C. R. Biologies 334 (2011) 620–628

Contents lists available at ScienceDirect

Comptes Rendus Biologies

www.sciencedirect.com

Evolution/Évolution

Yeasty clocks: Dating genomic changes in yeasts

Horloges tremblantes : datation des changements génomiques chez les levures

Thomas Rolland, Bernard Dujon [*]

Unité de génétique moléculaire des levures (CNRS URA2171 and University P.-M.-Curie UFR927), Institut Pasteur, 25, rue du Docteur-Roux, 75724 Paris cedex 15, France

ARTICLE INFO

Article history:
Received 12 November 2010
Accepted after revision 17 March 2011
Available online 1 July 2011

Keywords:
Evolution
Mutational rate
Polymorphism
Divergence times
Synteny conservation

Mots clés :
Évolution
Taux de mutation
Polymorphisme
Temps de divergence
Conservation de synténie

ABSTRACT

Calibration of clocks to date evolutionary changes is of primary importance for comparative genomics. In the absence of fossil records, the dating of changes during yeast genome evolution can only rely on the properties of the genomes themselves, given the uncertainty of extrapolations using clocks from other organisms. In this work, we use the experimentally determined mutational rate of *Saccharomyces cerevisiae* to calculate the numbers of successive generations corresponding to observed sequence polymorphism between strains or species of other yeasts. We then examine synteny conservation across the entire subphylum of *Saccharomycotina* yeasts, and compare this second clock based on chromosomal rearrangements with the first one based on sequence divergence. A non-linear relationship is observed, that interestingly also applies to insects although, for equivalent sequence divergence, their rate of chromosomal rearrangements is higher than that of yeasts.

RÉSUMÉ

L'étalonnage d'horloges moléculaires pour dater les changements évolutifs a une grande importance pour la génomique comparative. En l'absence de fossiles, la datation des changements durant l'évolution des génomes de levures ne peut se baser que sur les propriétés des génomes eux-mêmes, étant donnée l'incertitude des extrapolations à partir d'horloges d'autres organismes. Dans ce travail, nous utilisons le taux de mutation expérimentalement déterminé chez *Saccharomyces cerevisiae* pour calculer les nombres de générations successives correspondant aux degrés de polymorphisme de séquences observés entre souches ou espèces d'autres levures. Nous examinons ensuite la conservation de synténie à travers tout le sous-embranchement des levures *Saccharomycotina*, et comparons cette seconde horloge basée sur les réarrangements chromosomiques avec la première basée sur la divergence de séquence. Une relation non-linéaire est observée, qui s'applique également aux insectes bien que, pour une divergence de séquence équivalente, leur taux de réarrangements chromosomiques soit plus élevé que celui des levures.

---

[*] Corresponding author.
E-mail address: bdujon@pasteur.fr (B. Dujon).

1631-0691/$ – see front matter © 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.crvi.2011.05.010

1. Introduction

The concept of molecular evolutionary clocks is central to modern comparative genomics. From the pioneering work of Zuckerlandl and Pauling [1], it is commonly admitted that amino-acid substitutions between orthologous proteins accumulate with the time separating them from their common ancestor, and differences between aligned sequences are, therefore, used to build phylogenetic trees and to estimate the dates of separation between living species (or groups of species). With the increasing availability of genome sequence data, it became clear, however, that the rate at which protein sequences evolve varies among lineages [2], leading to the idea of relaxed molecular clocks [3–5], and raising the question of appropriate calibration to date major phylogenetic separations. In fungi, for example, this problem was remarkably illustrated by the work of Taylor and Berbee [6]: depending upon the reference used to calibrate the clock, the separation date between *Ascomycota* and *Basidiomycota* varies between 400 and 1800 Myr. Similarly, the origin of *Saccharomycotina* (budding yeasts) is dated, according to calibrations, at 250 Myr ago or 900 Myr ago, i.e. a range of uncertainty linking the Permian-Trias transition to deep precambrian times. Even when calibration is properly set, extrapolation of molecular clocks to large evolutionary scales can only give seemingly precise results if one takes the statistical limits of confidence into proper consideration [7]. Greater precision would require independent calibration points within short evolutionary timescales using increased taxon sampling or continuous fossil records, two conditions not always readily accessible. The identification of *Paleopyrenomycites devonicus* as the oldest fossil ascomycete dated to 400 Myr [8] played an important role to calibrate the fungal tree of life, but such fossils remain rare in fungi. Also, they are non-existent in yeasts, if one excepts amber inclusions which have received only limited attention so far [9,10] and are, anyway, too recent for setting clocks over long evolutionary times. Increasing taxon sampling is not easier for yeasts, since it is unlikely that living intermediates exist, given their very mode of propagation that creates constant bottlenecks.

