLOGENETIC APPROACHES TO MACROEVOLUTION: METHODS, MODELS, AND MARINE FISHES

By

CHAD DANIEL BROCK

A thesis submitted in partial fulfillment of
the requirements for the degree of

Master of Science in Zoology

WASHINGTON STATE UNIVERSITY
School of Biological Sciences

May 2009

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of CHAD DANIEL BROCK find it satisfactory and recommend that it be accepted.

_____
Michael E. Alfaro, Ph.D., Chair

_____
Luke J. Harmon, Ph.D.

_____
Michael S. Webster, Ph.D.

ACKNOWLEDGMENTS

CDB would like to acknowledge Amanda Donabauer and Alex Dornburg for help with figures, data acquisition, and analyses. I would also like to acknowledge the Alfaro lab and Devin Drown for comments on previous drafts of one or both manuscripts. Lastly I would like to thank my family and especially my wife Danielle for their continued love and support. This thesis is dedicated to the memory of Earl Dean Anderson, Huxley, Bud, and Nia Walz with love.

COMPARATIVE PHYLOGENETIC APPROACHES TO MACROEVOLUTION: METHODS, MODELS, AND MARINE FISHES

Abstract

by Chad Daniel Brock, MS.
Washington State University
May 2009

Chair: Michael E. Alfaro

Chapter 1: We used simulations to show that overdispersed non-random sampling substantially increases type-I error of the MCCR test. We develop two corrected MCCR tests (the POS and OSO tests) that alleviate some of this bias and behave well under the non-random sampling schemes investigated. We apply these new tests to a reinvestigation of the temporal patterns of diversification in the SE Asian radiation of homalopsid snakes. We find support for an early-burst of cladogenesis under the POS test, though not the OSO. Uncertainty in the degree of overdispersion when sampling may make it difficult to assess which correction is most appropriate, but if researchers have some knowledge of the sampling scheme used (i.e. that at least the bottom-most *n* nodes were sampled) one can fine-tune these approaches to better mimic the empirical sampling scheme.

Chapter 2: We present the most comprehensive investigation of temporal patterns of diversification to date for the marine radiation of *Sebastes* rockfishes. We find strong support for an early period of rapid cladogenesis and provide the first model-based evidence for density-dependent diversification in *Sebastes,* consistent with the idea of rockfish as an ancient adaptive radiation (Johns and Avise, 1998). Furthermore we provide the first comparative phylogenetic evidence for clade-specific rates of diversification within the genus. Lastly, our KMRI test and two-rate model-fitting identify the late Miocene as a period of exceptional diversification, lending support to species diversification in *Sebastes* being linked with levels of nearshore marine productivity.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Dedication

This thesis is dedicated to the memory of Earl Dean Anderson, Huxley, Bud, and Nia Walz who provided emotional support and comfort throughout. You are missed.

CHAPTER ONE: TESTING FOR TEMPORAL VARIATION IN DIVERSIFICATION RATES WHEN SAMPLING IS INCOMPLETE AND NONRANDOM

INTRODUCTION

The continued increase in molecular data coupled with more sophisticated methods for reconstructing time-calibrated molecular phylogenies (=chronograms) has led to an upsurge in the number of studies investigating patterns of lineage diversification (Weir and Schluter, 2004; Kozak et al. 2006; Rabosky, 2006; Alfaro et al. 2007a,b; Wiens, 2007). In particular, the investigation of temporal patterns of cladogenesis have been especially prominent and a number of studies have found evidence for elevated rates of early cladogenesis across a range of taxonomic groups (Pybus and Harvey, 2000; Ruber and Zardoya, 2005; Weir, 2006; Alfaro et al. 2007b; Price, 2007; Rabosky and Lovette, 2008a; Phillimore and Price, 2008). To date, the most common method employed when investigating temporal patterns of cladogenesis has been the constant-rates (CR) test of Pybus and Harvey (2000). This test evaluates the fit of a constant-rates pure birth (PB) process to the branching events in a tree, using a statistic, γ [1] (Pybus and Harvey, 2000; See also Zink and Slowinski, 1995):

[1] $$\gamma = \frac{\left(\frac{1}{n-2}\sum_{i=2}^{n-1}\left(\sum_{k=2}^{i} k g_k\right)\right) - \left(\frac{T}{2}\right)}{T\sqrt{\frac{1}{12(n-2)}}}, \quad T = \left(\sum_{j=2}^{n} j g_j\right)$$

where n is the number of lineages in the tree, and g_2, g_3, ...g_n are the internode distances. This statistic measures the standardized difference between the average sum of branch lengths between each internal node and the root and the midpoint of the tree (Pybus and Harvey, 2000; See also Cox and Lewis, 1966; Mooers et al. 2007) (Figure 1). Trees with a disproportionate

number of nodes towards the base of the tree (i.e. more rootward from the tree midpoint) will have a low average distance to the root, and thus result in negative γ values. Alternatively, trees with more tipwardly distributed nodes produce positive γ values. For a completely sampled pure birth process γ has a standard normal distribution with a mean of zero and thus, for completely sampled phylogenies, a γ-statistic < -1.645 is considered significant evidence for rejecting a constant-rate PB process of diversification in favor of declining rates of cladogenesis through time (Pybus and Harvey, 2000). Positive values of γ can result from either increases in diversification rate through time or extinction, and are generally ignored (Pybus and Harvey, 2000).

As recognized by Pybus and Harvey, empirical phylogenies often suffer from incomplete sampling, which could lead to an apparent slow-down in cladogenesis through time (and a negative γ value). Simulations confirm that as the proportion of sampled taxa in a tree decreases, the type-I error of the CR test increases (See Fig 4 in Pybus and Harvey, 2000). In order to account for incomplete-sampling, Pybus and Harvey (2000) suggested a parametric test known as the Monte Carlo constant-rates (MCCR) test. This test constructs the null distribution of γ as follows:

1) Simulate *N* PB trees to a clade size equal to that of all known extant species in your group.
2) Randomly prune taxa from each simulated tree until the number of sampled taxa is equal to that sampled for your empirical tree.
3) Calculate the γ-statistic for each simulated and pruned tree.

4) Calculate the MCCR corrected p-value as the proportion of the N γ-values that are smaller than the test statistic calculated from the empirical tree.

When its assumptions are met, the MCCR test, by definition, has an acceptable type-I error. Furthermore, the test is robust to extinction which increases γ by pruning proportionally more ancient lineages and leaving an excess of younger branching events (the so-called "pull-of-the-present", Pybus and Harvey, 2000; See also Weir 2006; Rabosky and Lovette, 2008a). However, the CR and MCCR tests make assumptions that may be violated in particular data sets, including that: 1) the phylogeny is known without error 2) node heights accurately represent relative branching times 3) rates of cladogenesis are equal at any one time slice across all lineages (the equal-rates Markov model, ERM) and 4) taxa are sampled randomly for incompletely sampled trees.

The first three of these assumptions have been considered in the literature. Phylogenetic uncertainty may not have a strong impact on the CR test (Pybus and Harvey, 2000). Uncertainty in both topology and node heights can be accommodated by repeating analyses over a sample of credible phylogenies (such as those within 2 log-likelihood units of a maximum-likelihood estimated phylogeny or a sample from the posterior of a Bayesian analysis; see Rabosky and Lovette, 2008b, see also Revell et al. 2006 and Ruber and Zardoya 2005 for potential bias stemming from incorrect models of molecular evolution and algorithms for reconstructing chronograms). The fit of an ERM model can be assessed using a number of different topological (Kirkpatrick and Slatkin, 1993; Moore et al. 2004) or temporal-based approaches (Nee et al. 1994; Rabosky et al. 2007) and a tree can be "linearized" by pruning out the groups that show significant rate-heterogeneity.

The final assumption, random-sampling, has not been investigated despite the fact that it is likely commonly violated in phylogenetic studies. Systematists studying higher level relationships will typically sample a small number of exemplar tips to capture major radiations. In addition, cryptic speciation (Coyne and Orr, 2004; Bickford et al. 2006; Figure 2) may cause recently diverged tips to go unrecognized and unsampled. Both of these practices differentially sample older splits in the tree and should thus bias the distribution of node ages for under-sampled trees towards the root. However, the influence of incomplete, biased sampling on the performance of the MCCR test is not currently known. We used simulations to assess the type-I error under two forms of nonrandom sampling. We find that there is a dramatic increase in type-I error under both modes of nonrandom sampling. We propose two potential modifications of the standard MCCR test to alleviate this artifact, and assess their behavior. We then apply both new approaches to a recent study of homalopsid snakes (Alfaro et al. 2007).

NON-RANDOM SAMPLING AND TYPE-1 ERROR OF THE MCCR TEST

To investigate the type-I error under conditions of non-random sampling (NRS) we simulated 1000 100-taxon PB trees (birth rate=1) using the constant-rate birth-death algorithm from the GEIGER package for R (Harmon et al. 2008) and sampled 25, 50, and 75 taxa under two biased sampling methods:

1) *Oldest Splits Only (OSO)*: For a given level of sampling, *n*, we pruned all but the n-1 bottom-most nodes (including the root) of the tree. This should dramatically inflate the

type-1 error of the MCCR test as one is sampling only the most rootward nodes of the tree.

2) *Proportionally Older Splits (PSO)*: We prune the number of missing taxa, *m*, selecting taxa with probability inversely proportional their tip branch length. This method will prune a higher proportion of taxa with shorter tip branches leaving a disproportionate amount of tips that attach to the rest of the tree at relatively deeper nodes. This sampling method is also expected to elevate type-I error of the MCCR test though to a lesser extent than the OSO approach.

The MCCR test had elevated type-I error rates (up to 99%) under both POS and OSO sampling approaches for all levels of incomplete sampling investigated (Figure 3). As expected, the OSO sampling approach has a larger impact on type-I error inflation than the POS approach.

ACCOUNTING FOR NONRANDOM SAMPLING IN THE MCCR TEST: THE POS AND OSO TESTS

We developed a correction for the MCCR test designed for cases where sampling of species in a phylogenetic study might be nonrandom. To do this, we incorporated the two methods of nonrandom sampling described above into the Monte Carlo simulation of the null distribution of γ. The procedure for the corrected MCCR tests are as follows:

1) Simulate 1000 PB trees

2) For each tree, prune the *m* missing taxa non-randomly according to either the OSO or POS approach
3) Calculate γ for each simulated pruned tree
4) Calculate the MCCR corrected p-value as the proportion of the *N* γ-values that are smaller than the test statistic calculated from the empirical tree.

We investigated the statistical behavior (type-I error and power) of these methods over three different levels of sampling (25%, 50%, 75%) and assessed power for three different magnitudes of rate-decrease through time (early rate = 2, 3, or 4 times the later rate). To calculate type-I error rates for each method we simulated 1000 100-taxon trees using the same constant-rate birth-death algorithm as above. Each tree was subsequently pruned to the sampled number of species either randomly, or by the POS and OSO nonrandom sampling schemes. We calculated γ for each tree and assessed statistical significance using both POS and OSO modified MCCR tests.

To assess power under each combination of sampling level and degree of rate-shift we used a birth-death tree simulation algorithm written by CDB for R that allows the birth rate to shift at some defined point in the tree. For each combination of parameter values we simulated 1000 100-taxon trees under a specific elevated early-rate of diversification until clade-size was 50, at which point the rate was decreased to 1 until a clade-size of 100 was reached. Each tree was then pruned to the respective level of sampling using both the OSO and POS approach and γ was calculated. Each of the corrected MCCR tests were used to produce the null distribution of the test statistic and assess the significance of γ for each simulated tree.

