

Systematic Palaeontology (Palaeobotany)

Temporal information, fossil record and phylogeny

Hervé Lelièvre^{a,*}, René Zaragüeta Bagils^b, Isabelle Rouget^c

^a *Muséum national d'histoire naturelle, UMR 5143, 8, rue Buffon, case 38, 75231 Paris cedex 05, France*

^b *Université Pierre-et-Marie-Curie (Paris-6), UMR 5143, laboratoire « Informatique et systématique », 12, rue Cuvier, 75005 Paris, France*

^c *Université Pierre-et-Marie-Curie (Paris-6), UMR 5143, « Paléobiodiversité et paléoenvironnements », laboratoire 'Informatique, systématique, recherche informatique et structuration des cladogrammes', 4, place Jussieu, case 104, 75252 Paris cedex 05, France*

Received 25 May 2007; accepted after revision 17 December 2007

Available online 11 February 2008

Presented by Philippe Taquet

Abstract

Fossils, i.e. remains of living organisms, provide critical data to reconstruct the history of life on Earth because they are the source of unique information concerning the past. Fossils allow statements to be made about morphological features that would otherwise remain unknown, that also help clarifying the phylogenetic relationships of organisms, which results in a better knowledge of the fossil record and, partly, of the stratigraphic record. The latter constitutes the basic source for palaeontologists in order to reconstruct the History of Life. Some palaeontologists consider that the stratigraphical record provides us with essential information necessary to propose phylogenetic hypotheses and to build evolutionary scenarios. However, this point of view is not generally accepted, some authors putting forward that the fossil record is too fragmentary to be accurate (see, e.g.: G.J. Nelson, N.I. Platnick, *Systematics and Biogeography: Cladistics and Vicariance*, Columbia University Press, New York, 1981; K.J. Peterson, A phylogenetic test of the calcichordate scenario, *Lethaia* 28 (1995) 25–38). We critically analyze the indices used to estimate the quality of the fossil record and the fit of phylogenetic hypotheses to stratigraphy. We use the concept of temporal paralogy to explain why all these measurements are biased. **To cite this article:** H. Lelièvre et al., *C. R. Palevol* 7 (2008).

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Résumé

Information temporelle, enregistrement fossile et phylogénie. Les données fossiles fournissent des informations sur des formes de vie éteintes, dont nous aurions beaucoup de difficultés à imaginer même l'existence à partir de l'étude des formes de vie actuelle. Les fossiles constituent ainsi la principale source d'information sur l'histoire de la vie sur Terre. Ces constatations mettent en évidence le fait que les fossiles, d'une part, apportent une information intrinsèque sur la connaissance de la morphologie des êtres vivants et sur leurs relations de parenté et, d'autre part, permettent la connaissance d'une succession d'événements dont l'ensemble constitue le registre fossile et, en partie, le registre stratigraphique. Ce dernier constitue la source de données élémentaires des paléontologues pour reconstruire l'histoire de la vie. Cependant, accorder une part prépondérante au registre fossile pour reconstruire des phylogénies et proposer des scénarios évolutifs n'est pas accepté unanimement, certains auteurs considérant que les archives fossiles sont trop fragmentaires et incomplètes pour représenter avec fiabilité la vie passée (voir par exemple : G.J. Nelson, N.I. Platnick, *Systematics and Biogeography: Cladistics and Vicariance*, Columbia University Press, New York, 1981; K.J. Peterson, A phylogenetic test of the calcichordate scenario, *Lethaia* 28 (1995) 25–38). Nous proposons une revue critique des indices publiés jusqu'ici pour estimer

* Corresponding author.

E-mail address: lelievre@mnhn.fr (H. Lelièvre).

la qualité de l'enregistrement fossile et leur adéquation avec les hypothèses de phylogénie. Le concept de paralogie temporelle est présenté comme une solution envisageable pour palier les biais des mesures connues jusque là. **Pour citer cet article : H. Lelièvre et al., C. R. Palevol 7 (2008).**

© 2008 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

Keywords: Fossil record; Geological scale; Cladistics; Hierarchy of time; Temporal paralogy

Mots clés : Registre fossile ; Échelle géologique ; Cladistique ; Hiérarchie temporelle ; Paralogie temporelle

Version française abrégée

Aucun domaine scientifique ne peut prétendre être fondé sur une information complète et, à ce titre, le registre fossile est comparable à beaucoup d'autres sources d'informations scientifiques : il est incomplet. Les raisons sont multiples et peuvent être attribuées à de nombreux facteurs. Sur un plan général, on peut invoquer les organismes eux-mêmes. Ainsi, s'il s'agit d'animaux à corps nu ou bien recouvert d'un squelette, leurs milieux de vie et les conditions de la fossilisation, et, enfin les méthodes de collecte et de conservation utilisées par les paléontologues [12,14,20,31,32]. Au-delà de la simple constatation de l'incomplétude du registre fossile, on peut chercher à mesurer la valeur que représente chaque donnée. Ainsi, nous affirmons que connaître la valeur de l'information temporelle dans les phylogénies de groupes étroitement apparentés présente un intérêt.

Dans le cadre de cette problématique, une série de questions peut être formulée. Comment estimer la qualité du registre fossile ? Quelle est la signification d'un ajustement entre la phylogénie et la stratigraphie ? Les indices qui mesurent cet ajustement sont-ils fiables ou pertinents ? Comment ajuster la phylogénie, qui présente une structure hiérarchisée, au temps linéaire ? Seule cette dernière question a reçu jusqu'ici une réponse [37] dont nous rappellerons l'importance.

Aux trois premières questions ont été apportées plusieurs réponses, notamment sous la forme de calcul d'indices, dont le but est de mesurer l'adéquation du registre fossile ou le calibrage stratigraphique de la phylogénie ; les tentatives pour ajuster des hypothèses de phylogénie à la stratigraphie sont plus rares [27–30]. En revanche, les propositions de calcul d'indices mesurant l'ajustement de la phylogénie à la stratigraphie, le plus souvent par une comparaison directe des deux grandeurs ont été assez nombreuses, par exemple au moyen de la comparaison entre les nœuds ou branchements d'un arbre phylogénétique avec l'ordre connu des taxons dans le registre fossile (Tableau 1). L'analyse de ces indices par certains auteurs a montré la récurrence de biais, parmi lesquels la forme des cladogrammes, pectinés vs symétriques, et la modi-

fication des données de la phylogénie sont les plus courants.

