

---

# 9 Epistemological Concern for Estimating Extinction

## *Introducing a New Model for Comparing Phylogenies*

*Prosanta Chakrabarty and Subir Shakya*  
Louisiana State University

### CONTENTS

9.1 Introduction .....	141
9.2 Methods .....	146
9.3 Results and Discussion .....	147
Acknowledgments.....	151
Appendix 9.1 Python Script for Automating the Process of Measuring the Extinction Index Calculated by the CTNM.....	151
Appendix 9.2 Draft (not Annotated) R Script Provided by an Anonymous Reviewer to Calculate the Extinction Index of the CTNM on Simulated Trees with Varying Extinction Rates (0, 0.025, 0.05, 0.1, 0.2, and 0.8) .....	154
References.....	155

*“Instead of choosing the best among a set of models, no matter how insufficient they all are, one could identify situations where there is no existing adequate model.”*

—Rabosky and Goldberg 2015

### 9.1 INTRODUCTION

Extinctions, being generally unobservable in the fossil record and in the lifetime of a scientist, are a mystery. Most species that have gone extinct left no trace of their existence. For most species that have lived on Earth, all that remains are their descendants, who themselves have evolved into other forms and perhaps have also gone extinct. Unfortunately, reliably measuring extinction by relying on the relationships among extant taxa and a scant fossil record is problematic, if not impossible (Rabosky 2010). The most widely used birth–death models estimate extinction rates that are close to zero (Nee 2006; Höhna et al. 2011) even though extinction

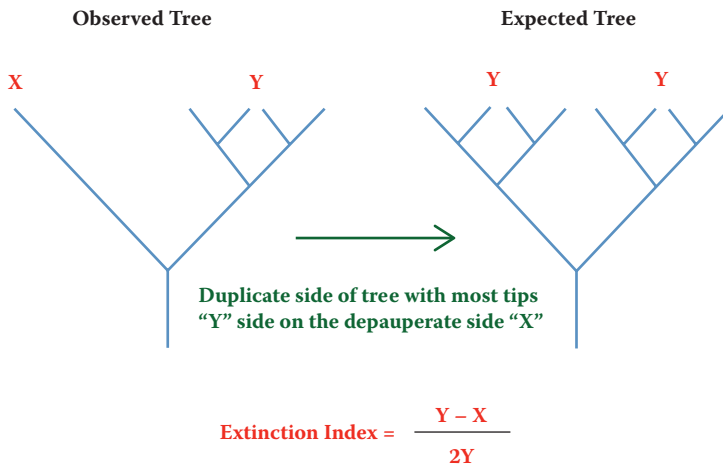
rates were more likely to be relatively high (Rosenblum et al. 2012). A lack of a robust measure of extinction is unfortunate because knowledge of the true amount of extinction that has taken place in a clade could be extremely informative (Simpson 1944). Extinction rates could help us better understand speciation and ecological turnover in a clade. Knowing the total number of extinct species in a group can tell us about the total diversity once contained within that clade (He and Hubbell 2011). Past extinction events can even provide relevant information on current ones (Régnier et al. 2015): for example, did past climate change result in a lineage more or less resilient to current levels of change? Are extinction events correlated across groups in space and time?

Here, we propose a new model for estimating extinction events using a simple formula that can be applied across any phylogeny. Like other models of extinction, this new model cannot be verified or falsified (because we cannot know the true amount of extinction) outside of a simulation. Our goal is for this new model to be a useful exercise to compare phylogenies and to help us rethink our assumptions about measuring extinction. We attempt with this method to harken back to the earliest days of studying biological diversification by relying on tree symmetry as a proxy for understanding diversification and extinction (see discussion in Pennell and Harmon 2013) and in using a sister clade comparison to investigate diversification within the tree (as previously proposed, perhaps first by Slowinski and Guyer 1993). The major assumption of this model is that a constant rate of origination without extinction will lead to a symmetrical tree. The pragmatic justification for this assumption is based on some empirical evidence that time-constant diversification tends to produce more balanced trees (Nee et al. 1994; Chan and Moore 2005; Ricklefs 2007; Morlon 2014), although never a symmetrical one. This assumption allows us to easily make comparisons between different kinds of asymmetrical trees versus the baseline of symmetry. The symmetrical “expected” tree will have two sides (sister clades), when viewed in two dimensions, with the same total number of species. As shown in Figure 9.1, one can then calculate the number of extinct taxa by subtracting the number of species in the “expected” tree from the number of taxa in the original “observed” tree.

The symmetrical, expected tree could be treated as the null hypothesis for any given phylogeny. This null hypothesis allows groups of different sizes and phylogenetic histories to be easily compared against their expected tree, which allows us to measure extinction as a fraction of the total number of expected species.

In this model, extinction is the only cause of asymmetry in a phylogeny (see Pearson 1998 for similar thoughts, also Heard and Mooers 2012 for a discussion). We note that other sources of noise, including uneven sampling, rate variation across lineages, and phylogenetic error (Huelsenbeck and Kirkpatrick 1996) may also contribute to tree asymmetry; however, these other sources of noise are beyond the scope of this manuscript, which focuses on the birth–death process assuming no other sources of error. However, we do propose a way that incomplete sampling, including from the fossil record, can be incorporated into our model of extinction.

