Amer. Malac. Bull. **27**: 25-45 (2009)

As time goes by: A simple fool's guide to molecular clock approaches in invertebrates*

Thomas Wilke, Roland Schultheiß, and Christian Albrecht

Department of Animal Ecology and Systematics, Justus Liebig University Giessen, Heinrich-Buff-Ring 26-32 IFZ, D-35392 Giessen, Germany

Corresponding author: tom.wilke@allzool.bio.uni-giessen.de

Abstract: Biologists have used a wide range of organisms to study the origin of taxa and their subsequent evolutionary change in space and time. One commonly used tool is the molecular clock approach, relating substitution rates of nucleotide or amino acid sequences to divergence times. The accuracy of the molecular clock, however, has long been subject to controversy, and numerous papers have addressed problems associated with estimating divergence times. Some workers pointed out a striking imbalance between sophisticated software algorithms used for molecular clock analyses on the one hand, and the poor data on the other hand. Moreover, there is often unease among workers relative to molecular clocks because of the controversy surrounding the approach, the complex mathematical background of many molecular clock tools, the still limited number of available, user-friendly software packages, the often confusing terminology of molecular clock approaches, and the general lack of reliable calibration points and/or external clock rates. The current review therefore briefly provides an overview of analytical strategies, covering approaches based on calibration points and/or bounds, approaches based on external clock rates, and approaches that attempt to estimate relative divergence times in the absence of information that can be used for estimating substitution rates. It also deals with major problems and pitfalls associated with data and analyses, including potential errors of calibration points and bounds, the performance of the gene(s) used, estimation of confidence limits, and misinterpretation of the results of clock analyses due to problems with sampling design. A substantial part of the review addresses the question of "universal" molecular clock rates and summarizes important biological and life history variables that account for deviations from rate constancy both between lineages and at different times within lineages. The usefulness of these factors is discussed within the framework of "trait-specific" molecular clock rates. One such clock rate is introduced here for the cytochrome *c* oxidase subunit I (COI) gene in small dioecious, tropical and subtropical Protostomia with a generation time of approximately one year. A flow chart is provided as a "simple fool's guide" to molecular clock analyses, together with a glossary of widely used terms in molecular clock approaches. Finally, step-by-step examples are provided for calculating divergence times in the caenogastropod subfamily Pyrgulinae based on both an internal calibration point and a "trait-specific" molecular clock rate, and it is demonstrated how a relative clock approach can be used for testing evolutionary hypotheses. Our review encourages a judicious use of molecular clock analyses in evolutionary studies of invertebrates by demonstrating their great potential on the one hand and (often-manageable) problems and pitfalls on the other hand.

Key words: molecular clocks, calibrating, Protostomia, Pyrgulinae

Over several decades, evolutionary biologists have used a wide range of organisms to study the origin of taxa from a common ancestor and their subsequent change and diversification in space and time. One commonly used tool is the molecular clock approach, relating number of fixed mutations (= substitutions) in nucleotide or amino acid sequences to divergence time of taxa.

The introduction of the molecular clock concept is attributed to Zuckerkandl and Pauling (1962), who found amino acid differences in mammalian α and β chains of hemoglobin to be roughly proportional to divergence times inferred from paleontological data. In 1965, these workers published a landmark paper (Zuckerkandl and Pauling 1965), naming the molecular clock and describing its stochastic nature as a Poisson process. It was also suggested that, if a molecular clock exists, amino acid changes must be limited almost exclusively to functionally nearly neutral changes—supporting the concept of near neutrality at the molecular level (Takahata 2007). In subsequent years, several workers attributed spontaneous mutations due to replication errors as a driving force of molecular evolution and suggested evolutionarily "neutral" changes in sequences be used to measure divergence times (*e.g.*, Sarich and Wilson 1967, Kimura 1968, Wilson and Sarich 1969).

The accuracy of the molecular clock, however, has long been subject to controversy and early on, workers noted that different proteins evolve at different rates. Over the years, numerous papers addressed further problems with molecular clock approaches like rate heterogeneity in different taxonomic groups or body size effects (see Takahata 2007 and references therein, also see below). Unfortunately, the controversy surrounding the molecular clock approach has

* From the symposium "Molluscs as models in evolutionary biology: from local speciation to global radiation" presented at the World Congress of Malacology, held from 15 to 20 July 2007 in Antwerp, Belgium.

divided the scientific community. Although an increasing number of workers perform judicious molecular clock analyses, there are still many "believers" who almost uncritically use statistical tools to estimate divergence times, often resulting in unrealistic estimates; "disbelievers" who reject molecular clock approaches as inappropriate; and "ignorant" workers who refrain from using molecular clock approaches due to the critics (right or not) of the method. No matter to what group people belong, there often is a feeling of unease among workers relative to molecular clock approaches. This is due to the controversy of the molecular clock approach (see above), the complex mathematical background of many molecular clock tools, often making it difficult for biologists to understand the statistics involved, the still limited number of user-friendly software packages, the often confusing terminology used in molecular clock approaches, and a general lack of reliable calibration points and/or rates.

As pointed out by Takahata (2007: 4) "It is now generally accepted that, although it is uncertain and rejected for a substantial proportion of proteins and genomic regions in comparisons of main taxonomic groups, the molecular clock can put a new timescale on the history of life, thereby allowing exploration of the mechanisms and processes of organismal evolution. Similarly, a molecular clock is an irreplaceable source of information in evolutionary biology and it would be foolish to abandon it altogether". He also states that the molecular clock needs not be exact and that an approximate clock can still be very useful. However, it often is exactly this approximation of the molecular clock (*i.e.,* estimations of meaningful errors) that is difficult to conduct (*e.g.*, Ayala 1997).

Moreover, there frequently is a striking imbalance between sophisticated algorithms used for molecular clock analyses and poor data (*e.g.*, Bandelt 2007). In fact, whereas many studies deal with optimizing the performance of molecular clock tools, there are relatively few publications addressing data and model selection. This is not trivial, as problems with data and/or misinterpretation of the results can account for divergences of molecular clock estimates in one and the same taxon by >1000% (*e.g.*, Wilke 2004, Pulquerio and Nichols 2007).

In the present review, we therefore attempt to: (1) give a brief overview of molecular clock approaches, (2) discuss major problems and pitfalls associated with data and analyses, (3) address the question of universal and trait-specific molecular clocks in invertebrates, (4) provide a conservative "simple fool's guide" to molecular clock analyses, (5) provide a step-by-step example for clock estimations in the caenogastropod subfamily Pyrgulinae based on both an internal calibration point and a trait-specific molecular clock rate, and (6) give definitions of some of the most widely used terms in molecular clock approaches in a glossary.

This review is intended neither to provide a full statistical background of the molecular clock hypothesis nor to cover all relevant methods and developments. Instead, we give basic information on principal molecular clock strategies, on approaches for mitigating problems commonly associated with molecular clock analyses, and on major pitfalls in molecular clock estimations.

Thus, we specifically target the "ignorant" people mentioned above, hoping to persuade them to look into the application of molecular dating. At the same time, we hope to make "believers" aware of major pitfalls in molecular clock approaches and to convince "disbelievers" that under a specific set of circumstances, the molecular clock can be a very useful tool for evolutionary analyses.

THE MOLECULAR CLOCK APPROACH

Molecular-dating methods, the estimation of divergence times of lineages from a common ancestor based on nucleotide or amino-acid sequences, can be broadly classified into population genetic and phylogenetic (*i.e.,* molecular clock) approaches. In population genetic approaches, a coalescent framework is used to estimate the 'age' of a most recent common ancestor (MRCA) of a number of alleles. The age of the MRCA is hereby measured in number of generations. This approach works backwards in time and is based on the assumption that a pair of alleles will coalesce, *i.e.*, find their MRCA, at some point in time in the past (see Edwards and Beerli 2000). Several models were developed to describe this process with respect to various parameters such as effective population size, gene flow, and changes in population size over time. These population genetic approaches are, as the name suggests, typically only applicable to estimation of divergence time within a species.

For estimating divergence times between species or between groups of species, several phylogenetic approaches have been suggested. Whereas in early studies, genetic distance matrices were used to estimate substitution rates for molecular clock estimations (*e.g.*, Nei 1987, Li and Graur 1991), today these substitution rates are typically derived from phylogenies or from sequence data in conjunction with tree topologies (Rutschmann 2006).

Given the scope of our review, we focus on these tree-based approaches that now appear to be the most widely accepted molecular-clock methodologies in phylogenetic studies. Tree-based molecular clock approaches typically use the branching topology of a phylogeny together with branch length information to estimate the node depth (d_N in number of substitutions per site). Together with a substitution (=

"molecular clock") rate (λ in number of substitutions per site and year), divergence time (t_D in years) can be calculated as following:

$$t_D = \frac{d_N}{\lambda}.$$

Some workers, however, use genetic distances not in terms of node depth but in terms of total branch length between two taxa for estimating divergence times. As total branch length equals node depth × 2, substitution rates have to be modified accordingly and in such cases, typically divergence rates (divergence rate = substitution rate × 2) are used as molecular clock rates.

Therefore, it is crucial to state whether molecular clock rates are based on substitution or divergence rates. Wilson and Sarich's (1969) controversial universal molecular clock rate for protein-coding mitochondrial genes of 2% My^{-1}, for example, is based on divergence rates and corresponds to a substitution rate of 1% My^{-1}. This ambiguity results from the fact that in early molecular clock approaches, workers used genetic distance matrices to calculate molecular clock rates. Today, most workers use tree-based approaches where divergence times are best estimated from node depths and known substitution rates. Thus, in this review all molecular clock rates are substitution rates, unless stated otherwise.

In order to calculate divergence times from node depths, substitution rate(s) have to be estimated from externally derived dates. This can be done with information from ancestral DNA, fossils, or biogeographical events (Bromham and Penny 2003). Alternatively, external molecular clock rates (see below) could be used to estimate divergence times (Fig. 1).

A critical point in molecular clock approaches is the question of rate constancy. Originally, it was believed that mutation rates, particularly in protein-coding genes, are largely constant across loci, species, and time (see introduction). This is also reflected in the term "molecular clock", which, in a strict sense, is associated with rate constancy. Several studies, however, indicated that in many cases the assumption of rate constancy is violated (*e.g.*, Arbogast *et al.* 2002, Pulquerio and Nichols 2007).

This has proved to be a major obstacle for molecular dating as the estimation of divergence times from branch length *per se* requires rate constancies. Whereas this is particularly important for molecular clock approaches employing external clock rates (Fig. 1C), there are several attempts to deal with problems of rate heterogeneity in trees utilizing calibration points or bounds (Figs. 1A-B). These statistical approaches focus on estimating rate variations across lineages ("relaxed molecular clock") in order to obtain more realistic divergence times. Whereas a strict molecular clock tree (Fig. 2A) assumes a single rate for all lineages in the tree, relaxed clocks assume different rates for different branches (Fig. 2C). As a special case of the relaxed clock, lineages within a clade may share the same evolutionary rates ("local clocks"). Note that from a formal standpoint, the strict clock also is a special case of the relaxed clock with the number of rates being one. However, for simplicity, we here only refer to relaxed clocks when the strict clock model is rejected. In the following section of this review, we will discuss the problem of rate heterogeneity in more detail together with other major problems that might affect molecular clock analyses.

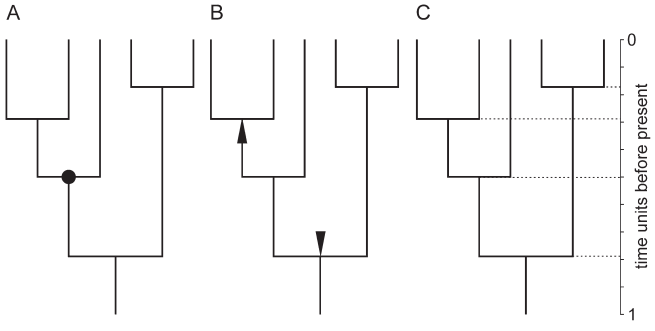

Figure 1. Methods for calibrating molecular clock trees. A, Calibration with point(s) from externally derived dates. B, Calibration with bounds from externally derived dates. C, Calibration with an external molecular clock rate.