Nous souhaitons faire le point sur cette question de l'ajustement entre stratigraphie et phylogénie, revenir sur la définition des indices proposés et sur les biais qui y ont été décelés, et enfin proposer des directions de réflexion qui permettront de formuler un indice qui évitera les pièges identifiés.

Quelques autres indices, différents par la nature de l'hypothèse qui les fonde, ont été proposés ; ils sont discutés dans Zaragüeta Bagils et Lelièvre [36]. Leur pertinence n'est pas retenue ici, dans la mesure où les prémisses qui les fondent ont été réfutées ; c'est le cas de la stratocladistique de Fisher [8,9] et de Clyde et Fisher [5]. En effet, les données stratigraphiques ne peuvent pas être utilisées dans les inférences phylogénétiques, parce que la position temporelle n'incorpore pas une hypothèse sur l'homologie des caractères. Les données stratigraphiques ne peuvent en elles-mêmes générer des hypothèses phylogénétiques, donc, a fortiori, elles ne peuvent servir pour tester une hypothèse de cette nature. Elles ne peuvent pas non plus établir une distribution de caractères [18,19].

Le Spearman Rank Correlation (SRC, Tableau 1) a reçu des critiques sévères qui conduisent à ne plus l'utiliser. L'usage du SRC, à moins d'être appliqué sur un cladogramme pectiné, modifie le nombre de nœuds de l'arbre phylogénétique en question en le réduisant [4]. L'indice de cohérence stratigraphique SCI [13] a été analysé en détail par Siddall [28]. Remarquons que l'hypothèse de départ nous semble erronée [36] ; en effet, une lignée ne peut donner naissance par spéciation à deux taxons et être considérée en même temps comme le groupe frère des deux taxons cités, car cette situation conduit tout au plus à une non-résolution des relations de parenté entre les trois taxons. Les remarques de Siddall portent sur d'autres points utiles à mentionner. Ainsi, il insiste sur la corrélation négative qui existe entre la valeur du SCI et le nombre de taxons et de nœuds, une critique déjà formulée, par exemple, pour l'indice de cohérence global d'un cladogramme. La Fig. 1 montre que la valeur du calcul d'un nœud particulier du cladogramme, et donc la cohérence du SCI, dépendra, d'une

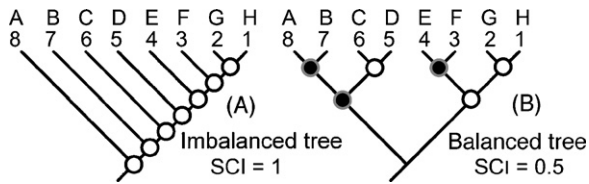


Fig. 1. Stratigraphic Consistency Index, SCI [13]. $SCI = C/N$ (C , number of coherent stratigraphical nodes, and N , total number of nodes). Considering the same number of taxa with the same age nested in a different shape of cladogram (A, B), then the value of SCI is different and depends on whether the tree is balanced (B) or not (A). A balanced tree (B) gives a lower value of SCI showing the bias introduced by the shape of cladogram. On a balanced tree (B), we will reach the conclusion that if a clade has a stratigraphical consistency, its sister-group could not have it.

Fig. 1. Indice de cohérence stratigraphique, SCI [1]. $SCI = C/N$ (C , nombre de nœuds cohérents stratigraphiquement ; N , nombre total de nœuds). Si deux solutions parcimonieuses (A, B) sont comparées, la valeur du SCI dépendra de la forme du cladogramme, pectiné (A) ou non (B). Dans ce dernier cas, la valeur du SCI sera plus faible que pour un arbre pectiné, ce qui démontre le biais entraîné par la forme du cladogramme. De plus, sur un arbre symétrique (B), si un clade a une cohérence stratigraphique, son groupe-frère ne l'aura pas.

part, de la forme du cladogramme et, d'autre part, de la distribution des âges des taxons. Ainsi, tout arbre qui présente au moins un nœud symétrique ne pourra avoir une cohérence stratigraphique maximale, indépendamment de la distribution stratigraphique des fossiles dont les relations sont analysées. Siddall [28] ajoute que tout indice présentant une corrélation avec le SRC dont le biais dû à la forme des arbres a été montré sera lui aussi biaisé (Fig. 1).

Les indices fondés sur le concept de lignée fantôme [16,17] reposent sur la mesure des lacunes d'enregistrement stratigraphique impliquée par une solution phylogénétique. Deux mesures, le RCI et le GER, appliquent cette définition (cf. Tableau 1). Le RCI a reçu des critiques, la mieux construite reposant sur la signification même de l'indice, puisque la cohérence maximale (RCI = 100 %) est obtenue seulement quand tous les taxons apparaissent au même moment [30].

Le calcul du *Gap Excess Ratio*, GER [35], revient à comparer l'arbre à ajuster par rapport à l'arbre qui minimise les lignées fantômes et à celui qui les maximise (Fig. 2). Bien que le GER soit réputé moins sensible aux biais liés à la forme ou à la taille des arbres, il est à noter que le GER est une version standardisée du RCI ; il varie entre 0 et 1, mais la variable mesurée est la même dans les deux indices. Le GER et le RCI sont par conséquent entièrement corrélés et les critiques formulées pour le RCI s'appliquent au GER.

La mesure stratigraphique Manhattan (MSM, [30]) est différente des précédentes, parce qu'elle est la seule qui propose une mesure de l'ajustement de la phylogé-

nie à l'information temporelle tirée de l'enregistrement fossile (Fig. 3). La MSM est fondée sur l'ajustement d'un caractère de Sankoff et Rousseau [26], qui représente les âges absolus des taxons. Son calcul est exprimé par le rapport entre deux longueurs qui représentent des durées relatives de lignées fantômes, l'une exprimant la longueur minimum des différences des valeurs absolues des âges des taxons et l'autre la longueur observée. La valeur de la MSM équivaut en fait à l'indice de cohérence du caractère de Sankoff [22]. Notons que l'indice de cohérence exprime une mesure de l'homoplasie contenue dans les matrices de caractères phylogénétiques. De ce fait, on peut adresser à la MSM la même critique que celle faite à l'indice de cohérence sur sa sensibilité au nombre de taxons. De plus, une relation entre la MSM et le GER a été démontrée (Fig. 4), suggérant la présence de biais dans la MSM [25].