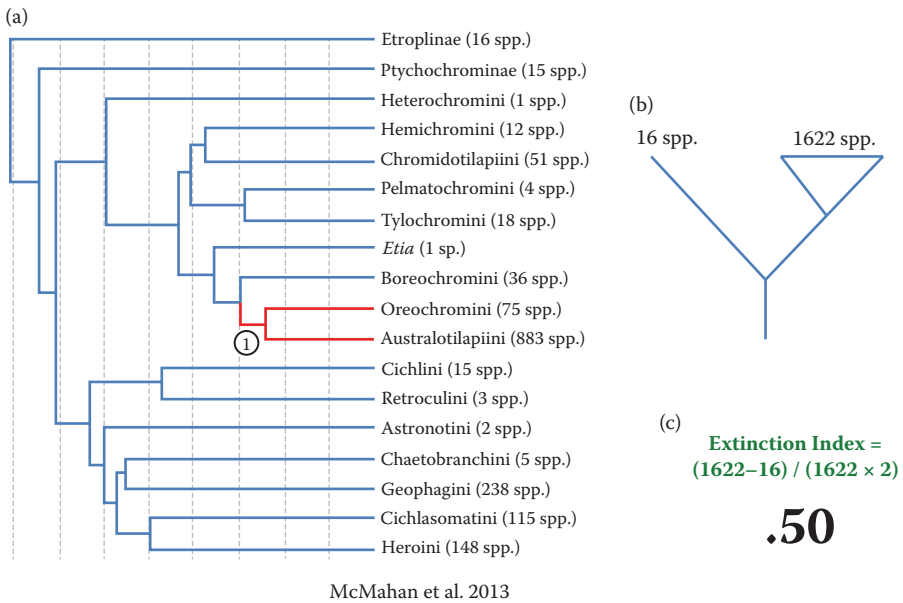
One advantage of this new method is that measures of extinction in trees with complete taxon sampling are less error prone than those with incomplete sampling



**FIGURE 9.1** *Using the Complete Tree Null Model:* The phylogeny obtained by the user, that is, the “observed tree,” should be pruned to the ingroup or section where extinction is to be measured. The phylogeny (which must not have a basal polytomy) is then made symmetrical by duplicating the species-rich side of the phylogeny (“Y” side) on to the species-poor side “X” side to obtain the “expected” (zero extinction) tree. The percentage of extinction is then measured by the equation  $(Y - X)/2Y$ . Where X is the number of species on the species-poor side of the tree, and Y is the number of taxa on the species-rich side of the tree.

(Rabosky 2010). Under this method, which we call the “Complete Tree Null Model” (CTNM), one need only know if missing (unsampled) taxa belong to the species-rich (“Y” clade or side) or species-poor clade (“X” clade or side) of a phylogeny to achieve a measure of complete taxon sampling as shown in Figure 9.1. In addition, Figure 9.2 shows an instance where a phylogeny of a family of organisms is represented by only one species of each genus in that family; by plugging in the numbers of species for each genus, one can then know the number of taxa expected on each sister clade of the tree. These “missing taxa” may also include fossils that can be added to the extinction total as shown in Figure 9.3. Even in the absence of major clades, if one knows the relative position of missing taxa within the tree (either right side or left), then estimates can be made about the total number of taxa in the ingroup. Maximum likelihood models for adding taxa to incomplete trees are also available for particularly difficult to place clades or extinct lineages (Revell et al. 2015) and these can easily be applied to the decision-making process for placing missing taxa.

This minimalist, or parsimonious approach (in terms of minimizing ad hoc assumptions) to measure extinction is an alternative to the many perhaps over-parameterized estimations of diversification. To our knowledge, no model currently exists that focuses solely on extinction. Instead, most models focus on speciation and extinction simultaneously to better understand diversification. In most of these cases, extinction is modeled strictly for the sake of understanding rates of speciation and overall diversification. Because speciation is as difficult to measure as extinction,



**FIGURE 9.2** Adding missing (unsampled) taxa to the CTNM: Cichlid phylogeny from McMahan et al. (2013) (a). Number of total species is included in the X and Y sides of the tree (b), even though they were not sampled in the original phylogeny missing taxa can be added simply by counting them in the totals for each side of the tree. The totals with the missing taxa can then be used to measure the percentage of extinction following the CTNM (c).

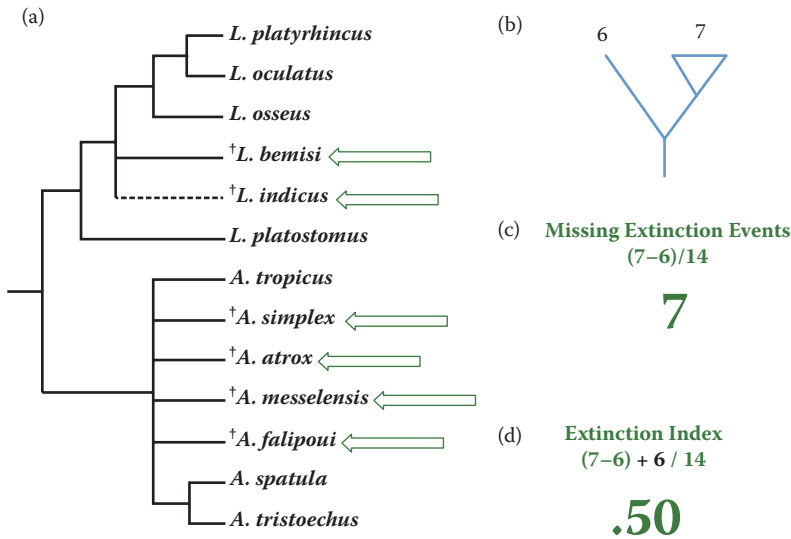


AU: Do we mean "... velocity and position of that of a Higgs Boson" or "... velocity and position of Higgs Boson"?

measuring both simultaneously could be thought of as impractical as simultaneously measuring the velocity and position of a Higgs **Boson**.

Extinction, perhaps because of the difficulty in finding evidence for it, typically plays second fiddle to speciation (although obviously extinction plays a major role in diversification). Extinction rates are often held constant (e.g., Heath et al. 2015) or if estimated, are typically underestimated, that is, close to zero (Nee 2006; Morlon et al. 2011; Höhna et al. 2011).

