Chapter 8
Use of Paleontological and Phylogenetic Data in Comparative and Paleobiological Analyses: A Few Recent Developments

Michel Laurin

Abstract Comparative biology has progressed tremendously but unevenly in the last decades, through incorporation of methodological progress in phylogenetics and in statistical methods that incorporate phylogenetic data into statistical analyses of character correlation or evolution. This review presents a few methods of general interest to comparative biologists, such as phylogenetic independent contrasts (PIC) and variance partition with phylogenetic eigenvector regression. In evo-devo, heterochrony detection has usually been done using event pairing, in the last decade. That method uses a topology, but does not exploit branch length information. A recently proposed method based on squared-change parsimony and PIC exploits both topology and branch lengths, and it outperforms event pairing. Molecular evolution can also benefit from a phylogenetic perspective, as shown in recent studies on genome size evolution. In paleobiology, phylogenies are still rarely and often incompletely incorporated in analyses. Recent developments facilitate time-tree compilations and the combination of paleontological and molecular age data, and new branch length transformation methods can help to standardize PIC, to determine if the characters evolved according to a Brownian motion model, and to deal with clades about which no age information is available.

8.1 Introduction

Comparative biology has roots extending at least into the early nineteenth century, through the works of Lamarck (1809), one of the very first evolutionists, and it could probably even be argued that some pre-evolutionary biologists did comparative

M. Laurin
Département Histoire de la Terre, UMR 7207, CNRS/MNHN/UPMC, Centre de Recherches sur la Paleodiversité et les Paléoenvironments, Muséum national d'Histoire naturelle, Bâtiment de Géologie, Case Postale 48, 43 rue Buffon, F-75231 Paris Cedex 05, France
e-mail: michel.laurin@upmc.fr

P. Pontarotti (ed.), *Evolutionary Biology – Concepts, Biodiversity, Macroevolution and Genome Evolution*, DOI 10.1007/978-3-642-20763-1_8,

biology, although in a different theoretical framework (Mayr 1982). The related field of paleobiology can be considered, to an extent, as a special form of comparative biology because one of the most reliable methods of paleobiological inference consists in demonstrating, in extant relatives of extinct taxa, a correlation between an attribute that usually fossilizes (e.g., skeletal characters in vertebrates) and another that is usually not observed in fossils but for which we need to infer the presence or value (such as a behavior or basal metabolic rate). Demonstrating such a correlation is a typical comparative biology problem, and paleobiologists rely extensively on work on extant taxa to draw their inferences (e.g., de Buffrénil and Rage 1993; Canoville and Laurin 2010).

Comparative biology has changed over time, as it incorporated new techniques and conceptual developments. The advent of cladistics (Hennig 1965) and later, of molecular phylogenetics and dating (Zuckerkandl and Pauling 1965) have greatly improved our knowledge of the tree of life, thus greatly facilitating the work of comparative biologists, to the extent that closely related taxa have to be compared to study transformation series. By now, most (but not all; see below) comparative biologists have integrated phylogenetics in their routine work.

Another very important development in comparative biology was the development of statistical methods that accounted for the statistical nonindependence of comparative data. Indeed, standard statistical methods assume that data about each point (terminal taxa, in the context of comparative analyses) are independent of each other. The very existence of the tree of life indicates that for many datasets, this assumption is violated; whether or not this happens depends mainly on the taxonomic sampling and the evolutionary rate of the characters, but empirical work (e.g., Freckleton et al. 2002; Laurin 2004; Cubo et al. 2005) and simulations (e.g., Martins et al. 2002; Laurin 2010a) show that this problem is pervasive. The first statistical method developed to solve this problem was the phylogenetic independent contrasts (Felsenstein 1985; abbreviated as PIC below), a method that has inspired most (Grafen 1989; Martins and Hansen 1997; Pagel 1997) but not all subsequent comparative methods (Gittleman and Kot 1990; Desdevises et al. 2003; Cubo et al. 2008).

This brief review shows that progress in comparative biology depends rather critically (but not exclusively) on the incorporation of phylogenetic data into the analysis, and on the use of comparative methods that adequately use these phylogenetic data. Below, I will first discuss briefly problems that may arise when this is not done. I will then present briefly some recent comparative methods, and show that most of them require phylogenies with estimated branch lengths. Getting these lengths remains difficult, despite recent progress in molecular and paleontological dating, and many authors still do not bother getting these. Thus, this topic deserves a discussion, which will lead into a short digression in a chronic problem in recent molecular dating studies, namely, the underuse of paleontological literature.

Throughout this discussion, I try to emphasize the most important points, but this review nevertheless emphasizes, to an extent, my own modest contributions to these fields, for the simple reasons that this is where my expertise lies, and that my

recent work on these topics is scattered in various journals and book chapters. Thus, a review summarizing these recent developments may be useful.

8.2 The Traditional Approach and Why It Is Being Abandoned

The most basic task of comparative biology, namely, showing a correlation between two features (e.g., body size and basal metabolic rate) used to be done using standard statistical methods, such as least-squares linear regressions. The problem with this approach is mostly that the statistical significance of the relationship is not assessed properly because the number of degrees of freedom is overestimated. Indeed, closely related species tend to resemble each other in most characters, so in a data matrix with n taxa, we do not have n independent data points. For instance, a standard simple linear regression is represented by equation (8.1):

$$y = ax + b \tag{8.1}$$

Two constants are estimated (a and b), and we should consequently expect to have n-2 degrees of freedom. In comparative biology, this is not true, but the number of degrees of freedom is difficult to estimate, and most comparative methods modify the data before performing regressions (see below). These problems are expected on the basis of theoretical considerations (Felsenstein 1985), and have been shown to occur in various situations represented by simulation parameters (Purvis et al. 1994; Martins et al. 2002; Laurin 2010a).

The importance of this problem in comparative biology cannot be overemphasized, especially because several studies are still conducted with inadequate comparative methods. Let us consider the classical problem of assessing the presence of evolutionary trends concerning some of the most basic questions about the history of life. For instance, did complexity of organisms increase over time (McShea 1996)? Did body size increase over time, a trend known as the Cope-Depéret rule (Laurin 2004)? Such studies are still being conducted with a great variety of methods, which hampers meaningful comparisons of results and of the reliability of analyses because the same data analyzed by different methods can yield contradictory results, for instance about the presence (Hone et al. 2008) or absence (Butler and Goswami 2008) of a trend of increasing body size in Mesozoic birds. In a recent simulation study attempting to remedy this situation (Laurin 2010a), I have shown that a simple, non-phylogenetic linear regression of body size vs. geological age of origin of terminal taxa, still used recently to assess evolutionary trends (e.g. Hone et al. 2008), has greatly inflated type I error rate, ranging from 0.12 to 0.18, at the 0.05 threshold. However, simple linear regression had good power, and yielded correct regression coefficient (slope) estimates (Laurin 2010a).