PROBLEMS ASSOCIATED WITH MOLECULAR CLOCK ESTIMATIONS

Rate heterogeneity

Testing the global molecular clock

The most fundamental assumption for a global molecular clock is that mutations occur at a single rate along all branches

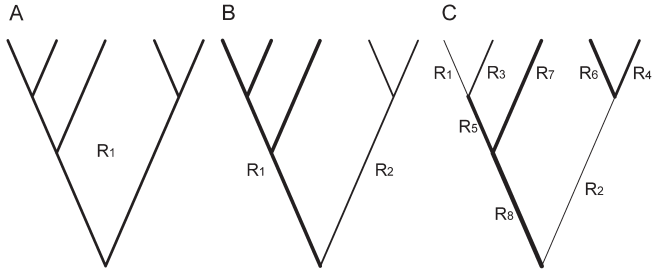

Figure 2. Sample phylogenies with different clock models based on different degrees of rate heterogeneity. A, Strict molecular clock tree with a single substitution rate (R_1). B, Molecular clock tree with two different substitution rates ("local clocks"; R_1, R_2). C, Relaxed clock tree with different substitution rates for different branches (R_1-R_8).

of a phylogenetic tree. This strict clock-like behavior, however, is rare and rates tend to vary. The question arises whether actual deviations are severe enough to bias dating approaches significantly. Several methods have been developed to test clock-like behavior in a data set. In the following paragraphs, we consider two of these approaches.

The likelihood ratio test (LRT) is a widely used approach for testing the acceptance of a global molecular clock. It is calculated as follows: $LR = 2(\ln L_1 - \ln L_0)$, where $\ln L_0$ is the maximized log likelihood of the null hypothesis (*i.e.,* the clock-like tree) and $\ln L_1$ of the alternative hypothesis (*i.e.*, the non clock-like tree). Under the assumption that the L_0 model is nested, *i.e.*, that it is a special case of the L_1 model, the LRT follows approximately a chi-squared distribution with q degrees of freedom (*df*), whereby $q = n\text{-}2$ with n being the number of taxa in the phylogeny (Felsenstein 1988). If the calculated ratio exceeds the critical value of the chi-squared distribution, the difference between the two models is considered to be significant and the global molecular clock is rejected. It has to be noted, however, that the chi-squared approximation applies only if the sample size is large enough (Posada and Buckley 2004). Under a given number of *df*, the critical value of the chi-squared distribution is a constant figure. In contrast, the calculated ratio of log likelihoods is roughly proportional to sequence lengths. Thus, the clock hypothesis is more often accepted in data sets with short sequences than in data sets with long sequences and/or multiple genes. To demonstrate this, let us assume two data sets of four sequences each (1: $GATC_n$, 2: $ATCG_n$, 3: $TCGA_n$, and 4: $CGAT_n$), with n = 100 ("short sequence data set") and 1000 ("long sequence data set"). Whereas the LRT value is only 0.48 under the HKY-model for the short sequence data set, it is 8.58 for the long one. With both data sets having 2 *df*, the clock is accepted for the short sequence data set (LR = 0.48; critical value of 5.99), but rejected for the long sequence data set (LR = 8.58; critical value of 5.99).

Another bias not accounted for by the LRT is the problem of identical haplotypes. Whereas data sets with identical haplotypes may not (or only slightly) affect the overall log likelihood of the phylogeny, they directly affect the number of *df* (*i.e.,* each duplicate haplotype raises the *df* by one). Thus, the clock is more often accepted in data sets with identical haplotypes than in data sets with unique haplotypes (Wilke, unpubl. data). However, if a deviant haplotype is represented by multiple copies in a data set, the clock may be rejected, whereas it may have been accepted when using unique haplotypes only. To avoid these problems, only data sets with unique haplotypes should be used in LRTs.

Alternatively to the LRT, Akaike's information criterion (AIC; Akaike 1974) can be used for testing the molecular clock. Although widely utilized in assessing models of sequence evolution for phylogenetic estimations, the AIC is still rarely utilized for testing the applicability of a global molecular clock (but see Thomas *et al.* 2006). Akaike's information criterion estimates the distance between a given model and the "truth" the model aims to approximate. The AIC for a certain model can be calculated with the following formula: $AIC = -2\ln L + 2K$, where lnL is the maximized log likelihood of the respective model and K the number of estimable parameters. Since the AIC is a relative value, it is crucial to calculate the difference of the AICs of the two competing models: $\Delta AIC = AIC_0 - AIC_1$, where AIC_0 is the value of the null model (*i.e.,* the clock-like tree) and AIC_1 the value of the alternative model (*i.e.,* the non clock-like tree). If ΔAIC is 10 or higher, the null model is likely to lack substantial support and the global clock may be considered as rejected (Burnham and Anderson 2002). Finally, parametric bootstrap approaches as well as the Bayesian information criterion (BIC) have also been suggested to test the clock-like behavior of a data set (Bollback 2005).

Molecular clock estimations when the global clock is rejected

There are several approaches to enable molecular clock analyses even if the applicability of the global clock is rejected. We want to stress, however, that utilizing these approaches requires a detailed understanding of the procedure itself as well as the underlying assumptions. Therefore, we will introduce only a small selection of possible approaches and suggest studying the primary literature of these and other respective approaches if dating in the absence of a global clock is intended.

Two possible strategies are "tree shopping" and "gene shopping" (Takezaki *et al.* 1995, Hedges *et al.* 1996). In principle, they are special cases of the above-mentioned global clock approach. In the case of "tree shopping" one would prune the original tree by removing clades and/or lineages that appear to increase rate heterogeneity (*e.g.*, extraordinary long branches/clades) and which are of minor importance for the dating of the nodes in question. Subsequently, the global molecular clock may be accepted. "Gene shopping" refers to individual clock tests for every gene previously used in a combined analysis. The rejection of the global clock in the test for the combined gene tree might have been due to considerable rate heterogeneity in a single gene but not necessarily in all fragments used. It has to be noted, however, that the arbitrary selection of lineages and/or genes and the implicit manipulation of the data set may be problematic and is controversial (Cutler 2000). Problems may be particularly severe when using LRTs for testing the global clock. If the clock is rejected in a multi-locus data set, it may be accepted in a single-locus data set not because the selected gene works "better", but simply because the sequence lengths get shorter and therefore the absolute difference in log likelihoods between the clock and non clock-like trees is smaller. As a

consequence, the calculated ratio may less often exceed the critical value of the chi-squared distribution (see above). However, if critical lineages are removed by tree shopping, the degrees of freedom for the chi-squared approximation also become lower and therefore the probability that the clock is accepted in the reduced data set.

In cases where the global clock is still rejected, or where tree and gene shopping is not possible, other approaches have been suggested. One strategy is to refrain from the application of a global clock (one rate for the whole phylogeny) but to apply different rates to individual clades within the phylogeny (the local clock approach, see Fig. 2B). However, this might be difficult because of the potentially large number of possibilities to assign different substitution rates along a given tree (Sanderson 1998). A potential way to bypass this arbitrary application of individual rates (yet allowing substitution rates to differ along a phylogeny) was developed by Sanderson in two key papers (Sanderson 1997, 2002). Whereas the first paper introduced a non-parametric method, the non-parametric rate smoothing (NPRS), the second paper presented a semi-parametric penalized likelihood approach. In principle, this method allows rates to change from an ancestral to descendant lineages but penalizes strong deviations from the ancestral rate. A similar approach was developed by Thorne *et al.* (1998) using a Bayesian framework (see also Kishino *et al.* 2001). In order to overcome the above-mentioned problem associated with the local clock approach, Yang (2004) developed a hybrid algorithm that groups branches using a clustering algorithm (a further development of this approach was presented by Aris-Brosou 2007).

For further information on variable-rate, molecular-dating methods see the excellent reviews of Welch and Bromham (2005) and Rutschmann (2006).

Calibration of the clock

In order to calculate divergence times from node depths within a given phylogeny, either calibration point(s) within the phylogeny or an external molecular clock rate (*e.g.*, universal, taxon-specific, or trait-specific clock rates) are necessary (Table 1). It should be noted that the validity and accuracy of both calibration points and external molecular clock rates are often controversial, and that calibrating the clock might be the most sensitive part of estimating divergence time. Errors and variability of calibrations often account for large discrepancies in molecular clock estimations using the same data set (Bromham *et al.* 1999, Bromham and Penny 2003, Pulquerio and Nichols 2007) and may outscore many other problems in molecular clock analyses.

Whereas molecular clock analyses based on calibration points are often less critically discussed than approaches based on external (*e.g.*, "universal") molecular clock rates, the latter approach might be more widely used in the literature. Approaches based on external molecular clock rates are not *per se* inferior to approaches using calibration points. Both approaches are based on a set of assumptions, which are often violated, and both approaches have their own advantages and disadvantages, which need to be assessed carefully for each individual analysis (Table 1).

Calibration via calibration points or bounds from externally derived dates

As mentioned above, substitution rates can be estimated from ancestral DNA, fossils, or biogeographical events with these dates either being point estimates, or upper and lower bounds (Table 2, Fig. 1). Calibration points can be used for interpolation of divergence times (*i.e.*, the event to be estimated falls within the calibration points or within the calibration point and the tip of the branch), for extrapolation of divergence times (*i.e.,* the event to be estimated falls beyond the calibration point(s)), or a combination of both. Due to the fact that extrapolation has fewer constraints, it is inferior over interpolations and uncertainties will increase with the distance between the calibration point and the estimated event (Bromham and Penny 2003). In other words, distant calibration

Table 1. Pros and cons of estimating divergence times from calibration points and bounds from externally derived dates or *via* external molecular clock rates (*e.g.*, universal, taxon-specific, or trait-specific clock rates). See text for details.

	Calibration points or bounds from externally derived dates	External clock rate
Pros	- locus-independent approach - multi-locus analyses possible	- no calibration point(s) required
Cons	- reliable calibration points rarely available - accuracy and error of the calibration point(s) often difficult to assess - often only bounds but no points available for calibration	- only applicable to data sets with strict molecular clock behavior - external clock rates are typically locus-specific and thus only applicable to single gene - only relatively few external clock rates available
Examples	- closure of the Isthmus of Panama - Mediterranean salinity crisis	- avian clock (see Weir and Schluter 2008) - trait-specific Protostomia COI clock (this paper)

points (*e.g.*, in only distantly related lineages) may affect the accuracy of the analyses (Bromham *et al.* 1999) because of lineage-specific effects or saturation (see below). Moreover, from a statistical point of view, the accuracy of molecular clock estimations can be increased by using multiple calibration points.

Ideally, one would use ancestral DNA/RNA of known age for directly estimating substitution rates in a given phylogeny (Drummond *et al.* 2002, 2003, 2006). However, with the exception of some viral and bacterial sequences, this source of information usually is not available (Table 2). Moreover, ancestral DNA isolated from, for example, subfossil or museum specimens (Lambert *et al.* 2002, Paxinos *et al.* 2002), is often relatively young and may not be adequate for estimating older phylogenetic events (*sensu* Ho 2007, also see below for ancestral polymorphism and the power gap).

More widely used approaches for estimating substitution rates involve indirect methods, that is, fossil and biogeographical data. However, these approaches may also have serious faults (Table 2) affecting molecular clock analyses. Fossils suffer from the fact that they can provide only minimum age estimates as a fossil is necessarily younger than the phylogenetic event that led to its existence and as there is an inherent problem with missing taxa (Marshall 1990, Cutler 2000, Hedges and Kumar 2004, Benton and Donoghue 2007, Ho 2007). Thus, fossil data are best used as lower constraints in molecular clock analyses. Moreover, fossils are often notoriously difficult to identify and to classify within a modern DNA-based phylogenetic framework. Doyle and Donoghue (1993) stressed the fact that fossils can rarely be assigned to branches in a phylogeny and that fossils are often not ancestors of extant taxa, but rather extinct sister groups (also see Cutler 2000)—problems that may also apply to ancestral DNA (see above). These problems might be responsible for large discrepancies of divergence times estimated from fossil data and those estimated from biogeographical data with the former one being, in part, an order of magnitude higher than the latter one (Luttikhuizen *et al.* 2003, Govindarajan *et al.* 2005, Pulquerio and Nichols 2007).