Many diversification models have so many parameters at play that it is difficult to measure any one parameter independently (Rabosky and Goldberg 2015) and certainly extinction cannot be accurately measured when extinction is assigned an *a priori* arbitrary constant rate. The point of many diversification models is not to measure extinction at all, perhaps because of the difficulty of measuring extinction with an incomplete fossil record and extrapolating the death of species solely on the pattern of diversification of extant species. Rabosky (2010) goes as far as to state that "extinction should not be estimated on molecular phylogenies" and that "extinction rates should not be estimated in the absence of fossil data"—a stand at odds with other researchers (see Nee et al. 1994; Paradis 2003). Morlon et al. (2011) who have noted that many researchers purported to find an extinction rate of zero based on likelihood-based molecular phylogenies despite the frequency of known extinction



Grande 2010 (image from Wright et al. 2012)

**FIGURE 9.3** Adding fossil information to the CTNM: Phylogeny of Lepisosteidae from Grande 2010 (image from Wright et al. 2012) that includes placement of fossil gars (green arrows) (a). The fossil species are included in the “Expected” tree just as with extant taxa (b). The percentage of missing extinct taxa is then calculated in the same way as previously discussed (c) with the exception that the fossil taxa are then added to the numerator again because they too are extinct species (d). Note that groups with a large fossil record can have extinction over 0.50 or even over 1 depending on the number of fossil species known.

events. Höhna et al. (2011:2586) show that the commonly used birth–death models often underestimate extinction and that they “... often show an estimated death rate close to zero even though it is well known that species go extinct during evolution.”

The greatest source of empirical evidence of extinction and speciation is the fossil record (Rosenblum et al. 2012). Yet, fossils also only provide an incomplete record of extinction. We continue to struggle with incorporating evidence from the fossil record into our understanding of the phylogenetic history of a group (Rabosky 2010; Morlon et al. 2011; Pennell and Harmon 2013). Unfortunately, many groups lack a known fossil record altogether, but that does not mean their extinction rate should be considered zero as many birth–death models estimate. In the CTNM method, fossils can easily be incorporated into our measure of extinction.

In the CTNM fossil taxa are at first treated like extant taxa; that is, they are placed as tips in the phylogeny similar to how missing (unsampled) taxa were added as described above and as shown in Figure 9.3. The percentage of missing taxa is then calculated as before with the exception that these fossil taxa are then re-added to the numerator of the extinction equation. Therefore, the CTNM provides an easy and

intuitive way to add evidence of extinct species from the fossil record to a phylogeny and incorporates that evidence to estimate the amount of extinction in a group.

How overall extinction is determined in many studies is not always clear. Many models of speciation and divergence assume a constant rate of extinction (see above) or one that changes based on a fixed set of branching points based on a likelihood function and/or fossil evidence (Morlon et al. 2011). In the CTNM, we assume a constant rate for speciation in order to measure the variable of interest: extinction.

## 9.2 METHODS

To carry out the complete tree null model approach, the user obtains an empirically derived phylogeny (from maximum likelihood, parsimony, simulations, etc.) that should be pruned to the ingroup taxa of interest. Extinction will be calculated on the remaining (nonpruned) section of the tree; this section is treated as the “observed tree.” The two sides of the observed tree are then made symmetrical by duplicating the species-rich side of the phylogeny (“Y-side”) on to the opposite species-poor side (“X-side”) to create the “expected tree,” which under the CTNM is assumed to be the result of zero extinction. The index of extinction is then calculated by the fraction:  $(Y - X)/2Y$ ; where Y equals the number of species on the species-rich half of the tree and X equals the number of species on the species-poor side of the tree. A schematic representation of this method is shown in Figure 9.1. The result from the calculation, although presented as a proportion, should not be considered a literal proportion of extinct taxa (again this is unknowable). Rather we ask the user to interpret this value as an “index of extinction” without units that can be used to compare across groups.

We examined several independent phylogenies and used our CTNM to compare the index of extinction across groups. We sampled phylogenies with either robust taxon sampling (Near et al. 2011; Collins et al. 2015) or with a tree structure that permits easy addition of unsampled taxa. For instance, Thompson et al. (2014) sampled less than 40% of the known species of piranhas (Serrasalminidae) but included examples of all genera. Assuming the unsampled species remain with their congeners, it was relatively easy to add the missing taxa to these trees by simply adding the number of unsampled taxa to those existing branches. After adding those unsampled taxa the side of the phylogeny (either right or left) with the most taxa was again designated “Y” and the side with the least species was designated “X.”  $Y - X$  is the number of species “missing” from the X side of the tree (given the assumption of constant origination);  $2Y$  is the total number of expected species under the CTNM. A schematic representation of this method is shown in Figure 9.2. Similarly, one can add information about known fossils in the same way unsampled taxa were added. The fossil taxa are added to the expected tree just as extant taxa are and the extinction index is calculated in the same manner as discussed above with the exception that the number of fossil taxa are added to the numerator ( $Y - X$ ) again (because they too are extinct species). A schematic representation of this method is shown in Figure 9.3.

We also created a custom Python script, in Appendix 9.1, to run the CTNM and calculate an extinction index automatically. This Python script will accept a newick file for any phylogenetic tree and provide an extinction index as calculated by the CTNM. This script also calculates a  $p$ -value for a chi-squared test assuming the

degrees of freedom equals 1 (because there are two variables, the X and Y side of the tree). We remind the user to input the newick file with the ingroup taxa of interest only. An anonymous reviewer also generously provided a draft R script (Appendix 9.2) to examine the distribution of the CTNM on simulated trees. Each analysis ran a different extinction rate to generate 1,000 simulated trees to calculate the extinction index on each tree.