Similar problems pervade comparative biology and extend to the assessment of evolution of qualitative (discrete) characters. This applies also, for instance, to the field of evo-devo, in which a rigorous comparative phylogenetic framework is

unfortunately still often lacking. A good example of this is provided by the classical work on *Hox* gene expression patterns in developing vertebrate appendage buds. Sordino et al. (1995) showed that the teleost *Danio rerio* lacked the discrete third phase of *Hox* D-10 to D-13 gene expression pattern (Fig. 8.1d, f) that characterizes the tetrapod limb bud (Fig. 8.1c, e), then documented at least in the mouse, but since then demonstrated in other tetrapod taxa, such as in the chick. The territory where that third expression phase is located (Fig. 8.1e) corresponds more or less with the autopod (hand and foot). Furthermore, *Hox* A-11 is expressed at the apex of the fin in *D. rerio* (Fig. 8.1h), but in a territory proximal to the future autopod in tetrapod limb buds (Fig. 8.1g). Therefore, Sordino et al. (1995) concluded that the presence of a third phase of *Hox* gene expression pattern in tetrapod limbs supports the conclusion that the autopod is a neomorph. The problem with this interpretation is that with data only on one actinopterygian (the teleost *Danio rerio*), few tetrapods, and no other taxa, the polarity of the change in *Hox* gene expression patterns could not be established (the condition in the ancestral osteichthyan could not be determined unambiguously). Furthermore, the teleost *Danio rerio* has a diminutive paired fin endoskeleton lacking a metapterygial axis (usually considered to be homologous with the main axis of tetrapod limbs) that is probably reduced from that of the earliest actinopterygians, judging by the more developed fin endoskeleton (with a metapterygial axis) found in more basal actinopterygians, and this raises the possibility that *D. rerio* lost the third *Hox* gene expression phase when its paired

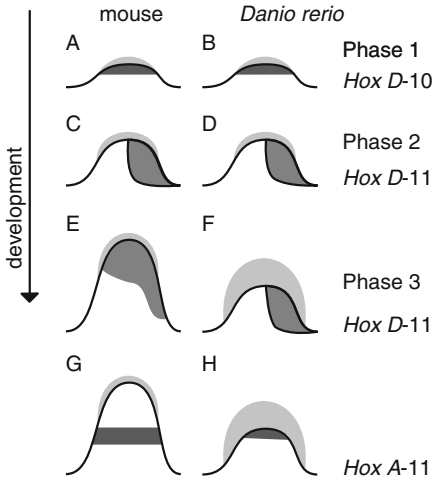


Fig. 8.1 *Hox* gene expression pattern in actinopterygian and tetrapod appendages. *Hox* gene expression pattern in mouse limb buds (*left*) and in fin buds of the teleost *Danio rerio* (*right*). Proximal is below, and cranial is to the left, in all figure parts. The zones of various *Hox* gene expressions are shaded dark gray; the apical ectodermal ridge, in which fin rays (dermal skeleton) develop, is in light gray. Note that in *Danio*, the third expression phase (**f**) is not distinct from the second one (**d**; the same expression pattern prevails), contrary to the pattern displayed by the mouse. (Redrawn from Sordino et al. 1995; modified version from Laurin 2010b)

fin endoskeleton was simplified. I pointed out this problem (Laurin et al. 2000) in a review paper, and raised doubts about these conclusions, but my concerns were ignored. Nevertheless, the subsequent discovery of a tetrapod-like third phase of *Hox* gene expression pattern in the basal actinopterygian *Polyodon spathula* (Davis et al. 2007) that retains a metapterygial axis in its paired fins confirmed my alternative interpretation.

The few examples mentioned above illustrate the need for a rigorous analytical method in comparative or evolutionary biology, if attributes of taxa are compared. Although this contribution focuses on quantitative characters, similar conclusions apply to all kinds of data, from nucleotides to ecology and behavior.

8.3 Modern Comparative Methods

8.3.1 Phylogenetic Independent Contrasts

Felsenstein (1985) laid the foundation for statistical analysis of comparative data by proposing the method of phylogenetic independent contrasts (PIC), a method that will feature prominently in this paper because of its widespread use. Its popularity is shown by the fact that on August 8, 2007, the ISI reported 2,382 citations for the paper that presented it (Felsenstein 1985). This method works by making comparisons between sister-groups (the most closely related taxa on a tree, terminal taxa, or higher taxa, represented by nodes). Thus, for n terminal taxa, if the tree is fully resolved (dichotomous), *n*-1 contrasts can be taken (Fig. 8.2). These contrasts are based on the difference in character value between taxa because despite the phylogenetic relationships of taxa, differences in character value should be statistically independent, if measured between taxa, and if no path linking contrasted taxa overlaps another such path. Thus, the raw (unstandardized) contrast between taxa 1

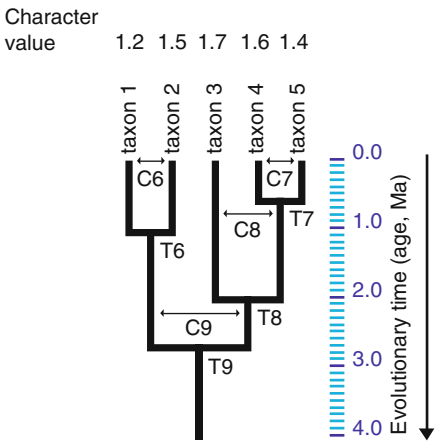
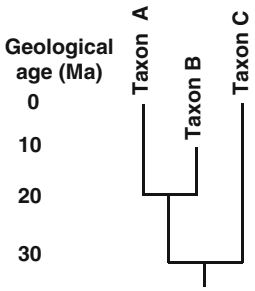