Another important problem for dating using molecular data is that substitution rates also vary between the different genes of a same organism. In yeasts, for example, a dispersion of nearly three orders of magnitude exists in the rate of non-synonymous substitutions per site (dN) between the fastest and the slowest evolving proteins [11]. The dispersion is lower in organisms with smaller genetically effective population sizes such as *Drosophila* and mammals [12], hence the necessity to compare homogeneous groups of organisms sharing similar life style and mode of propagation to properly date evolutionary changes. Yeasts offer such a case with more than three dozens of species fully sequenced [13] and population genomic studies now available for a few of them [14,15]. These fungi proved particularly meaningful to elucidate the mechanisms of unicellular eukaryotic genome evolution by allowing us to easily confront hypotheses based on comparative genome analysis with the results of direct experimental approaches [16]. Most yeasts whose genomes have been fully sequenced so far belong to the *Saccharomycotina* (also called hemiascomycetes), a large subphylum of *Ascomycota* that includes *Saccharomyces cerevisiae*. Despite the conservation of their unicellular mode of life with bud formation, these yeasts cover a very broad evolutionary range, and very important degrees of sequence divergence exist between orthologous genes of distinct yeast species, even those belonging to the same clade [17,18]. Dating major evolutionary changes in yeast genomes, such as the change of codon assignation in the CTG group [19], the triplication of mating cassettes in *Saccharomycetaceae* [13], or the whole-genome duplication in the ancestry of *Saccharomyces sensu stricto* and related clades [20], remains, therefore, highly imprecise. Phylogenetic interpolation within the fungal tree of life has been attempted [21–23], but the specific mode of propagation of yeasts with rapid clonal expansions raises the question of the validity of the comparisons with multicellular organisms having obligate sexual reproduction and possibly distinct evolutionary rates. A specific calibration of the molecular clock of yeasts is, therefore, desirable. But, besides the genomic changes themselves, no independent piece of information such as fossils records, is available to cover their very large evolutionary range.

In this work, we have addressed this question from two different viewpoints. Starting from the mutation rates that have been precisely measured by experiments in *S. cerevisiae* [24–26], we have computed the minimal number of successive generations separating distinct lineages in this yeast, and extrapolated similar calculations to the separation of species within clades. This clock is appropriate for short evolutionary timescales but gradually loses precision with increasing evolutionary range. We have, therefore, looked for a second clock more appropriate to larger evolutionary timescales by examining the relationship between sequence divergence and degrees of chromosomal rearrangements. This relationship has been quantitatively established over the entire evolutionary range of *Saccharomycotina*, and compared to a similar relationship established for insects.

2. Calibrating sequence divergence in terms of the minimal number of successive generations

The spontaneous mutation rate has recently been determined with precision in *S. cerevisiae* by three independent approaches. A per-base-pair mutation rate (μ) was established for two genes using the classical Luria-Delbrück fluctuation assays [24]. Figures of 3.80×10^{-10} and 6.44×10^{-10} mutations per nucleotide per generation were obtained for the *URA3* and the *CAN1* genes, respectively, indicating that, even if not entirely uniform across the genome, the mutation rate shows a limited variation range (ca. two times). An independent estimation of the per-base-pair mutation rate (μ) along the entire genome was obtained using novel sequencing technology in mutation-accumulation experiments [25]. Partial resequencing (ca. 40% genome coverage) of four independent cultures of *S. cerevisiae* grown in rich medium for a total of ca. 4800 generations after 200 successive single-cell