Ciências da Terra (UNL) | Lisboa | Nº 18 | pp. 53-56 | 2013

Phylogeny and Palaeobiodiversity

Michel Laurin
UMR 7207, Muséum National d'Histoire Naturelle, Bâtiment de Géologie, Case postale 48, 43 rue Buffon,
F-75231 Paris cedex 05, France

Summary

The development of the phylogenetic diversity index now allows, at least in theory, a close integration of of phylogeny and paleobiodiversity, but in practice, this unification is only incipient. It has been hampered by the lack of generally-applicable methods to determine the actual appearance dates of taxa (rather than the younger dates at which they become visible in the fossil record). Most recent works on producing timetrees have emphasized molecular data, a suboptimal situation because the most direct evidence of the chronology of taxonomic diversification is the fossil record, which is now under-exploited. Fortunately, progress is being made in using the fossil record to produce timetrees. Two complementary approaches have been pursued: total evidence dating (based on molecular and morphological data), and the use of birth and death models.

Keywords: phylogeny, timetrees, birth and death processes, total evidence dating, molecular dating

The interactions between phylogeny and paleobiodiversity used to be tenuous at best. Until very recently, most paleobiodiversity estimates were simple counts of taxa of a given rank (species, genera, families, or even orders, in decreasing order of precision thus obtained), and such simple taxonomic counts continue to be widely used in paleontological research (e.g., Peters, 2008; Sahney *et al*., 2010). This is very unfortunate because the limitations of using taxon counts (especially supra-specific taxa) in comparative biology are well known (Bertrand *et al*., 2006; Laurin, 2010). It could be argued that an explicit framework linking both was proposed only in the 1990s, when Faith (1992) proposed a biodiversity index that fully uses phylogenetic information, in the form of a topology with branch lengths. However, the potential of this method has been exploited so far in very few studies (e.g., Marjanović & Laurin, 2008), although a few other paleontological studies have integrated the phylogeny into paleobiodiversity studies through other techniques (e.g., Ruta *et al*., 2007; Benson *et al*., 2012).

Paleontologists make increasing use of phylogenies to assess paleobiodiversity and its evolution through time, although in many cases, intermediate approaches combining use of rank-based nomenclature with phylogeny are used (e.g., Ruta & Benton, 2008). The most sophisticated methods to study the evolution of biodiversity remain to be integrated into paleontological research. Thus, the method of Nee *et al*. (1994a, b) to study speciation and extinction rates through time using molecular timetrees, along with other methods based on such timetrees of extant taxa, has been used extensively to assess macroevolutionary patterns (Morlon *et al*., 2010). Because of the proliferation of such studies based on molecular timetrees, fossil data have been directly exploited in only a minority of studies on the evolution of biodiversity, which is a suboptimal situation because the fossil record provides the most direct source of information about paleobiodiversity. Among the many problems raised by this situation is that estimates of extinction rates based on such methods are known to be fairly inaccurate and under-estimated (Paradis, 2004). This unfortunate situation prevails because the recent, sophisticated timetree-based methods used to study the evolution of biodiversity require substantial modification to be able to incorporate data from the fossil record. Progress in this direction was recently made on several fronts. For instance, Morlon *et al.* (2011) presented a more sophisticated timetree-based method, applied it to a cetacean dataset, and showed that it produced results compatible with data from the fossil record. Didier *et al.* (2012) modified the method of Nee *et al.* (1994a, b) to directly incorporate fossils into timetrees and estimate speciation, extinction, and fossilization rates from the tree. Their simulations showed that the incorporation of fossil data significantly reduces error on parameter estimates, even with a fairly sparse fossil record in the order of 10% of species present in the record. However, this method remains to be applied to an empirical dataset.

