C. R. Palevol xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Comptes Rendus Palevol

www.sciencedirect.com

General palaeontology, systematics and evolution (Phylogenetic analysis)

A practical guide to molecular dating

Guide pratique de la datation moléculaire

Hervé Sauquet

Université Paris-Sud, Laboratoire Écologie, Systématique, Évolution, CNRS UMR 8079, 91405 Orsay, France

ARTICLE INFO

Article history:
Received 7 May 2013
Accepted after revision 18 July 2013
Available online xxx

Written on invitation of the Editorial board

Keywords:
Age constraint
Confidence interval
Fossil calibration
Molecular dating
Relaxed clock
Uncertainty

ABSTRACT

Molecular dating has now become a common tool for many biologists and considerable methodological improvements have been made over the last few years. However, the practice of estimating divergence times using molecular data is highly variable among researchers and it is not straightforward for a newcomer to the field to know how to start. Here I provide a brief overview of the current state-of-the-art of molecular dating practice. I review some of the important choices that must be made when conducting a divergence time analysis, including how to select and use calibrations and which relaxed clock model and program to use, with a focus on some practical aspects. I then provide some guidelines for the interpretation of results and briefly review some alternatives to molecular dating for obtaining divergence times. Last, I present some promising developments for the future of the field, related to the improvement of the calibration process.

Mots clés :
Contrainte d'âge
Intervalle de confiance
Calibration fossile
Datation moléculaire
Horloge relâchée
Incertitude

RÉSUMÉ

La datation moléculaire est devenue un outil commun pour de nombreux biologistes et des progrès méthodologiques considérables ont été apportés ces dernières années. Cependant, l'estimation des temps de divergence à partir de données moléculaires demeure très variable dans sa pratique et sa mise en œuvre est délicate pour le novice. Cet article propose une revue concise de l'état de l'art de la pratique de la datation moléculaire. Plusieurs décisions importantes sont nécessaires, notamment la sélection et l'implémentation des calibrations, ainsi que le choix d'un modèle d'horloge relâchée et le logiciel pour l'analyse. Après une discussion de certains aspects pratiques de la conduite de ces analyses, plusieurs règles sont proposées pour l'interprétation des résultats. Les alternatives possibles pour obtenir des temps de divergence sont également discutées. Enfin, plusieurs développements prometteurs pour l'avenir de la discipline sont soulignés, en particulier dans le cadre de l'amélioration du processus de calibration.

1. Introduction

Estimating divergence times among species or lineages using molecular sequence data, commonly referred to as molecular dating, has now become a commonplace tool for

E-mail address: herve.sauquet@u-psud.fr

1631-0683/$ – see front matter © 2013 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.crpv.2013.07.003

many evolutionary biologists and ecologists. Although the idea was laid more than 50 years ago, the field of molecular dating has now become more mature, largely thanks to considerable methodological improvements brought over the last 15 years. However, many questions remain and significant improvements and discoveries are still being made on various aspects of the method. Yet, the practice of molecular dating varies considerably in quality. Numerous reviews have been written on the topic (Donoghue and Benton, 2007; Forest, 2009; Kumar, 2005; Laurin, 2012; Magallón, 2004; Pulquério and Nichols, 2007; Renner, 2005; Rutschmann, 2006). In this paper, I do not intend to provide yet another review of the field or a hands-on tutorial based on a specific example. Instead, my aim is to provide practical guidelines to design, conduct, and interpret a molecular dating experiment according to the state-of-the-art, citing extensive reviews and key studies where appropriate.

The idea of molecular dating was originally proposed in a paper by Zuckerkandl and Pauling (1962), who suggested that the divergence time between two species could be measured by the number of differences between two molecular sequences (in their case, protein sequences). This was centered on the assumption of a molecular clock, whereby the rate of molecular evolution remains constant through time. Although this was initially envisioned as an alternative and independent method to using the fossil record, it is widely accepted today that fossils should be used to calibrate the molecular clock (except in special cases such as virus phylogenies when virus sequences can be sampled through recent time). Indeed, molecular sequence divergences can only provide a relative time scale. Calibration from another source of information is always required in order to convert relative into absolute divergence times. At about the same time when technological progress allowed biologists to routinely obtain molecular sequences and reconstruct phylogenies using them, important methodological advances were made in molecular dating methods. These advances largely focused on relaxing the assumption of a molecular clock. Indeed, in many studies, this assumption appeared to be contradicted by lineages with molecules evolving particularly fast or slow. Therefore, methods were developed that allowed molecular rates to vary through time and across lineages (Rutschmann, 2006; Sanderson et al., 2004). These so-called *relaxed clock* methods are now used preferentially by most researchers. Finally, increased attention has been given recently in improving calibration of molecular dating studies. It is now commonly accepted that not only one, but instead multiple fossil calibrations are required for accurate estimation of divergence times (Ho and Phillips, 2009; Magallón et al., 2013; Sauquet et al., 2012). New standards are being proposed for documenting and justifying these calibrations as well as implementing them in the statistical framework of Bayesian relaxed clocks.

In this paper, we will first ask whether a molecular dating study is needed or not to answer a particular biological question. We will then go through the various stages of designing a molecular dating analysis, from choosing genes, taxa, and calibrations to selecting a relaxed clock model and software to compute the analysis. After giving some practical advice about the analysis itself, I will suggest some guidelines for interpreting the results. Finally, we will look at some alternative ways to obtain divergence times quickly, and I will briefly outline some exciting developments to expect in the field over the next few years.

2. To date or not to date: is molecular dating essential for my study?

First, it is essential to state that paleontological dating is an alternative to molecular dating only for very few branches of the Tree of Life and for a restricted set of applications. Indeed, even for the most fossil-rich taxa with well-known phylogenetic relationships to extant taxa, a fossil date cannot be given for every single divergence in a given phylogeny (except in exceptional circumstances). Molecular dating, calibrated with multiple fossils, must thus be seen as a reasonable alternative to obtain divergence time estimates for all nodes of a phylogeny, when such information is required to conduct further analyses.

Molecular dating has been used to answer a wide range of questions. Many of these applications fall in the following categories.

2.1. Biogeography

In order to compare the evolutionary history of a group with global processes of the Earth, an absolute time scale is required. For example, molecular dating studies have severely challenged the commonly assumed role of vicariance (due to plate tectonics) in shaping the biogeographic distribution of many clades (for reviews, see Crisp et al., 2011; Renner, 2005). Furthermore, current probabilistic approaches to reconstructing biogeographic history typically rely on dated trees, such as the dispersal-extinction-cladogenesis (DEC) model implemented in Lagrange (Ree and Smith, 2008). Another example where molecular dating studies have played an important role is whether past climate change had an influence on the diversification of large clades at a regional scale (Linder, 2003).

2.2. Diversification rates

Over the last few years, considerable interest has grown in estimating speciation and extinction rates from molecular phylogenies, with the continuous development of increasingly complex models accounting for variable rates through time and across lineages (Alfaro et al., 2009; FitzJohn et al., 2009; Morlon et al., 2011; Stadler, 2011). All of these methods require an absolute or relative time scale as a prerequisite.

2.3. Comparative methods

A very wide range of contemporary methods in comparative biology rely on dated trees, for instance to infer ancestral states or to test the correlation between two traits (for reviews, see O'Meara, 2012; Pagel, 1999). However, for many methods, nonultrametric trees obtained from phylogenetic analyses (i.e., phylograms, with branch lengths

Please cite this article in press as: Sauquet, H., A practical guide to molecular dating. C. R. Palevol (2013), http://dx.doi.org/10.1016/j.crpv.2013.07.003