Calibrations based on biogeographical data, however, suffer from the fact that phylogenetic events may or may not be associated with major biogeographical events. The separation of a pair of Pacific/Caribbean geminate species, for example, may be the result of the closure of the Isthmus of Panama some 3.0-2.5 million years ago (Mya). However, it also could have separated an unknown period of time prior to the closure of the isthmus. This is why in relevant phylogenetic studies workers suggest using the youngest phylogenetic event(s) from a set of events to calibrate the clock (*e.g.*, Knowlton and Weigt 1998). However, gene flow between trans-isthmian subpopulations may even have persisted until after the final closure of the isthmus due to passive dispersal, and thus the youngest phylogenetic event may not always reflect the time of closure of the isthmus either.

However, even if we assume that a phylogenetic event is directly linked to a geological event, several problems persist. One problem is the exact age of geological events to be used for calibration. Whereas some events like the beginning and end of the Mediterranean salinity crisis are dated with high accuracy (Krijgsman *et al.* 1999, also see "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below), the timings of other major events like the closure of the Isthmus of Panama are subject to ongoing controversy. Whereas Cronin and Dowsett (1996) suggested major hydrological impacts of the barrier starting some 3.5 Mya, with leakages between 3.1 and 2.8 Mya (also see the discussion in Knowlton and Weigt 1998), newer data indicate an initial decrease of circulation through the Panama Strait 4.5-4.0 Mya with temporary re-openings near 3.8 and 3.4-3.3 Mya, and a final closure 3.0-2.5 Mya (reviewed in Bartoli *et al.* 2005). Another problem is that during a certain geological event, species with different biology and ecology may have different divergence times. Taking the example of the closure of the Isthmus of Panama, benthic species adapted to specific coastal habitats may have diverged much earlier than, for example, small pelagic taxa (also see Schubart *et al.* 1998). Thus, timing of calibration points for molecular clocks ideally involves a specific assessment of a geological event in the context of the biology of the respective species.

Whereas dating uncertainties of vicariance events (such as the Isthmus) are often manageable, dispersal events linked to major geological changes are even more difficult to assess. The geological origin of newly evolved oceanic islands, for example, cannot typically be used to calibrate the clock for distinct phylogenetic groups from isolated islands. This is because the colonization of these islands occurred via dispersal an often unknown period of time after the geological origin of the islands and the geological age of an island may serve as lower bound for molecular clock calibrations, at best. In general, the utilization of dispersal events for molecular clock approaches is very controversial and vicariance events are more suitable.

The example above already indicates the complexity and difficulties associated with using calibration points and bounds (no matter whether ancestral DNA/RNA, fossil, or biogeographical data) for molecular clock analyses. Unfortunately, in the literature poor calibration "points" (which are, in fact, frequently only bounds or ranges at best; also see section on external clock rates below) are often subjected to sophisticated molecular clock algorithms. This striking imbalance already noted by Bandelt (2007) is also in our opinion one of the most critical points in molecular clock approaches. Typically, errors in calibration points

Table 2. Pros and cons of estimating divergence times from ancestral DNA/RNA, fossils, or biogeographical data (see text for details).

	Ancestral DNA/RNA	Fossils	Biogeographical data
Pros	- allows for direct calibration	- often readily available	- comparative analyses of different taxa possible
Cons	- rarely available - difficulties to place on phylogeny - ancestral DNA often cannot be assumed to originate from an ancestors	- difficulties to place on phylogeny - fossils cannot be assumed to be ancestors - fossils can provide only minimum age estimates for divergence events - sometimes uncertainties in dating fossils (both for absolute values and errors of date estimation) - often difficult to identify and to classify (problems with homoplasies) - often severe problems with missing taxa	- often uncertainties in dating biogeographical events - potential mis-linkage of phylogenetic and biogeographical events - typically only one or few calibration points in a given phylogeny - often taxon specific differences - biogeographical events can often provide only upper or lower bounds

out-compete errors of the actual molecular clock analyses (*e.g.*, Wilke 2004).

Calibration via external clock rates

Problems associated with estimating divergence times utilizing external clock rates are listed (Table 1) and discussed in more detail in the section "External molecular clock rates in invertebrates" below.

Molecular dating without calibration points or external clock rates

For many taxa and genes, neither calibration points nor bounds are available and no external clock rates might be applicable. This, however, may not prevent molecular clock analyses. Although absolute divergence times might not be assessable from phylogenies without calibration, the molecular clock approach also allows for estimation of relative divergence times. In other words, it is possible to compare different MRCA in a given phylogeny and to test whether their divergence times are significantly different. Such an approach, may, for example, be of interest in analyses of radiation patterns.

In a recent study, Geyer *et al.* (2006) used relative divergence time estimates for members of a solute carrier protein family in different taxa to infer the temporal evolution of its protein-subfamilies. In order to test whether two specific protein subfamilies represent the youngest members of this family, the variance of node depths of selected splits was calculated for the 100 best trees from a Bayesian search. Pairwise comparisons of relative divergence times of selected MRCA using paired Student's *t*-tests indicated that the depths of the nodes (and therefore their ages) differed significantly, which, in turn, allowed for the identification of the youngest MRCA in the given phylogeny (also see "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below).

Ancestral polymorphism

The amount of polymorphism present in an ancestral population prior to the separation of the descending species introduces a bias into molecular clock analyses if not corrected for. This is because calculating the depth of a node in a gene tree yields not the age of the divergence of the two descending species (t_{sps} in Fig. 3) but the age of the split of the analyzed genetic lineages (t_{MRCA} in Fig. 3). The latter event predates the former by an unknown amount of time (*ap* in Fig. 3), leading to a temporal overestimation of the age of the species divergence. Potential solutions for this problem are discussed in the literature (*e.g.*, Edwards and Beerli 2000, Arbogast *et al.* 2002).

In practice, ancestral polymorphism is often difficult to estimate because it typically requires detailed information of population sizes. In phylogenetic studies, however, the sampling design is usually optimized to reflect species level or higher-level relationships and not polymorphism at the population level. Thus, in many cases authors refrain from calculating and correcting for ancestral polymorphism. Instead, they consider the estimated uncorrected values to be maximum values, with the actual values being smaller by an unknown figure.

Ancestral polymorphism may vary among species. In molluscs, for example, we have found typical values for ancestral polymorphism corresponding to a time frame ranging from 0.1 to 0.4 million years (My) (Wilke *et al.*, unpubl. data, also see section "Examples of molecular clock estimations subfamily Pyrgulinae" below). Using these values simply to demonstrate the effect of ancestral polymorphism, an uncorrected phylogenetic event of, for example, 1.0 Mya, may actually be only 0.6-0.9 My old (the rate of overestimation thus is approx. 10-70%; also see Edwards and Beerli 2000). For older events of, for example, 5.0 Mya this bias would be much smaller, *i.e.,* 2-9%. Although this bias should not be

neglected, ancestral polymorphism for older phylogenetic events is often an order of magnitude smaller than the overall confidence limit of clock estimations.

Appropriate time frame

The accuracy of molecular clock analyses crucially depends on the performance of the gene(s) used, *i.e.*, how well the observed mutation rate corresponds to the actual mutation rate. Phylogenetic events in closely related taxa, corresponding to low divergence times, are characterized by a major problem, here called the "power gap". For these young phylogenetic events, the number of mutational differences is very low, causing stochastic effects to severely bias molecular clock analyses (also see Walsh *et al.* 1999).

Walsh *et al.* (1999) suggested a power test to determine whether a specific data set is sufficient for resolving possible polytomies among clades resulting from divergence events within a relatively short time interval. This test statistic might also serve as a basis for calculating the extant of the power gap in a given data set. Based on equation 3 in Braun and Kimball (2001) and equation 2 in Walsh and Friesen (2001), the length of an internode (t) in years that can be detected with a given data set becomes:

$$t = \frac{-\ln(\beta)}{L\lambda}$$

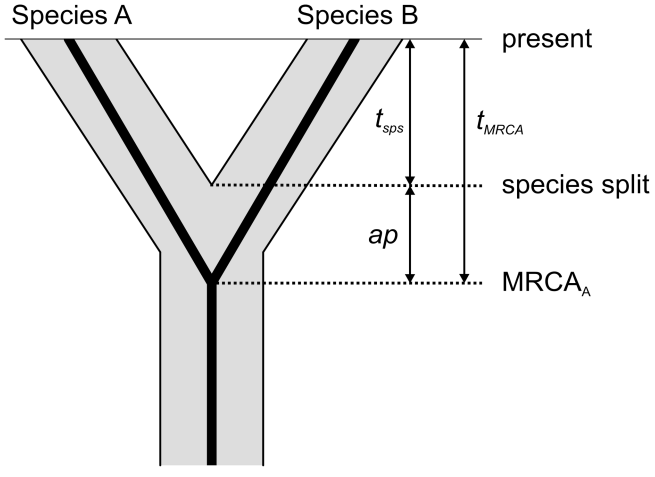

Figure 3. Species tree with an underlying gene tree showing the effect of ancestral polymorphism (*ap*) on the estimation of divergence times. Calculating the age of the most recent common ancestor (t_{MRCA}) of the genetic lineages (black lines) predates the age of the split of the species A and B (t_{sps}) by an unknown amount of time (*ap*). The fraction of this ancestral polymorphism of the total time estimate depends on the age of the species split: the older the split, the smaller the ancestral polymorphism bias (modified from Edwards and Beerli 2000).

where β is the type II error (*e.g.*, 0.05 is the 95% probability of observing one or more substitutions, and 0.20 is the 80% probability), L is the sequence length in bp, and λ is the substitution rate (substitutions per site and year).

We will provide an example of the power test utilizing the widely used COI fragment of approx. 658 bp length defined by the universal primers of Folmer *et al.* (1994). Based on a mean substitution rate of 1.3% My^{-1} (= 1.3 10^{-8} y) for the COI gene for selected invertebrate taxa (see Table 3) and a power of 80% ($\beta = 0.20$) set by convention (see Walsh and Friesen 2001), this fragment should be sufficient to resolve internodes with an age of about 200,000 years or more. A more conservative estimate of t based on a power of 95% ($\beta = 0.05$) yields an internode length of about 360,000 years. In other words, this fragment should not be used to date phylogenetic events with an age of <200,000 years because in such short time periods, the clock does not "tick" often enough in order to calculate reliable divergence times.

Distantly related taxa, however, are often characterized by saturational effects. Mutational saturation occurs when multiple mutations at a given site lead to a randomization of the phylogenetic signal with the number of observed differences being lower than the expected number of differences (Fig. 4). This, in turn, leads to an underestimation of observed divergence times, particularly for older phylogenetic events (also see Arbogast *et al.* 2002). The problem of saturation has been raised early on in molecular clock analyses. Brown *et al.* (1979: fig. 3) suggested in their landmark paper on the temporal evolution of mitochondrial genes saturational effects for divergence times >10 Mya. Problems with saturation may, to a certain extent, be mitigated through the application of sophisticated models of sequence evolution (see Kelchner and Thomas 2007). Moreover, relaxed molecular clocks, which incorporate rate variations across lineages (Fig. 2), may be less prone to saturational effects than strict molecular clock approaches. Nonetheless, tests for saturation, which are implemented in some phylogenetic software packages (*e.g.*, DAMBE, Xia and Xie 2001) should be performed for all molecular clock data sets under the respective model of sequence evolution, as data sets with significant levels of saturation are not suitable for molecular clock estimations. This is particularly important for molecular clock approaches utilizing external molecular clock rates as they are typically based on the strict clock model, thus not allowing for rate variation throughout time.

Arbogast *et al.* (2002) pointed out that estimating divergence times between both distantly and closely related taxa are challenging due to the problems discussed above. Whereas many workers are aware of problems in distantly related taxa (*i.e.,* saturation), problems with closely related taxa (*i.e.,* ancestral polymorphism, power gap), are more frequently neglected.