Cette revue rapide de certains indices d'ajustement montre qu'ils sont affectés par des biais, parmi lesquels la taille ou la symétrie des cladogrammes revient le plus souvent. Cependant, certains auteurs [4] ont proposé une utilisation conjointe de ces indices pour pondérer l'influence des biais et exploiter les mesures comme critères de sélection d'arbres, ou comme un moyen d'estimer la qualité de l'enregistrement fossile. Tel n'est pas notre point de vue.

Nous avons signalé que l'une des questions que nous posions avait reçu une réponse : il s'agit de l'identification du biais qui lie la forte valeur des indices d'ajustement aux cladogrammes ne présentant que des nœuds orthologues, c'est-à-dire des nœuds qui ne mènent qu'à un seul autre nœud, en un terminal. Cette question est à rapprocher d'une observation faite par Siddall [27], qui remarque que les données stratigraphiques, linéaires, et les arbres phylogénétiques, ou hiérarchiques, ne sont pas directement comparables, compte tenu de leur représentation. Il ajoute qu'un arbre pectiné, qui ne présente que des nœuds orthologues, peut s'ajuster parfaitement, car ses nœuds admettent un arrangement linéaire, comme les rangs recherchés par certains indices (SRC, SCI).

Une proposition de Hennig [11 (Fig. 48)] peut être considérée, d'une part, comme la clef de la compréhension entre l'information temporelle, le registre fossile et la phylogénie, et, d'autre part, comme la meilleure réponse à l'observation de Siddall [27]. Selon Hennig, dans un cladogramme (Fig. 5) les nœuds définissent une hiérarchie de taxons pour laquelle l'âge d'origine du groupe monophylétique étudié est identique à l'âge d'origine de son groupe-frère. Ainsi, l'âge des clades est un attribut des nœuds, et cela a pour conséquence que le temps sur un cladogramme est distribué selon

Table 1

Main indices proposed to fit phylogenetic data with the fossil record

Tableau 1

Principaux indices proposant un ajustement des données entre l'enregistrement fossile et la phylogénie

Indices	Temporal data	General computation	Values
Spearman Rank Correlation Coefficient SRC [19]	Order of appearance of taxa	Branching order <i>vs</i> Appearance order	$0 \leq \text{SRC} \leq 1$
Stratigraphic Consistency Index SCI [13]	Order of appearance of taxa	Branching order <i>vs</i> Appearance order	$0 \leq \text{SCI} \leq 1$
Relative Completeness Index RCI [2]	Absolute age	Relative length of ghost range	$-\infty < \text{RCI} \leq 100$
Gap Excess Ratio GER [35]	Absolute age	Relative length of ghost range	$0 \leq \text{GER} \leq 1$
Manhattan Stratigraphic Measure MSM [29]	Absolute age	Relative length of ghost range	$0 < \text{MSM} \leq 1$

une hiérarchie. Il a été montré qu'une classification des intervalles temporels admet une représentation hiérarchique. Dans cette représentation, toute hiérarchie temporelle est un arbre pectiné. Zaragüeta Bagils et al. [37] ont utilisé cette propriété pour proposer une hiérarchisation de l'échelle stratigraphique afin de rendre la temporalité de l'hypothèse phylogénétique et celle des datations stratigraphiques directement comparables (Fig. 6). La distribution hiérarchique du temps sur les cladogrammes a permis d'identifier un nouveau cas du phénomène, connu sous le nom de paralogie [10]. Dans sa version génétique, la paralogie est repérable par une duplication des taxons terminaux. De façon analogue, quand une période géologique apparaît plusieurs fois dans un cladogramme, on parlera alors de paralogie temporelle [37]. Les nœuds qui révèlent cette paralogie sont en fait les nœuds qui ne sont pas pectinés, c'est-à-dire ceux qui, parce qu'ils contiennent deux nœuds, forment les parties symétriques de certains cladogrammes, comme nous l'avons remarqué dans la présentation des indices d'ajustement. Les nœuds d'un cladogramme qui portent une information temporelle sur l'ordre d'apparition des taxons qu'ils contiennent sont uniquement ceux orthologues. Seuls ces nœuds permettent d'identifier un conflit entre l'âge des fossiles et la hiérarchie temporelle (Fig. 7)

Une conséquence surprenante de cette comparaison est le fait qu'une meilleure connaissance de l'enregistrement fossile ne permet pas toujours d'améliorer la résolution temporelle d'un cladogramme. En effet, l'ajout d'un taxon nouvellement découvert peut transformer la forme d'un nœud orthologue en un nœud paralogue (Fig. 7B). Dans ce cas, l'information sur l'ordre d'apparition des taxons, présente avant la découverte, est perdue. Cette remarque s'accompagne de la constatation du fait que le nombre de nœuds paralogues lors de l'ajustement temporel d'une phylogénie ne peut décroître.

Nous considérons que les tentatives précédentes pour mesurer l'ajustement de la phylogénie à la stratigraphie sont réfutées parce qu'elles n'ont pas identifié les

exemples de paralogies temporelles et n'ont pas expliqué l'effet de la forme lors de l'ajustement de la phylogénie à la stratigraphie.

1. Introduction

In the context of a debate on the general topics relative to the evaluation of the congruence between phylogeny and stratigraphy, some questions are often raised [12,14,15,20,21,31,32]. These questions, mainly addressed to systematists and palaeontologists, concern the adequacy of the fossil record and the value of the temporal information in phylogenies. These questions are formulated as following: What is a fossil record of quality? What is the meaning of a fit between phylogeny and stratigraphy? Are the different indices (see Table 1) produced during the last decades to measure the phylogenetic fit to stratigraphy good enough to estimate this relationship? How to proceed in adjusting a phylogeny, structured as a mathematical hierarchy of nested taxa, to inherently linear and more constrained, stratigraphic data [27]? We consider that this last question is the only one to have received a clear answer [37], an answer central to our discussion.

