

**Synopsis.**

3iahifi -v *filein*

**Description.**

3iahifi procède à l'analyse Hierarchical Fit Index (HiFi) des cladogrammes fournis dans le fichier *filein*. Le résultat de l'analyse est produit sur *stdout*.

Le fichier *filein* a un format semblable au format *.3ia*, spécialisé pour l'analyse de cohérence stratigraphique. Le travail effectué par 3iahifi consiste, pour chaque cladogramme, à calculer les horogrammes H+ et H- associés, ainsi que la liste des 3is générés par chacun de ces trois arbres. Enfin, le rapport HiFi est calculé tenant compte des intersections entre les ensembles de 3is générés par ces trois arbres. Le fichier d'entrée donne :

- Une liste de taxons.
- Une relation d'ordre entre ces taxons ; cette relation est censée représenter l'ancienneté du taxon dans une interprétation temporelle.
- Une liste d'arbres (cladogrammes) exprimés sur les taxons.

Les étapes suivantes sont effectuées. Si le fichier *filein* contient plusieurs cladogrammes, l'ensemble de ces traitements est effectué successivement pour chacun d'eux de façon indépendante ; il n'y a pas d'interactions entre les résultats obtenus pour les différents cladogrammes.

**1/ Cladogrammes de base.**

Le cladogramme de base, fourni dans *filein*, est traité de façon à en supprimer les polytomies, comme suit.

Pour un nœud non terminal, le tableau ci-dessous donne toutes les combinaisons possibles pour la liste de ses fils :

		Nombre de terminaux			
		0	1	2	3 ou +
Nombre de non terminaux	0	C	D	-	B
	1	-	-	A	A
	2 ou +	-	-	-	-

- Rien de spécial.
- A On ne garde que le terminal (taxon) le plus âgé (\*).
- B *idem* (\*) (\*\*).
- C Cas pathologique qui ne peut survenir en principe. On ne fait rien. Ce nœud pourrait être supprimé sans conséquences.
- D Cas pathologique qui ne peut survenir en principe, mais qui est produit par le cas B. On ne fait rien (\*\*).

Tous les nœuds non-terminaux du cladogramme sont traités ainsi. Dans les cas A et B, le cladogramme est remanié par suppression d'un ou plusieurs nœuds terminaux. En fin de traitement, c'est le nouveau cladogramme remanié qui sera utilisé d'une part pour calculer la liste des 3is, d'autre part pour calculer les horogrammes H+ et H-.

(\*) (\*\*) Voir § 3.

(\*\*) Dans les cas B et D, un non-terminal porte un taxon unique, ce qui se traduit en notation parenthésée par la forme "(X)". Bien que sa signification soit douteuse, cette structure est conservée. La solution consistant à supprimer le non-terminal et à rattacher le taxon à son père n'a pas été retenue.

**2/ Calcul de l'horogramme H+.**

Le remaniement opéré sur le cladogramme initial consiste à réordonner chaque chaîne de nœuds orthologues de façon à placer les taxons les plus âgés vers la racine. Ce remaniement est réalisé comme suit.

L'arbre est parcouru depuis la racine vers les feuilles. Pour chaque nœud non-terminal, le tableau ci-dessous donne toutes les combinaisons possibles pour la liste de ses fils :

		Nombre de terminaux			
		0	1	2	3 ou +
Nombre de non terminaux	0	-	-	-	
	1	-	A		
	2 ou +	-	-	-	-

- On ne fait rien.
- A Nœud orthologue → traitement ci-dessous.
- hachuré* Impossible suite à l'élimination des polytomies du cladogramme de base.

On ne considère que les nœuds N1 du type A (nœuds orthologues). Sa chaîne de nœuds orthologues associée est alors remontée comme suit :

- Le nœud N1 étant orthologue, il possède un et un seul non-terminal N2 parmi ses fils.
- On considère ce nœud N2 ; il est lui-même de l'un des différents types ci-dessus en ce qui concerne ses fils.
- Si N2 est lui-même orthologue (cas A), on continue de même avec son fils non-terminal N3.
- etc.
- La montée s'arrête donc forcément sur un nœud de type autre que A.

Pour les nœuds parcourus, les différents cas possibles sont les suivants :

		Nombre de terminaux			
		0	1	2	3 ou +
Nombre de non terminaux	0	D	E	F	
	1	C	A		
	2 ou +	B	B	B	B

(note : d'après ce qui précède, le cas A concerne forcément un nœud intermédiaire dans la chaîne, et les autres cas concernent forcément le nœud qui termine la chaîne).

Pour chaque nœud de la chaîne, on considère son âge de la façon suivante :

- A Nœud orthologue : on considère l'âge du taxon porté par l'unique terminal.
- B Nœud paralogue : on considère l'âge maximal (\*) de tous les taxons dans la descendance de ce nœud.
- C, D Cas pathologiques ignorés.
- E On considère l'âge du taxon.
- F On considère l'âge du taxon le plus âgé (\*\*).

Dans toute la chaîne ainsi explorée, on considère :

- l'âge ainsi calculé pour le nœud N1 : AGE1.
- l'âge ainsi calculé pour le nœud Nx pour lequel cet âge AGE<sub>x</sub> est le plus élevé (\*\*).

