

The « Evolutionary Signal » of Homoplasy in Protein-Coding Gene Sequences and its Consequences for a priori Weighting in Phylogeny.

Le « signal évolutif » de l'homoplasie dans les séquences de gènes codants et ses implications pour la pondération en phylogénie.

Alexandre Hassanin^{1*,2}, Guillaume Lecointre^{1,3}, and Simon Tillier^{1,4}.

RESUME

Afin d'analyser indépendamment l'homoplasie pour les six types de substitutions (i.e. A-G, C-T, A-C, A-T, C-G et G-T) à chaque position de codon du gène du cytochrome b, deux approches ont été utilisées : l'une est fondée sur l'indice de cohérence qui mesure la quantité d'homoplasie, l'autre repose sur une analyse de la saturation qui décrit graphiquement la répartition de l'homoplasie dans l'échantillonnage taxinomique. Les résultats obtenus sur une matrice comprenant 32 séquences d'*Artiodactyla* indiquent que le cytochrome b est soumis à des contraintes différentielles: (1) entre les six types de substitutions, (2) entre les trois positions de codon et (3) entre les deux brins d'ADNmt. Nos analyses révèlent également que les sites non synonymes peuvent être plus homoplasiques que les sites synonymes lorsque les possibilités de substitutions sont fortement contraintes par l'hydrophobicité. La plupart des schémas de pondération appliqués sur les gènes codants apparaissent élaborés sur des présupposés injustifiés. Nous proposons de pondérer chaque type de substitution à chaque position de codon en fonction inverse de son contenu en homoplasie estimé soit par l'indice de cohérence, soit par un indice traduisant le degré de saturation mutationnelle.

ABSTRACT

To analyze independently homoplasy for the six possible types of substitution (i.e. A-G, C-T, A-C, A-T, C-G, and G-T) at each of the three codon-positions of the cytochrome b gene, two approaches were used: the first is based on the consistency index which measures the amount of homoplasy, and the second is based on the saturation analysis which describes graphically the distribution of homoplasy within the taxonomic sampling. The results obtained from a data set of 32 sequences of *Artiodactyla* indicate that evolution of the cytochrome b is governed by differential constraints : (1) between the six substitutions-types, (2) between the three codon-positions and (3) between the two mtDNA strands. Moreover, we find that nonsynonymous sites can be more homoplastic than synonymous sites when the possibilities of substitutions are severely restricted because of the functional requirements of hydrophobicity. Most weighting schemes applied to protein-coding genes are shown elaborated from unjustified assumptions. We propose to weight each substitution-type at each codon-position according to its homoplasy content evaluated either with the consistency index or with an index representing the level of mutational saturation.