Concerning the first three questions, several answers have been proposed; they have produced different measures of adjustment between the fossil record and phylogenies. Attempts to fit phylogenetic hypotheses to stratigraphy are less common, as only one measure has been proposed, by Siddall [30]. Various critical analyses of these indices have been proposed by Siddall [27–29], Pol and Norell [22], Benton et al. [4], Zaragüeta Bagils and Lelièvre [36]. They demonstrated the existence of recurring problems in these indices, among which the most common are: (1) a bias linked to the shape of cladograms, unbalanced *vs* balanced trees, and (2) the modification of phylogenetic results in order to adjust stratigraphical data.

We re-examine some of the criticisms that have been addressed to these indices, and we evaluate their relevance.

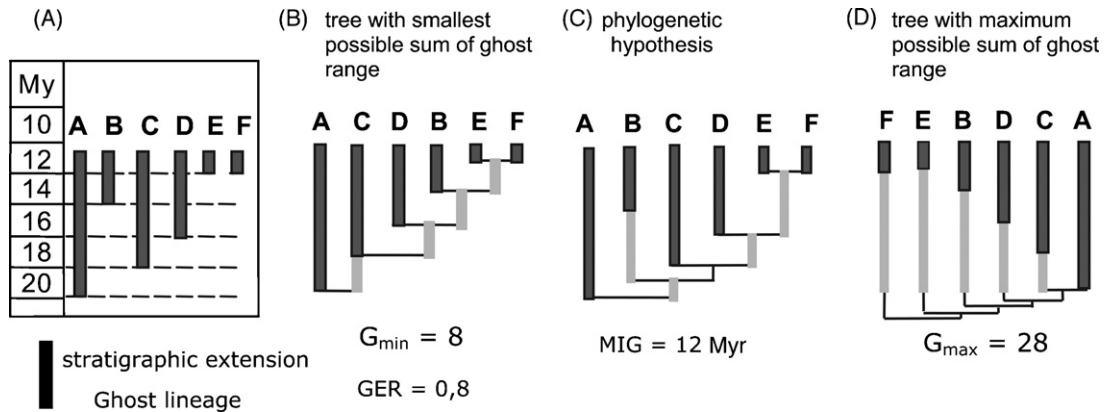


Fig. 2. Calculation of the Gap Excess Ratio, GER [35] for randomized data (A); $GER = 1 - (MIG - G_{\min}/G_{\max} - G_{\min})$. (B, D) Cladograms that express respectively the smallest and maximum possible sum of ghost ranges. (C) Value of GER for the estimated phylogeny. A sensitivity of GER to absolute difference in age can be seen.

Fig. 2. Calcul du *Gap Excess Ratio*, GER [35] pour un jeu de données quelconque (A); $GER = 1 - (MIG - G_{\min}/G_{\max} - G_{\min})$. (B, D) Cladogrammes exprimant respectivement la plus petite et la plus grande lacune d'enregistrement ou la plus petite et la plus grande somme des lignées fantômes. (C) Calcul du GER pour la phylogénie retenue; on voit que le GER est sensible à la différence d'âge absolu entre taxons.

2. Previous attempts to measure stratigraphic fit to phylogenies

Proposals of stratigraphic indices differ in the nature of the hypothesis on which they rely. They are discussed in Zaragüeta Bagils and Lelièvre [36], who have refuted the premises on which they are based. Here we shortly introduce some of the indices and their criticism. Stratocladistics of Fischer [8,9] and Clyde & Fischer [5] introduce stratigraphic data, under the form of a pseudo-character, into a parsimony-based phylogenetic analysis. An obvious inaccuracy of the stratocladistic method stems from stratigraphic information being different from phylogenetic inferences in that the temporal position or age of taxa is totally free of any hypothesis of homology. In other words, stratigraphic information cannot give rise to any phylogenetic hypothesis and consequently cannot provide a test for any phylogenetic inference (see [18,19]).

It has been shown [27,28,36] that the Spearman Rank Correlations (SRC) [19] and the Relative Completeness Index (RCI) [2] suffer from severe criticisms, leading the authors of these measurements to recommend their abandonment [3]. The SRC and the Relative Completeness Index (RCI) use the non-parametric Spearman rank test. The nodes are arranged following their rank in the hierarchy of taxa. A node is deemed consistent if its inferred age is older than the age of lower-ranked nodes. In trees that are not totally imbalanced, the use of SRC reduces the number of nodes of the cladogram.

The Stratigraphical Coherence Index (SCI) [13] was reviewed in detail by Siddall [28]. This author empha-

sizes different points, among which are the negative correlation between values of SCI and the number of taxa and nodes. He adds that if the SCI and the RCI are unbiased measures of the same phenomena measuring the fit of fossil data on a cladogram, and if the SRC could be considered as flawed, then a positive correlation between it and the SCI will appear as a bias for the SCI. This bias is demonstrated by Siddall [28], and is shape-dependent, as presented in Fig. 1. This example shows that the consistency of a particular node depends on the shape and the distribution of age across taxa. If a clade has a stratigraphic consistency, then its sister groups will not have it (Fig. 1B). Consequently, balanced trees will never reach the maximum values of stratigraphical consistency. We can also add that the initial hypothesis by Huelsenbeck [13] seems wrong to us [36]. An ancestral lineage, which gives rise to two different lineages through a branching speciation event, cannot be at the same time considered as the sister-group of its descents. Following cladistic principles, a cladogram showing a relationship between an ancestral taxon and its descendants cannot be resolved. Siddall [27,28] includes also other critical points. They concern the negative correlation that links the SCI value and the number of nodes and taxa in a way that reminds the criticisms addressed to the consistency index of parsimony analysis.

Two indices (RCI and GER), based on the ghost lineage concept [16,17], rely on the measure of stratigraphic record gaps implied by a phylogenetic result (Table 1).