### 9.3 RESULTS AND DISCUSSION

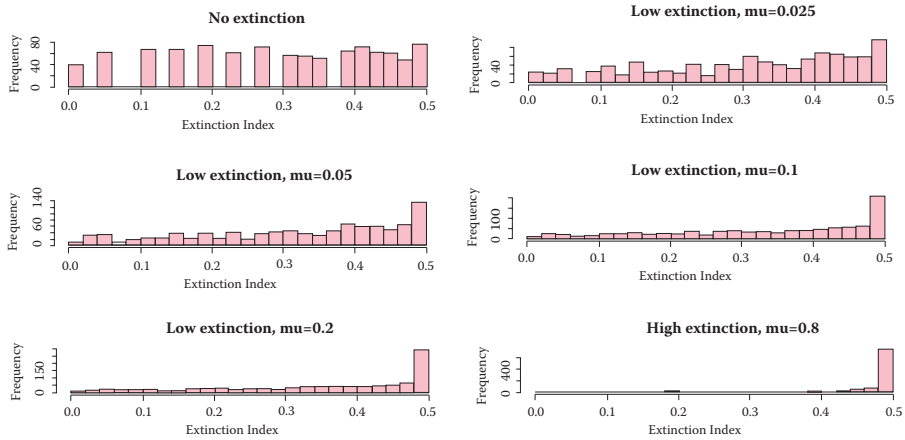
Our comparison of multiple groups using the CTNM is shown in Table 9.1. We find a somewhat narrow range of extinction indices across groups. The highest value achievable is 0.50 (unless additional fossils can be added as shown in Figure 9.3), the lowest 0. This narrow range was also achieved under simulations from a custom R script as shown in Figure 9.4, simulating trees with extinction rates from 0 to 0.8 and showing a heavy skewing towards an extinction index of 0.50, the higher the extinction rate. The results from the simulation also differ from the conclusions of Slowinski and Guyer (1989) in showing that tree probabilities do change with random extinction. Notably, some variation was found with our empirical data particularly when the contribution from the fossil record was included.

The percentage of total extinction for a group as measured by the CTNM is essentially a measure of tree imbalance or asymmetry. The more asymmetrical (many species on one side versus the other) the greater the amount of extinction that will be assessed. The most symmetrical lineage we discovered belonged to the Crocodylia (Oaks 2011), a group that was found to have an extinction index of 0.26.

**TABLE 9.1**

**List of Groups Compared Using the CTNM. An \* denotes the use of fossils in measuring the % of total extinct in a group**

Taxon	Reference	Number of Living Species	Extinction Index
Crocodylia	Oaks (2011)	23	0.26
Etheostomatinae	Near et al. (2011)	247	0.36
Monarchidae	Andersen et al. (2014)	99	0.37
Polycentridae	Collins et al. (2015)	5	0.38
Cotingidae	Berv and Plum (2014)	66	0.41
Serrasalmidae	Thompson et al. (2014)	99	0.42
Falconidae	Fuchs et al. (2015)	64	0.43
Leiognathidae	Chakrabarty et al. (2011)	43	0.45
Vireonidae	Slager et al. (2014)	52	0.49
Hirundinidae	Sheldon et al. (2005)	83	0.49
Cichlidae	McMahan et al. (2013)	1638	0.50
Lepisosteiformes	Grande (2010)	6	0.50*
Notothenioidei	Near et al. (2015)	114	0.50
Ostariophysi	Fink and Fink (1981)	10,237	0.50



**FIGURE 9.4** Results from R script from Appendix 9.2 showing the frequency distribution of the extinction index as calculated by CTNM) with 1,000 simulated trees with varying extinction rates.

We recovered several groups with very high asymmetry and thus a very high extinction index (0.49). These groups include: Vireonidae (Vireos), Hirundinidae (swallows), Cichlidae (cichlids), Lepisosteiformes (gars), Notothenioidei (icefishes), and Ostariophysi (the speciose superorder that includes catfishes, minnows, tetras, electric knifefishes, and among other freshwater fishes).

This simple method to estimate extinction may be too simple for the palate of many modern researchers. We know, for instance, that extinction and speciation rates vary among clades (Jetz et al. 2011; Rabosky et al. 2013) and also that a constant speciation rate will not necessarily lead to a symmetric tree (Losos and Adler 1994). Our analysis does not address any rates of extinction but shows only an overall fraction of extinction (which we refer to an index), a more conservative approach. This CTNM may even be of use to those that prefer more complex models; for instance, the conservative estimate of extinction derived from the CTNM can be used in choosing a simulated rate of extinction being applied in another method.

The CTNM examines just one set of sister-clade contrasts, the one between the left and right side of the tree within the ingroup. More complex models could be developed in which each clade in a tree is examined from the base to the tips to get a “rate” based on the CTNM approach. Slowinski and Guyer (1993) proposed a similar method using multiple sister-clade contrasts, except these contrasts were modeled with random speciation and extinction. More recent constant rate models, such as the equal rate Markov (ERM), allow stochasticity and will result in some random tree imbalance (Mooers and Heard 1997). These stochastic models make adding missing (unsampled) taxa more complex than in the CTNM.

Knowledge of the diversification rate of a clade is highly sought after, potentially informing us about adaptive radiations, ecological shifts, and key adaptations, among other evolutionary phenomena. Unfortunately, knowledge about diversification shifts may be misleading due to assumptions about extinction. Differences



in extinction rate can lead to two groups with the same diversification rate leaving behind phylogenetic trees of different (Nee et al. 1994b; Rabosky 2006b). A popular model by Magallón and Sanderson (2001) that searches for evidence of exceptional diversification rates uses estimates of extinction between 0% and 99%: if the speciation rate falls beyond the error bars of an assumed 0–99% extinction, the diversification is assumed to have a significant shift in rate. This broad estimate of extinction (0–99%) likely keeps us from capturing many of the other possible significant shifts. The focus of the CTNM on extinction specifically is not novel, but is intended to shift the discussion back to simpler criteria or assumption sets of measuring extinction.