Fig. 8.2 Phylogenetic independent contrasts. Five terminal taxa (1–5) and four higher taxa (T6–9) showing the four phylogenetically independent contrasts that can be taken (C6–C9), each of which corresponding to a higher taxon

and 2 (C6) is $1.5 - 1.2 = 0.3$. However, contrasts should be standardized because they are expected to be greater between distantly related taxa (e.g., C8, between taxon 3 and T7) than between closely related ones (e.g., C7, between taxa 4 and 5), and PIC is based on parametric linear regressions. If characters evolve according to a Brownian motion model (the basic assumption of PIC), variance in the characters is expected to increase linearly with time. Thus, standardization is performed by dividing the raw contrasts by the square root of the path length involved in the contrast. This path length is the sum of lengths of branches connecting the contrasted taxa. For contrasts between terminal taxa, this length is easy to calculate; for instance, for C6, given that the contrasted taxa (1 and 2) are contemporary (extant) and that their last common ancestor dates from 1 Ma, each branch measures 1 Ma, and the path length is 2 Ma. Thus, the standardized contrasts C6 would be $0.3/2^{0.5}$. For contrasts involving higher taxa (e.g., C8, C9), the branches have to be lengthened because the nodal values (of higher taxa) are estimated, not measured; therefore, this introduces error and the variance is expected to be greater. We need not dwell further on how to compute the PIC; the above explanation should suffice for our purpose. Nevertheless, the adequacy of the standardization can be verified using various statistical tests, four of which are available in the PDAP module of Mesquite (Midford et al. 2008). It is important to check that the contrasts are adequately standardized, because otherwise, results will not be reliable. Inadequate standardization may have several causes: the characters may not have evolved according to a Brownian motion model (rather frequent), or errors may be present in the topology, branch lengths (both of which are rather common, if not the rule), or character value measurement. When character data are reliable (fairly common) and the when the phylogeny is considered also reliable (much less common), such tests yield information about the evolutionary model because they can suggest that the characters did not evolve according to a Brownian model. Some methods, such as phylogenetic regression (Grafen 1989) or PGLS (Martins and Hansen 1997; Pagel 1997), can actually yield more detailed data about the probable model of character evolution, if we assume no errors in data measurements and in the phylogeny.

8.3.2 Variance Partition with Phylogenetic Eigenvector Regression (PVR)

Another comparative method, variance partition with PVR, is based on an entirely different principle. It works on values of terminal taxa (not contrasts), but tries to control for phylogenetic effects by using a phylogenetic distance matrix (Fig. 8.3). That matrix simply shows the phylogenetic distances (sum of branch lengths on the path linking the compared taxa). It cannot be used directly; instead, a principal coordinate analysis (a technique related to principal component analysis) is performed to extract coordinates that reflect the distances between taxa. Given the structure of the tree, for n terminal taxa, n-1 axes are necessary to represent the position of all taxa without distortion, but not all these axes can be used because no



Phylogenetic distance matrix

Taxa	A	B	C
A	-	30	60
B		-	50
C			-

Fig. 8.3 Phylogentic tree, branch lengths, and the corresponding phylogenetic distance matrix (used in variance partition with PVR, among other comparative techniques)

degrees of freedom would be left to compute the statistical significance of regression coefficients. Thus, axes are selected, either using a broken stick model (Diniz-Filho et al. 1998), that selects the first few axes that explain more phylogenetic variance than expected by chance alone, or by regressing the coordinates of these axes against the dependent character, to determine which axes have a significant effect (Desdevises et al. 2003). After that, regressions allow determining the portion of the variance in the dependent character reflecting the independent characters, the phylogenetic signal, and the covariance between both (and some variance remains unexplained). The statistical significance of the effect of the independent characters and of the phylogenetic effects can also be tested.

Regressing the dependent character on the independent characters and on the selected principal coordinate axes representing the phylogeny allows estimating the total explained variance. Partial regressions are then used to establish the portion of variance explained only by the independent characters, only by the phylogeny (the statistical significance of both of these can be established), the portion explained by covariance between the independent characters and the phylogeny (whose statistical significance cannot be assessed), and the residual variance.

There is no need to delve further into the mathematics involved in variance partition with PVR because the purpose of this brief review is to show how all recent comparative methods for continuous (quantitative) characters require a phylogeny with estimated branch lengths.

8.3.3 The Use of Phylogenies in Evo-Devo

Phylogenies can also be useful to analyze other types of data (such as discrete data) or to assess other types of problems, such as heterochrony, rather than character

correlation. For instance, several recent studies have dealt with how to analyze developmental sequence data (typically based on relative or absolute time data on the position of various events in ontogeny of several species) to detect heterochronies. This problem is complex because there is no universal developmental time metric, and ontogenies can differ drastically between species, by the number of events that they include, by rather extensive differences in sequences, etc. (Jeffery et al. 2002, 2005). Thus, Smith (1997) developed a method called "event pairing" perfected in subsequent studies (e.g., Jeffery et al. 2002, 2005) to circumvent these problems and analyze developmental data on several species simultaneously. The method, initially developed to compare the craniofacial development of marsupials and placentals (Smith 1997), relies on coding the relative time (before, simultaneous, or after) between two events. All events are inserted into a C by C table (where C represents the individual events) and the table gives the relative order between the events listed in the various rows and those listed in the columns. A separate table is made for each taxon, and the data are subsequently treated to see the relative timing (the heterochronies). A full explanation of the method would require considerable developments that are beyond the scope of this chapter (see Smith 1997; Jeffery et al. 2002, 2005), but the point to remember is that analyzing developmental data to detect heterochronies using this method requires a topology (but no branch lengths).

An alternative method using both topology and branch lengths was recently proposed. That method, called the "continuous analysis," relies on squared-change parsimony to infer ancestral (nodal) values and PIC to calculate 95% confidence intervals (CIs) for these ancestral (nodal) values (Germain and Laurin 2009). The method consists in estimating the sequence position (or standardized time, if such data are available) of an event in a given ancestor along with the 95% CI on this value. Then, the observed or inferred sequence position (or standardized time) of the same event in the descendant is compared; if it lies outside the 95% CI of the ancestor, the heterochrony is statistically significant. This method was used to infer the ancestral cranial ossification sequence for urodeles with that of a potential sister-group (the Permo-Carboniferous branchiosaur *Apateon*). This method showed that contrary to previous claims (Schoch and Carroll 2003), *Apateon* was significantly different from the reconstructed ancestral urodele sequence (Germain and Laurin 2009). In any case, the shared similarities turn out to be mostly primitive, as shown by an event pairing analysis (Schoch 2006).