As expected, the OSO corrected MCCR test was the more conservative of the two tests when sampling was random (data not shown). When the empirical tree was nonrandomly sampled using the POS method for 25% of taxa, only ~1% of the trees simulated under a model of two-fold decline in rate were found significant. Similarly, only ~12% and 37% of the trees for the three- and four-fold decline simulations were found significant (Figure 4C). Increased taxon-sampling significantly improved the power of the OSO test, though overall it remained conservative when rate changes were small (i.e. $\lambda_0/\lambda_1 = 2$). When the OSO sampling method was used, the OSO test had acceptable type-I error (≤ 0.05) and reasonable power, especially with more complete sampling (Figure 4D).

The POS corrected MCCR test was also conservative though it had more power than the OSO test when sampling was random (data not shown). When the empirical tree was nonrandomly sampled using the OSO method, type-I error was significantly inflated (Figure 4A), though substantially less-so than the standard MCCR test (Figure 3). When the POS sampling method was used, type-I error was acceptable (≤ 0.05)

EMPIRICAL EXAMPLE: THE HOMALOPSID SNAKES

Alfaro et al. (2007b) constructed a chronogram for 21 of the 34 extant species of the Southeast Asian colubroid family homalopsidae (Figure 6). Alfaro et al. (2007) used the standard MCCR test to evaluate the fit of a constant-rates PB process to the branching events corrected for the number of unsampled taxa (MCCR-corrected p-value=0.001). However, sampling was non-random, as the inter-generic relationships were of primary interest, and thus the significantly negative γ may be an artifact of overdispersed sampling. In an attempt to address this, we applied

both the OSO and POS corrected MCCR tests to assess the significance of the γ calculated for the homalopsid tree. The corrected p-value for the homalopsid γ under the OSO test was nonsignificant (p-value=0.186), while the POS-corrected p-value remained significant (p-value=0.0386).

CONCLUSIONS

Overdispersed non-random sampling can substantially increase type-I error of the MCCR test. Our corrected MCCR tests help to alleviate some of this bias and behave well under the non-random sampling schemes investigated. Uncertainty in the degree of overdispersion when sampling may make it difficult to assess which correction is most appropriate. While a conservative approach may be to only reject a constant-rates PB process if both POS and OSO MCCR tests are significant this may be too extreme under circumstances where sampling is known to be less-biased towards the rootward nodes than the OSO. For the homalopsid example above, a number of closely related species were included (e.g. *Cerberus "rhynchops"* and *C. microlepis*) and hence it's unlikely that sampling was similar to the OSO, and instead may fall between the OSO and POS approaches. If a researcher has some knowledge of the sampling scheme used (i.e. that at least the bottom-most *n* nodes were sampled) one can even fine-tune these approaches to first sample the bottom-most *n* nodes of the tree and then randomly or non-randomly (e.g. POS) sample the rest (code available from CDB by request). Alternatively, a researcher could investigate the robustness of a significant γ to different degrees of non-random sampling by investigating the degree, if any, of biased sampling necessary for γ to lose significance.

Any comparative method that is dependent on the distribution of nodes in a tree, including likelihood-based model fitting approaches to diversification (Rabosky, 2006; Rabosky et al. 2007; Rabosky and Lovette, 2008) will be affected by nonrandom sampling. As here, it may be necessary to consider the sampling scheme used when analyzing these data (Rabosky, 2006; Rabosky and Lovette, 2008).

LITERATURE CITED

Alfaro, M. E., F. Santini, and C. D. Brock. 2007a. Do Reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies. Evolution. 61(9): 2104-2126.

Alfaro, M. E., Karns, D. R., Voris, H. K., Brock, C. D., and B. L. Stuart 2007b. Phylogeny and evolutionary history of Oriental-Australian rear-fanged water snakes (Colubridae:Homalopsidae) inferred from mitochondrial and nuclear DNA sequences. Mol. Phylo. Evol. 46: 576-593.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. I. Ingram, and I. Das. 2006. Cryptic species as a window on diversity and conservation. Trends Ecol. Evol. 22:3.

Cox, D. R. and P. A. W. Lewis. 1966. Statistical analyses of series of events. Chapman & Hall, London.

Coyne, J. A. and H. A. Orr. 2004. Speciation. Sinauer Assoc., Sunderland, MA, USA.

Harmon, L. J., Wier, J., Brock, C., Glor, R. E., and Challenger, W. 2007. Geiger: A statistical package for investigating evolutionary radiations in a comparative context. Bioinformatics. 24:129-131.

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

Kozak, K. H., D. W. Weisrock, and A. Larson. 2006. Rapid lineage accumulation in a non-adaptive radiation: Phylognetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). Proc. R. Soc. Lond. B. 273. 539-546.

Mooers, A., L. J. Harmon, M. G. B. Blum, D. H. J. Wong, and S. B. Heard. 2007. Some models

of phylogenetic tree shape. Pages 149-170 in Reconstructing evolution: New mathematical and computational advances (O. Gascuel and M. Steel, eds.). Oxford Univ. Press, New York, NY, USA.

Moore, B. R., K. M. A. Chan, and M. J. Donoghue. 2004. Detecting diversification rate variation in supertrees. Pages 487-533 in Phylogenetic supertrees: Combining information to reveal the tree of life (O. R. P. Bininda-Edmonds ed.). Kluwer Academic, Dordrecht, The Netherlands.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B 344:305-311.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B. 267:2267-2272.

Phillimore, A. P. and T. Price. Density-dependent cladogenesis in birds. PLOS Biol. 6:e71.

Price. T. 2007. Speciation in birds. Roberts and Company, Greenwood Village, CO.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60: 1152-1164.

Rabosky, D. L., S. C. Donellan, A. L. Talaba, and I. J. Lovette. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's largest vertebrate clade. Proc. R. Soc. Lond. B. 274:2915-2923.

Rabosky, D. L. and I. J. Lovette. 2008a. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62: 1866-1875.

Revell, L. J., L. J. Harmon, and R. E. Glor. 2005. Underparameterized model of sequence evolution leads to a bias in the estimation of diversification rates from molecular phylogenies. Sys. Biol. 54:973-983.

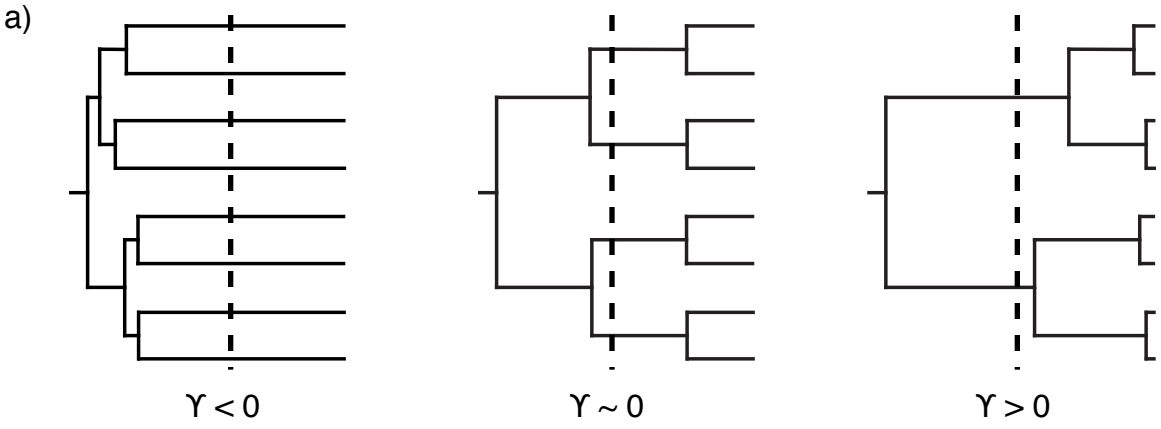
Ruber, L. and R. Zardoya. 2005. Rapid cladogenesis in marine fish revisited. Evolution 59: 1119-1127.

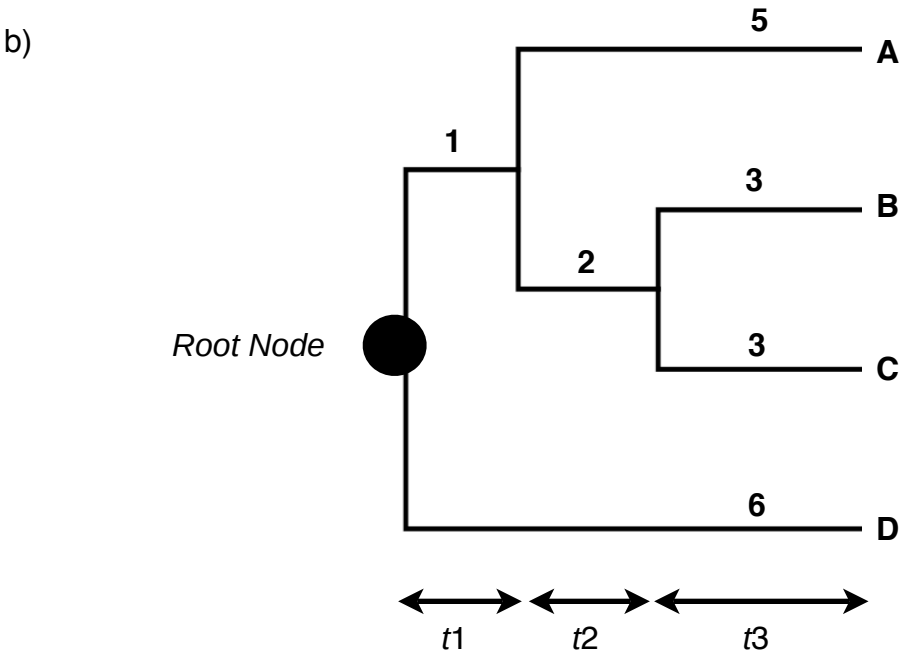
Weir, J. T. 2006. Divergent timing and patterns of species accumulation in lowland and highland neotropical birds. Evolution 60:842-855.

Weir, J. T. and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B. 271:1881-1887.

Wiens, J. J. Global patterns of species richness and diversification in amphibians. Am. Nat. 170:S86-S106.

Zink, R. M. and J. B. Slowinski. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene epoch. Proc. Natl. Acad. Sci. USA 92:5832-5835.





$$\gamma = \frac{[(6*1 + 6*2 + 4*3)/(4-2)] - [2*1 + 3*2 + 4*3)/2]}{[2*1 + 3*2 + 4*3)/2]*[\sqrt{(1/4-2)}]} =$$

Figure 1: **a)** Visual depiction of phylogenies producing a respective value of γ. The tree midpoint is represented by the dashed line. **b)** Depiction of the relevant internode intervals used for calculating γ and an example calculation of γ. Numbers above branches represent branch lengths in units of time.

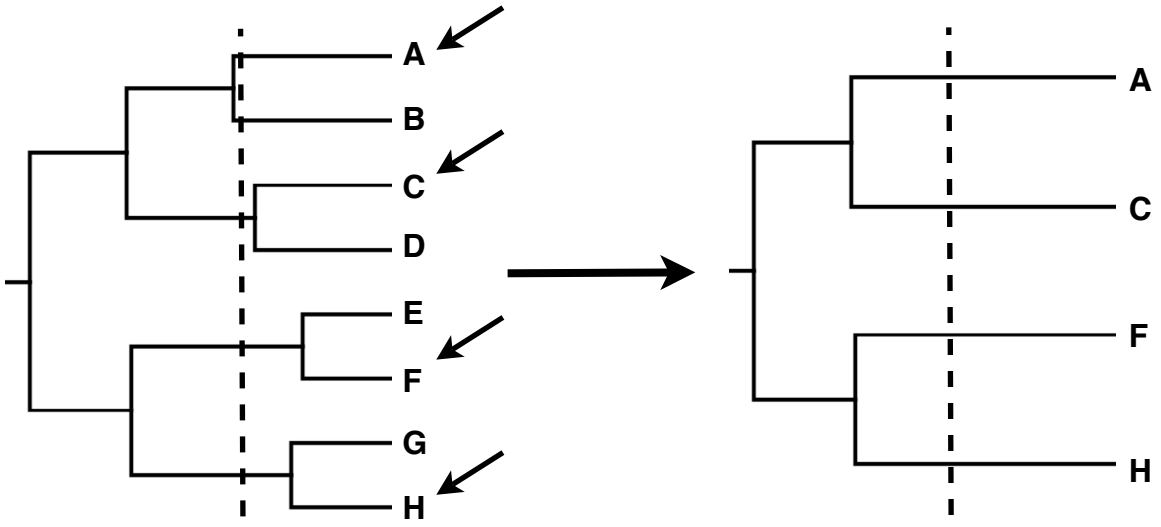


Figure 2: An example of overdispersed, non-random sampling and its effects on the relative distribution of the remaining nodes in the tree. The tree on the left is the phylogeny for all extant members of the clade. Arrows indicate taxa sampled in the reconstructed phylogeny, which is shown on the right. The dashed line represents the midpoint of the trees.