Given that timetrees are so essential to assess paleobiodiversity and its evolution through time, a review of recent progress in methods used to date the tree of Life is relevant. Timetrees were based mostly, if not exclusively, on the fossil record for much of the 20th century, to the point that some authors maintained that the evolutionary history of a taxon lacking a fossil record could not be known (Gingerich, 1979, p. 454). The advent of molecular systematics drastically changed this situation; most recent timetrees have been produced using molecular dating techniques in which little data from the fossil record was used to establish time constraints, most of which are only miminal age constraints, on the ages of various nodes (e.g., Roelants *et al.*, 2007). This strategy is suboptimal because using the fossil record mostly to provide minimal ages can be expected to inflate molecular ages (Marjanović & Laurin, 2007), and this may explain many disagreements between paleontologists and molecular biologists about the age of various taxa, such as Lissamphibia (Marjanović & Laurin, 2007; Roelants *et al.*, 2007) or Metazoa (Hug & Roger, 2007). The best strategy to improve our age estimates for taxa would be to collect more fossil data and produce more sophisticated methods to analyze these because even the shape of the probability distribution of priors of node ages (for Bayesian methods) has a strong impact on molecular ages, as shown by Ho & Phillips (2009) for birds. Fortunately, much progress has been made recently on methods that can extract most probable taxon ages (as opposed to minimal ages) from the fossil record (Laurin, 2012). Early attempts in this direction focused on obtaining confidence intervals on the maximal ages of clades (Strauss & Sadler, 1989). However, these methods are difficult to apply to large clades because they require much data, because they require a uniform fossil recovery potential through time (Strauss & Sadler, 1989; Marshall, 2008) or require a fossil recovery potential curve obtained without reference to the fossil record, but able to explain the latter well (Marshall, 1997). The latter can be poorly constrained near the ends of stratigraphic ranges of taxa, which results in similarly poorly constrained confidence intervals (Marjanović & Laurin, 2008).

Two promising strategies to make better use of the fossil record to date the tree of Life have been explored recently. One, called “total evidence dating”, consists in incorporating extinct taxa from the fossil record directly into “total evidence” data matrices, which incorporate molecular sequences from extant taxa, as well as morphological data from extant and extinct taxa. Phylogenetic analyses and dating are then performed on the combined dataset, with node ages linked to extinct taxa dated using the ages of the fossils considered to be at the tips of branches, as well as the length of these branches, estimated (in the case of extinct taxa) from their length (number of inferred transitions) and inferred evolutionary rates. This method is attractive because it automatically incorporates uncertainty about the affinities of fossils (which is considered in the phylogenetic analysis) and avoids having to select prior ages for various nodes. However, this method requires much data and given the complexity of the calculations involved, is applicable only to moderately-sized datasets. So far, it has been used to date Lissamphibia (Pyron, 2011) and Hymenoptera (Ronquist *et al.*, 2012a). It is now implemented in the MrBayes 3.2 (Ronquist *et al.*, 2012b).

The second promising strategy is based on birth and death processes. Using such processes, which, in the simplest situation, assume constant speciation and extinction rates (these assumptions can be relaxed), data about biodiversity at various times in the history of a clade can be used to get a probability distribution about the origin of this clade. This method has been used to assess the probable time of origin of anthropoids and of primates

(Wilkinson & Tavaré, 2009). In the implementation used by Wilkinson & Tavaré (2009), the data used were the paleobiodiversity of primates and anthropoids from 14 time bins plus the extant biodiversity of these clades, along with the age of the oldest fossil of both clades. This method thus makes use of very little phylogenetic data. The analyses suggested that the 95% confidence on the age of primates extends up to 99 Ma, in the middle of the Cretaceous, which illustrates that the discrepancy between the minimal age and the most probable age of a taxon, estimated from fossil data, can differ drastically, even when its fossil record is fairly extensive. However, it is possible that incorporating more phylogenetic data into such analyses might lead to narrower confidence intervals; this question should be tackled soon, as it may be one of the most promising strategies to better date the tree of Life.

References

BENSON R. B. J., EVANS M. & DRUCKENMILLER P. S. (2012) – High diversity, low disparity and small body size in plesiosaurs (Reptilia, Sauropterygia) from the Triassic–Jurassic boundary. *PLoS ONE* 7 (3), 1–15.

BERTRAND Y., PLEIJEL F. & ROUSE G. W. (2006) – Taxonomic surrogacy in biodiversity assessments, and the meaning of Linnaean ranks. *Systematics and Biodiversity* 4 (2), 149-159.

DIDIER G., ROYER-CARENZI M. & LAURIN M. (2012) – The reconstructed evolutionary process with the fossil record. *Journal of Theoretical Biology* 315, 26–37.