The measure of the Gap Excess Ratio (GER) [35] implies the comparison of a phylogenetic tree to both

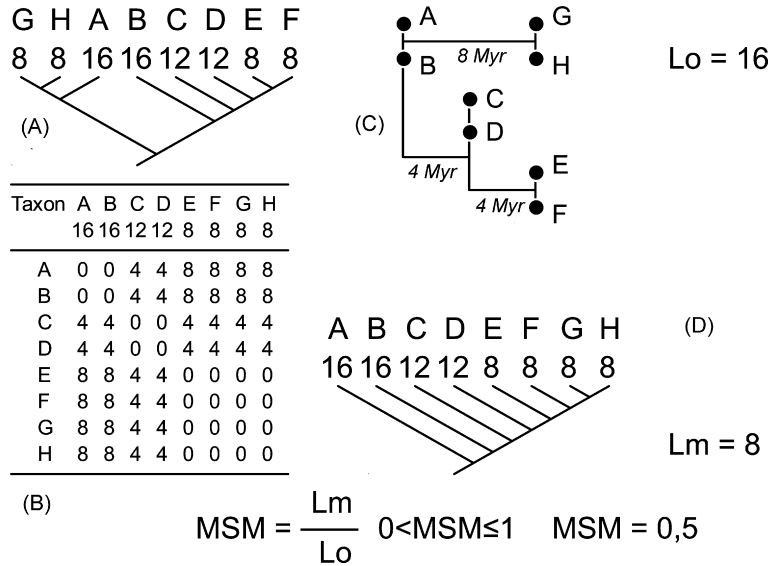


Fig. 3. The Manhattan Stratigraphic Measure, MSM [30], $MSM = L_m/L_o$ (L_m , minimal length calculated from a Sankoff matrix; L_o , observed length or sum of the differences between the ages of taxa. (A) Parsimonious tree on which the MSM will be measured. (B) Stratigraphic Manhattan matrix established from the distance of stratigraphic first appearance age of taxa. (C) The cost of transformations between taxa is calculated by the equivalence of the Manhattan distance between states or ages. (D) Tree obtained from a Sankoff matrix in order to get the minimal length (L_m). It can be noted that, in order to treat states as Sankoff characters as irreversible (ordered), the taxa are rearranged according to their age.

Fig. 3. La mesure stratigraphique Manhattan, MSM, [30], $MSM = L_m/L_o$. (A) Solution parcimonieuse sur laquelle la MSM sera mesurée. (B) Matrice stratigraphique de Manhattan établie d'après l'âge des taxons. (C) Le coût des transformations entre les taxa sera calculé selon l'équivalence de la distance Manhattan entre les états (= âges). (D) Arbre obtenu par une matrice de Sankoff afin de calculer la longueur minimale (L_m). On observe que la succession des taxons est réarrangée, afin de ne pas obtenir de réversion de certains âges.

a tree that minimizes the amount of ghost lineages, and another that maximises them (Fig. 2). Although the GER is supposed to be less biased than the RCI, GER is just a standardised measure of the RCI (both measure the observed ghost lineage duration implied by a tree) and should probably be sensitive to the same bias that affects the RCI.

The Manhattan Stratigraphic Measure [30] follows a different rationale from the former because it aims to estimate a fit of the phylogeny to the temporal information contained in the fossil record. The MSM is based on the length of a Sankoff character [26], which represents the absolute ages of taxa (Fig. 3). Its value is expressed by the ratio of the minimal length of the age differences to the observed length. In other terms, the MSM is the consistency index (CI) of the Sankoff character. Thus, it suffers from the bias that affects the CI [6,7].

More recently, Pol et al. [24], and Pol and Norell [23] have shown that three indices (SCI, GER and MSM) are sensitive to the size and shape of the trees. They also added that the difference between ages of appearance of fossil taxa used in the stratigraphic consistency analysis depends on the age of the associated sediments. There is an uncertainty issue concerning the ages of

first appearance. According to these authors, the MSM is less sensitive to parameters such as tree shape than other indices (GER). If we examine the formulation of both GER and MSM, the only variables implied in the

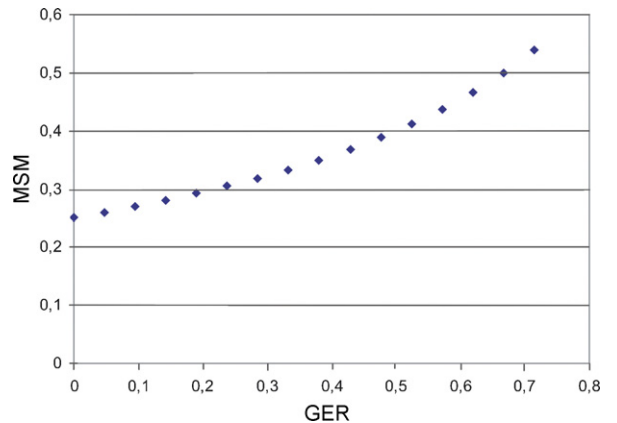


Fig. 4. Relation between GER and MSM measured on a theoretical data set of eight taxa with different first-appearance ages comprised between 1 to 8 Myr. The values are given for the first 500 trees, using exhaustive search in PAUP 4.b10.

Fig. 4. Relation entre le GER et la MSM, mesurée sur un jeu de données théoriques de huit taxons présentant tous des âges d'apparition différents (de 1 à 8 Ma). Les valeurs sont données pour les 500 premiers arbres calculés par le logiciel Paup 4.b10 (analyse exhaustive).