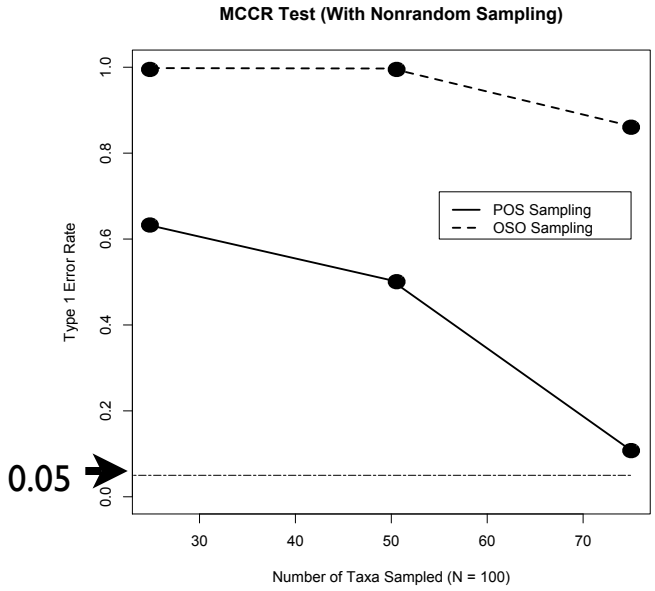


Figure 3: Type-I error of the standard MCCR test under different levels of nonrandom sampling. The solid line is the results when the tree is pruned using POS sampling. The dashed line are the results when trees were pruned using OSO sampling.

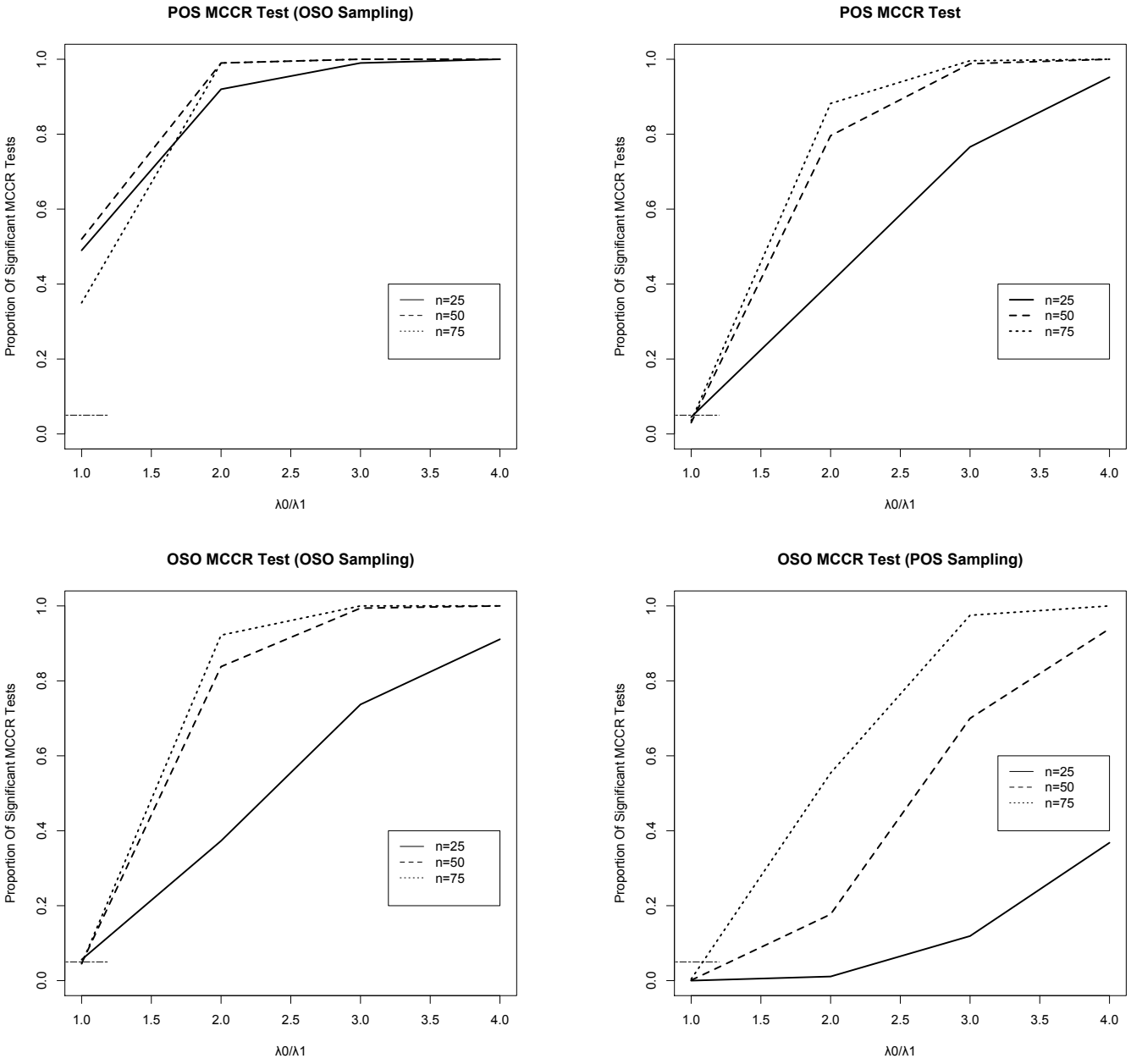


Figure 4: Plots of type-I error ($\lambda_0/\lambda_1 = 1$) and power ($\lambda_0/\lambda_1 > 1$) for both the POS (A and B) and OSO (C and D) MCCR tests. The hatched line in the lower left corner of each plot delineates 0.05 cut-off for type-I error.

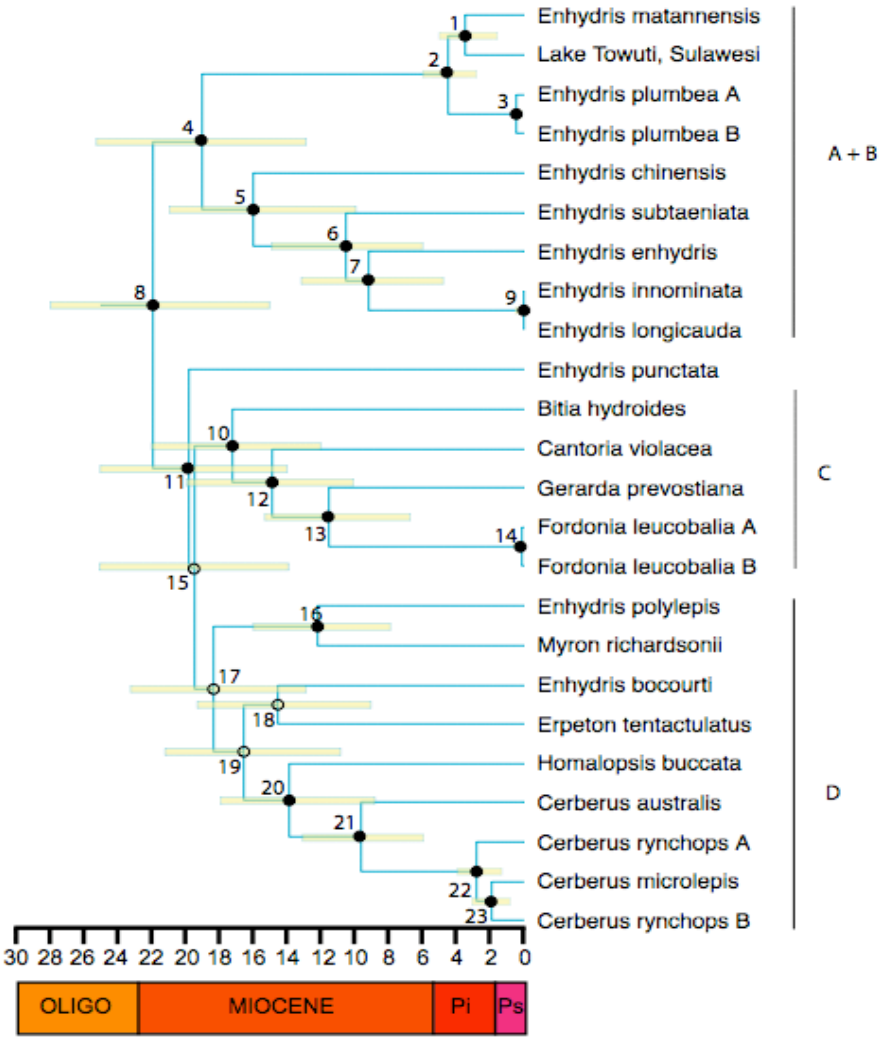


Figure 5: Chronogram of the homalopsidae. From Alfaro et al. 2007.

CHAPTER TWO: Rockfish of Ages: Temporal Patterns of Diversification in *Sebastes* Rockfishes

INTRODUCTION

Identifying the tempo and mode of cladogenesis has been an important goal for biologists and paleontologists since Simpson's pivotal contributions to the NeoDarwinian Synthesis (Simpson, 1944; 1953; Mayr and Provine, 1998). Until recently, quantification of these patterns was limited primarily to data from the paleontological record (e.g. Sepkoski 1981; Gould et al. 1977; Raup et al. 1973). However, the continued increase in available molecular data coupled with more sophisticated methods for reconstructing time-calibrated molecular phylogenies (=chronograms) has led to an upsurge in the number of studies investigating patterns of lineage diversification (Weir and Schluter, 2004; Kozak et al. 2006; Rabosky, 2006; Alfaro et al. 2007a,b; Rabosky et al. 2007, 2008; Wiens, 2007). In particular, the investigation of temporal patterns of cladogenesis have been especially prominent. The most common finding of these studies is the tendency for rates of cladogenesis to decline through time (Pybus and Harvey, 2000; Ruber and Zardoya, 2005; Weir, 2006; Alfaro et al. 2007b; Price, 2007; McPeek, 2008; Rabosky and Lovette, 2008a,b; Phillimore and Price, 2008).

This pattern is consistent with the expectations of adaptive radiations driven by ecological factors (sensu Schluter, 2000), with diversification rates peaking early-on during periods of maximal available niche space (Schluter, 2000; Phillimore and Price, 2008; Rabosky and Lovette, 2008a; Alfaro et al. In Review) and decreasing as niche space becomes filled. While this pattern of rapid early diversification has been found repeatedly, this has been predominantly through the use of summary statistics (e.g. the γ-statistic of Pybus and Harvey, 2000) that provide little insight into the actual underlying processes responsible for the empirical patterns of cladogenesis. While these statistics provide a useful heuristic for identifying overall patterns of lineage diversification for a clade of interest, an ideal approach would compare the fit of different cladogenetic processes to phylogenetic data.