The need for a simpler model is clear. In some commonly used programs, even arbitrary characters (like the length of a taxon name) can be found to have a (obviously spurious) correlation with a given rate of speciation under a constant birth–death assumption (see Rabosky and Goldberg 2015); also see discussion of birth–death models in Pennell and Harmon 2013). Models to estimate tree balance exist for both maximum likelihood (Chan and Moore 2005) and Bayesian approaches (Moore and Donoghue 2009) that can be compared with results from our parsimony approach. Despite some concerns, birth–death models remain a useful and popular tool in estimating diversification (Nee 2006; Höhna 2011; Frost 2014; Heath 2015). Here, we propose an alternative that may also be of value and that is philosophically rooted in Hennig’s (1966) model of speciation.

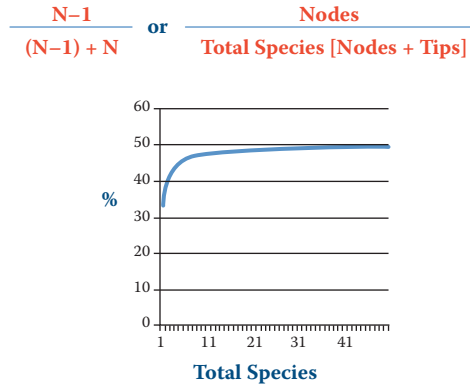
In Hennig’s (1966) model, every speciation event results in the origin of two sister taxa with the simultaneous extinction of the ancestor. In this model, every node in a phylogenetic tree represents an ancestor and also an extinction event. The number of nodes in a resolved tree is always equal to  $n-1$ , with the number of total taxa in the phylogeny =  $n$ . Therefore, the index of extinction events is always  $(n-1)/[n + (n-1)]$ ; or, in other words, the index of extinction events is equal to the total number of nodes in a tree  $(n-1)$  divided by the total number of extinct and extant taxa  $[(n-1) + n]$ . In a phylogenetic tree with 25 taxa represented as tips, there will be 24 internal nodes; the background extinction in this case is then  $24/(25 + 24)$  or 0.49. As shown in Figure 9.4, for most trees the number will approach 0.50. This baseline index of 0.50 extinction, paired with a constant rate of speciation, is effectively a high turnover birth–death model (see more on that in Höhna et al. 2011) that some argue is improbable (Rabosky and Lovette 2008). In the CTNM, the extinction index will never go above 0.50 in the absence of additional evidence from the fossil record, because extinction is effectively only measured on one side of the tree (the depauperate side). As shown in the case of Figure 9.3, the addition of fossils can increase this percentage infinitely.

In addition to Hennig, others have proposed that ancestral species are more likely to become extinct than their descendants (the so-called “Simpsonian step-series”; Simpson 1953; Pearson 1998). Rosenblum et al. (2012) argue that measuring the persistence of species is more important than speciation itself, since speciation rates appear to be higher than the number of taxa that persist. These “failed speciations” of incipient species discussed in Rosenblum et al. (2012) might explain the high background extinction rates of the Hennigian model (but see their caveats about high turnover in birth–death models) (Figure 9.5).



AU: Please confirm the in-text citation of Figure 9.5 okay here.

# of nodes is  $(N-1)$ ; where  $n$  is total number of species



**FIGURE 9.5** *The Hennig background extinction rate graphed:* The Hennig (1966) speciation model predicts that every extinct ancestor is a node on a phylogeny. The number of nodes is equal to “ $n-1$ ,” where “ $n$ ” is the number of tips or species that are sampled in a phylogeny. The percentage of nodes (or extinctions) approaches but never reaches 50% as the number of species sampled increases. This 50% baseline from the Hennig model is a philosophical basis for the extinction bar set in the CTNM with the exception that evidence from the fossil record will allow the percentage to increase above this 50% bar.

MEDUSA (Alfaro et al. 2009) and BAMM (Rabosky 2014) are two of the most widely used programs to examine shifts in diversification rate using more dynamic birth–death models. MEDUSA uses a likelihood-based step-wise Akaike information criterion to find the best-fit rate shifts in birth and death models but assumes a constant rate of speciation and extinction through clades. BAMM 2.0, a Bayesian approach, allows time-constant and time-varying diversification modes and varies extinction rates across clades. The CTNM proposed here can be considered the parsimony alternative to these Bayesian and likelihood approaches, and one that focuses solely on extinction (and not diversification as a whole) by keeping speciation constant.

The assumption that a constant rate of speciation will result in species arising equally across all branches is certainly wrong (empirically proven so by Losos and Adler 1994). And as a rough estimate of extinction, the CTNM is certainly underestimating extinction on the speciose side of the tree (“Y” side—where it is zero) and overestimating it on the depauperate side of the tree (“X” side—where all the extinction is counted). Despite that, averaged over the entire tree, the number of extinct taxa calculated might be closer to reality than in the constant rate models that find extinction to be near zero or that are based on arbitrary or assumption-laden parameters. The CTNM model has the advantage of easily incorporating missing taxa (the largest source of error for measuring extinction) and for incorporating the known fossil record for a group (the only empirical evidence of extinction that we have). Certainly better models to explain imbalance in empirical trees are still needed. This new approach is far from a perfect measure of extinction, but it can be a first step in rethinking how we measure the most enigmatic of biological processes: extinction.

## ACKNOWLEDGMENTS

This work was inspired by Donn Rosen, whose clear thinking approach often found the simplest ways to explain complex patterns in natural history. Although we are obviously incapable of similar clear thinking, we hope Dr. Rosen would have considered our approach interesting if not useful. We would like to thank the 2015 Systematics Discussion Group at Louisiana State University that helped hammer out the details of this idea through tough questions, parallel dialogue, perplexed looks, and a sprinkling of mockery and ridicule—often over beers. Also to those that commented on both our poster presentation in the 2015 Evolution meetings in Brazil and the oral presentation at the 2015 Joint Meetings of Ichthyologists and Herpetologists in Reno and two anonymous reviewers of this chapter. PC also thanks the Albert lab at the University of Louisiana, Lafayette, for discussing this topic between trawls on the Amazon. Lastly, we thank the organizers of the JMIH Rosen (“Donn Rosen and the Assumptions that Inhibit Scientific Progress in Comparative Biology”) Symposium, Lynne Parenti and Brian Crother, for the invitation to present and publish our work. This work was supported by NSF grant DEB 1354149 to PC.