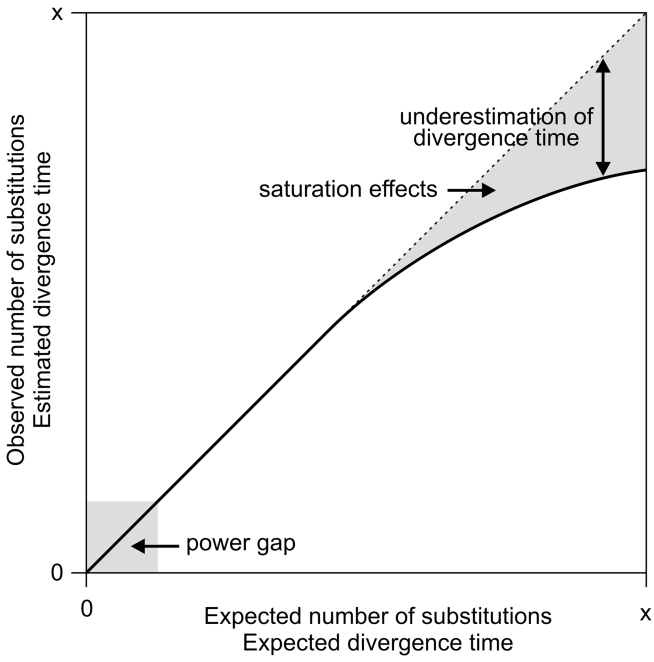

Figure 4. Effect of saturation on estimating divergence times. For genetically divergent taxa, saturational effects lead to a randomization of the phylogenetic signal with the number of observed mutations (solid line) being lower than the actual number of differences (dotted line). This causes an underestimation of divergence times.

Estimation of confidence limits of clock estimations

Calculation of confidence limits is a crucial aspect of clock estimations (Hillis *et al.* 1996, Wilke 2004). This is because confidence limits can be very large (Bromham *et al.* 1998, Bromham and Penny 2003), often making estimates without considering variability meaningless.

Confidence limits are largely affected by two major groups of errors: (a) molecular clock variations and (b) uncertainties of calibration points or external molecular clock rates.

Causes of molecular clock variations within and between lineages can have two major sources (reviewed in Bromham and Penny 2003). First, the molecular clock is probabilistic and ticks at irregular intervals. This behavior, commonly described by a Poisson process, potentially causes large confidence intervals. Second, there might be differences in substitution rates within and between lineages (for biological variables that account for deviations from rate constancy, see below).

Confidence limits of calibration points and external molecular clock rates include uncertainties in the timing of ancestral DNA, fossils, or geological events, time lags of biogeographical and phylogenetic events as well as the variation of external molecular clock rates. In early molecular clock studies, assessments of clock variations were largely neglected. This has led to numerous, partly conflicting molecular clock estimates, which in turn raised general criticism of the molecular clock approach.

Today, estimation of molecular clock confidence limits is standard procedure and many molecular clock software packages incorporate algorithms or approaches for quantification of uncertainties utilizing, for example, bootstrapping, Bayesian posterior distribution, or Poisson distribution (see reviews of Welch and Bromham 2005, Rutschmann 2006). Whereas most publicly available packages account for variation of the clock itself, not all consider uncertainties of calibration points or external molecular clock rates. In this case, the total error of the clock can be calculated using, for example, a propagation of uncertainty analysis (see "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below).

Sampling design and interpretation of data

Molecular clock approaches allow for a dating of the MRCA of extant lineages. In this regard, the clock approach is unambiguous; split I in Fig. 5, for example, shows the age of the MRCA of taxa 1+2 and taxon 3. The interpretation of the phylogenetic and taxonomic relevance of such events, however, might be affected by sampling design (*e.g.*, missing taxa) and may be subject to misinterpretation. Wilke (2004), for example, compared the results of two molecular clock analyses of similar sets of rissooidean snails that differed in their age estimates by more than an order of a magnitude. He showed that the difference was largely due to missing taxa together with misinterpretations of the results. Thus, these problems, though rarely discussed in the literature, may affect molecular clock analyses more severely than many other problems discussed in the present paper.

To demonstrate possible adverse effects of missing taxa, we provide a sample data set with ten species (Fig. 5) with the complete phylogeny (including extinct and unsampled) taxa to the left and a phylogeny with four missing taxa to the right. As mentioned above, molecular clock analyses estimate the age of MRCAs in a given phylogeny. In a complete phylogeny, that is, a phylogeny that contains all species of a given taxon, these age estimates of MRCAs can directly be used for phylogenetic interpretations. The age of node IV, for example, represents the age of genus A, and node II the onset of the intra-generic diversification within genus C. In an incomplete phylogeny such interpretations, however, may be erroneous. Node *IV* in the right tree is the MRCA of three taxa of genus A and three taxa of genus C. It does, however, not correspond to the age of genus C because that genus is much younger and represents the sister to the extinct genus B (see node III in the complete phylogeny). Also, node *II* in the sampled phylogeny does not represent the age of the onset of diversification in

genus C because taxon 5 (not sampled) is older than any of the sampled taxa (see node II in the complete phylogeny).

In most cases we do not know whether a phylogeny is complete or not. Thus, molecular clock estimates are best explained within the phylogenetic concept of the MRCA. If taxonomic interpretations were to be made, then this should be done within the context of minimum and maximum ages. In a given phylogeny, the observed **onset of** the **diversification** within a supra-specific taxon (genus and beyond) should be expressed as **minimum** age. Node *II* in the sampled phylogeny thus is the minimum age of diversification within genus C. Assuming a robust phylogeny, additional taxa previously not sampled cannot render this age younger; they only can render it older (see node II in the complete phylogeny). In contrast, the **age** of a given supra-specific taxon should be expressed as **maximum** age. Node *IV* in the sampled phylogeny thus is the maximum age of genus C. Additional taxa cannot render the age of genus C older; they only can render it younger (see node III in the complete phylogeny).

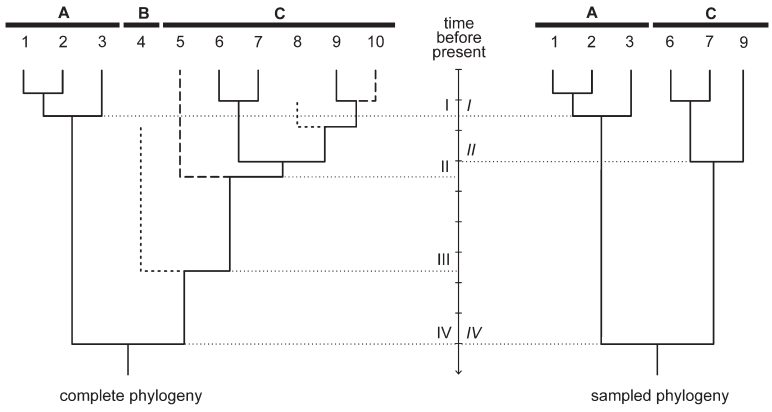

Figure 5. Effects of unsampled (dashed lines) and extinct (dotted lines) species on inferring the timing of phylogenetic events within three putative genera (A-C). Left, complete phylogeny; right, phylogeny with only six out of ten taxa sampled. Major nodes are marked with Roman numerals. See text for details.

EXTERNAL MOLECULAR CLOCK RATES IN INVERTEBRATES

Problems with external molecular clock rates

During our survey of molecular clock analyses reported in the literature, we noted an interesting bias. Whereas putative calibration points are often uncritically unitized for clock approaches (often simply as "pers. comm."), external molecular clock rates (particularly "universal" molecular clock rates) are typically dismissed outright as invalid. In fact, there are several theoretical and practical studies demonstrating that molecular clock rates can largely vary among genes and organisms (*e.g.*, Thomas *et al.* 2006). While we agree on these findings, we also would like to raise a cautionary note. An external molecular clock rate can only be as good as the individual rates upon which it is based. Given the uncertainties of many calibration points, conflicts between fossil and biogeography-based data as well as general clock problems such as ancestral polymorphism, missing taxa, and saturation (see above), external molecular clock rates (particularly those applicable to a larger taxon) are hard to establish. Thus, a universal, taxon-specific or trait-specific molecular clock rate may be rejected because there is no such rate, but they may also be rejected because of potential errors in individual rates on which they are based.

Nonetheless, there is a strong and compelling theoretical background suggesting that there is no clock universal for all genes and taxa, and that mutation rates *per se* vary among genes and broad taxonomic groups (*e.g.*, Ayala 1997, Drake *et al.* 1998, Bromham and Penny 2003, Takahata 2007).

If we ask whether the concept of a "universal" clock can possibly be saved, we have to understand which biological factors affect mutation rates. Ayala (1999) invoked five biological variables that account for deviations from rate constancy within and between lineages:

(1) generation time (shorter generation time "accelerates the clock" as it shortens the time for fixing new mutations, particularly if DNA replication-dependent errors are the major source of mutations; also see Takahata 2007),

(2) population size (larger population sizes will "slow the clock" because of increased times for fixing new mutations),

(3) species-specific differences in properties that affect DNA replication (different species may, for example, have different DNA polymerases with different error rates; Bromham and Penny 2003),

(4) changes in the function of a protein as evolutionary time proceeds, and

(5) stochasticity of natural selection.

Other factors might include:

(6) body size (smaller species tend to have faster rates of molecular evolution, Gillooly *et al.* 2005, Lanfear *et al.* 2007, but see Thomas *et al.* 2006),

(7) body temperature, including ectothermy *vs.* endothermy (body temperature affects metabolic rates, which in turn affects production of free radicals causing mutations, Gillooly *et al.* 2005), and

(8) life history, particularly reproductive traits (mutation rates are presumably higher in hermaphrodites compared with gonochorists, Davison 2006, also see Foltz *et al.* 2004).

According to Gillooly *et al.* (2005), many biological factors can be linked to two major hypotheses explaining rate heterogeneity: the metabolic rate hypothesis (higher metabolic rates are related to higher production of mutation-causing free radicals) and the generation time hypothesis (most mutations are caused by DNA replication errors during division in germ cell lines) (but see Lanfear *et al.* 2007).

It should be noted that most of these effects are largely hypothetical and have only rarely been tested in the context of the molecular clock. Moreover, the relationship between these variables and substitution rates might not be universal, but gene and taxon specific, and the underlying mechanisms often are still poorly understood (*e.g.*, Lanfear *et al.* 2007).

In fact, an extensive study conducted by Thomas *et al.* (2006) could not find a relationship between body size and mutation rates and most workers simply attribute lineage specific differences to the fickle process of natural selection (*e.g.*, Ayala 1999, Takahata 2007).

Gillooly *et al.* (2005), however, introduced a controversial model accounting for body size and temperature effects on metabolic rates, which supposedly could explain rate heterogeneity in different genes, taxa, and environments. Moreover, the authors argue that this model suggests a single molecular clock that ticks according to mass-specific metabolic energy.

Although this model was recently rejected by Lanfear *et al.* (2007), the effect of these and other biological factors could explain why the existence of an universal molecular clock had to be rejected for larger and biologically diverse groups, such as invertebrates (Thomas *et al.* 2006), but appears to be valid in some smaller groups with similar biology and life history like birds (Weir and Schluter 2008).

In fact, knowledge of the relevant factors affecting the clock could help reduce deviation from rate constancy within larger sets of taxa and lead to the establishment of a series of gene-specific molecular clock rates for groups of species that share similar biological and life history traits. An example for a potential "trait-specific" molecular clock rate in invertebrates is given in the following section.

A potential trait-specific molecular clock rate for the Protostomia

As outlined above, several biological and life history properties of animals might affect the tick rate of the clock. At the same time, clock rates can vary considerably among genes, and the performance of a given gene might be poor for relatively young (power gap) and relatively old (saturational effects) divergence events.

Acknowledging that invertebrates are a paraphyletic and highly diverse group, Wilke (2003) first attempted to establish a specific COI clock for the Protostomia, a clade of bilateral animals including the three major groups Ecdysozoa (*e.g.*, Arthropoda and Nematoda), Lophotrochozoa (*e.g.*, Mollusca and Annelida), and Platyzoa (*e.g.*, Platyhelminthes and Rotifera).