In fact, recently developed model-based approaches have allowed researchers to evaluate the fit of different models of diversification to a set of branching events in a molecular chronogram (Nee et al. 1994; Rabosky, 2006; Rabosky et al. 2007; Rabosky and Lovette, 2008). These methods provide a more rigorous statistical framework for evaluating various macroevolutionary phenomena, such as key evolutionary innovations (Maddison et al. 2007) and ecological adaptive radiations (Schluter, 2000; Rabosky and Lovette, 2008). For example, conceptual models of adaptive radiation that assume ecological opportunity as a primary determinant of rates of cladogenesis predict niche-filling models of diversification (e.g. Schluter, 2000; Rabosky and Lovette, 2008). Assuming the number of niches is relatively constant, as these become filled (through the formation of new, ecologically divergent, species) rates of cladogenesis are expected to decrease, leading to diversification that is density-dependent. Such density-dependence has in fact been found in a number of empirical systems (Passeriformes and Ciconiformes: Nee et al. 1992; Birds: Phillimore and Price, 2008; *Dendroica* warblers: Rabosky and Lovette, 2008) and theoretical studies provide some evidence for an association between niche-filling models of diversification and early-bursts of cladogenesis (Walker and Valentine, 1984; Gavrilets and Vose, 2005; McPeek, 2007; 2008). As such, model-fitting approaches provide a potentially powerful means of assessing the fit of models of ecologically-driven adaptive radiation to empirical data.

Model-based approaches also allow researchers to identify clades with exceptional levels of species richness, providing a first step towards the identification of potential correlates of elevated rates of speciation/extinction (Maddison et al. 2007; Rabosky et al. 2007; Alfaro et al. In Review). For instance, a recent study evidence for a correlation between genome-duplication events and elevated rates of cladogenesis in teleost fish (Santini et al. In Review), supporting the results of previous work (e.g. Hoegg et al. 2004) and providing the first statistically rigorous assessment of the "genome-duplication" hypothesis (sensu Hoegg et al. 2004) using comparative phylogenetic approaches.

Patterns of diversification in Sebastes *rockfishes*

The diverse (~110 species) marine genus *Sebastes* is one clade for which hypotheses of both adaptive radiation (e.g. Johns and Avise, 1998) and clade-specific rates of diversification (e.g. Rocha-Olivares et al. 1999; Hyde and Vetter, 2007) have been inferred, though to date neither has been tested using the model-based framework outlined above. The majority of extant *Sebastes* species are found in the northeast pacific (~70 species) and diversification in this group has often been tied to fluctuating levels of upwelling (and thus productivity) throughout the late Miocene/early Pliocene in this region (Johns and Avise, 1998; Jacobs et al. 2004). Johns and Avise (1998) applied the method of Wollenberg et al (1996) to a molecular phylogeny of 26 *Sebastes* species. This analysis provided evidence for an early-burst of cladogenesis in *Sebastes*, and the inferred time of this burst (~ 5 MYA) was argued to correlate well with late Miocene peaks in upwelling. The authors argued that *Sebastes* was an example of an "ancient species flock" and other researchers have largely followed suit (e.g. Alesandrini and Bernardi, 1999; Burford and Bernardi, 2008). However, as pointed out by Pybus and Harvey (2000), the null employed in the method of Wollenberg et al. (1996) is inappropriate for rejecting a constant-rates process, since it employs model in which speciation = extinction as the null. Rejection of the Wollenberg et al. null only implies that speciation and extinction rates are not equivalent, while both may still be constant throughout the clades history. A more appropriate null would be a constant-rates PB process, as employed by Pybus and Harvey (2000) in their γ-statistic. This latter approach was used by Ruber and Zardoya (2005) to investigate patterns of cladogenesis in *Sebastes* (among other marine fish clades) and provided strong support for a decrease in the net rate of cladogenesis through time, consistent with the hypothesis of Johns and Avise (1998).

Heterogeneity in the rate of cladogenesis between *Sebastes* subclades has also been inferred, and potential explanations for this variation have included habitat, life history, and biogeography (e.g. Hyde and Vetter, 2007). However, statistical assessment of the fit of a rate-heterogeneous model has yet to be accomplished, with previous discussions being based on

either poor nodal support at the deeper portions of subclade trees (e.g. *Sebastomus*: Rocha-Olivares et al. 1999) or long waiting times between sister-species (*Sebastodes*: Hyde and Vetter, 2007). Rigorous evaluation of rates of diversification across subclades would provide a initial step towards understanding the potential correlates of species richness in *Sebastes*.

With the availability of a near-complete species-level phylogeny for the genus *Sebastes* (Hyde and Vetter, 2007), we wanted to readdress the temporal patterns of diversification within the genus using recently developed model-based phylogenetic comparative methods (e.g. Rabosky and Lovette 2008; Alfaro et al. In Review). Specifically, we addressed four general questions: 1) Is the net rate of diversification elevated in the early history of *Sebastes*? 2) Is diversification in *Sebastes* density-dependent, consistent with a model of adaptive radiation? 3) Are rates of diversification within *Sebastes* heterogeneous? 4) Is the period of peak upwelling during the late Miocene a time of exceptional diversification in *Sebastes*? We find strong evidence for an early burst of cladogenesis in *Sebastes* that is most consistent with a model of density-dependent diversification. Furthermore we identify four outlier clades for which rate-shifts are identified (three increases and a single decrease). Lastly, we find the late Miocene was a period of exceptional diversification in *Sebastes*, consistent with the hypothesis that diversification in this group has been tied (at least partly) to local levels of marine productivity.

METHODS

Sequence alignment and chronogram reconstruction

Nucleotide data for seven genes (two nuclear and five mitochondrial) were downloaded from GenBANK for 103 of the 110 extant species (See Hyde and Vetter 2007 for details). Alignments were done using the Clustal X module in Mesquite (Maddison, and Maddison, 2008) and adjusted by eye. Structural gene alignments (e.g. 12S rRNA) were further refined by

referencing secondary-structure models for fish and ambiguously aligned regions were removed before analyses (Orti and Meyer, 1997). Alignments of protein-coding loci were checked by translating the sequences and confirming the alignment of the corresponding amino acids. The seven genes were concatenated into a single dataset of 5,541 nucleotides.

In order to identify an appropriate partitioning scheme, the program MrBayes (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003) was used to calculate the harmonic mean of the marginal likelihood under five partitioning strategies: 1) A single partition. 2) Partitioning into coding and non-coding regions. 3) Partitioning by gene. 4) Partitioning by gene with stems and loops partitioned separately. 5) Partitioning by gene, as well as stems, loops, and codon position within protein-coding loci. The resulting harmonic means were inputed into the program Tracer (Rambaut and Drummond 2007) and BayesFactors were calculated. This identified the most complex partitioning strategy (#5, 14 partitions) as the best fit to the data and this scheme was used for the all analyses.

The bayesian relaxed-clock method BEAST v1.4.8 (Drummond and Rambaut 2007) was used to reconstruct a timetree using the aligned dataset. To help alleviate the effects of a truncated prior distribution (Rannala, 2007; Yang and Rannala 2006) we assigned exponential priors to constrain the node ages of two clades where the 95% upper limit on the prior was our best estimate for the maximum age of the clade. Two nodes were assigned priors based on the following information: 1) *Sebastes* fossils from late Miocene diatomite deposits of Lompoc, California (Barsukov, 1989) were used to assign a minimum and maximum age for the crown-*Sebastes* clade of 5.3 MYA and 23.7 MYA, respectively. 2) The MRCA of *Sebastes alutus* and *S. norvegicus* was assigned a minimum and maximum age of 1.8 MYA and 5.3 MYA, respectively (sensu Rocha-Olivares et al. 1999; Hyde and Vetter, 2007).

We ran BEAST for 150 Million generations sampling every 1000 generations. Convergence was assessed visually using the program TRACER to plot likelihood vs. generation as well as estimate effective sample size (ESS) of all parameters. Initial problems with convergence were seen using the 14-partition scheme with the GTR + Γ + I model of evolution. Inspection of the transition rates showed a number of them had values approaching zero, hinting at overparameterization (A. Rambaut, personal communication). Two modifications were developed to address this. First, we changed the model of evolution from GTR + Γ + I to HKY + Γ + I for all partitions and reran the BEAST analysis. Second, we also reran the BEAST analysis using the GTR + Γ + I but reducing our number of partitions to seven (by gene and by stem/loop). Results were qualitatively similar using both approaches and thus we present only the results using the first correction (14-partition scheme with HKY + Γ + I).

Temporal patterns of diversification in Sebastes

The packages GEIGER (Harmon et al. 2007) and LASER (Rabosky 2006) in the computing language R, in addition to code written by CDB were used to carry out all diversification analyses. A lineage-through-time (LTT) plot for the *Sebastes* chronogram was plotted as the log(number of lineages) extant at a given time working from the root of the tree toward the present. The maximum-clade-credibility (MCC) tree from the BEAST analysis was used to produce the LTT-plot, and 95% quantiles were estimated for the log-number of lineages through time based on the tree posterior distribution of the BEAST analysis. As a first assessment of rate-constancy, the γ-statistic of Pybus and Harvey (2000) was used to statistically assess the distribution of nodes in the chronogram. The γ-statistic for a completely sampled pure birth

process has a standard normal distribution and a significantly negative γ is indicative of a slowdown in the net rate of cladogenesis toward the present. However, as sampling for *Sebastes* is incomplete, the Monte Carlo Constant-Rates (MCCR) test was used to evaluate the significance of the empirical γ (Pybus and Harvey, 2000). Briefly, 1000 110-taxon trees were simulated under a constant-rate birth-death process using the estimated speciation and extinction rates (see below). These trees were then randomly pruned of the 7 missing taxa and the γ-statistic was calculated for each. The p-value was calculated as the proportion of the simulated trees with a γ value ≤ the empirical γ. Corrections for overdispersed sampling (Brock et al. In Preparation) were also employed and produced qualitatively similar results (data not shown). In order to account for uncertainty in both topology and branch-length estimates we calculated γ for each tree from the posterior distribution of our BEAST analysis (Figure 6, Inset (b)).

In addition to γ, we also employed recently developed model-based approaches to investigate temporal patterns of diversification. Specifically we were interested in assessing whether patterns of diversification within the *Sebastes* clade fit a model of density-dependence, consistent with the prediction of *Sebastes* as an adaptively radiating clade. Four models of diversification were fit using maximum likelihood to both the MCC tree and 1000 random trees drawn from the posterior distribution of the BEAST analysis: a constant-rates pure birth model; logistic and exponential density-dependent models; as well as a linear decline model in which net diversification rate decreases through time at a rate that is independent of the number of extant species. Akaike Information Criterion (AIC) were used to compare models (Akaike, 1973; Burnham and Anderson, 2002), where the model with the lowest AIC score is the one that best-fits the data of the models compared. ΔAIC (difference between the AIC of the focal model and

the model with the lowest AIC score) and Akaike weights were calculated for each model (Burnham and Anderson, 2002).

Likelihood-based approaches were also used to evaluate the hypothesis of rate-heterogeneity for subclades within the genus *Sebastes*. To identify clades with aberrant levels of species richness we used MEDUSA, a recently developed comparative method that combines phylogenetic and taxonomic information to estimate rate shifts on a chronogram (Alfaro et al. In Review). Missing taxa were assigned to tip clades based on the taxonomy of Hyde and Vetter (2007). These clades were then pruned down to a single extant representative to which a species richness value was assigned. A compound phylogenetic and taxonomic likelihood was then optimized sequentially using a stepwise AIC approach. Briefly, the likelihood in the first iteration is maximized assuming the tree has a single speciation and extinction rate. Next, the AIC score for this constant-rate model is compared to a model where both speciation and extinction were allowed to shift on the optimum branch (i.e. the one with the highest likelihood). This process is repeated until the addition of further rate-shifts results in improvements of < 4 AIC units. This optimum model was also used to estimate the overall speciation and extinction rates. These estimates were employed for the parametric simulations of γ (see above).