FAITH D. P. (1992) – Conservation evaluation and phylogenetic diversity. *Biological Conservation* 61 (1), 1–10.

GINGERICH P. D. (1979) – Paleontology, phylogeny, and classification: an example from the mammalian fossil record. *Systematic Zoology* 28 (4), 451–464.

HO S. Y. W. & PHILLIPS M. J. (2009) – Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* 58 (3), 367–380.

HUG L. A. & ROGER A. J. (2007) – The impact of fossils and taxon sampling on ancient molecular dating analyses. *Molecular Biology and Evolution* 24 (8), 1889–1897.

LAURIN M. (2010) – The subjective nature of Linnaean categories and its impact in evolutionary biology and biodiversity studies. *Contributions to Zoology* 79 (4), 131–146.

LAURIN M. (2012) – Recent progress in paleontological methods for dating the Tree of Life. *Frontiers in Genetics* 3 (130), 1–16.

MARJANOVIC D. & LAURIN M. (2007) – Fossils, molecules, divergence times, and the origin of lissamphibians. *Systematic Biology* 56 (3), 369–388.

MARJANOVIĆ D. & LAURIN M. (2008) – Assessing confidence intervals for stratigraphic ranges of higher taxa: the case of Lissamphibia. *Acta Palaeontologica Polonica* 53 (3), 413–432.

MARSHALL C. R. (1997) – Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiology* 23 (2), 165–173.

MARSHALL C. R. (2008) – A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. *The American Naturalist* 171 (6), 726–742.

MORLON H., PARSONS T. L. & PLOTKIN J. B. (2011) – Reconciling molecular phylogenies with the fossil record. *Proceedings of the National Academy of Sciences of the United States of America* 108 (39), 16327–16332.

MORLON H., POTTS M. D. & PLOTKIN J. B. (2010) – Inferring the dynamics of diversification: a coalescent approach. *PLoS Biology* 8 (9), 1–13.

NEE S., HOLMES E. C., MAY R. M. & HARVEY P. H. (1994a) – Extinction rates can be estimated from molecular phylogenies. *Philosophical Transactions of the Royal Society of London,* Series B 344 (1307), 77–82.

NEE S., MAY R. M. & HARVEY P. H. (1994b) – The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society of London,* Series B 344 (1309), 305–311.

PARADIS E. (2004) – Can extinction rates be estimated without fossils? *Journal of Theoretical Biology* 229, 19-30.

PETERS S. E. 2008. Environmental determinants of extinction selectivity in the fossil record. *Nature* 454 (7204), 626–629.

PYRON R. A. (2011) – Divergence-time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* 60 (4), 466–481.

ROELANTS K., GOWER D. J., WILKINSON M., LOADER S. P., BIJU S. D., GUILLAUME K., MORIAU L. & BOSSUYT F. (2007) – Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences of the United States of America* 104 (3), 887-892.

RONQUIST F., KLOPFSTEIN S., VILHELMSEN L., SCHULMEISTER S., MURRAY D. L. & RASNITSYN A. (2012a) – A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* 61 (6), 973–999.

RONQUIST F., TESLENKO M., VAN DER MARK P., AYRES D. L., DARLING A., HÖHNA S., LARGET B., LIU L., SUCHARD M. A. & HUELSENBECK J. P. (2012b) – MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61 (3), 539–542.

RUTA M. & BENTON M. J. (2008) – Calibrated diversity, tree topology and the mother of mass extinctions the lesson of temnospondyls. *Palaeontology* 51 (6), 1261–1288.

RUTA M., PISANI D., LLOYD G. T. & BENTON M. J. (2007) – A supertree of Temnospondyli: cladogenetic patterns in the most species-rich group of early tetrapods. *Proceedings of the Royal Society of London,* Series B 274 (1629), 3087–3095.

SAHNEY S., BENTON M. J. & FALCON-LANG H. J. (2010) – Rainforest collapse triggered Carboniferous tetrapod diversification in Euramerica. *Geology* 38 (12), 1079–1082.

STRAUSS D. & SADLER P. M. (1989) – Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Mathematical Geology* 21 (4), 411–427.

WILKINSON R. D. & TAVARE S. (2009) – Estimating primate divergence times by using conditioned birth-and-death processes. *Theoretical Population Biology* 75 (4), 278–285.