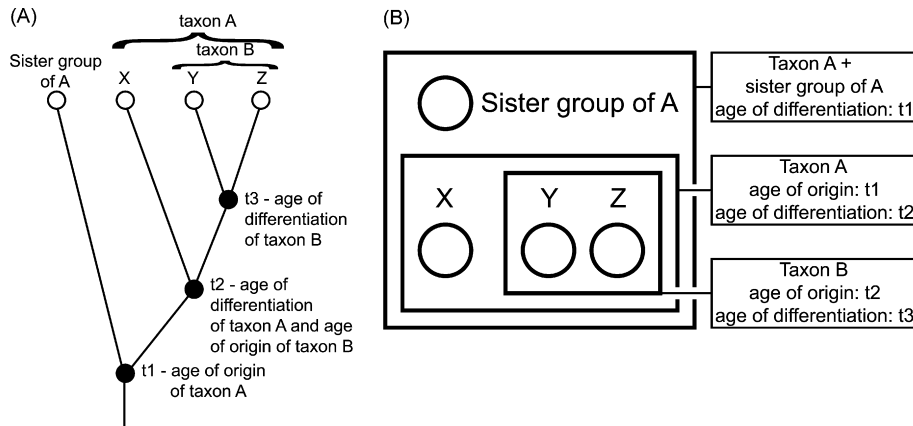


Fig. 5. Cladogram and time. (A) Temporal information in a cladogram according to Hennig [11 (Figure 48)], modified from Hennig, in Tassy [34]. (B) The same information treated as a Venn diagram to show that temporal information is an attribute of the node and not of the branches. These relationships between temporal data and cladogram demonstrate that time could be expressed as a hierarchy on a cladogram.

Fig. 5. Le cladogramme et le temps. (A) Relation entre l'information temporelle et un cladogramme selon Hennig [11 (Figure 48)], modifié d'après Hennig, in Tassy [34]. (B) Même information qu'en (A), mais traitée dans un diagramme de Venn, afin de montrer que l'information temporelle est l'attribut d'un nœud et non pas celui des branches.

calculation are respectively the Minimum Implied Gap (MIG), and the observed length (L_0). Both correspond to the amount of ghost ranges in phylogenetic trees. Thus, both indices may be strongly correlated. The difference between the values of each one of these indices for the same data is explained by the standardisation of the calculation. When the value of GER is 0, it means that a maximum inconsistency is observed between the branching order of taxa and their age in the fossil record (the youngest taxa are in the most inclusive part on the tree). At the opposite end, a value equal to 1 corresponds to the reverse temporal situation, where the order of appearance of taxa follows their relative age and the inclusiveness of tree nodes. For the MSM, this last value has the same meaning as that for GER, but the zero value cannot be reached because the use of this index implies always the presence of fossils, or differently aged taxa.

An instance is given in Fig. 4, where it develops a calculation from theoretical data with eight taxa (a full-balance tree shape can be obtained) with eight different ages of appearance (from 1 to 8 Myr). The correlation between the values of GER and MSM is established for the first 500 trees, computed with PAUP 4.1 [33], and in relation with tree-shape balance [25].

Tree-shape dependence, in particular for balanced trees, appears as being the most common bias in the development of the stratigraphic or phylogenetic indices. Even if identified relatively earlier in the development of this topic [28,36], authors have not paid attention to the comprehension of the analysis between the structures of a cladogram and time. Siddall [18,27] noticed that stratigraphic data are linear, whereas phylogene-

tic trees represented by a hierarchy of nested taxa are not so constrained. Following this, he adds that a pectinate tree easily fits ages because its internal nodes are linearly arranged. Note that stratigraphic indices reach their maximum value only when measured on pectinate trees. Previous analysed indices are good examples of the incompatibility between hierarchical (non-linear) data applied to stratigraphic linear time.

3. Temporal hierarchy as a solution

Hennig's [27 (Fig. 48)] concept of 'age of origin' and 'age of differentiation' are key clues for our understanding of the link between temporal information, fossil record, and phylogeny. According to Hennig, the inferred age of origin of a monophyletic group is the same as that of its respective sister-group; therefore, the age of clades is an attribute to node, not to branches (Fig. 5). Consequently, ages in a cladogram are distributed hierarchically. When temporal intervals are represented this way, they define a pectinate hierarchical tree (Fig. 6E).

Zaragueta Bagils et al. [37] applied Hennig's hierarchical representation of ages [34] to the stratigraphical scale in order to make the comparison between phylogeny and time possible (Fig. 6). The stratigraphical events, on which the origin of each geological division is based, are represented in a hierarchical framework (a mathematical hierarchy in which the classification of temporal intervals defines a pectinate tree).

The hierarchical representation of time in a cladogram led to the discovery of a new class of paralogy [10]; it is easily recognizable as a duplication of terminal taxa.

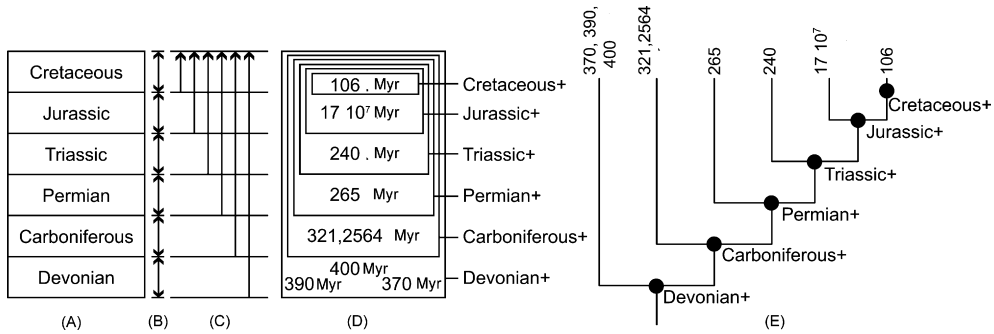


Fig. 6. Time is hierarchical. (A) At the geological scale, time is linear and organized as an ordered partition. (B) In the linear representation of time, each division is defined as an event where the beginning is also the end of the previous one. In order to define the temporal hierarchy, each event defines only the beginning of a period. (C) Geological periods are defined by their temporal extensions. (D) Temporal hierarchy described in (C) and represented as a Venn diagram. (E) Temporal hierarchy as a pectinate tree where nodes correspond in our example to some of the geological systems. The age of each fossil specimen represents its terminal age and is also the least inclusive temporal class that contains it.