Finally, to evaluate whether the late Miocene period of upwelling and high productivity was a period of exceptional diversification for *Sebastes* we developed a novel parametric method, the Kendall-Moran Rate Interval (KMRI) test. The test calculates the Kendall-Moran estimate of speciation-rate (See Nee 2001) over the interval in question and then compares this estimate to a simulated null distribution. For the null, the overall speciation and extinction-estimates are used to simulate 1000 trees with *n* number of taxa, where *n* is the number of extant species. Any missing taxa are then pruned from these trees and the KM rate estimate is

calculated over the same interval as the empirical tree. The p-value is then calculated as the proportion of the simulated trees with a rate estimate ≤ the empirical estimate. This method was used to investigate whether the period of high productivity between the origin of the *Sebastes* crown-group (8.25 MYA) and the end of the Miocene (5.3 MYA) was a period of exceptional diversification as inferred by Jacobs et al. (2004). We also evaluate the fit of a two-rate model in which the late Miocene and post-Miocene periods are allowed to have there own rates of diversification to evaluate whether this model is an improved fit over the four previous models.

RESULTS

Sebastes *chronogram and the time-frame of* Sebastes *diversification*

Results from the BEAST analysis are largely concordant with previous work (e.g. Kai et al. 2003; Hyde and Vetter, 2007). Specifically, we find a similar age for the origin of crown-group *Sebastes* (8.5 MYA), and this falls within the late Miocene timeframe of high upwelling, as predicted by Jacobs et al. (2004). There are minor topological differences at the base of the tree, though these primarily involve poorly supported nodes. These discrepancies should have a negligible affect on the results of our diversification analyses, as topology has little effect on the distribution of watiting-times across the phylogeny (e.g. Pybus and Harvey, 2000). However, as noted above, we accounted for phylogenetic uncertainty (i.e. uncertainty in both topology and branch-lengths) by repeating analyses over the posterior distribution from our BEAST analysis.

Temporal patterns of diversification in Sebastes

The LTT plot for the BEAST MCC tree shows a distinct slowdown in the rate of lineage accumulation through time (Figure 6, Inset (a)). The null expectation under a constant-rate pure birth process falls well below the 95% quantiles from the posterior distribution for the entire history of the clade, consistent with an early burst of cladogenesis in crown *Sebastes*. Figure 6, Inset (b) shows the distribution of γ from the posterior of the BEAST analysis (Black) plotted with the null distribution of γ derived through simulation (Gray). The consistently negative value of γ across the posterior, the minimal overlap (<0.05) between the empirical and null distribution, and the highly significant value for the MCC tree (γ = -3.05, MCCR corrected p-value < 0.002) are all indicative of a strong burst of cladogenesis early in the history of this group. Furthermore, this early burst is temporally concordant with the period of intense upwelling during the late Miocene (see below).

Our model-fitting results provide weak support for density-dependent diversification in *Sebastes* (Table 1). While 75% of the Akaike weight is on the density-dependent models, the linear decline model falls within the 95% credible interval (ΔAIC = 1.01; Akaike weight = 0.249) and thus cannot be excluded. The pure birth model, however, falls well outside of the 95% CI of models and has a minuscule fraction of the Akaike weight (= 0.001). These results provide further support for elevated early rates of cladogenesis and provide some weak signal of DD diversification, consistent with niche-filling models of diversification (Schluter 2000; Rabosky and Lovette 2008; McPeek 2008). Figure 8 shows the posterior distribution of ΔAIC of the best fitting DD model (the exponential) and the linear-decline model. Again, we see the exponential DD model consistently fitting better the the LD model (μ = 2.378, 95% CI = [0.76, 4.82]).

However, Rabosky and Lovette (2008) report a tendency for DD models to overfit with respect to the LD model, and our distribution overlaps somewhat with that seen in their Figure 4. Given this, and the fact that the ΔAIC between the exponential DD model and LD model is <2.0 (sensu Burnham and Anderson, 2002) for the MCC tree, we cannot reject with confidence the latter model in favor of the former and thus have only weak evidence for DD in *Sebastes*.

Our MEDUSA results provide evidence for four outlier clades (Figure 8): three rate decreases and a single rate increase. Some of these rate-shifts were at least partly consistent with previously (untested) hypotheses (e.g. decrease in epibenthic *Sebastodes* clade: Hyde and Vetter 2007; Increase in some *Sebastomus* subclades: Hyde and Vetter, 2007; Rochas-Olivares et al. 1999) while others were entirely novel (decrease in Kiyomatsui-Scythropus and MacDonaldi-Ruberrimus clades). Furthermore, previous hypotheses of significant rate heterogeneity between the major subclades of *Sebastes* (i.e. *Sebastomus, Pteropodus, Sebastocles*) were not supported by our analyses.

Estimates for the clade-level average rate of speciation and extinction were 0.39 and 0.0004, respectively (Figure 9). This speciation rate estimate approaches that estimated for other inferred "adaptively radiating" clades, including Hawaiian silverswords (0.56; Baldwin and Sanderson, 1998) and some Neogene horses (lower estimates = 0.50; Hulbert, 1993) though it is substantially lower than others, including some fish groups (upper estimate = 1.49 for Lake Tanganyika cichlids; McCune, 1997). A broader database of rate estimates across groups is needed before concrete conclusions can be drawn as to which clades are, overall, showing exceptional rates of diversification.

Lastly, results from the KMRI tests provide strong support (p-value < 0.01) for the late Miocene period of intense upwelling as a time of exceptional diversification in *Sebastes* (Figure

10). This is consistent with results from previous studies (e.g. Johns and Avise, 1998) while the methods we employed corrected shortcomings of these previous approaches by using a more appropriate null expectation (constant-rates BD model vs. constant-rates B = D model; See above). When the test is repeated over the Pliocene or Plio-Pleistocene we found no evidence for elevated rates of cladogenesis (p-value = 0.613, and 0.928, respectively). Our two-rate PB model estimated separate rates for the period before and after the Mio-Pliocene boundary, and led to a slight improvement in the AIC score over the DD exponential model (AIC = -313.004 vs. -311.395). The estimated speciation rate across the late Miocene time interval was >2 times that for the subsequent Plio-Pleistocene period (0.697 vs. 0.311), and the later rate is only slightly lower than the average rate for *Sebastes* overall (0.39).

DISCUSSION

We present the most comprehensive investigation of temporal patterns of diversification to date for a clade of marine fish. We find further support for an early period of rapid cladogenesis and provide the first model-based evidence, albeit weak, for density-dependent diversification in *Sebastes,* consistent with the idea of rockfish as an adaptively radiating clade (Johns and Avise, 1998; Ruber and Zardoya, 2005). Furthermore we provide the first comparative phylogenetic evidence for clade-specific rates of diversification within the genus. Lastly, our KMRI test and two-rate model-fitting identify the late Miocene as a period of exceptional diversification, lending support to species diversification in *Sebastes* being linked with levels of nearshore marine productivity.

Productivity and patterns of temporal diversification in Sebastes

The early-burst pattern of cladogenesis in *Sebastes* is consistent with the hypothesis that the clade radiated adaptively during the late Miocene. This radiation could be due to a number of potential factors, including: an increase in ecological opportunity due to colonization of a relatively uninhabited environment (Schluter, 2000), the acquisition of a key innovation that allowed the exploitation of a novel resource base (Simpson, 1953), or the increase in available resources due to various abiotic events (e.g. periods of peak upwelling and increases in nearshore marine productivity; Jacobs et al. 2004). While the former two explanations remain plausible, the latter has been inferred as a driver of diversification in a number of marine groups, including *Sebastes* (Jacobs et al. 2004), and our KMRI test and two-rate model-fitting results provide strong evidence for the late Miocene period of peak upwelling as a time of exceptional diversification in *Sebastes* rockfishes. This suggests that diversification in *Sebastes* may have proceeded at a relatively unfettered pace in the late Miocene, until levels of nearshore marine productivity decreased during the Pliocene.

Previous authors have identified productivity as a primary factor driving the diversification of a number of marine clades (Vermeij, 1989; Leigh and Vermeij, 2002; Reaka et al. 2008), though the exact relationship between productivity and species diversity, as well as its generality across different spatial scales (and taxa), remains contentious (Rosenzweig, 1995; Rosenzweig and Sandlin, 1997; Mittelbach et al. 2001; Chase and Leibold, 2002; Whittaker and Heegaard, 2003; Mittelbach et al. 2003). Productivity could potentially impact rates of lineage accumulation in a number of ways. Energy, a positive correlate of productivity in terrestrial systems (e.g. Storch et al. 2005), has been inferred to translate into more trophic levels and

increased abundance of rare resources (Evans et al. 2005). If productivity is the primary factor driving these effects, than one may expect similar impacts of increased productivity in marine systems. Thus high productivity may lead to the development of a diverse resource base (e.g. increased ecological opportunity) that could facilitate diversification through ecologically driven diversification (Schluter, 2000; Vose and Gavrilets, 2005; McPeek, 2007; 2008). High energy/ productivity has also been inferred to increase a population's niche width (e.g. Evans et al. 2005) and recent work suggests that this niche expansion may occur through within population diversification through individual specialization (e.g. Bolnick et al. 2007). Theoretical models suggest this could facilitate lineage-splitting through disruptive selection (Dieckmann and Doebeli, 1999; Bolnick, 2006; Gavrilets et al. 2007; But see Bolnick and Doebeli, 2003; Coyne and Orr, 2004; Kirkpatrick and Nuismer, 2004) though well-supported empirical examples are largely lacking (But see Barluenga et al. 2006; Gavrilets et al. 2007). Furthermore, high productivity may sustain larger population sizes, and this could potentially lead to lower rates of extinction and thus higher net diversification (Stanley, 1979). However, correlated decreases in speciation rate are also predicted under some verbal models of diversification when population sizes increase, potentially offsetting decreases in extinction rates and diminishing changes in the net rate of lineage accumulation through time (Stanley, 1979; Stanley, 1986). While we find high productivity seems to facilitate lineage diversification in *Sebastes*, the exact mechanisms responsible for this effect are unknown. An increased understanding of the community-level effects of productivity should help guide expectations as to the probable mechanisms by which productivity may act to alter rates of diversification in marine clades.

Clade-specific rates of diversification in Sebastes

Our MEDUSA results provide evidence for clade-specific rates of diversification in four *Sebastes* subclades. The epibenthic *Sebastodes* clade showed a rate-decrease, in support of the hypothesis of Hyde and Vetter (2007) that this group was diversifying more slowly than other *Sebastes* clades. However, while these authors highlight the general habitat preferences for this clade, they do not provide a mechanistic explanation as to why inhabiting epibenthic environments would hinder rates of lineage diversification nor do they find similar patterns in other epibenthic clades. Two other clades, the Kiyomatsui-Scythropus and MacDonaldi-Ruberrimus clades also show rate-decreases, though to our knowledge these are novel results that have not been hypothesized previously.

A single rate-increase was found in the Cortezi-group, a clade of four species within the subclade *Allosebastes* that are found in the Pacific Ocean along Baja California, and in the Gulf of California. Hyde and Vetter (2007) noted that this clade showed rapid lineage accumulation over the last 0.5-1 MY and our results provide statistical support for this. The BEAST analysis provides weak support for monophyly in the Gulf of California clade, contrary to Hyde and Vetter (2007), which could imply that there was a single colonization event of the Gulf of California with *S. cortezi* and *S. peduncularis* subsequently diverging. However, the low support for the Cortezi-Peduncularis node lends equal weight to the possibility of multiple colonization events leading to speciation through sequential allopatric divergence.