## APPENDIX 9.1 PYTHON SCRIPT FOR AUTOMATING THE PROCESS OF MEASURING THE EXTINCTION INDEX CALCULATED BY THE CTNM

 [link: https://github.com/subirshakya/Epi\\_Con\\_Est\\_Extinction/blob/master/Ext\\_index\\_estimate.py](https://github.com/subirshakya/Epi_Con_Est_Extinction/blob/master/Ext_index_estimate.py)

```

__from__future__ import (division, print_function)
import numpy
from scipy.stats import chisquare as chi

class Node:

    """This holds the tree data and creates node class"""
    def __init__(self, name="", parent=None, children=None,
branchlength = 0):
        self.name = name #Name of node
        self.parent = None #Name of parent (initially set to
None, but can be changed)
        if children is None:
            self.children = [] #List to hold children nodes
        else:
            self.children = children #List to append children nodes
        self.brl = branchlength #Branch lengths

class Tree:

    """
    Defines a class of phylogenetic tree, consisting of linked
    Node objects.
    """

    def __init__(self, data):
        self.root = Node("root") #Define root

```

```

self.newicksplicer(data, self.root) #Splice newick data

defnewicksplicer(self, data, base):
    """
    Splices newick data to create a node based tree. Takes a
    base argument which is the root node.
    Make sure the file has a single line of newick code. Any
    miscellaneous characters will throw the program off. Also make
    sure it does not end in ";"
    Newick should start with ( and end in ). Can modify
    script to do more.
    Note: The script handles every node as if there is a
    bifurcation. Any polytomy will not be correctly programmed.
    """

    data = data.replace(" ", "")[1: len(data)] #Get rid of
    all spaces and removes first and last parenthesis
    n = 0
    if data.count(",") != 0: #While there is no more comma
    separated taxa
        for key in range(len(data)): #Find the corresponding
    comma for a given parenthesis (n will be 0 for the correct
    comma)
            if data[key] == "(":
                n += 1 #Increase index of n by 1 for 1 step into
    new node
            elif data[key] == ")":
                n -= 1 #Decrease index of n by 1 for 1 step out node
            elif data[key] == ",":
                if n == 0: #To check for correct comma
                    vals = (data[0:key], data[key+1:len(data)-1])
    #Break newick into left and right datasets
                    for unit in vals: #For each entry of dataset
                        if ":" in unit: #For cases with branch lengths
                            data = unit[0:unit.rfind(":")] #get rid of
    trailing branchlength if provided
                            node_creator = Node(data, parent = base)
    #Create node entry
                            node_creator.br1 = float(unit[unit.
    rfind(":")+1:]) #Append branch length of that branch
                            base.children.append(node_creator) #Create
    children
                                    self.newicksplicer(data, node_creator)
    #Recursive function
                                else: #For case with no branch lengths
                                    data = unit
                                    node_creator = Node(data, parent = base)
                                    base.children.append(node_creator)
                                    self.newicksplicer(data, node_creator)
                                break #Terminate loop, we don't need to look any
    further

```

```

def countsym(self, node):
    """
    Breaks tree into two halves at the root node and
    passes a command to each node.
    """
    list = [] #List to hold output values
    for child in node.children:
        val = self.countsym2(child)
        nodes = self.nodcount (child)
        list.append((val, nodes))
    return list

def nodcount(self, node):
    """
    Count number of nodes on tree
    """
    start = 0
    if node.children == []:
        return 0 #Terminal node returns 0 as there
are no nodes above it
    else:
        start += 1 #Any non-terminal node will add
1 to the total
        for child in node.children:
            start += self.nodcount(child)
        return start

def length(self, node):
    """
    Count length to each bifurcation
    """
    list = []
    if node.children == []: #Terminal branch returns
branch length
        return node.br1 #Not sure about this part,
but without it the numbers are skewed
    else:
        try:
            list.extend(node.br1)
        except:
            list.append(node.br1)
        for child in node.children:
            try:
                list.extend(self.
length(child))
            except:
                list.append(self.
length(child))
        list = filter(lambda x:x != 0, list)
#Filter to remove zeros from the list
    return list

```

```


defdivcount (self):
    """
    Tallies the lengths to give total diversification
rate
    """
    list = self.length(self.root)
    val = numpy.mean(list)
    returnval

def summarize(self):
    """
    Summarize data
    """
    list = self.countsym(self.root)
    obs = list[0][0]+list[1][0]
    exp = max(list[0][0],list[1][0])*2
    print ("Number of observed species: ", obs)
    print ("Number of extinct species: ", exp-obs)
    print ("% extinct: ", (exp-obs)/(exp)*100)
    print ("chi: ", chi(f_obs = obs, f_exp = exp,
ddof = -(obs-1)))

    """
    data =          #Provide newick here
    sim = Tree(data) #Loads tree into program
    sim.summarize() #Prints summary stats for the data
    """

```

**APPENDIX 9.2 DRAFT (NOT ANNOTATED) R SCRIPT PROVIDED BY AN ANONYMOUS REVIEWER TO CALCULATE THE EXTINCTION INDEX OF THE CTNM ON SIMULATED TREES WITH VARYING EXTINCTION RATES (0, 0.025, 0.05, 0.1, 0.2, AND 0.8)**

 link: [https://github.com/subirshakya/Epi\\_Con\\_Est\\_Extinction/blob/master/Graph\\_test.R](https://github.com/subirshakya/Epi_Con_Est_Extinction/blob/master/Graph_test.R)

```

### function to compute asymmetry of root node
library(ips)
library(ape)
library(TreeSim)

# root asymmetry (ra) function
# 'tr' is a object of class 'phylo' from the ape library
ra<- function(tr)
{
  N<- Ntip(tr)
  rn<- N+1
  ee<- tr$edge
  ees<- ee[ee[,1]==rn, ] # extract edges that stem from root
node
  desc.node<- ees[,2]