The relevance of the continuous analysis to this contribution is that like PIC and PVR, it uses branch length information whenever it is available, contrary to event-pairing analyses. It is thus not surprising that simulations show that the continuous analysis has a lower Type I error rate, and that it is more powerful (Germain and Laurin 2009).

8.3.4 Phylogenies and Paleontological Data in Paleogenomics

Given the increasing popularity of molecular biology, a brief illustration of how branch length data can contribute to genomics may be relevant. In addition to their

widespread use in studies on the evolution of gene expression patterns and of the genes themselves, phylogenies can be used to study genome size (and any other quantitative molecular character) evolution. Thus, Organ et al. (2011) recently took advantage of a correlation between genome size and osteocytic lacuna size to infer the size of genomes of early tetrapods and thus better constrain scenarios on genome size evolution. It has long been known that among extant tetrapods, urodeles have the largest genomes, and that birds have the smallest genomes. However, the polarity of change was difficult to assess from extant taxa alone and at least three main scenarios could explain the observed distribution: (1) the ancestral tetrapod genome was large, as in urodeles, and shrank to various extents in all taxa except for urodeles; (2) the ancestral tetrapod genome had a moderate size, as found in extant placental mammals, and it expanded in amphibians and shrank in birds; or (3) the ancestral tetrapod genome was small, as in birds, and it increased in all other taxa to various extents. Discriminating between the three scenarios with data from extant taxa alone is very difficult because evolutionary trends usually require temporally spread data, as shown by simulations (Laurin 2010a). Thus, the finding that all studied early tetrapods (some amphibians and amniotes) had mid-sized genomes like extant mammals shows that the second scenario is the correct one (Fig. 8.4). This study incorporated a time-calibrated tree at various steps of the analysis, namely, in the assessment of the correlation between genome size and osteocyte lacuna size in extant taxa in which both are known, and in the inference on the evolution of genome size in extant and extinct taxa, using a Bayesian method (Organ et al. 2011).

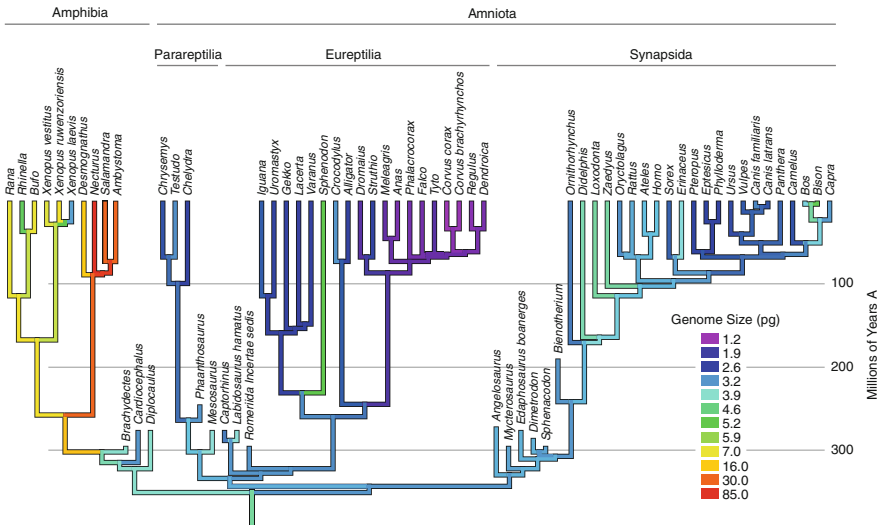


Fig. 8.4 Evolution of genome size in tetrapods based on observed values in extant taxa, and inferred values for extinct taxa based on osteocyte lacuna size. (Reproduced from Organ et al. 2011)

This survey illustrates the usefulness of time-calibrated trees and paleontological data in a wide array of fields in comparative biology. Similar examples could have been taken from evolutionary physiology (Careau et al. 2007), functional morphology (Pouydebat et al. 2008), ecology (Canoville and Laurin 2010), or conservation biology (Faith 1992).

8.3.5 Phylogenies in Paleobiology

Most paleobiological studies have exploited phylogenetic data little, if at all, until recently. This is the case of most studies based on an observed correlation between bone microanatomy and lifestyle (aquatic to terrestrial) to infer the lifestyle of various extinct tetrapods, as was done on extant and extinct snakes (de Buffrénil and Rage 1993). Even the latest studies in this field (e.g., Canoville and Laurin 2010) use the phylogeny only to assess the relationship between bone microanatomy and lifestyle and to build general paleobiological inference models. Thus, the linear discriminant models used by Canoville and Laurin (2010) assess the probability that an extinct taxon was aquatic, amphibious, or terrestrial by using the distribution of quantitative long bone microanatomical characters of taxa of known lifestyle. A graphical representation can also be obtained and shows the relative position of taxa of known lifestyles (along with polygons that encompass all taxa of each given lifestyle) and of the extinct taxa of unknown lifestyle (Fig. 8.5). Ironically, in this particular case, all extinct taxa of unknown lifestyle fit outside the polygons representing the distribution of extant taxa, so the inferences must be viewed with caution, but they can nevertheless be made based on the distance to the centroid and the variance. Thus, the early Permian amniote *Mesosaurus* and Triassic diapsid *Neusticosaurus* were certainly aquatic.

More inference methods are available for quantitative characters, and some of these use classical, well-known statistical methods. Thus, Pouydebat et al. (2008) used multiple linear regression models to infer the grasping behavior of three extinct primate taxa (the Pliocene hominid *Australopithecus afarensis* and the Miocene hominoids *Oreopithecus bambolii* and *Proconsul africanus*) based on the hand proportion, which is itself correlated with frequency of use of various such behaviors. Again, these inferences did not take into consideration the systematic position of these taxa. However, it is possible to incorporate the phylogeny into such inferences, either using the Bayesian method used, among others, by Organ et al. (2011), namely, the program BayesTraits (Pagel and Meade 2006), or by using a program (PhyloPars) originally designed to estimate missing values in comparative datasets (Bruggeman et al. 2009). Both of these methods use character correlation and the systematic position of extinct taxa to make their inferences, which should result in more reliable estimates. Unfortunately, the interface of BayesTraits is not very user-friendly, and PhyloPars works only with quantitative characters.