A recent review of 12 species with disjunct marine populations across the Baja California Peninsula found evidence for both vicariance (8 species) and continuous dispersal and panmixia (4 species), providing support for the Peninsula as an effective barrier to gene flow for some

species (e.g. *Hypsoblennius jenkinsi*), but not others (e.g. *Sebastes macdonaldi*). The authors found that dispersal potential was a poor predictor of population divergence across the Peninsula, but that ecological characteristics (intertidal habitat spawners, e.g. *Leuresthes* and *Girella* spp.) were positively correlated with levels of population subdivision. Given the ecology and life-history characteristics of *S. cortezi* and *S. peduncularis* are poorly known (Love et al. 2002), comparison to those of *S. macdonaldi* prevent us from making reasonable inferences as to the mode by which species formation may have occurred across the Baja Peninsula. More detailed information on the ecology of the Gulf of California *Sebastes* as well as the implementation of recently developed likelihood-based approaches for biogeographic inference (Ree and Smith, 2008) may help to uncover the processes responsible for species-formation in this geographic isolate of *Sebastes*.

While we found statistical support for rate heterogeneity, potential correlates of these rate-shifts have yet to be statistically evaluated. Future work should employ recently developed phylogenetic comparative methods to investigate whether ecological, life-history, or biogeographic factors have promoted or hindered diversification in *Sebastes*.

CONCLUSIONS

Using a near-complete species-level phylogeny (Hyde and Vetter, 2007) we provide strong evidence for an early-burst pattern of cladogenesis in the *Sebastes* rockfish. Furthermore, we provide the first evidence, albeit weak, for density-dependent diversification in a clade of marine fishes. This result is consistent with crown-group *Sebastes* being an ancient adaptive radiation as inferred previously (Johns and Avise, 1998) though the inability to exclude a

density-independent model of rate-decrease (and the slight improvement in AIC in our two-rate model) prevents us from making too strong of conclusions as to the nature of *Sebastes* diversification. We also find evidence for cladogenetic-rate heterogeneity in *Sebastes*, with evidence for three rate decreases and a single rate increase with respect to the background rate of speciation and extinction. These results provide evidence for potential correlates of lineage diversification (e.g. epibenthic habitat; Hyde and Vetter, 2007) though a statistical association between species traits and rates of cladogenesis is lacking. Lastly, we provide the first rigorous statistical assessment and support for the late Miocene upwelling events as a period of exceptional diversification in a *Sebastes* rockfishes. These results provide further evidence and insight into the temporal patterns of diversification in *Sebastes* rockfishes and help bolster support for the impact of productivity on patterns of species richness in marine systems.

LITERATURE CITED

Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle, in B. N. Petrov and F. Csaki (eds), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary.

Alesandrini. S, and G. Bernardi. 1999. Ancient species flocks and recent speciation events: what can rockfish teach us about cichlids (and vice-versa)? J. Mol. Evol. 49: 814-818.

Alfaro, M. E., F. Santini, and C. D. Brock. 2007a. Do Reefs drive diversification in marine teleosts? Evidence from the pufferfishes and their allies. Evolution. 61(9): 2104-2126.

Alfaro, M. E., Karns, D. R., Voris, H. K., Brock, C. D., and B. L. Stuart 2007b. Phylogeny and evolutionary history of Oriental-Australian rear-fanged water snakes (Colubridae:Homalopsidae) inferred from mitochondrial and nuclear DNA sequences. Mol. Phylo. Evol. 46: 576-593.

Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, and L. J. Harmon. Eleven exceptional radiations plus high turnover explain species diversity in jawed vertebrates. In Review

Baldwin, B. G., and M. J. Sanderson. 1998. Age and rate of diversification of the hawaiian silversword alliance (Compositae). PNAS 95: 9402-9406.

Barluenga, M., K. N. Stolting, W. Salzburger, M. Muschick, and A. Meyer. Sympatric speciation in Nicaraguan crater lake cichlid fish. Nature 439: 719-723.

Barsukov, V. V. 1989. The upper Miocene rockfishes (Scorpaenidae, Sebastinae) from California. Proc. Zool. Inst. Len. 201: 73-109

Bernardi, G., L. Findley, and A. Rocha-Olivares. 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. 57: 1599-1609.

Bolnick, D. I. 2006. A multi-species outcome of a common model of sympatric speciation. J. Theor. Biol. 241: 734-744.

Bolnick, D. I., R. Svanback, M. Araujo, and L. Persson. 2007. More generalized populations are also more heterogeneous: comparative support for the niche variation hypothesis. PNAS 104: 10075-10079.

Burford, M. O. and G. Bernardi. 2008. Incipient speciation within a subgenus of rockfish (Sebastostomus) provides evidence of recent radiations within an ancient species flock. Mar. Biol. 154: 701-717.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information theoretic approach 2nd edition. New York:Springer-Verlag.

Chase, J. M. and M. A. Leibold. 2002. Spatial scale dictates the productivity-biodiversity relationship. Nature 416: 427-430.

Dieckmann, U. and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400: 354-357.

Drummond AJ, Rambaut A (2007) "BEAST: Bayesian evolutionary analysis by sampling trees." *BMC Evolutionary Biology* **7**:214.

Evans, K. L., P. H. Warren, and K. J. Gaston. 2005. Species-energy relationships at the macroecological scale: a review of the mechanisms. Biol. Rev. 80: 1-25.

Harmon, L. J., Wier, J., Brock, C., Glor, R. E., and Challenger, W. 2007. Geiger: A statistical package for investigating evolutionary radiations in a comparative context. Bioinformatics. 24:129-131.

Hoegg, S., H. Brinkmann, J. S. Taylor, and A. Meyer. 2004. Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J. Mol. Evol. 59(2).

Hueslsenbeck, J. P. and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogeny. Bioinformatics 17:754-755.

Hulbert, R. C. The rise and fall of an adaptive radiation. Paleobiology 19: 216-234.

Gavrilets, S. and A. Vose. Dynamic patterns of adaptive radiation. PNAS 102: 18040-18045.

Gavrilets, S., A. Vose, M. Barluenga, W. Salzburger, and A. Meyer. 2007. Case studies and mathematical models of ecological speciation. 1. Cichlids in a crater lake. Mol. Ecol. 16: 2893–2909

Gould, S. J., D. M. Raup, J. J. Jr. Sepkoski, T. J. M. Schopf, and D. S. Simberloff. The shape of evolution: a comparison of real and random clades. Paleobiology 3: 23-40.

Hyde, J. R. and R. D. Vetter. 2007. The origin, evolution, and diversification of rockfishes of the genus *Sebastes* (Cuvier). Mol. Phy. Evol. 44: 790-811.

Jacobs, D. K., T. A. Haney, and K. D. Louie. 2004. Genes, diversity, and geologic process on the Pacific coast. Ann. Rev. Earth. Planet. Sci. 32: 601-652.

Johns, G. C. and J. C. Avise. 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of Sebastes rockfishes and other marines fishes. Evolution. 52(4): 1135-1146.

Kai, Y., K. Nakayama, and T. Nakabo. 2003. Molecular phylogenetic perspective on speciation in the genus *Sebastes* (Scorpaenidae) from the Northwest Pacific and the position of *Sebastes* within the subfamily Sebastinae. Ichthyol. Res. 50: 239-244.

Kozak, K. H., D. W. Weisrock, and A. Larson. 2006. Rapid lineage accumulation in a non-

adaptive radiation: Phylognetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). Proc. R. Soc. Lond. B. 273. 539-546.

Leigh, E. G. and G. J. Vermeij. Does natural selection organize ecosystems for the maintenance of high productivity and diversity? Philos. Trans. R. Soc. London Ser. B 357(1421): 709-718.

Love, M. S., M. Yoklavich, and L. Thorsteinson. 2002. The rockfishes of the northeast Pacific. Univ. of California Press, Los Angeles, CA, USA.

Maddison, W. P. and D. R. Maddison. 2008. Mesquite: a modular system for evolutionary analysis. Version 2.5 http://mesquiteproject.org.

Mayr, E. and W. B. Provine. 1998. The evolutionary prospective: perspectives on the unification of biology. Harvard Univ. Press, Cambridge, MA, USA.

McCune, A. R. 1997. How fast is speciation? Molecular, geological, and phylogenetic evidence from adaptive radiation of fishes. In Molecular evolution and adaptive radiation (ed. T. J. Givnish & K. J. Sytsma), pp. 585-610. Cambridge. Univ. Press, New York, NY, USA.

McPeek, M. A. 2007. The macroevolutionary consequences of of ecological differences among species. Palaeontology 50: 111-129.

McPeek, M. A. 2008. The ecological dynamics of clade diversification and community assembly. Am. Nat. 172:6 E270-E284.

Mittlebach. G. G., C. F. Steiner, S. M. Scheiner, K. L. Gross, H. L. Reynolds, R. B. Waide, M. R. Willig, S. I. Dodson, and L. Gough. 2001. What is the observed relationship between species richness and productivity? Ecology, 82: 2381-2396.

Mittelbach, G. G., S. M. Scheiner, and C. F. Steiner. 2003. What is the observed relationship between species richness and productivity? Reply. Ecology, 84(12): 3390-3395.

Nee, S. 2001. Inferring speciation rates from phylogenies. Evolution. 55: 661-668.

Nee, S., A. O. Mooers, and P. H. Harvey. 1992. Tempo and mode of evolution revealed from molecular phylogenies. PNAS. 89(17): 8322-8326.

Nee, S., R. M. May, and P. H. Harvey. 1994. The reconstructed evolutionary process. Philos. Trans. R. Soc. Lond. B 344:305-311.

Orti, G. and A. Meyer. 1997. The radiation of characiform fishes and the limits of resolution of mitochondrial ribosomal DNA sequences. Sys. Biol. 46: 75-100.

Pybus, O. G. and P. H. Harvey. 2000. Testing macro-evolutionary models using incomplete molecular phylogenies. Proc. R. Soc. Lond. B. 267:2267-2272.

Phillimore, A. P. and T. Price. Density-dependent cladogenesis in birds. PLOS Biol. 6:e71.

Rabosky, D. L. 2006. Likelihood methods for detecting temporal shifts in diversification rates. Evolution 60: 1152-1164.

Rabosky, D. L., S. C. Donellan, A. L. Talaba, and I. J. Lovette. 2007. Exceptional among-lineage variation in diversification rates during the radiation of Australia's largest vertebrate clade. Proc. R. Soc. Lond. B. 274:2915-2923.

Rabosky, D. L. and I. J. Lovette. 2008a. Explosive evolutionary radiations: decreasing speciation or increasing extinction through time? Evolution 62: 1866-1875.

Rabosky, D. L. and I. J. Lovette. 2008b. Density-dependent diversification in North American wood warblers. Proc. R. Soc. B. 275: 2363-2371.

Rambaut, A. and A. J. Drummond. 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer

Rannala, B. 2007. Inferring speciation times under an episodic molecular clock. Sys. Biol. 56: 453-466.

Raup, D. M., S. J. Gould, T. J. M. Schopf, and D. S. Simberloff. 1973. Stochastic models of phylogeny and the evolution of diversity. J. Geol. 81: 525-542.

Reaka, M. L., P. J. Rodgers, and A. U. Kudla. 2008. Patterns of biodiversity and endemism on Indo-West Pacific coral reefs. PNAS 105: 11474-11481.

Ree, R. H. and S. A. Smith. 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Sys. Biol. 57(1):4-414.

Rocha-Olivares, A., C. A. Kimbrell, B. J. Eitner, and R. D. Vetter, 1999. Evolution of mitochondrial cytochrome b gene sequence in the species-rich genus *Sebastes* (Teleostei, Scorpaenidae). Can. J. Fish. Aquat. Sci. 56: 803-813.

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19.