Based on published and his own estimates of molecular clock rates for the COI gene in taxa separated by less than 10 Mya (*i.e.,* the presumed time frame in which the COI gene is not saturated), Wilke found relatively coherent rates ranging from 0.7 to 1.2% My^{-1} (uncorrected substitution rates) or 1.4 to 2.4% My^{-1} (uncorrected divergence rates). These published individual rates were later reanalyzed and refined within a tree-base approach by Albrecht *et al.* (2006) utilizing Kimura's two-parameter model (K2P) model.

In this paper, we build upon the studies of Wilke (2003) and Albrecht *et al.* (2006) in order to establish a preliminary trait-specific molecular clock rate for the COI gene in the Protostomia.

The basic idea of this trait-specific clock rate is to find (within a larger taxon) groups of species:

(1) that share biological and life history characteristics supposedly affecting rate heterogeneity (*e.g.*, mode of reproduction, generation time, body size and temperature, population size),

(2) which individual clock rates can be calibrated with robust calibration points, and

(3) where molecular clock estimations are not affected by the power gap or significant degrees of saturation.

These individual rates can be assessed for dispersion and, if applicable, average trait-specific clock rates could be established together with their errors. The trait-specific clock rate for the COI gene in the Protostomia suggested here involves data from a total of 12 pairs of species from several higher taxonomic groups within the Protostomia with the following characteristics:

- they are aquatic,
- they are dioecious (see point 8 under variables that account for deviations from rate constancy between lineages above),
- they have a generation time of approximately one year (see point 1 above),
- they are ectothermic and live in tropical or subtropical waters (see point 7 above), and
- they are relatively small with body sizes differing by not more than an order of magnitude (see point 6 above).

In order to calculate an average trait-specific clock rate for these taxa, we obtained the COI sequences used in the original publications (see Table 3), tested the applicability of a strict molecular clock, and analyzed substitution rates under the assumption of the molecular clock and under different models of sequence evolution in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003).

Note that we were unable to correct for ancestral polymorphism in the clock data sets because of missing

Table 3. Molecular clock (substitution) rates for the COI gene under different models of sequence evolution based on a total of 12 pairs of sister taxa from 5 groups of Protostomia separated by biogeographical events. These taxa share major biological and life history traits. Mean trait-specific molecular clock rates (R_{MC}) together with their standard deviations are given for each model. All values are uncorrected for ancestral polymorphism (JC, Jukes-Cantor model; K2P, Kimura's two-parameter model; F81, Felsenstein 1981 model; HKY, Hasegawa-Kishino-Yano model; GTR, general time reversible model; I, invariable sites; and Γ, gamma distribution).

Taxon	Reference	# Taxon pairs	Mean body size in mm	Divergence time in My	Molecular clock rates (in % My^{-1})for selected models of sequence evolution						
					JC	K2P	F81	HKY	HKY+ I+Γ	GTR	GTR+ I+Γ
Salenthydrobia/Peringia (Gastropoda)	Wilke (2003)	1	3-4	5.64*	1.33	1.29	1.36	1.32	1.60	1.29	1.96
Chlorostoma (= *Tegula*) spp. (Gastropoda)	Hellberg and Vacquier (1999)	1	10-20	2.75**	1.37	1.40	1.37	1.37	1.48	1.37	2.06
Alpheus spp. (Decapoda)	Knowlton and Weigt (1998)	7***	20-50	2.75**	1.21	1.23	1.17	1.18	1.89	1.03	1.94
Sesarma spp. (Decapoda)	Schubart *et al.* (1998)	2	15-30	2.75**	1.01	1.03	1.03	1.10	1.61	1.02	1.59
Alvinella/Paralvinella (Annelida)	Chevaldonné *et al.* (2002)	1	2-3	2.75**	1.20	1.21	1.20	1.21	1.26	1.19	1.25
			Mean (n = 5)		**1.22**	**1.23**	**1.23**	**1.24**	**1.57**	**1.18**	**1.76**
			95% confidence interval (n = 5)		**0.27**	**0.26**	**0.28**	**0.22**	**0.45**	**0.31**	**0.66**

* Mediterranean Salinity Crisis, timing for its climax taken from Krijgsman *et al.* (1999).

**Closure of the Isthmus of Panama, timing based on average estimates suggested by Bartoli *et al.* (2005).

***From 15 pairs of geminate sister species suggested by Knowlton and Weigt (1998), the seven most closely related pairs of species were used here (also see Albrecht *et al.* (2006)).

population-level data (see section on ancestral polymorphism). Thus, the trait-specific clock rate suggested here might be slightly overestimated, but because of the relatively old biogeographical events used for calibration (see Table 3), we would expect the bias to be <10%.

Nonetheless, the trait-specific clock rates suggested here for several major models of sequence evolution are surprisingly coherent. The 95% confidence intervals for individual models are typically around 20%. Even among models, the average clock rates are very similar, indicating that they are relatively robust against model misspecifications. The only exceptions are the models with gamma distribution and invariables sites (Γ+I), which show, as expected, an elevated clock rate and elevated confidence intervals.

Of course, the trait-specific COI clock suggested here would need further refinement involving more taxa and more independent calibration events in order to assess its validity for a large set of taxa. Nonetheless, this example already indicates that deviations from rate constancy can be mitigated with relatively simple means. Moreover, if the validity of this trait-specific clock would be confirmed, it may not only be applicable to many different species; it also could help to establish a general predictive model for substitution rates taking saturation as well as specific life history, biological, and biochemical characteristics into account.

A simple fool's guide to molecular clock approaches

The flowchart given (Fig. 6) intends to provide relatively simple and conservative, yet sound, guidance through crucial steps of molecular clock analyses. It is simple because it is based on a series of straightforward tests and tools readily available, and it is conservative because it does not attempt to deal with problems the solution of which would require expert knowledge (*e.g.*, estimation of divergence times from saturated data).

The guide is applicable to molecular dating of data sets for which (a) calibration points or bounds exist, (b) an external clock rate is available, or (c) for which no such information exists. In the latter case, however, only estimations of relative divergence times (see section "Molecular dating without calibration points or external clock rates" above) would be possible.

The initial information required is whether the data set in questions shows significant levels of saturation. The appropriate model of sequence evolution (typically the best

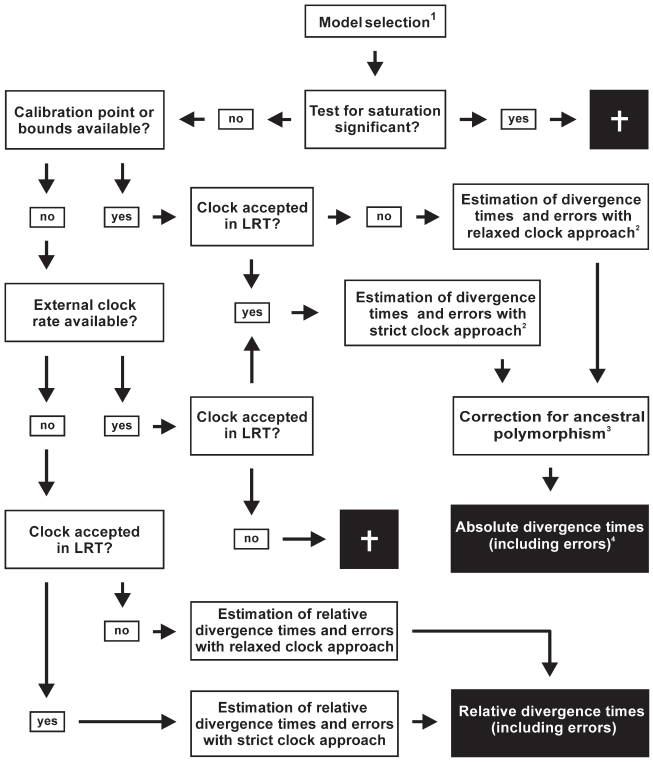

Figure 6. Simple fool's guide for molecular clock analyses. See text for details. [1]Typically, the best fit model of sequence evolution is used; approaches utilizing external clock rates may, however, require a pre-defined model. [2]Error estimations should include both errors of molecular clock variations as well as uncertainties of calibration points or external molecular clock rates. [3]If correction for ancestral polymorphism is not possible, divergence times should be treated as maximum divergence times. [4]For young phylogenetic events, the power gap should be tested.

fit model) has to be selected using, for example, the program Modeltest (Posada and Crandall 1998), MrModeltest (Nylander 2004), or similar software tools. Then the data set can be tested under the chosen model for substantial nucleotide saturation (*e.g.*, using the program DAMBE, Xia and Xie 2001). If the data set is saturated, molecular dating is not advisable.

If the test reveals no substantial saturation, then the question arises whether at least one calibration point or at least one lower and one upper bound for estimating substitution rates exist. If so, the tree can be calibrated with those points (see below). If not, the guide asks for the existence of an external molecular clock rate. If available, this external rate could be utilized to calibrate the clock. Otherwise, only estimations of relative divergence times might be possible.

In cases were both calibration point(s) and an external clock rate are available, we suggest using both approaches and comparing the results for consistence (*e.g.*, Wilke *et al.* 2007, also see section "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below).

Calibration point or bounds available

If at least one calibration point or two bounds are available, the clock has to be tested (*e.g.*, using LRT) in order to decide whether a strict molecular clock approach can be used for estimating divergence times or whether methods that account for rate heterogeneity ("relaxed clocks") have to be applied. If the clock is accepted, one global rate of substitution can be assumed and several packages are available to estimate divergence times and to provide error estimates (reviewed in Rutschmann 2006). Note that not all of these packages can deal with multiple calibration points.

If the clock is rejected, then methods should be applied that either correct for or incorporate rate heterogeneity (reviewed in Rutschmann 2006). However, as with strict clock approaches, not all of these methods can deal with multiple calibration points.

Whether strict or relaxed clock approaches are used, it should be checked if error estimations also account for uncertainties in the calibration point(s). If not, this error should be incorporated via, for example, propagation of uncertainty (see section "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below). Finally, divergence times should be corrected for ancestral polymorphism, if possible. If the data set does not allow for correction of ancestral polymorphism, divergence times should be treated as maximum times. Moreover, if divergence times are very low, it should be tested whether the data set is sufficient for resolving such young phylogenetic events (see section "Appropriate time frame" above).

External molecular clock rate available

If an external clock rate is available, the data set has to be tested for clock like behavior. If the clock is rejected, molecular dating is not advisable, as the applicability of an external rate typically requires one global rate of substitution (strict clock). If the clock is accepted, estimation of divergence times is as described above.

Some external clock rates are available for a specific model of sequence evolution only. In this case, the same model would need to be applied to the data set in question. Such model misspecifications are, however, controversially discussed (particularly for data sets with distantly related taxa) and should be used with caution.

Another consideration is that external clock rates may or may not be corrected for ancestral polymorphism. If uncorrected rates are used, it might be difficult to estimate the bias in the data set in question, and dates for young

divergence events should be treated with particular caution (see section "Examples of molecular clock estimations in the caenogastropod subfamily Pyrgulinae" below).

No calibration point or bounds and no external molecular clock rate available

If no information is available for calibrating the tree, relative divergence times can be estimated. Depending on whether the clock is accepted, strict or relaxed clock approaches should be used. Relative clock approaches are best implemented by setting the age of the root to 1 by default. Then, after testing for normality of node depth distribution, either paired Mann-Whitney *U*-tests or Student's *t*-tests can be used to study whether two specific MRCA significantly differ in their age utilizing replicates from the phylogenetic search. Although still rarely applied in phylogenetic studies, such relative clock approaches are powerful tools for testing hypotheses in evolutionary biology in the absence of specific divergence times.

EXAMPLES OF MOLECULAR CLOCK ESTIMATIONS IN THE CAENOGASTROPOD SUBFAMILY PYRGULINAE

To demonstrate different molecular clock approaches in our model taxon, the subfamily Pyrgulinae, we here use the two hydrobiid data sets of Wilke *et al.* (2007) for three largely independent clock analyses. The full data set A contains combined fragments of the mitochondrial COI gene, the mitochondrial LSU rRNA gene, and the nuclear SSU rRNA gene. The reduced data set B only contains COI sequences. The data set B is used for clock estimations utilizing the trait-specific COI Protostomia clock introduced above. Data set A serves as the basis for estimating divergence times based on an available calibration point as well as for relative clock estimations. The following descriptions are based on the flow chart (Fig. 6) presented above.