The phylogeny is often used in paleobiology to infer character history, often based on extant and/or extinct taxa. When extinct taxa are included, the character

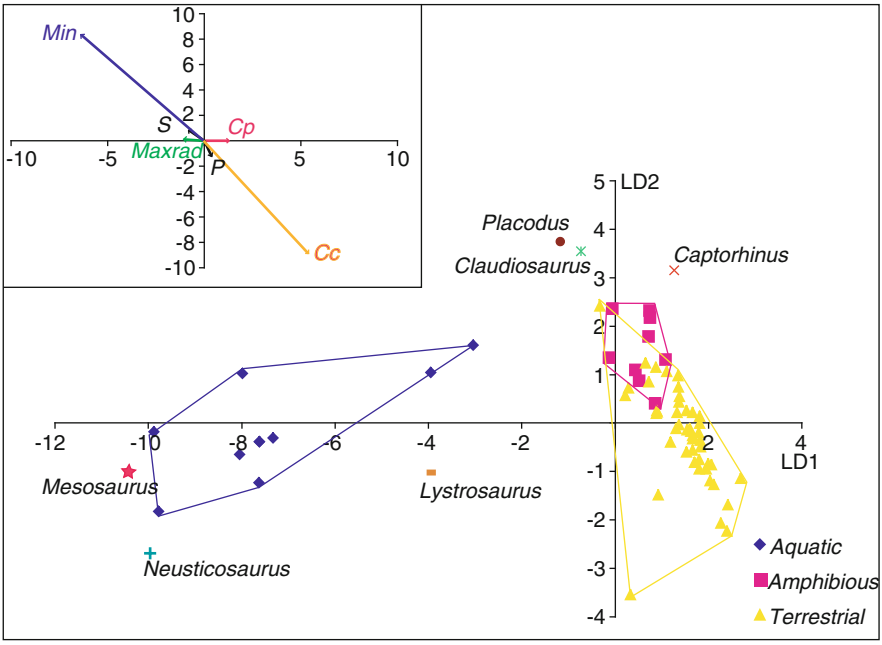


Fig. 8.5 Distribution of taxa of known (in the *polygons*) and of unknown lifestyle, according to their bone microanatomical and body size characters. (Reproduced from Canoville and Laurin 2010)

states of terminal taxa are often inferred before the optimization is carried out. Thus, Canoville and Laurin (2010) used parsimony optimization of the inferred or observed lifestyle of 28 terminal taxa to reconstruct the history of the conquest of land by vertebrates. However, as expected, some parts of the tree are ambiguous. Thus, parsimony indicates that the first amniote was probably amphibious or terrestrial (Fig. 8.6), a conclusion that gives some support to Romer's (1958) suggestion that the first amniotes retained the amphibious lifestyle of their distant ancestors. However, it is easy to use the phylogeny to better investigate this question. Canoville and Laurin (2010) suggested that in this case, the value of the quantitative characters that are used in the linear discriminant inference models can be inferred on the node of interest (here Amniota) and that these inferences could be used to infer the lifestyle. Confidence intervals on all the quantitative characters can also be computed using PIC, allowing a sensitivity analysis of the lifestyle inference. This analysis suggests that the first amniote was amphibious (Canoville and Laurin 2010: supplementary online material 9). Of course, Bayesian methods would allow for a more complete and more rigorous incorporation of various sources of uncertainty into this analysis, but remain impractical because of the software limitations evoked above. Nevertheless, the various phylogeny-informed methods recently developed open exciting perspectives in paleobiology.

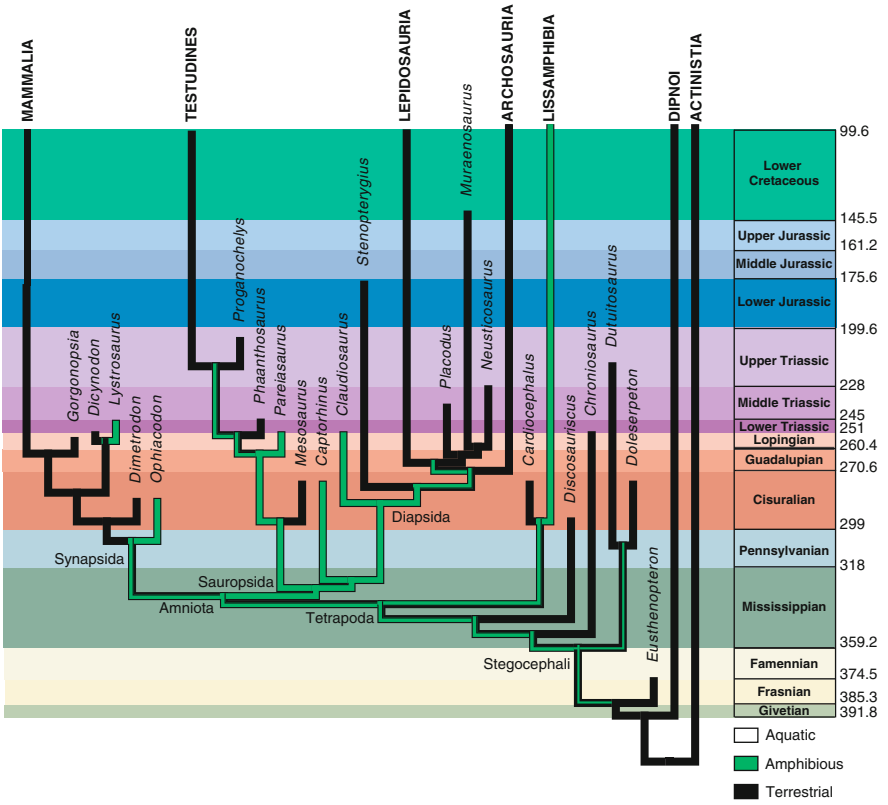


Fig. 8.6 Optimization of the lifestyle (considered as an ordered character) on a phylogeny including extinct taxa whose lifestyle was inferred using microanatomical data and various inference models (especially, linear discriminant models). (Reproduced from Canoville and Laurin 2010)