Rosenzweig, M. L. 1995. Species diversity in space and time. Cambridge Univ. Press, Cambriidge.

Rosenzweig, M. L. and Sandlin. 1997. Species diversity and latitudes: listening to area's signal. Oikos, 80: 172-176.

Ruber, L. and R. Zardoya. 2005. Rapid cladogenesis in marine fish revisited. Evolution 59: 1119-1127.

Santini, F., L. J. Harmon, G. Carnevale, and M. E. Alfaro. First evidence for a ling between the fish-specific genome duplication and the origin of teleosts. In Review.

Schluter, D. 2000. The ecology of adaptive radiations. Oxford Univ. Press, USA.

Sepkoski, J. J. Jr. 1981. A factor analytic description of the Phanerozoic marine fossil record. Paleobiology 7: 36-53.

Simpson, G. G. 1944. Tempo and mode of evolution. Columbia Univ. Press, New York, NY, USA.

Simpson, G. G. 1953. The major features of evolution. Columbia Univ. Press, New York, NY, USA.

Stanley, S. M. 1979. Macroevolution: pattern and process. Freeman, San Francisco, CA, USA.

Stanley, S. M. 1986. Population size, extinction, and speciation: the fission effect in Neogene bivalvia. Paleobiology 12(1): 89-110.

Storch, D., K. L. Evans, K. J. Gaston. 2005. The species-area-energy relationship. Ecology Letters, 8(5): 487-492.

Vermeij, G. J. 1989. Interoceanic differences in adaptation: effects of history and productivity. Mar. Ecol. Progr. Ser. 57: 293-305.

Walker, T. D. and J. W. Valentine. 1984. Equilibrium models of evolutionary species diversity and the number of empty niches. Am. Nat. 124: 887-899.

Weir, J. T. and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. Proc. R. Soc. Lond. B. 271:1881-1887.

Whittaker, R. J. and E. Heegaard. 2003. What is the observed relationship between species richness and productivity? Comment. Ecology, 84(2): 3384-3390.

Wiens, J. J. Global patterns of species richness and diversification in amphibians. Am. Nat.

170:S86-S106.

Wollenberg, K., J. Arnold, and J. C. Avise. 1996. Recognizing the forest for the trees: testing temporal patterns of cladogenesis using a null model of stochastic diversification. Mol. Biol. Evol. 13(6): 833-849.

Yang, Z. and B. Rannala. 2006. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol. Biol. Evol. 23: 212-226.

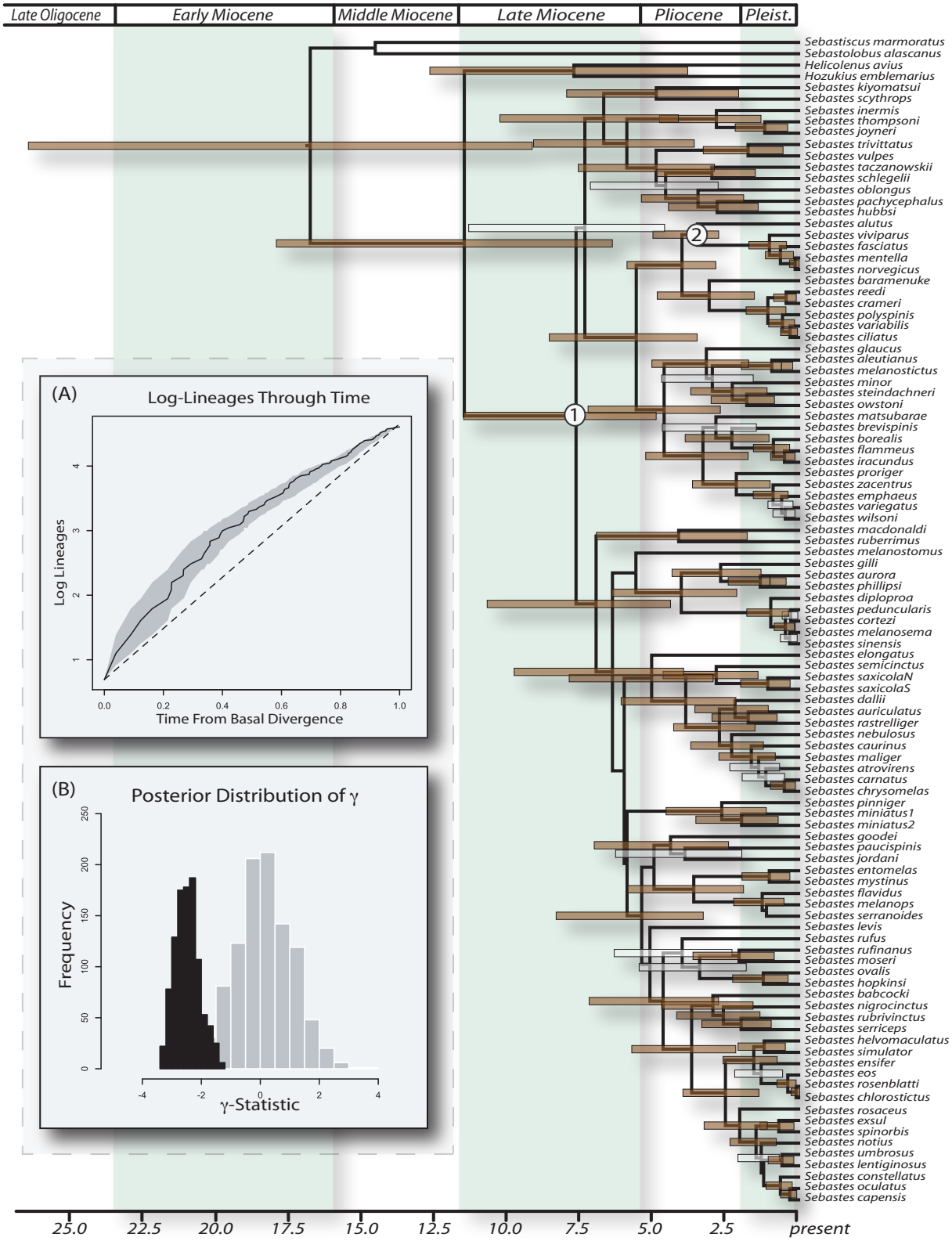

Late Oligocene
Early Miocene
Middle Miocene
Late Miocene
Pliocene
Pleist.
Sebastiscus marmoratus
Sebastolobus alascanus
Helicolenus avius
Hozukius emblemarius
Sebastes kiyomatsui
Sebastes scythrops
Sebastes inermis
Sebastes thompsoni
Sebastes joyneri
Sebastes trivittatus
Sebastes vulpes
Sebastes taczanowskii
Sebastes schlegelii
Sebastes oblongus
Sebastes pachycephalus
Sebastes hubbsi
Sebastes alutus
Sebastes viviparus
Sebastes fasciatus
Sebastes mentella
Sebastes norvegicus
Sebastes baramenuke
Sebastes reedi
Sebastes crameri
Sebastes polyspinis
Sebastes variabilis
Sebastes ciliatus
Sebastes glaucus
Sebastes aleutianus
Sebastes melanostictus
Sebastes minor
Sebastes steindachneri
Sebastes owstoni
Sebastes matsubarae
Sebastes brevispinis
Sebastes borealis
Sebastes flammeus
Sebastes iracundus
Sebastes proriger
Sebastes zacentrus
Sebastes emphaeus
Sebastes variegatus
Sebastes wilsoni
Sebastes macdonaldi
Sebastes ruberrimus
Sebastes melanostomus
Sebastes gilli
Sebastes aurora
Sebastes phillipsi
Sebastes diploproa
Sebastes peduncularis
Sebastes cortezi
Sebastes melanosema
Sebastes sinensis
Sebastes elongatus
Sebastes semicinctus
Sebastes saxicolaN
Sebastes saxicolaS
Sebastes dallii
Sebastes auriculatus
Sebastes rastrelliger
Sebastes nebulosus
Sebastes caurinus
Sebastes maliger
Sebastes atrovirens
Sebastes carnatus
Sebastes chrysomelas
Sebastes pinniger
Sebastes miniatus1
Sebastes miniatus2
Sebastes goodei
Sebastes paucispinis
Sebastes jordani
Sebastes entomelas
Sebastes mystinus
Sebastes flavidus
Sebastes melanops
Sebastes serranoides
Sebastes levis
Sebastes rufus
Sebastes rufinanus
Sebastes moseri
Sebastes ovalis
Sebastes hopkinsi
Sebastes babcocki
Sebastes nigrocinctus
Sebastes rubrivinctus
Sebastes serriceps
Sebastes helvomaculatus
Sebastes simulator
Sebastes ensifer
Sebastes eos
Sebastes rosenblatti
Sebastes chlorostictus
Sebastes rosaceus
Sebastes exsul
Sebastes spinorbis
Sebastes notius
Sebastes umbrosus
Sebastes lentiginosus
Sebastes constellatus
Sebastes oculatus
Sebastes capensis
1
2
(A)
Log-Lineages Through Time
Log Lineages
Time From Basal Divergence
(B)
Posterior Distribution of γ
Frequency
γ-Statistic
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
present

Figure 6: BEAST chronogram for *Sebastes* rockfishes. Colored bars represent 95% credible interval for node ages. Node bars are color-coded based on the posterior probability (PP) of the node in question: Dark brown bars represent PP = 1, Light brown bars represent PP = 0.95, White bars represent PP = 0.75, and Clear bars represent PP < 0.75. Calibrated nodes are numbered (See text fro details). The timescale displayed on the bottom is in million-years. Inset (A) shows the ltt-plot for the MCC (black line) with 95% quantiles for the number of lineages at a specific time calculated from the psoterior distribution of the BEAST analysis (shaded region). The light dashed line represent the PB expectation. Inset (B) shows the distribution of γ derived from posterior distribution (black) compared to the MCCR simulated null distribution (gray).

Model	LnL	ΔAIC	Akaike Weight
DDX	157.698	0	0.41
DDL	157.515	0.366	0.34
LD	157.193	1.01	0.249
PB	152.282	10.829	0.001

Table 1: Model-fitting results. Model abbreviations are as follows: DDX = density-dependent exponential; DDL = density-dependent lognormal; LD = linear-decline; PB = pure birth. ΔAIC is this difference between the model in question and the model with the lowest AIC score.

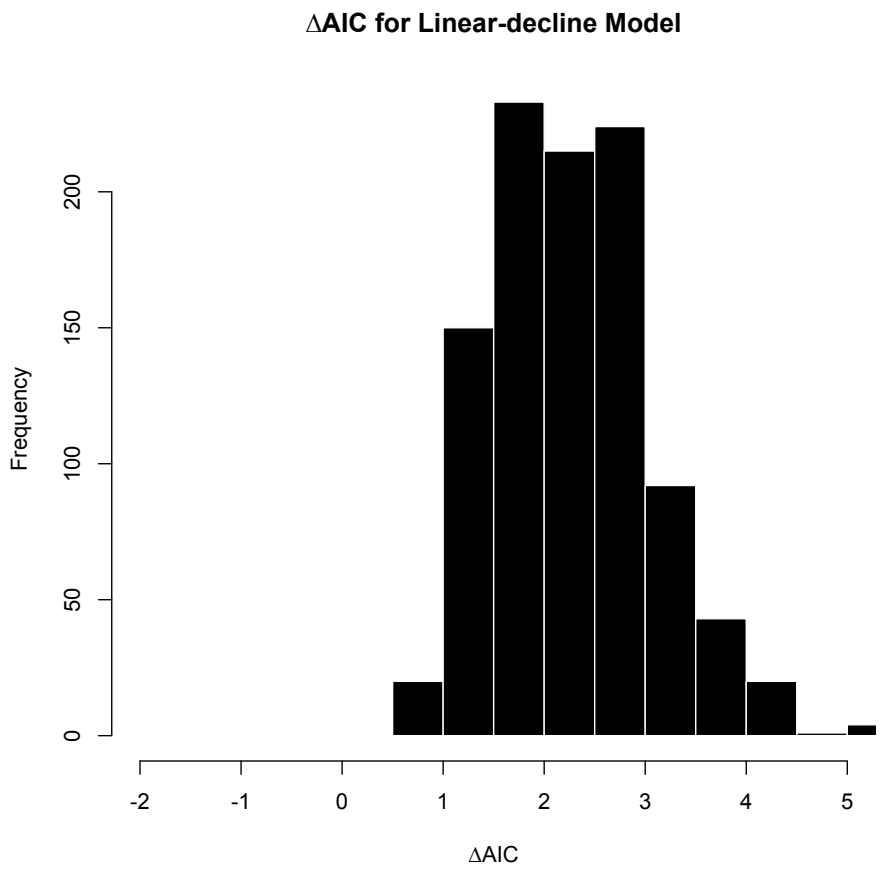


Figure 7: Histogram showing the distribution of ΔAIC between the best-fitting DD model and the linear-decline model derived 1000 randomly drawn trees from the posterior distributionof the BEAST analysis.