I. Time estimation with calibration point (data set A)

Model selection

The data set was analysed in MrModeltest 2.3. The models suggested were HKY+I+Γ (Hasegawa-Kishino-Yano model with invariable sites and Γ distribution), GTR+I+Γ (general time reversible model with invariable sites and Γ distribution), and TrNef+I (Tamura-Nei model with equal base frequencies and invariable sites) for the COI, LSU rRNA, and SSU rRNA fragments, respectively.

Test for saturation significant?

In order to test whether the individual partitions show significant levels of saturation, the test of Xia *et al.* (2003), as implemented in the software package DAMBE 4.2.13 (Xia and Xie 2001), was used with the proportion of invariable sites suggested by MrModeltest (*i.e.,* 0.6082 for COI, 0.5805 for LSU rRNA, and 0.9406 for SSU rRNA). The test did not reveal a significant degree of saturation even under the very conservative assumption of an extremely asymmetrical tree for any of the three data partitions. Therefore, the data set is considered to be suitable for further molecular clock analyses.

Calibration point or bounds available?

For the Pyrgulinae, the known phylogenetic age of the monotypic genus *Salenthydrobia* Wilke, 2003 (see Wilke 2003, 2004) could be used as the calibration point for estimating timing of evolutionary events (see node A in Fig. 7). Ecological and biogeographical data strongly indicate that *Salenthydrobia* originated during the Messinian salinity crisis (MSC), that is, between 5.96 and 5.33 Mya (see Krijgsman *et al.* 1999 for the dating of the MSC). As *Salenthydrobia* belongs to the potential sister subfamily of the Pyrgulinae (*i.e.,* the Hydrobiinae), and as their relationships do not show signs of saturation, it is here assumed that the substitution rates in these taxa are similar.

Clock accepted in LRT?

For the LRT, we first ran two analyses in MrBayes 3.1.2 (clock enforced and clock not enforced) with the partitioned model suggested by MrModeltest (see above) until the chains converged on similar results (*i.e.,* <0.01 after 1.000.000 generations). Then, the best tree from each analysis was used for the LRT. With $\log L_0 = -5493.43$, $\log L_1 = -5488.52$, $-2\log \Lambda = 9.82$, and $df = 19$, the clock hypothesis was not rejected ($P < 0.05$).

Estimation of divergence times and errors with strict clock approach

For calculation of the age of the split between the Black Sea/Asia Minor pyrgulinids and the Pyrgulinae from the Balkan (node B in Fig. 7) as well as the split between the Lake Ohrid pyrgulinids from their sister taxon (node C in Fig. 7), we are using all trees generated from the (clock-enforced) Bayesian search, except for those that fall within a predefined burn-in of 10%.

For each individual tree, we calculated the age of nodes B and C using the rule of three with the known variables being the age of node A (*i.e.,* the climax of the MSC with 5.64 Mya), the node depth of node A, and the depth of nodes B or C.

Separately for nodes B and C, we then averaged their ages over all trees and calculated the respective 95% confidence intervals (resulting in 7.72 ± 2.37 Mya for node B and 3.00 ± 1.04 Mya for node C). Finally, the error of the calibration point has to be added to this error of the clock.

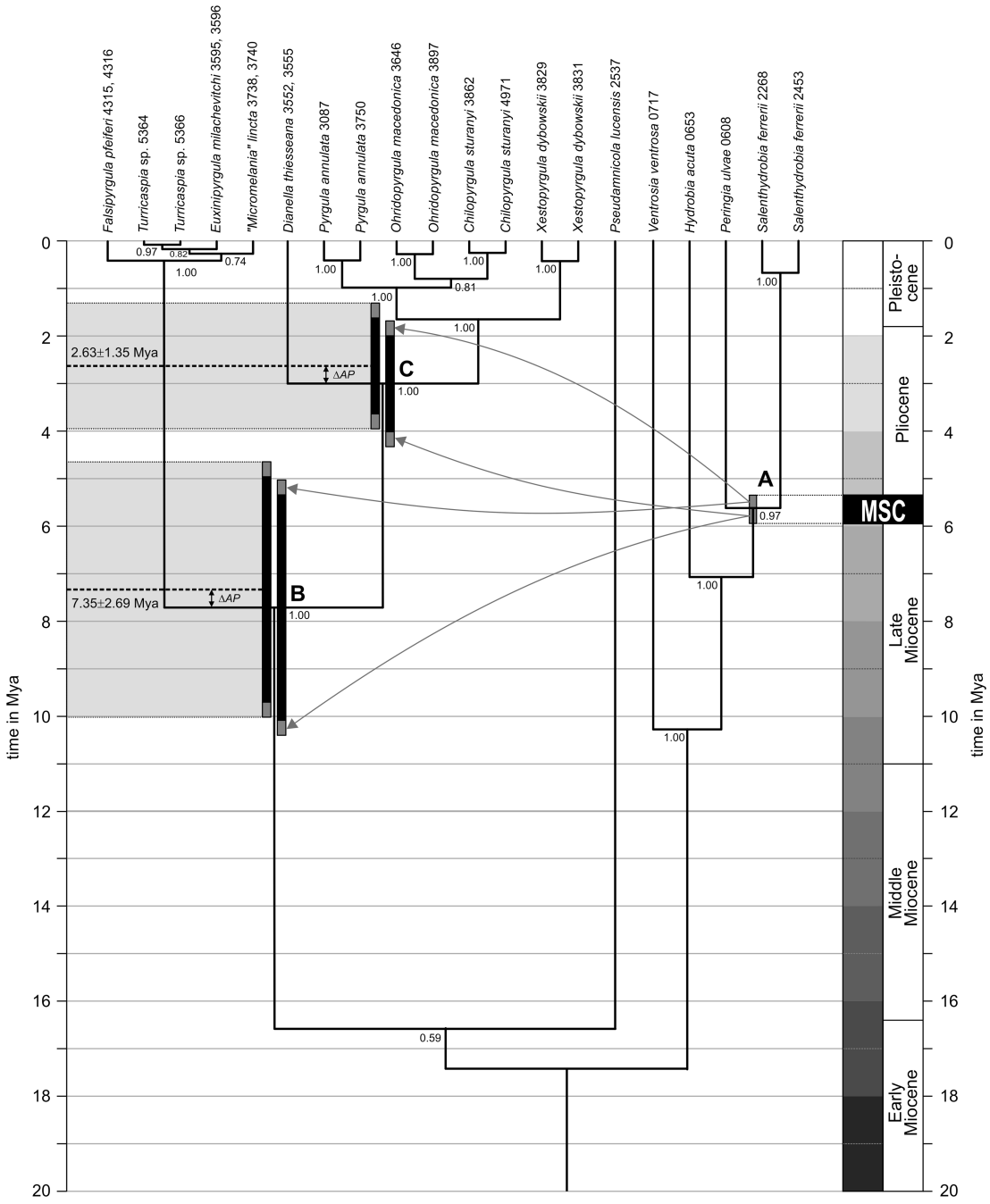

Figure 7. Estimation of divergence times utilizing a calibration point according to approach I. Shown is the best Bayesian tree under the clock criterion based on three gene fragments for representatives of the nominal subfamilies Pyrgulinae, Pseudamnicolinae, and Hydrobiinae (modified from Wilke *et al.* 2007). The outgroup taxon was removed *a posteriori*. The calibration point (*i.e.,* the origin of the genus *Salenthydrobia* that is associated with the Messinian salinity crisis [MSC; marked A]) and the time frame of major phylogenetic events (*i.e.,* the split of the Ponto-Caspian/Asia Minor taxa from the Balkan taxa [B], and the split of Lake Ohrid pyrgulinids from their sister taxon [C]) are shown on the tree. Branch length variations are plotted as black bars at nodes B and C. For estimating the total error of our clock calculations, we also incorporated the error of the calibration point (gray bars, see text for details). Bars right to the node are not corrected for ancestral polymorphism (ΔAP); bars to the left constitute corrected values. The timing of the events in question and their confidence intervals are shown as light gray bands. Posterior probabilities are given for all nodes.

Unfortunately, we do not have any information as to when the split between *Salenthydrobia* and its sister taxon occurred within the MSC. Thus, each point between the onset of the MSC (*i.e.*, 5.96 Mya) and its end (*i.e.,* 5.33 Mya) is equally likely. As we used the climax of the MSC for our clock estimation, we here suggest a pragmatic approach for calculating an approximate total error of the clock. We simply add the difference from the climax to both the beginning and end of the MSC (*i.e.*, ± 0.315 My) to the confidence interval of the clock estimations. The values for node B and C thus would be 7.72 ± 2.69 Mya and 3.00 ± 1.35 Mya, respectively (see Fig. 7)

Correction for ancestral polymorphism

To infer the amount of ancestral polymorphism (Fig. 3) within our data set in the absence of sufficient population genetics data, we here use an approach based on a suggestion of Edwards and Beerli (2000). We assume that the effective population sizes in extant species reflect the population sizes of the ancestral species. Hence, averaging the sequence diversity within the four descendant species of node C (Fig. 7) might provide a rough estimate of the ancestral polymorphism that was present when splits C and B occurred. The node depth under the chosen HKY+I+Γ model is 0.0039 for *Pyrgula annulata* (Linnaeus, 1767), 0.0029 for *Ohridopyrgula macedonica* (Brusina, 1896), 0.0028 for *Chilopyrgula sturanyi* Brusina, 1896, and 0.0042 for *Xestopyrgula dybowskii* Polinski, 1929 (mean node depth: 0.0034). Using node A as calibration point, the amount of ancestral polymorphism corresponds to approx. 0.37 My.

Estimation of absolute divergence times including errors

Deducing the ancestral polymorphism corresponding to 0.37 My from our uncorrected divergence time estimations results in corrected divergence times and confidence intervals of **7.35 ± 2.69 Mya** for node B and **2.63 ± 1.35 Mya** ago for node C (see Fig. 7).

II. Time estimation with external trait-specific COI clock (data set B)

Model selection

The COI data set was analysed in MrModeltest 2.3, which suggested the HKY+I+Γ model based on the Akaike information criterion.

Test for saturation significant?

Utilizing the test of Xia *et al.* (2003) and the proportion of invariable sites suggested by MrModeltest (*i.e.,* 0.6082), the test did not reveal a significant degree of saturation, even under the very conservative assumption of an extremely asymmetrical tree. The COI data set is therefore considered to be suitable for further molecular clock analyses.

Calibration point or bounds available?

We here ignore the *Salenthydrobia*-calibration point used above and continue with the flow chart assuming that there is no calibration point available.

Clock accepted in LRT?

The COI data set was used to run two analyses in MrBayes 3.1.2 (one with the clock enforced and one without enforced clock) under the HKY+I+Γ model suggested by MrModeltest (see above) until the chains converged on similar results (*i.e.,* <0.01 after 1.000.000 generations). Then, the best tree from each analysis was used for the LRT. The clock hypothesis was not rejected ($-2\log \Lambda = 10.66$, $df = 19$, $P < 0.05$).

Estimation of divergence times and errors with strict clock approach

For calculating the age of the split between the Black Sea/Asia Minor pyrgulinids and the Pyrgulinae from the Balkan (node B in Fig. 7) as well as the split between the Lake Ohrid pyrgulinids from its sister taxon (node C in Fig. 7), we use the external trait-specific COI clock rate suggested in the present paper. Based on a rate and 95% confidence interval of 1.57 ± 0.45% My^{-1} under the HKY+I+Γ model (see Table 3), we used all trees generated from the Bayesian search (under the clock criterion), except for those that fall within a predefined burn-in of 10%.

For each individual tree (in our case 90,000), we calculated the depths of nodes B and C and their respective standard deviations from the tree files with an R-routine (available upon request), resulting in mean node depths of 11.13 ± 1.90% and 4.46 ± 0.80% for nodes B and C, respectively. Alternatively this calculation can be carried out utilizing the program TreeAnnotator1.4.8 from the BEAST package (Drummond and Rambaut 2007).

Using the rule of three with the known variables being the depth of nodes B (11.13%) and C (4.46%) and the external clock rate of 1.57% My^{-1}, we calculated the mean age resulting in 7.11 Mya for node B and 2.85 Mya for node C.