8.4 Branch Lengths in Comparative Methods

As mentioned above, a frequent problem when applying PIC is that contrasts are not adequately standardized. Data transformation may solve this problem, but when it does not, transforming the branch lengths is the next logical step. Unfortunately, many methods of branch length transformations are not particularly biologically meaningful, and obscure (or discard) the relationship between branch lengths and time. For instance, an exponential or natural log transformation makes subsequent calculation of evolutionary rates difficult. Setting all branches of equal lengths, often done to save time (actually used mostly when no attempt was made to collect branch length data), precludes any meaningful calculation of evolutionary rate. Other methods are more useful and sophisticated, such as Grafen's (1989) rho transform, which consists in a power transformation of branch lengths that distorts

the tree to change the relative lengths of terminal and more basal branches. This is useful to adjust the analysis to reflect the amount of phylogenetic signal present in the data; if it is large, the internal branches are rather long; if it is small, internal branches are short, and terminal branches are long. However, when the rho transform needs to be used, the investigator will normally conclude that the analyzed characters did not evolve according to a Brownian motion model. This may, in some cases, represent overinterpretation of the results (see below), especially considering that when Grafen (1989) proposed his method, branch length data were usually unavailable, so he proposed his method to adapt "artificial" branch lengths to better fit the data.

All branch length transformation methods discussed above assume that if divergence time data were used to build the tree, the resulting branch lengths were more or less correct. This may not be so, and it is conceivable that in at least some cases, it is precisely branch lengths that cause the lack of adequate standardization of contrasts because they are wrong (and the characters really evolved according to a Brownian motion model). Two new simple branch length transformation methods were developed by Josse et al. (2006) to facilitate paleontological tree construction (Marjanović and Laurin 2007), and they can be used to check if slight modifications of the initial lengths yield adequate standardization. A few empirical tests on bone microanatomical data demonstrate that in many cases, adequate PIC standardization can be obtained by thus manipulating initial branch lengths (Laurin et al. 2009; Canoville and Laurin 2010). The resulting lengths may remain plausible estimates of evolutionary time (because the exact length of each branch is usually only moderately well-constrained), which allows determination of absolute evolutionary rates of characters. This is essential if evolutionary rates of characters need to be compared between studies in which the taxonomic sample overlaps partly, if at all. This method can test, to an extent, the hypothesis that the characters have evolved according to a Brownian motion model, even if the initial branch lengths were slightly inaccurate. Given that the Brownian model is the simplest model of character evolution (Martins et al. 2002), it should not be discarded in favor of more complex models needlessly.

8.5 Getting Branch Length Data

The last section assumes that branch length data reflecting evolutionary time can be obtained. This is becoming increasingly frequent, but only a minority of comparative studies incorporates such lengths because they are time-consuming to collect, as my recent experience as an editorial board member of the *Journal of Evolutionary Biology* (2008–current) has made me realize. Paleontological data can be used to estimate minimal and (with less precision) maximal divergence dates, but such data are typically scattered in the literature, although a few recent compilations (Benton 1993; Benton and Donoghue 2007; Marjanović and Laurin 2007) ease this process. Molecular divergence dates are becoming increasingly common and their

reliability is improving as more taxa and genes are being sequenced, and as the analytical methods become more sophisticated (Sanderson 1997, 2002; Thorne and Kishino 2002). Contrary to the earliest attempts at molecular dating that assumed a single evolutionary rate over the whole tree (Zuckerkandl and Pauling 1965), all modern molecular dating methods rely on a "relaxed molecular clock" that allows for each branch to have its own rate. Recent compilations of molecular timetrees (Hedges and Kumar 2009) should prove invaluable for comparative biologists, especially for taxa in which the fossil record is patchy or inexistent (as in taxa that lack mineralized body parts, such as nematodes, or those in which the morphology is not very informative, as in many eubacteria).

However, combining paleontological and molecular ages is not straightforward. The fossil record provides mostly minimal divergence dates that usually underestimate the actual divergence dates by an unknown amount, whereas molecular ages represent attempts at dating the actual divergences. For reasons explained below, molecular ages are usually older (sometimes much older) than paleontological dates and may often tend to be overestimated, for at least two reasons. First, a methodological factor tends to inflate such ages because molecular divergence age is an "asymetrically bounded random variate" (Rodríguez-Trelles et al. 2002), and random variations around the actual age scale divisively forward (to the present) but multiplicatively backward (to the past). Therefore, the arithmetic means of such age estimates (which are used by molecular dating software to estimate the actual divergence date from several genes and/or portions of the evolutionary tree) are upwardly biased. Second, most molecular phylogeneticists tend to enforce a single (or very few) maximal age constraint in a tree, but several minimal age constraints (e.g., Roelants et al. 2007). This also creates an upward bias, as recently shown empirically (Marjanović and Laurin 2007). This second factor is especially important because the greatest part in the variance in molecular age estimates appears to result from calibration choice, rather than the algorithm used to analyze the molecular sequence data (Marjanović and Laurin 2007).

Thus, it may not be appropriate to simply mix paleontological and molecular divergence dates into a timetree because they appear often fundamentally different. For this reason, Laurin et al. (2009) suggested to use the lower bound (minimal age) of the 95% CI of molecular ages along with paleontological ages to compile an initial timetree. These ages are entered as real (paleontological data) or virtual (molecular age data) taxa into the tree. In most cases, this timetree will underestimate true ages, but if it fails to adequately standardize PICs, the stratigraphic tools (Josse et al. 2006) can be used to lengthen the branches while keeping all time constraints at their minimal age. The result is to push back in time some or all nodes. This procedure can be repeated a few times to check if such transformations adequately standardize PICs while retaining plausible branch lengths. Given that the uncertainty on actual divergence dates is substantial (the 95% CI often represents the data $\pm$ 10–50%), there is usually ample room to perform such branch-length manipulations. Thus, in the case of lissamphibians, the initial time-calibrated tree compiled by Laurin et al. (2009) implied a divergence data between urodeles and anurans near the Permo/Triassic boundary (251 Ma), which is barely