Fig. 6. La hiérarchisation du temps. (A) Le temps représenté dans l'échelle géologique est linéaire et structuré comme une partition ordonnée. (B) Dans le temps linéaire, le même événement sert à définir à la fois le début d'une période et aussi la fin de la précédente. (C) Les systèmes géologiques sont représentés par leur extension temporelle. (D) Représentation de la hiérarchie temporelle de (C) sous forme d'un diagramme de Venn. Des âges précis sont définis et appartiennent à chaque système ou classe. (E) La hiérarchie temporelle est représentée sous forme d'un arbre pectiné dans lequel les nœuds sont, dans l'exemple choisi, des systèmes géologiques. L'âge de chaque fossile est un nœud terminal, et donc aussi la classe temporelle la moins inclusive qui le contient.

The distribution of temporal data on a cladogram does not escape to that general phenomenon and a new class of paralogy, termed 'temporal paralogy' has been erected [37]. Thus, it is not surprising that the tree-shape representation of the temporal hierarchy is a pectinate or balanced tree. This offers to our knowledge the explanation of the fact that measurements of fits of cladograms to stratigraphy are often shape dependent, and explains why ambiguous results are obtained when the cladogram is not strictly pectinate. Temporal paralogy leads to distinguish between two categories of nodes (Fig. 7). Orthologous nodes, i.e. nodes that lead to a single less

inclusive internal node, convey temporal information. Paralogous nodes do not provide temporal information (Fig. 7B); they correspond to the duplication of temporal hierarchies, hence the term paralogous. The ambiguity introduced in the calculation of stratigraphical indices with more or less balanced cladograms has been pointed out in some papers by Siddall [27,28,30], Pol and Norell [22–24], and Wills [35], but none of them has offered any explanation concerning the structure of the cladogram. These indices are applied to a pectinate tree; so, they give acceptable results. However, none of these authors has identified temporal paralogy to explain why results are

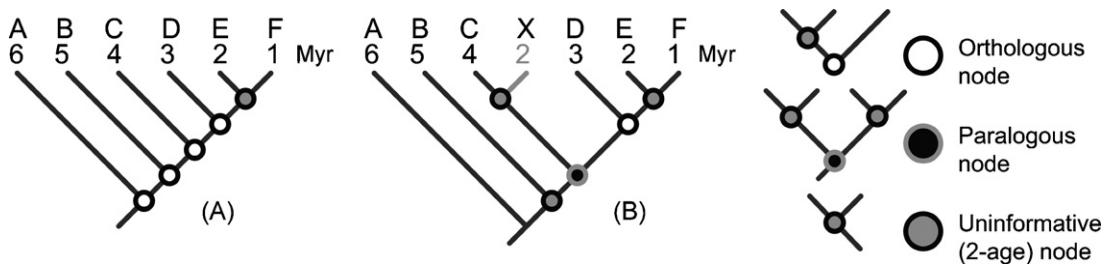


Fig. 7. Temporal information and cladogram. (A) Six taxa and their relative ages organised as a maximally informative tree. All nodes are orthologous and the age of fossils are consistent or not with the temporal hierarchy. (B) The effect of a better knowledge of the fossil record by addition of a taxa (2 Myr) leads to a decrease of the temporal resolution. The number of informative or orthologous nodes decreases as temporal uninformatives nodes are generated. The black node is a new paralogous node that indicates a temporal paralogy. The ages of sister-clades (C) and (D (E,F)) are temporal paralogues.

Fig. 7. L'information temporelle et le cladogramme. (A) Six taxons avec chacun son âge relatif, arrangés selon un arbre d'information temporelle maximale. Ainsi, tous les nœuds sont orthologues et l'âge des fossiles est cohérent avec la hiérarchie temporelle. (B) L'effet d'une meilleure connaissance du registre fossile par addition d'un taxon (2 Ma) entraîne une décroissance de la résolution temporelle. Le nombre de nœuds informatifs ou orthologues décroît, tandis que des nœuds ne portant pas d'information temporelle sont générés. Les nœuds cerclés de noir sont de nouveaux nœuds paralogues qui indiquent une paralogie temporelle. Les âges des groupes-frères (C) et (D (E,F)) sont des paralogues temporels.

more relevant on a pectinate tree than on a more balanced one. We consider that the relevance of measurements that do not take into account temporal paralogy have to be questioned, if not simply abandoned.

Finally, the temporal paralogy has shown that some of the concepts employed in comparing phylogeny to stratigraphy, or dating phylogeny, must be left or revised. Among them, the palaeontological calibration may be problematic because it uses the first-appearance datum instead of using a sister-group context and the fact that the age of a clade is that of its sister-group [11,34]. This has a serious consequence on the age calibration of the molecular phylogenetic trees when they are only based on fossil data. This could explain why sometimes the ‘origin’ of a group becomes suddenly older than usually thought. We have also pointed out [37] that in the Cladestore database coordinated by Benton [1], more than 63.5% of the nodes are paralogous and so are uninformative as regards the order of appearance of temporally paralogous taxa. The fossil record appears reasonably well documented for understanding the history of life, only when the relationships of extinct taxa are represented by a pectinate cladogram or, more locally, by the occurrence of orthologous nodes. However, some paradoxical results have to receive a clearer answer. Among these is the fact that the improvement of the fossil record (increasing the number of taxa) can lead to a decrease in the temporal information of the cladograms (appearance of more paralogous nodes).

Acknowledgements

This paper is a contribution of the team ‘Système-rique, recherche informatique et structuration des cladogrammes’ of the UMR 5143 MNHN–UPMC–CNRS. We would like to thank warmly the organizing committee of the International Congress *Palaeobotany and Evolution of Plants* for having invited us to give a talk on the new concept of temporal paralogy. We specially thank Profs J. Broutin and A. de Ricqlès. We are grateful to G. Edgecomb (Natural History Museum, London) for useful comments and to Bob Carr (University of Ohio, USA) for review of the English text and comments.