Finally we calculated the total error of these estimates by combining the error of node depth (standard deviations of 1.90% and 0.80% for nodes B and C, respectively) with the error of the trait-specific clock (the standard deviation of 0.23% My^{-1} corresponds to a confidence interval of 0.45% My^{-1}) by utilizing the method of error propagation (note that this method is based on standard deviations rather than confidence intervals):

$$\Delta G = \sqrt{\left(\frac{1}{y}\Delta x\right)^2 + \left(\frac{\bar{x}}{y^2}\Delta y\right)^2}$$

with ΔG being the total error, Δx the error of the node depth, Δy the error of the external clock rate, $\bar{x}$ the mean node depth, and y the external clock rate.

Based on a mean node depth of $\bar{x} = 11.13\%$, a relative node depth error of $\Delta x = 1.90\%$, a relative error of the external clock rate of $\Delta y = 0.23\%$ My^{-1}, and an external clock rate of $y = 1.57\%$ My^{-1}, the total error (as standard deviation) for node B in My is:

$$\Delta G = \sqrt{\left(\frac{1}{1.57}1.9\right)^2 + \left(\frac{11.13}{(1.57)^2}0.23\right)^2} = 1.59$$

Multiplying this standard deviation of 1.59 My with 1.96 results in a 95% confidence interval of 3.10 My. Hence, the age and confidence intervals for node B would be 7.11 ± 3.10 Mya (the corresponding values for node C are 2.85 ± 1.29 Mya).

Correction for ancestral polymorphism

Correction for ancestral polymorphism in the present pyrgulinid data set is not possible. This is because the external clock rate used here is not corrected due to the lack of knowledge of intraspecific diversities within the taxa used for establishing this trait-specific clock rate. However, assuming that the extent of ancestral polymorphism in the latter data sets is similar to the one in our pyrgulinid data set, some approximate information on a potential bias can be given. If, for example, phylogenetic events to be estimated in the pyrgulinid data set have an age similar to the average age of those events used for establishing the trait specific-clock (here approx. 3 Mya, see Table 3), then the bias might be small. If the event to be estimated is older, then we likely will see an underestimation of time. However, if the events to be estimated are younger than those used to establish the trait-specific clock, then we will see an overestimation of divergence times.

Estimation of absolute divergence times including errors

Without correcting for ancestral polymorphism, the values presented above would have to serve as approximate final values. The time estimate of **2.85 ± 1.29 Mya** for node C might represent a relatively unaffected value. The estimate of **7.11 ± 3.10 Mya** for node B, however, is likely underestimated by an unknown value.

III. Relative time estimations (data set A)

Model selection

The best-fit models of sequence evolution suggested by MrModeltest 2.3 were HKY+I+Γ, GTR+I+Γ, and TrNef+I for the COI, LSU rRNA, and SSU rRNA fragments, respectively (see section "Time estimation with calibration point" above).

Test for saturation significant?

The test of Xia *et al.* (2003) did not reveal a significant degree of saturation for any of the three data partitions (see section "Time estimation with calibration point" above). Therefore, the data set is considered to be suitable for further molecular clock analyses.

Calibration point or bounds available?

We here assume that no calibration points and no external rates are available.

Clock accepted in LRT?

The LRT did not reject the clock hypothesis ($-2\log \Lambda = 9.82$, $df = 19$, $P < 0.05$, see section "Time estimation with calibration point" above).

Estimation of relative divergence times including errors

Calculation of relative divergence times of phylogenetic events can be done using all trees generated from the Bayesian search under the clock criterion, except for those that fall within the burn-in. For each individual tree, the relative age of a given node is estimated by either setting the node depth of the root to one or by simply using absolute node depth as relative divergence time. In most cases, relative divergence times, however, are meaningless. Instead there often is an interest in testing whether a specific split occurred simultaneously with another split in the phylogeny.

In Fig. 7, for example, node A (the split of *Salenthydrobia* from its sister taxon) appears to be younger than node B (the split of the BlackSea/Asia Minor pyrgulinids from the Balkan pyrgulinids). In order to test this assumption, the following statistics can be used. For each individual tree of the Bayesian search (with the trees from the burn-in ignored), the depth of node A is compared to the depth of node B either using a paired Student's *t*-test or a paired Mann-Whitney *U*-test, depending on whether the data are distributed normally.

As a Shapiro-Wilk test did not reject normal distribution of the data ($P > 0.05$), a paired Student's *t*-test was used to test whether depths of the nodes A and B (and therefore their ages) differ significantly. As the test was significant ($P < 0.01$), it can be assumed that node A is significantly younger than node B and that these phylogenetic events do not coincide in time.

CONCLUSIONS

Over the past few decades, molecular clock approaches have become increasingly popular, and it is now widely accepted that the molecular clock is an important source of information in evolutionary biology. Although not exact, it can provide useful information on divergence times, and a number of approaches are being developed to mitigate problems associated with the clock. This concerns both major clock strategies—the application of calibration points or

bounds and the application of external molecular clock rates.

For approaches based on calibration points and bounds, we show that they have the advantage of being locus- (gene-) independent, but that reliable calibration points are rarely available, that the accuracy and the error of calibration points often are difficult to assess, and that some suggested calibration points can only serve as upper or lower bounds, at best. We also show that data typically used for calibrating trees (*i.e.,* ancestral DNA/RNA, fossils, or biogeographical information) all suffer from specific problems that might severely bias molecular clock estimations.

Approaches based on external clock rates, however, have the advantage of not requiring such calibration points or bounds. They, however, typically call for a strict molecular clock behavior of the data set in question and are usually locus-specific. Moreover, relatively few external clock rates are available, and many of them are controversial. Whereas universal clocks are often dismissed outright, newer studies suggest that there might be a single substitution rate (including error) for a range of taxa that share biological and life history characteristics supposedly affecting rate heterogeneity, *i.e.,* a trait-specific molecular clock. One such trait-specific clock within the Protostomia is introduced in the present paper.

Common to all the approaches above is that they crucially depend on the performance of the gene(s) used, with the lower end of the performance (corresponding to low divergence times) being affected by the "power gap" and the upper end (corresponding to high divergence times) being affected by saturation. In addition, ancestral polymorphism may cause overestimation of divergence times, particularly affecting young phylogenetic events.

Another problem in molecular clock approaches is the estimation of confidence limits of the clock. Whereas many available software tools account for the stochastic nature of the clock, not all tools can account for the uncertainties of calibration points or external molecular clock rates.

Finally, we show that the arguably single most important source of errors in molecular clock estimates is not the underlying statistics, but misinterpretation of the results of clock analyses due to problems with sampling design (missing and extinct taxa).

Nonetheless, the examples presented here for two largely independent molecular clock strategies (calibration point vs. external trait-specific clock) yielded concurrent results, differing by less than 10% (*i.e.,* 7.35 ± 2.69 and 2.63 ± 1.35 Mya vs. 7.11 ± 3.10 and 2.85 ± 1.29 Mya). Although being an isolated case, it adds to the increasing evidence that many problems with molecular clocks and associated data are manageable and that the estimation of meaningful confidence intervals is crucial for a judicious interpretation of the results.

GLOSSARY

Akaike's information criterion (AIC): A method of model selection developed by Akaike (1974). The AIC suggests the best model out of a candidate set of models based on the differences between the individual AIC values of each model. These values are calculated using the likelihood estimator and a penalizing term that increases with the number of model parameters. By doing so, AIC provides a measure of uncertainty of each model rather than a significance value (for further information on model selection see Burnham and Anderson 2002).

Ancestral polymorphism: The amount of heterogeneity that is present in an ancestral population prior to the separation of the descending species. As a consequence, genetic divergence predates species divergence by a certain amount of time. This amount corresponds to the coalescent analogue of the polymorphism in the ancestral species. It averages $2N_e$ generations (with N_e being the effective population size) in a random mating population (see Arbogast *et al.* 2002, also see Fig. 3).

Coalescent theory: A population genetics approach that models the history of gene copies backwards in time. The theory provides a mathematical framework describing the characteristics of coalescent events, *i.e.*, lineages finding their most recent common ancestor (MRCA). The theory was developed by Kingman (1982).

Divergence rate: Substitution rate × 2.

Divergence time: Time since separation of descendent taxa from a most recent common ancestor (MRCA).

Effective population size: Wright (1938) defined the term as "the number of breeding individuals in an idealized population that would show the same amount of dispersion of allele frequencies under random genetic drift or the same amount of inbreeding as the population under consideration". To date, there are several definitions for effective population size (see Ewens 2004).

Generation time effect: Assuming that DNA replication-dependent errors constitute a major fraction of the overall number of mutations, taxa with shorter generation times accumulate more mutations per unit time than taxa with longer generation times. Consequently, the substitution rate of the former would be higher (see Takahata 2007).

Global clock: A global clock assumes a single substitution rate along all branches of a given phylogeny. It is also termed as "strict clock". Note that some workers, however, use this term synonymously with "universal clock".

Likelihood ratio test (LRT): A model testing approach based on the difference in likelihood estimators of two nested models. The LRT approximately follows a chi-squared distribution with q degrees of freedom. If the likelihood ratio exceeds the critical value of the chi-squared distribution with

q degrees of freedom, the difference between both models is considered to be significant.

Local clock: A substitution rate for a specific clade within a given phylogeny. Following this concept, several different rates may be assigned to different clades of a given phylogeny. This approach requires *a priori* information justifying the subdivision in rate-specific clades. Note that some workers, however, use this term for a clock that is applicable to a set of closely related taxa, *i.e.*, a taxon-specific clock.

Molecular clock: A concept that correlates the number of substitutions to time, assuming that (a) the mutations are selectively neutral (or nearly neutral) and (b) the substitution rate is uniform. Consequently, the number of substitutions that separate two gene copies would be a function of the elapsed time since their most recent common ancestor (see Hillis *et al.* 1996).

Molecular clock rate: Number of substitutions per site and year and given as a fraction of one or in percent. Molecular clock rates are either based on substitution or divergence rates (divergence rate = substitution rate × 2). Also see the term "Substitution rate".

Nested models: A model is considered to be nested if it constitutes a special case of a more general, *i.e.*, a more parameter-rich model.

No-clock model: A term for the general model in molecular clock testing approaches against which the nested model (*e.g.*, a global clock) is tested.

Node depth: The distance between a specific node and the tip of an ultrametric phylogenetic tree in number of substitution per site. The value is either given as a fraction of one or as a percent.

Poisson process: Describes the accumulation of discrete, independent events (such as mutations) over time. The waiting times between the events are exponentially distributed.

Relative clock: A concept that correlates number of substitutions per site to relative time. This concept allows for comparing the relative age of MRCAs in a given phylogeny.

Relaxed clock: A dating approach that relaxes the assumption of a single substitution rate within a phylogeny and allows rates to vary. These approaches often assume that rate variation is comparatively small between an ancestral and a descending lineage, *i.e.*, that rates are auto-correlated (see Fig. 2, but see Drummond *et al.* 2006).

Saturation: Difference between the expected and observed number of mutations in a given gene due to multiple (hence "invisible") mutations at one or more sites (Fig. 4).

Strict clock: see "Global clock"

Substitution rate: Here used in terms of number of fixed mutations per site and time unit. In molecular clock approaches, the term "substitution rate" is often used synonymously with the term "molecular clock rate" and given in number of substitution per site and year. As the number of substitutions per site can either be given as a fraction of one or in percent, a substitution rate of, for example, 0.01 My^{-1} equals a rate of 1.0% My^{-1}.

Trait-specific clock: A single molecular clock rate (including error) of a specific gene that can be assigned to a range of taxa that share similar biological and life history characteristics supposedly affecting rate heterogeneity.

Universal clock: A single molecular clock rate (including error) of a specific gene (or group of closely related genes) for all taxa of a broader taxonomic group.

LITERATURE CITED

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716-723.

Albrecht, C., S. Trajanovski, K. Kuhn, B. Streit, and T. Wilke. 2006. Rapid evolution of an ancient lake species flock: Freshwater limpets (Gastropoda: Ancylidae) in the Balkan lake Ohrid. *Organisms, Diversity & Evolution* **6**: 294-307.