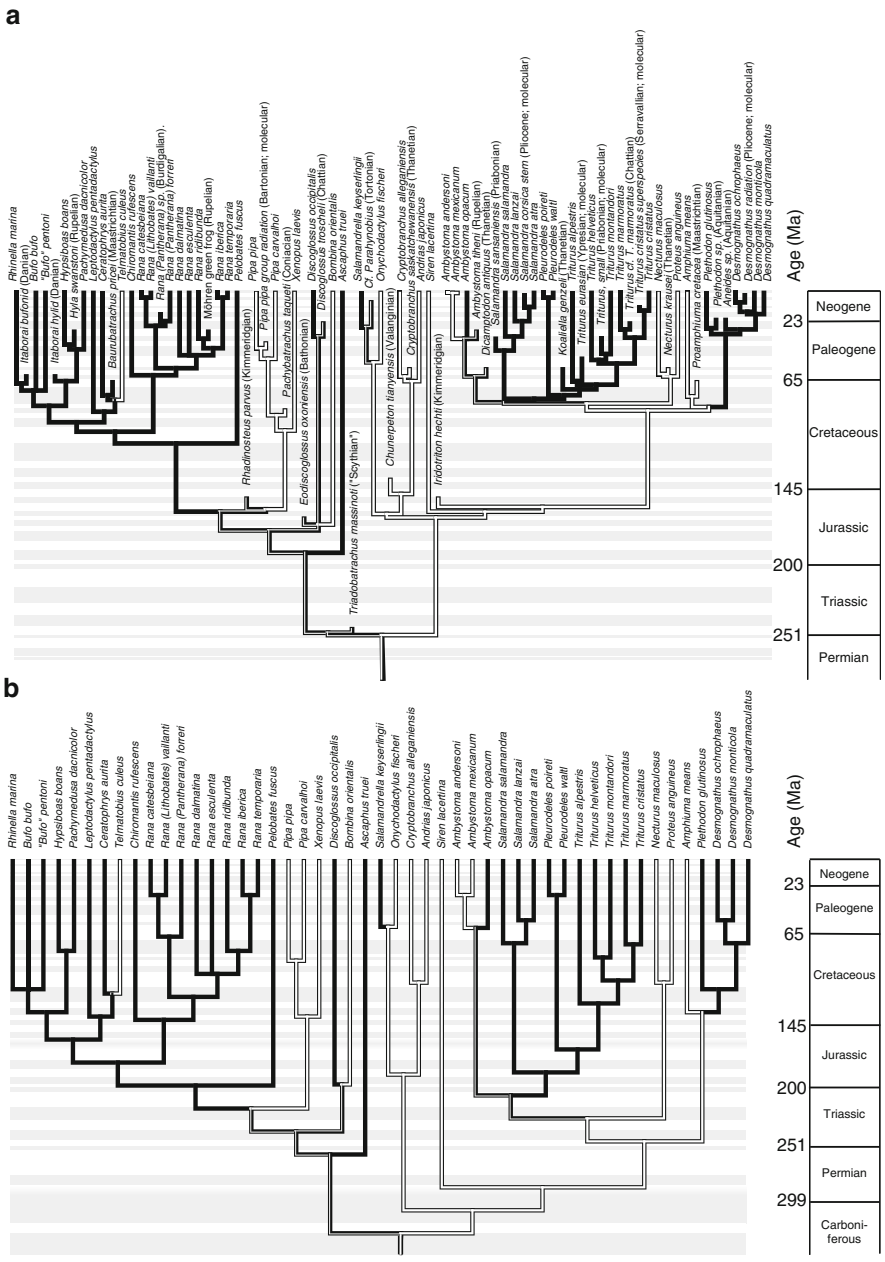


Fig. 8.7 Initial time-calibrated tree of lissamphibians (**a**) incorporating minimal divergence dates (lower bounds of 95% CIs on molecular dates or oldest fossil of each clade), along with the transformed tree (**b**) that adequately standardized most PICs for bone microanatomical and body size data analyzed by Laurin et al. (2009). Branches in white denote aquatic taxa; branches in black denote amphibious or terrestrial taxa

older than *Triadobatrachus,* the oldest known lissamphibian (Fig. 8.7a). This tree failed to adequately standardize most PICs, but a transformed tree (Fig. 8.7b) implying a divergence between anurans and urodeles in the Carboniferous (325 Ma) adequately standardized most PICs. This tree thus implies much greater ages of most nodes, but it remains biologically plausible (branch lengths may reflect evolutionary time) to the extent that most molecular dating studies propose a Carboniferous (or even Devonian) age for this divergence (e.g., San Mauro et al. 2005; Zhang et al. 2005).

The method outlined above eases somewhat the burden of compiling divergence time data because it allows paleontological and molecular data to be combined in a coherent way. Furthermore, given the branch manipulation methods implemented in the stratigraphic tools (Josse et al. 2006), if no data are available for some nodes, some minimal branch lengths can be inserted and lengthened (along with all other branches in the tree) to achieve adequate PIC standardization, thus minimizing the problem created by missing data (when neither molecular ages nor fossil data can be used to date a node, a common situation).

This method could clearly be pushed further. Developing an automated method to adjust minimal branch lengths using the algorithms implemented in the stratigraphic tools to obtain the shortest tree that adequately standardizes the PICs of a given dataset would be very useful, as the procedure of testing various settings and the resulting standardization on all characters of a dataset can be time-consuming. This has not been performed so far because of lack of time, but it could have widespread applications in comparative biology. Soon, most comparative biologists may be able to use time-calibrated trees in their analyses, rather than using trees with arbitrary branch lengths. This change will be facilitated by the growing number of molecular dating studies (e.g., San Mauro et al. 2005; Zhang et al. 2005; Hedges and Kumar 2009), although the meager level of funding of paleontological research, required for these fields to progress (especially, molecular dating), will be the limiting factor.

Acknowledgments I thank Pierre Pontarotti for inviting me to participate in this symposium and in this volume, and for his patience while waiting for this draft. Eli Amson provided comments that improved the draft.