```

```

    if(any(desc.node< N)) { RS<- 1; LS<- N-1 # if either
    descendant of root is a terminal tip
    } else {
      RN<- descendants(tr, node=desc.node[1], type="t") # get
    descendants
      LN<- descendants(tr, node=desc.node[2], type="t")
      RS<- length(RN)
      LS<- length(LN)
    }

    ext<- abs(RS - LS) / (2* max(c(RS, LS))) # compute asymmetry
    metric
    res<- c(ext, max(c(RS, LS)), min(c(RS, LS)))
    names(res)<- c("asym", "right", "left")

    # side with more tips is considered the "right" side of tree
    return(res)

  }

nrep<- 1000
layout(1:3)

# no extinction
tr<- sim.bd.taxa(n=32, numbsim=nrep, lambda=1, mu=0)
rav<- t(sapply(tr, ra))
hist(rav[,1], 25, col="pink", main="No extinction")

# low extinction
tr<- sim.bd.taxa(n=32, numbsim=nrep, lambda=1, mu=0.2)
rav<- t(sapply(tr, ra))
hist(rav[,1], 25, col="pink", main="Low extinction, mu=0.2")

# high extinction
tr<- sim.bd.taxa(n=32, numbsim=nrep, lambda=1, mu=0.8)
rav<- t(sapply(tr, ra))
hist(rav[,1], 25, col="pink", main="High extinction, mu=0.8")

```

## REFERENCES

- Alfaro, M. E., F. Santini, C. Brock, H. Alamillo, A. Dornburg, D. L. Rabosky, G. Carnevale and L. J. Harmon. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proceedings of the National Academy of Sciences* 106: 13410–13414.
- Andersen, M. J., P. A. Hosner, C. E. Filardi and R. G. Moyle. 2015. Phylogeny of the monarch flycatchers reveals extensive paraphyly and novel relationships within a major Australo-Pacific radiation. *Molecular Phylogenetics and Evolution* 83: 118–136.
- Beaulieu, J. M. and B. C. O'Meara. 2015. Extinction can be estimated from moderately sized molecular phylogenies. *Evolution* 69: 1036–1043.

- Berv, J. S. and R. O. Prum. 2014. A comprehensive multilocus phylogeny of the Neotropical cotingas (Cotingidae, Aves) with a comparative evolutionary analysis of breeding system and plumage dimorphism and a revised phylogenetic classification. *Molecular Phylogenetics and Evolution* 81: 120–136.
- Chakrabarty, P., M. P. Davis, W. L. Smith, Z. H. Baldwin and J. S. Sparks. 2011. Is sexual selection driving diversification of the bioluminescent ponyfishes (Teleostei: Leiognathidae)? *Molecular Ecology* 20: 2818–2834.
- Chan, K. M. A. and B. R. Moore. 2005. SYMMETREE: Whole-tree analysis of differential diversification rates. *Bioinformatics* 21: 1709–1710.
- Fink, S. V. and W. L. Fink. 1981. Interrelationships of the ostariophysan fishes (Teleostei). *Zoological Journal of the Linnean Society* 72: 297–353.
- Frost, S. D. W., O. G. Pybus, J. R. Gog, C. Viboud, S. Bonhoeffer and T. Bedford. 2015. Eight challenges in phylodynamic inference. *Epidemics* 10: 88–92.
- Fuchs, J., J. A. Johnson and D. P. Mindell. 2015. Rapid diversification of falcons (Aves: Falconidae) due to expansion of open habitats in the Late Miocene. *Molecular Phylogenetics and Evolution* 82: 166–182.
- Grande, L. 2010. An empirical synthetic pattern study of gars (Lepisosteiformes) and closely related species, based mostly on skeletal anatomy. The resurrection of Holostei. *Copeia* 10(2A): 1.
- Heard, S. B. and A. O. Mooers. 2002. Signatures of random and selective mass extinctions in phylogenetic tree balance. *Systematic Biology* 51: 889–897.
- Heath, T. A., J. P. Huelsenbeck and T. Stadler. 2015. The fossilized birth–death process: A coherent model of fossil calibration for divergence time estimation. *Proceedings of the National Academy of Sciences* 111: 2957–2966.
- He, F. and S. P. Hubbell. 2011. Species–area relationships always overestimate extinction rates from habitat loss. *Nature* 473: 368–371.
- Hennig, W. 1966. *Phylogenetic Systematics*. Urbana, Illinois: University of Illinois Press.
- Höhna, S., T. Stadler, F. Ronquist and T. Britton. 2011. Inferring speciation and extinction rates under different sampling schemes. *Molecular Biology and Evolution* 28: 2577–2589.
- Huelsenbeck, J. P. and M. Kirkpatrick. 1996. Do phylogenetic methods produce trees with biased shapes? *Evolution* 50: 1418–1424.
- Jetz, W., G. H. Thomas, J. B. Joy, K. Hartmann and A. Mooers. 2012. The global diversity of birds in space and time. *Nature* 491: 444–448.
- Losos, J. B. and F. R. Adler. 1994. Stumped by trees? A Generalized Null Model for patterns of organismal diversity. *American Naturalist* 145: 329–342.
- Magallón, S. and M. J. Sanderson. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55: 1762–1780.
- McMahan, C., P. Chakrabarty, W. L. M. Smith, J. S. Sparks and M. P. Davis. 2013. Temporal patterns of diversification across global cichlid biodiversity (Acanthomorpha: Cichlidae). *PLoSOne* 8(e711162): 1–9.
- Mooers, A. O. and S. B. Heard. 1997. Inferring evolutionary process from phylogenetic tree shape. *Quarterly Review of Biology* 72: 31–54.
- Moore, B. R. and M. J. Donoghue. 2009. A Bayesian approach for evaluating the impact of historical events on rates of diversification. *Proceedings of the National Academy of Sciences* 106: 4307–4312.
- Morlon, H. 2014. Phylogenetic approaches for studying diversification. *Ecology Letters* 17: 508–525.
- Morlon, H., T. L. Parsons and J. B. Plotkin. 2011. Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences* 108: 16327–16332.
- Near, T. J., C. M. Bossu, G. S. Bradburd, R. L. Carlson, R. C. Harrington, P. R. Hollingsworth Jr., B. P. Keckand and D. A. Etnier. 2011. Phylogeny and temporal diversification of darters (Percidae: Etheostomatinae) *Systematic Biology* 60: 565–595.