Arbogast, B., S. V. Edwards, J. Wakeley, P. Beerli, and J. B. Slowinski. 2002. Estimating divergence time from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics* **33**: 707-740.

Aris-Brosou, S. 2007. Dating phylogenies with hybrid local molecular clocks. *PLoS ONE*: **2**: e879.

Ayala, F. J. 1997. Vagaries of the molecular clock. *Proceedings of the National Academy of Sciences of the U.S.A.* **94**: 7776-7783.

Ayala, F. J. 1999. Molecular clock mirages. *BioEssays* **21**: 71-75.

Bandelt, H. J. 2007. Time dependency of molecular rate estimates: Tempest in a teacup. *Heredity* **100**: 1-2.

Bartoli, G., M. Sarnthein, M. Weinelt, H. Erlenkeuser, D. Garbe-Schönberg, and D. W. Lea. 2005. Final closure of Panama and the onset of northern hemisphere glaciation. *Earth and Planetary Science Letters* **237**: 33-44.

Benton, M. J. and P. C. J. Donoghue. 2007. Palaeontological evidence to date the Tree of Life. *Molecular Biology and Evolution* **24**: 26-53.

Bollback, J. P. 2005. Posterior mapping and predictive distributions. *In*: R. Nielsen, ed., *Statistical Methods in Molecular Evolution*. Springer Verlag New York, Inc., New York. Pp. 439-462.

Braun, E. L. and R. T. Kimball. 2001. Polytomies, the power of phylogenetic inference, and the stochastic nature of molecular evolution: A comment on Walsh *et al.* (1999). *Evolution* **55**: 1261-1263.

Bromham, L. and D. Penny. 2003. The modern molecular clock. *Nature Reviews Genetics* **4**: 216-224.

Bromham, L., M. J. Phillips, and D. Penny. 1999. Growing up with dinosaurs: Molecular dates and the mammalian radiation. *Trends in Ecology and Evolution* **14**: 113-118.

Bromham, L., A. Rambaut, R. Fortey, A. Cooper, and D. Penny. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proceedings of the National Academy of Sciences of the U.S.A.* **95**: 12386-12389.

Brown, W. M., M. George, and A. C. Wilson. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the U.S.A.* **76**: 1967-1971.

Burnham, K. P. and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd Edition. Springer, New York.

Chevaldonné, P., D. Jollivet, D. Desbruyères, R. A. Lutz, and R. C. Vrijenhoek. 2002. Sister-species of eastern Pacific hydrothermal-vent worms (Ampharetidae, Alvinelidae, Vestimentifera) provide new mitochondrial clock calibration. *Cahiers de Biologie Marine* **43**: 367-370.

Cronin, T. M. and H. J. Dowsett. 1996. Biotic and oceanographic response to the Pliocene closing of the Central American isthmus. *In*: J. B. C. Jackson, A. F. Budd, and A. G. Coates, eds., *Evolution and Environment in Tropical America*. University of Chicago Press, Chicago. Pp. 76-104.

Cutler, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. *Molecular Biology and Evolution* **17**: 1647-1660.

Davison, A. 2006. The ovotestis: An underdeveloped organ of evolution. *BioEssays* **28**: 642-650.

Doyle, J. A. and M. J. Donoghue. 1993. Phylogenies and angiosperm diversification. *Paleobiology* **19**: 141-167.

Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow. 1998. Rates of spontaneous mutation. *Genetics* **148**: 1667-1686.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Drummond, A., O. G. Pybus, and A. Rambaut. 2002. Inference of viral evolutionary rates from molecular sequences. *Advances in Parasitology* **54**: 331-358.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* **4**: e88.

Drummond, A. J., O. G. Pybus, A. Rambaut, R. Forsberg, and A. G. Rodrigo. 2003. Measurably evolving populations. *Trends in Ecology and Evolution* **18**: 481-488.

Edwards, S. V. and P. Beerli. 2000. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* **54**: 1839-1854.

Ewens, W. J. 2004. *Mathematical Population Genetics*. Springer, New York.

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. *Annual Review of Genetics* **22**: 521-565.

Folmer, O., M. Black, W. Hoeh, R. A. Lutz, and R. C. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* **3**: 294-299.

Foltz, D. W., A. W. Hrincevich, and A. Rocha-Olivares. 2004. Apparent selection intensity for the cytochrome oxidase subunit I gene varies with mode of reproduction in echinoderms. *Genetica* **122**: 115-125.

Geyer, J., T. Wilke, and E. Petzinger. 2006. The solute carrier family SLC10: More than a family of bile acid transporters regarding function and phylogenetic relationships. *Naunyn-Schmiedeberg's Archives of Pharmacology* **372**: 413-431.

Gillooly, J. F., A. P. Allen, G. B. West, and J. H. Brown. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences of the U.S.A.* **102**: 140-145.

Govindarajan, A. F., K. M. Halanych, and C. W. Cunningham. 2005. Mitochondrial evolution and phylogeography in the hydrozoan *Obelia geniculata* (Cnidaria). *Marine Biology* **146**: 213-222.

Hedges, S. B. and S. Kumar. 2004. Precision of molecular time estimates. *Trends in Genetics* **20**: 242-247.

Hedges, S. B., P. H. Parker, C. G. Sibley, and S. Kumar. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**: 226-229.

Hellberg, M. E. and V. D. Vacquier. 1999. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Molecular Biology and Evolution* **16**: 839-848.

Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Applications of molecular systematics: The state of the field and a look to the future. *In*: D. M. Hillis, C. Moritz, and B. K. Mable, eds., *Molecular Systematics*. Sinauer Associates, Sunderland, Massachusetts. Pp. 515-543.

Ho, S. Y. W. 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. *Journal of Avian Biology* **38**: 409-414.

Kelchner, S. A. and M. A. Thomas. 2007. Model use in phylogenetics: Nine key questions. *Trends in Ecology and Evolution* **22**: 87-94.

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* **217**: 624-626.

Kingman, J. F. C. 1982. The coalescent. *Stochastic Processes and Their Applications* **13**: 235-248.

Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution* **18**: 352-361.

Knowlton, N. and L. A. Weigt. 1998. New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society London* (B) **265**: 2257-2263.

Krijgsman, W., F. J. Hilgen, I. Raffi, F. J. Sierro, and D. S. Wilson. 1999. Chronology, causes and progression of the Messinian salinity crisis. *Nature* **400**: 652-655.

Lambert, D. M., P. A. Ritchie, C. D. Millar, B. Holland, A. J. Drummond, and C. Baroni. 2002. Rates of evolution in ancient DNA from Adélie penguins. *Science* **295**: 2270-2273.

Lanfear, R., J. A. Thomas, J. J. Welch, T. Brey, and L. Bromham. 2007. Metabolic rate does not calibrate the molecular clock. *Proceedings of the National Academy of Sciences of the U.S.A.* **104**: 15388-15393.

Li, W.-H. and D. Graur. 1991. *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Luttikhuizen, P. C., J. Drent, W. van Delden, and T. Piersma. 2003. Spatially structured genetic variation in a broadcast-spawning bivalve: Quantitative versus molecular traits. *Journal of Evolutionary Biology* **16**: 260-272.

Marshall, C. R. 1990. The fossil record and estimating divergence times between lineages: Maximum divergence times and the importance of reliable phylogenies. *Journal of Molecular Evolution* **30**: 400-408.

Nei, M. 1987. *Molecular Evolutionary Genetics*, 1st Edition. Columbia University Press, New York.

Nylander, J. A. A. 2004. *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University, Sweden.

Paxinos, E. E., H. F. James, S. L. Olson, M. D. Sorenson, J. Jackson, and R. C. Fleischer. 2002. mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (*Branta canadensis*). *Proceedings of the National Academy of Sciences of the U.S.A.* **99**: 1399-1404

Posada, D. and T. Buckley. 2004. Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793-808.

Posada, D. and K. A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Pulquerio, M. J. F. and R. A. Nichols. 2007. Dates from the molecular clock: How wrong can we be? *Trends in Ecology and Evolution* **22**: 180-184.

Ronquist, F. and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572-1574.

Rutschmann, F. 2006. Molecular dating of phylogenetic trees: A brief review of current methods that estimate divergence times. *Diversity and Distributions* **12**: 35-48.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218-1231.

Sanderson, M. J. 1998. Estimating rate and time in molecular phylogenies: Beyond the molecular clock? *In*: P. Soltis, D. Soltis, and J. Doyle, eds., *Plant Molecular Systematics*, II. Kluwer, Boston. Pp. 242–264.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* **19**: 101-109.

Sarich, V. M. and A. C. Wilson. 1967. Immunological time scale for hominid evolution. *Science* **158**: 1200-1203.

Schubart, C. D., R. Diesel, and S. B. Hedges. 1998. Rapid evolution to terrestrial life in Jamaican crabs. *Nature* **393**: 363-365.

Takahata, N. 2007. Molecular Clock: An Anti-neo-Darwinian Legacy. *Genetics* **176**: 1-6.

Takezaki, N., A. Rzhetsky, and M. Nei. 1995. Phylogenetic test of the molecular clock and linearized trees. *Molecular Biology and Evolution* **12**: 823-833.

Thomas, J. A., J. J. Welch, M. Woolfit, and L. Bromham. 2006. There is no universal molecular clock for invertebrates, but rate variation does not scale with body size. *Proceedings of the National Academy of Sciences of the U.S.A.* **103**: 7366-7371.

Thorne, J. L., H. Kishino, and I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Molecular Biology and Evolution* **15**: 1647-1657.

Walsh, H. E. and V. L. Friesen. 2001. Power and stochasticity in the resolution of soft polytomies: A reply to Braun *et al. Evolution* **55**: 1264-1266.

Walsh, H. E., M. G. Kidd, T. Moum, and V. L. Friesen. 1999. Polytomies and the power of phylogenetic inference. *Evolution* **53**: 932-937.

Weir, J. T. and D. Schluter. 2008. Calibrating the avian molecular clock. *Molecular Ecology* **17**: 2321-2328.

Welch, J. J. and L. Bromham. 2005. Molecular dating when rates vary. *Trends in Ecology and Evolution* **20**: 320-327.

Wilke, T. 2003. *Salenthydrobia* n. gen. (Rissooidea: Hydrobiidae): A potential relict of the Messinian salinity crisis. *Zoological Journal of the Linnean Society* **137**: 319-336.

Wilke, T. 2004. How dependable is a non-local molecular clock? A reply to Hausdorf *et al.* (2003). *Molecular Phylogenetics and Evolution* **30**: 835-840.

Wilke, T., C. Albrecht, V. V. Anistratenko, S. K. Sahin, and M. Z. Yildirim. 2007. Testing biogeographical hypotheses in space and time: Faunal relationships of the putative ancient lake Egirdir in Asia Minor. *Journal of Biogeography* **34**: 1807-1821.

Wilson, A. C. and V. M. Sarich. 1969. A molecular time scale for human evolution. *Proceedings of the National Academy of Sciences of the U.S.A.* **63**: 1088-1093.

Wright, S. 1938. Size of population and breeding structure in relation to evolution. *Science* **87**: 430-431.

Xia, X. and Z. Xie. 2001. DAMBE: Software Package for Data Analysis in Molecular Biology and Evolution. *The Journal of Heredity* **92**: 371-373.

Xia, X., Z. Xie, M. Salemi, L. Chen, and Y. Wang, Y. 2003. An index of substitution saturation and its application. *Molecular Phylogenetics and Evolution* **26**: 1-7.

Yang, Z. 2004. A heuristic rate smoothing procedure for maximum likelihood estimation of species divergence times. *Acta Zoologica Sinica* **50**: 645-656.

Zuckerkandl, E. and L. Pauling. 1962. Molecular disease, evolution, and genic heterogeneity. *In*: M. Kasha and B. Pullman, eds., *Horizons in Biochemistry*. Academic Press, New York. Pp. 189-225.

Zuckerkandl, E. and L. Pauling. 1965. Evolutionary divergence and convergence in proteins. *In*: V. Bryson and H. J. Vogel, eds., *Evolving Genes and Proteins*. Academic Press, New York. Pp. 97-166.

Submitted: 9 April 2009; **accepted:** 15 April 2009; **final revisions received:** 1 May 2009