References

Benton MJ (ed) (1993) Fossil record 2. Chapman & Hall, London

Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24:26–53

Bruggeman J, Heringa J, Brand BW (2009) PhyloPars: estimation of missing parameter values using phylogeny. Nucleic Acids Res 37:W179–W184

Butler RJ, Goswami A (2008) Body size evolution in Mesozoic birds: little evidence for Cope's rule. J Evol Biol 21:1673–1682

Canoville A, Laurin M (2010) Evolution of humeral microanatomy and lifestyle in amniotes, and some comments on paleobiological inferences. Biol J Linn Soc 100:384–406

Careau V, Morand-Ferron J, Thomas D (2007) Basal metabolic rate of canidae from hot deserts to cold arctic climates. J Mammal 88:394–400
Cubo J, Ponton F, Laurin M, de Margerie E, Castanet J (2005) Phylogenetic signal in bone microstructure of sauropsids. Syst Biol 54:562–574
Cubo J, Legendre P, de Ricqlès A, Montes L, de Margerie E, Castanet J, Desdevises Y (2008) Phylogenetic, functional, and structural components of variation in bone growth rate of amniotes. Evol Dev 10:217–227
Davis MC, Dahn RD, Shubin NH (2007) An autopodial-like pattern of Hox expression in the fins of a basal actinopterygian fish. Nature 444:473–477
de Buffrénil V, Rage J-C (1993) La "pachyostose" vertébrale de *Simoliophis* (Reptilia, Squamata): données comparatives et considérations fonctionnelles. Ann Paléontol 79(4):315–335
Desdevises Y, Legendre P, Azouzi L, Morand S (2003) Quantifying phylogenetically structured environmental variation. Evolution 57:2467–2652
Diniz-Filho JAF, de Sant'Ana CER, Bini LM (1998) An eigenvector method for estimating phylogenetic inertia. Evolution 52:1247–1262
Faith DP (1992) Conservation evaluation and phylogenetic diversity. Biol Conserv 61:1–10
Felsenstein J (1985) Phylogenies and the comparative method. Am Nat 125:1–15
Freckleton RP, Harvey PH, Pagel M (2002) Phylogenetics analysis and comparative data: a test and review of evidence. Am Nat 160:712–726
Germain D, Laurin M (2009) Evolution of ossification sequences in salamanders and urodele origins assessed through event-pairing and new methods. Evol Dev 11:170–190
Gittleman JL, Kot M (1990) Adaptation: statistics and a null model for estimating phylogenetic effects. Syst Zool 39:227–241
Grafen A (1989) The phylogenetic regression. Philos Trans R Soc Lond B 326:119–157
Hedges SB, Kumar S (eds) (2009) The timetree of life. Oxford University Press, New York
Hennig W (1965) Phylogenetic systematics. Annu Rev Entomol 10:97–116
Hone DWE, Dyke GJ, Haden M, Benton MJ (2008) Body size evolution in Mesozoic birds. J Evol Biol 21:618–624
Jeffery JE, Richardson MK, Coates MI, Bininda-Emonds ORP (2002) Analyzing developmental sequences within a phylogenetic framework. Syst Biol 51:478–491
Jeffery JE, Bininda-Emonds ORP, Coates MI, Richardson MK (2005) A new technique for identifying sequence heterochrony. Syst Biol 54:230–240
Josse S, Moreau T, Laurin M (2006) Stratigraphic tools for Mesquite. http://mesquiteproject.org/packages/stratigraphicTools/
Lamarck JB (1809) Philosophie zoologique. Flammarion, Paris
Laurin M (2004) The evolution of body size, Cope's rule and the origin of amniotes. Syst Biol 53:594–622
Laurin M (2010a) Assessment of the relative merits of a few methods to detect evolutionary trends. Syst Biol 59:689–704
Laurin M (2010b) How vertebrates left the water. University of California Press, Berkeley
Laurin M, Girondot M, de Ricqlès A (2000) Early tetrapod evolution. Trends Ecol Evol 15: 118–123
Laurin M, Canoville A, Quilhac A (2009) Use of paleontological and molecular data in supertrees for comparative studies: the example of lissamphibian femoral microanatomy. J Anat 215:110–123
Marjanović D, Laurin M (2007) Fossils, molecules, divergence times, and the origin of lissamphibians. Syst Biol 56:369–388
Martins EP, Hansen TF (1997) Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. Am Nat 149: 646–667
Martins EP, Diniz-Filho JAF, Housworth EA (2002) Adaptative constraints and the phylogenetic comparative method: a computer simulation test. Evolution 56:1–13
Mayr E (1982) The growth of biological thought—diversity, evolution, and inheritance. The Belknap Press of Harvard University Press, Cambridge

McShea DW (1996) Metazoan complexity and evolution: Is there a trend? Evolution 50:477–492
Midford P, Garland TJ, Maddison WP (2008) PDAP package for Mesquite. http://mesquiteproject.org/pdap_mesquite/index.html
Organ CL, Canoville A, Reisz RR, Laurin M (2011) Paleogenomic data suggest mammal-like genome size in the ancestral amniote and derived large genome size in amphibians. J Evol Biol 24:372–380
Pagel M (1997) Inferring evolutionary processes from phylogenies. Zoolog Scr 26:331–348
Pagel M, Meade A (2006) BayesTraits. http://www.evolution.rdg.ac.uk/BayesTraits.html
Pouydebat E, Laurin M, Gorce P, Bels V (2008) Evolution of grasping among anthropoids. J Evol Biol 21:1732–1743
Purvis A, Gittleman JL, Luh H-K (1994) Truth or consequences: effects of phylogenetic accuracy on two comparative methods. J Theor Biol 167:293–300
Rodríguez-Trelles F, Tarrío R, Ayala FJ (2002) A methodological bias toward overestimation of molecular evolutionary time scales. Proc Natl Acad Sci USA 99:8112–8115
Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju SD, Guillaume K, Moriau L, Bossuyt F (2007) Global patterns of diversification in the history of modern amphibians. Proc Natl Acad Sci USA 104:887–892
Romer AS (1958) Tetrapod limbs and early tetrapod life. Evolution 12:365–369
San Mauro D, Vences M, Alcobendas M, Zardoya R, Meyer A (2005) Initial diversification of living amphibians predated the breakup of Pangaea. Am Nat 165:590–599
Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol Biol Evol 14:1218–1231
Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol 19:101–109
Schoch RR (2006) Skull ontogeny: developmental patterns of fishes conserved across major tetrapod clades. Evol Dev 8:524–536
Schoch RR, Carroll RL (2003) Ontogenetic evidence for the Paleozoic ancestry of salamanders. Evol Dev 5:314–324
Smith KK (1997) Comparative patterns of craniofacial development in eutherian and metatherian mammals. Evolution 51:1663–1678
Sordino P, van der Hoeven F, Duboule D (1995) Hox gene expression in teleost fins and the origin of vertebrate digits. Nature 375:678–681
Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689–702
Zhang P, Zhou H, Chen Y-Q, Liu Y-F, Qu L-H (2005) Mitogenomic perspectives on the origin and phylogeny of living amphibians. Syst Biol 54:391–400
Zuckerkandl E, Pauling L (1965) Molecules as documents of evolutionary history. J Theor Biol 8:357–366