- Near, T. J., A. Dornburg, R. C. Harrington, C. Oliveira, T. W. Pietsch, C. E. Thacker, T. P. Satoh, E. Katayama, P. C. Wainwright, J. T. Eastman and J. M. Beaulieu. 2015. Identification of the notothenioid sister lineage illuminates the biogeographic history of an Antarctic adaptive radiation. *BMC Evolutionary Biology* 15: 109.
- Nee, S., R. M. May and P. H. Harvey. 2004. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B* 344: 305–311.
- Nee, S. 2006. Birth–death models in macroevolution. *Annual Review of Ecology Evolution and Systematics* 37: 1–17.
- Nee, S., E. C. Holmes, R. M. May and P. H. Harvey. 1994a. Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London B* 344: 77–82.
- Nee, S., R. M. May and P. H. Harvey. 1994b. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London B* 344: 305–311.
- Oaks, J. R. 2011. A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. *Evolution* 65: 3285–3297.
- Paradis, E. 2003. Analysis of diversification: combining phylogenetic and taxonomic data. *Proceedings of the Royal Society of London B* 270: 2499–2505.
- Pearson, P. N. 1988. Speciation and extinction asymmetries in paleontological phylogenies: Evidence for evolutionary progress. *Paleobiology* 24: 305–335.
- Pennell, M. W. and L. J. Harmon. 2013. An integrative view of phylogenetic comparative methods: Connections to population genetics, community ecology, and paleobiology. *Annals of the New York Academy of Sciences* 1289: 90–105.
- Rabosky, D. L. 2006a. LASER: A maximum likelihood toolkit for detecting temporal shifts in diversification rates. *Evolutionary Bioinformatics Online* 2: 247–250.
- Rabosky, D. L. 2006b. Likelihood methods for detecting temporal shifts in diversification rates. *Evolution* 60: 1152–1164.
- Rabosky, D. L. 2010. Extinction rate should not be estimated from molecular phylogenies. *Evolution* 64: 1816–1824.
- Rabosky, D. L. and I. J. Lovette. 2008. Density dependent diversification in North American wood-warblers. *Proceedings of the Royal Society of London B* 276: 995–997.
- Rabosky, D. L., F. Santini, J. T. Eastman, S. A. Smith, B. Sidlauskas, J. Chang and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4: 10.1038.
- Rabosky, D. L. and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Systematic Biology* 64: 340–355.
- Rabosky, D. L. 2014. Automatic detection of key innovations, rate shifts, and diversity dependence on phylogenetic trees. *PLoS One* 9: e89543.
- Régnier, C., G. Achaz, A. Lambert, R. H. Cowie, P. Bouchet and B. Fontaine. 2015. Mass extinction in poorly known taxa. *Proceedings of the National Academy of Sciences* 112: 7761–7766.
- Revell, L. J., D. L. Mahler, R. G. Reynolds and G. J. Slater. 2015. Placing cryptic, recently extinct, or hypothesized taxa into an ultrametric phylogeny using continuous character data: A case study with the lizard *Anolis roosevelti*. *Evolution* 69: 1027–1035.
- Ricklefs, R. E. 2007. Estimating diversification rates from phylogenetic information. *Trends in Ecology and Evolution* 22(11): 601–610
- Rosenblum, E. B., B. A. J. Sarver, J. W. Brown, S. D. Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell and L. J. Harmon. 2012. Goldilocks meets Santa Rosalia: An ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology* 39: 255–261
- Simpson, G. G. 1944. *Tempo and Mode of Evolution*. New York, NY: Columbia University Press.
- Simpson, G. G. 1953. *The Major Features of Evolution*. New York: Columbia University Press.



AU: Please provide the page range for reference "Rabosky et al., 2013."

- Sheldon, F. H., L. A. Whittingham, R. G. Moyle, B. Slikas, and D. W. Winkler. 2005. Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* 35: 254–270.
- Slager, D. L., C. J. Battey, R. W. Bryson, G. Voelker and J. Klicka. 2014. A multilocus phylogeny of a major New World avian radiation: The Vireonidae. *Molecular Phylogenetics and Evolution* 80: 95–104.
- Slowinski, J. B. and C. Guyer. 1993. Testing whether certain traits have caused amplified diversification—An improved method based on a model of random speciation and extinction. *American Naturalist* 142: 1019–1024.
- Thompson, A. W., R. R. Betancur, H. Lopez-Fernandez and G. Orti. 2014. A time calibrated, multi-locus phylogeny of piranhas and pacus (Characiformes: Serrasalminidae) and a comparison of species tree methods. *Molecular Phylogenetics and Evolution* 81: 242–257.
- Wright, J. J., S. R. David and T. J. Near. 2012. Gene trees, species trees, and morphology converge on a similar phylogeny of living gars (Actinopterygii: Holostei: Lepisosteidae), an ancient clade of ray-finned fishes. *Molecular Phylogenetics and Evolution* 63: 848–856.