References

- [1] M.J. Benton Cladestore. An electronic source of data matrices from published cladograms, University of Bristol, 2002.
- [2] M.J. Benton, G.W. Storrs, Testing the quality of the fossil record: paleontological knowledge is improving, *Geology* 22 (1994) 111–114.
- [3] M.J. Benton, P.C.J. Donoghue, Palaeontological evidence to date the tree of life, *Mol. Biol. Evol.* 24 (2007) 26–53.
- [4] M.J. Benton, R. Hitchin, M.A. Wills, Assessing congruence between cladistic and stratigraphic data, *Syst. Biol.* 48 (1999) 581–596.
- [5] W.C. Clyde, D.C. Fischer, Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis, *Paleobiology* 23 (1997) 1–19.
- [6] J.S. Farris, The retention index and the rescaled consistency index, *Cladistics* 5 (1989) 417–419.
- [7] J.S. Farris, The retention index and homoplasy excess, *Syst. Zool.* 38 (1989) 406–407.
- [8] D.C. Fischer, Stratigraphic parsimony, in: W.P. Maddison, D.R. Maddison (Eds.), *MacClade: Analysis of Phylogeny and Character Evolution*, version 3, Sinauer Associates, Sunderland, MA, USA, 1992, pp. 124–129.
- [9] D.C. Fischer, Stratocladistics: morphological and temporal patterns and their relation to phylogenetic process, in: L. Grande, O. Rieppel (Eds.), *Interpreting the Hierarchy of Nature. From Systematic Patterns to Evolutionary Process Theories*, Academic Press, London & New York, 1994, pp. 133–171.
- [10] W.M. Fitch, Distinguishing homologous from analogous proteins, *Syst. Zool.* 19 (1970) 99–113.
- [11] W. Hennig, *Phylogenetic systematics*, University of Illinois Press, Chicago, Illinois, 1966.
- [12] S.M. Holland, The Stratigraphic distribution of fossils, *Paleobiology* 21 (1995) 92–109.
- [13] J.P. Huelsenbeck, Comparing the stratigraphic record to estimates of phylogeny, *Paleobiology* 20 (1994) 470–483.
- [14] E. Nardin, I. Rouget, P. Neige, Tendencies in paleontological practice when defining species, and consequences on biodiversity studies, *Geology* 33 (2005) 969–972.
- [15] G.J. Nelson, N.I. Platnick, *Systematics and Biogeography: Cladistics and Vicariance*, Columbia University Press, New York, 1981.
- [16] M.A. Norell, Taxic Origin and Temporal Diversity: The Effect of Phylogeny, in: M.J. Novacek, Q.D. Wheeler (Eds.), *Extinction and Phylogeny*, Columbia University Press, New York, 1992, pp. 89–118.
- [17] M.A. Norell, Ghost taxa, ancestors, and assumptions: a comment on Wagner, *Paleobiology* 22 (1996) 453–455.
- [18] M.A. Norell, M.J. Novacek, The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history, *Science* 255 (1992) 1690–1693.
- [19] M.A. Norell, M.J. Novacek, Congruence between superpositional and phylogenetic patterns: comparing cladistic patterns with fossil records, *Cladistics* 8 (1992) 319–337.
- [20] C. Patterson, Significance of fossils in determining evolutionary relationships, *Annu. Rev. Ecol. Syst.* 12 (1981) 195–223.
- [21] K.J. Peterson, A phylogenetic test of the calcichordate scenario, *Lethaia* 28 (1995) 25–38.
- [22] D. Pol, M.A. Norell, Comments on the Manhattan Stratigraphic Measure, *Cladistics* 17 (2001) 285–289.
- [23] D. Pol, M.A. Norell, Uncertainty in the age of fossils and the stratigraphic fit to phylogenies, *Syst. Biol.* 55 (2006) 512–521.
- [24] D. Pol, M.A. Norell, M.E. Siddall, Measures of stratigraphic fit to phylogeny and their sensitivity to tree size, tree shape, and scale, *Cladistics* 20 (2004) 64–75.
- [25] I. Rouget, Reconstruction phylogénétique chez les ammonites : confrontations des approches cladistique et stratigraphique. Le cas des Dayiceras (Ammonitina, Eodeoceratoidea), École doctorale « Images et modélisation des objets naturels », Université de Bourgogne, Dijon, France, 2002, p. 286.

- [26] D.D. Sankoff, P. Rousseau, Locating the vertices of a Steiner tree in arbitrary metric space, *Math. Program.* 9 (1975) 240–246.
- [27] M.E. Siddall, Stratigraphic consistency and the shape of things, *Syst. Biol.* 45 (1996) 111–115.
- [28] M.E. Siddall, Stratigraphic indices in the balance: a reply to Hitchin and Benton, *Syst. Biol.* 46 (1997) 569–573.
- [29] M.E. Siddall, *Nature debates*, 1998, www.nature.com/nature/debates/fossil.
- [30] M.E. Siddall, Stratigraphic fit to phylogenies: a proposed solution, *Cladistics* 14 (1998) 201–208.
- [31] P.W. Signor, Real and apparent trends in species richness through time, Phanerozoic diversity patterns—profiles, in: J.W. Valentine (Ed.), *Macroevolution*, Princeton University Press, Princeton, NJ, USA, 1985, p. 452.
- [32] A.B. Smith, Large-scale heterogeneity of the fossil record: implications for Phanerozoic biodiversity studies, *Philos. Trans. R. Soc., Sect. B* 356 (2001) 351–367.
- [33] D.L. Swofford, PAUP*. *Phylogenetic Analysis Using Parsimony (*and other Methods)*, Version 4, Sinauer Associates, Sunderland, MA, USA, 1998.
- [34] P. Tassy, Willi Hennig et l'objet paléontologique, *Geodiversitas* 21 (1999) 5–23.
- [35] M.A. Wills, Congruence between phylogeny and stratigraphy: randomization tests and the Gap excess ratio, *Syst. Biol.* 48 (1999) 559–580.
- [36] R. Zaragüeta Bagils, H. Lelièvre, Âges et noms versus caractères et phylogénie : le problème de l'ajustement entre la stratigraphie et la phylogénie, in: P. Tassy, A. de Ricqlès (Eds.), *Biosystema 19. Systématique et Paléontologie*, Société Française de systématique, Paris, 2001, pp. 169–185.
- [37] R. Zaragüeta Bagils, H. Lelièvre, P. Tassy, Temporal paralogy, cladograms, and the quality of the fossil record, *Geodiversitas* 26 (2004) 381–389.