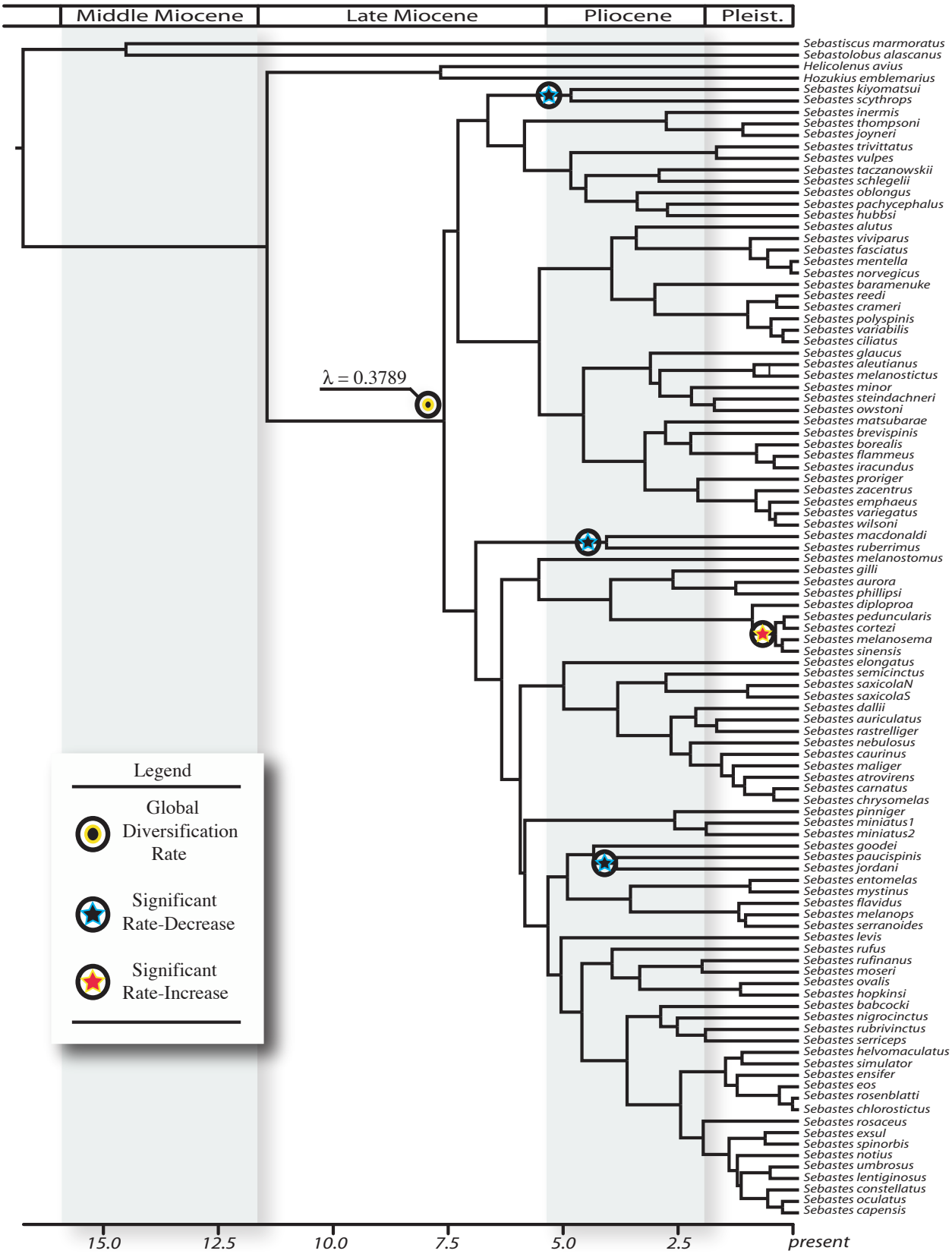

Middle Miocene
Late Miocene
Pliocene
Pleist.
Sebastiscus marmoratus
Sebastolobus alascanus
Helicolenus avius
Hozukius emblemarius
Sebastes kiyomatsui
Sebastes scythrops
Sebastes inermis
Sebastes thompsoni
Sebastes joyneri
Sebastes trivittatus
Sebastes vulpes
Sebastes taczanowskii
Sebastes schlegelii
Sebastes oblongus
Sebastes pachycephalus
Sebastes hubbsi
Sebastes alutus
Sebastes viviparus
Sebastes fasciatus
Sebastes mentella
Sebastes norvegicus
Sebastes baramenuke
Sebastes reedi
Sebastes crameri
Sebastes polyspinis
Sebastes variabilis
Sebastes ciliatus
Sebastes glaucus
Sebastes aleutianus
Sebastes melanostictus
Sebastes minor
Sebastes steindachneri
Sebastes owstoni
Sebastes matsubarae
Sebastes brevispinis
Sebastes borealis
Sebastes flammeus
Sebastes iracundus
Sebastes proriger
Sebastes zacentrus
Sebastes emphaeus
Sebastes variegatus
Sebastes wilsoni
Sebastes macdonaldi
Sebastes ruberrimus
Sebastes melanostomus
Sebastes gilli
Sebastes aurora
Sebastes phillipsi
Sebastes diploproa
Sebastes peduncularis
Sebastes cortezi
Sebastes melanosema
Sebastes sinensis
Sebastes elongatus
Sebastes semicinctus
Sebastes saxicolaN
Sebastes saxicolaS
Sebastes dallii
Sebastes auriculatus
Sebastes rastrelliger
Sebastes nebulosus
Sebastes caurinus
Sebastes maliger
Sebastes atrovirens
Sebastes carnatus
Sebastes chrysomelas
Sebastes pinniger
Sebastes miniatus1
Sebastes miniatus2
Sebastes goodei
Sebastes paucispinis
Sebastes jordani
Sebastes entomelas
Sebastes mystinus
Sebastes flavidus
Sebastes melanops
Sebastes serranoides
Sebastes levis
Sebastes rufus
Sebastes rufinanus
Sebastes moseri
Sebastes ovalis
Sebastes hopkinsi
Sebastes babcocki
Sebastes nigrocinctus
Sebastes rubrivinctus
Sebastes serriceps
Sebastes helvomaculatus
Sebastes simulator
Sebastes ensifer
Sebastes eos
Sebastes rosenblatti
Sebastes chlorostictus
Sebastes rosaceus
Sebastes exsul
Sebastes spinorbis
Sebastes notius
Sebastes umbrosus
Sebastes lentiginosus
Sebastes constellatus
Sebastes oculatus
Sebastes capensis
λ = 0.3789
Legend
Global Diversification Rate
Significant Rate-Decrease
Significant Rate-Increase
15.0
12.5
10.0
7.5
5.0
2.5
present

Figure 8: *Sebastes* chronogram showing the results from the MEDUSA analysis. The yellow-ringed bullseye displays the the global diversification rate at the base of the Sebastes radiation. The red and green stars denote lineages for which there has been a speciation rate increase and decrease, respectively. Timescale is the same as Figure 6.

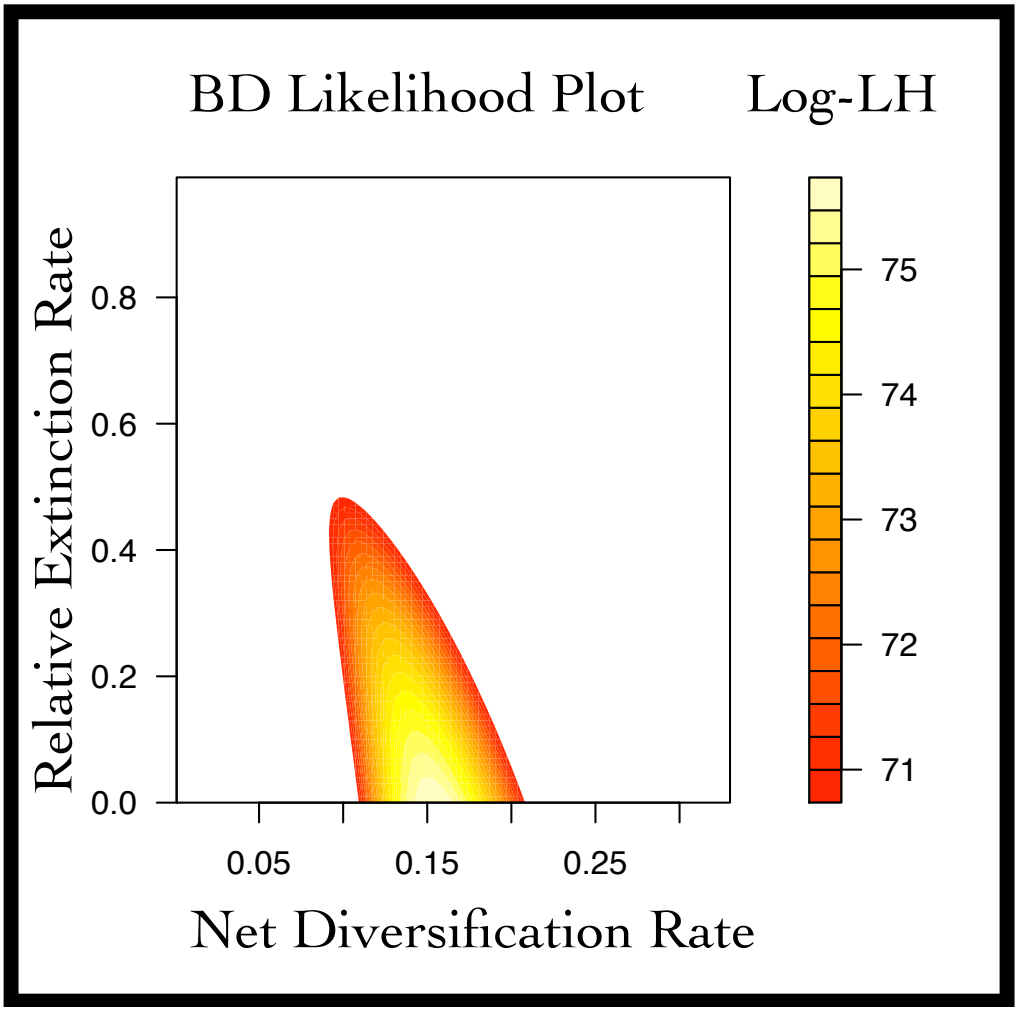


Figure 9: Log-likelihood plot for the global speciation (0.387) and extinction (0.0004) rate estimates for *Sebastes* rockfishes derived from the MEDUSA analysis. The log-likelihood color scale is shown to right, with hotter colors representing higher log-likelihoods.