Si Nx est différent de N1 et si de plus AGE<sub>x</sub> > AGE1 (\*\*), alors on procède à la permutation entre N1 et Nx, comme suit selon le cas de Nx :

- A On permute le taxon unique porté par N1 et le taxon unique porté par Nx.
- B On permute le taxon unique porté par N1 et le nœud Nx lui-même, avec toute sa descendance.
- E On permute le taxon unique porté par N1 et le taxon unique porté par Nx.
- F On permute le taxon unique porté par N1 et celui des deux taxons portés par Nx qui a l'âge AGE<sub>x</sub>.

Sinon, on ne fait rien.

Remarque : on a permuté ici (éventuellement) le nœud N1 avec l'un des nœuds de sa chaîne orthologue. En continuant à explorer l'arbre de la racine vers les feuilles, le reste de la chaîne se trouvera ordonné à son tour.

A l'issue de ce traitement de réordonnement complet de l'arbre, un dernier traitement consiste à fusionner en une polytomie les nœuds orthologues successifs portant des taxons de même âge. L'arbre ainsi obtenu constitue l'horogramme H+.

### 3/ Calcul de l'horogramme H-.

L'horogramme H- est calculé à partir du cladogramme de base de la même manière que l'horogramme H+, sauf que les taxons sont ordonnés de façon à placer les moins âgés vers la racine.

Le traitement est le même que pour l'horogramme H+. Les points où l'âge importe sont notés dans les § 1 et 2 par (\*) ou (\*\*):

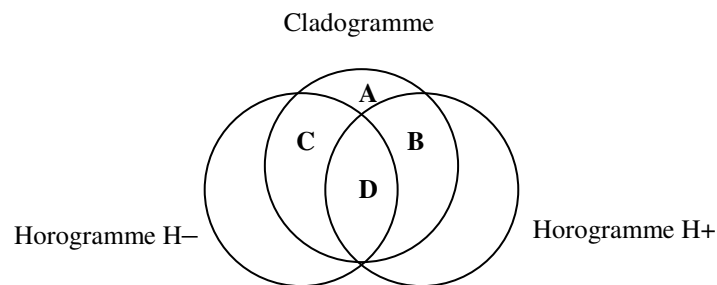
- pour les points marqués (\*), le traitement est le même pour H+ et pour H-.
- pour les points marqués (\*\*), le traitement est inversé pour H- : on considère alors le taxon d'âge minimal.

### 4/ Calcul des 3is et du rapport HiFi.

La liste des 3is générés par chacun des trois arbres obtenus (cladogramme, horogrammes H+ et H-) est calculée. Chaque 3is de cette liste peut appartenir à un seul des trois arbres, à deux d'entre eux ou aux trois. Pour chacun des 3is générés par le cladogramme de base est également calculée sa pondération fractionnaire PF.

Le rapport HiFi est calculé comme le rapport entre :

- Les 3is communs au cladogramme et à l'horogramme H+, mais absent de l'horogramme H-.
- Les 3is du cladogramme, à l'exception de ceux communs avec les deux horogrammes.



$$\text{HiFi} = B / (A + B + C)$$

Deux rapports sont calculés :

- L'un utilisant simplement le nombre de 3is dans chacun de ces deux ensembles.
- L'autre utilisant la somme des PF des 3is de chacun de ces deux ensembles.

Auteur : J. Ducasse, juillet 2006.

### Options.

- ∇ En plus de l'affichage par défaut, affiche :
  - Toutes les opérations individuelles de modification des arbres.
  - La liste des 3is du système avec, pour chaque 3is, son appartenance au cladogramme ou aux horogrammes, ainsi que sa pondération fractionnaire.

### Clefs de configuration.

Le programme utilise le fichier de configuration `tts.cfg`. Ce fichier est recherché en priorité dans le répertoire courant ou, à défaut, dans le répertoire référencé par la variable d'environnement `TTS_PKG`. Les entrées suivantes sont utilisées :

FORMAT            La valeur est une série de mots clefs spécifiant des options de format en sortie. Les valeurs suivantes sont acceptées :

p2/p4    Format des 3is : avec p2, les 3is seront produits au format "a (b c) " ; avec p4, ils seront produits au format " (a (b c) ) ". Par défaut : p2.

### Format du fichier d'entrée *filein*.

Ce fichier a un format semblable au format `.3ia` (voir le manuel du programme `3ia`). Les sections suivantes sont prises en compte :

- Taxons : *idem* `3ia`.
- Descriptions : *idem* `3ia`. On peut avoir plusieurs cladogrammes, qui seront traités l'un après l'autre mais de façon tout à fait indépendante.
- HiFi : section spécifique au programme `3iahifi`, donnant la relation d'ordre entre les taxons.

Exemple de section HiFi :

```
HiFi
A B C
D
E F
;
```

Cette section a la syntaxe suivante :

- Elle donne la relation d'ordre entre les taxons. Les taxons sont spécifiés par leur code ; tous les taxons cités dans la section `Taxons` doivent y figurer une et une seule fois. La relation va du plus jeune au plus ancien.
- Dans une ligne, deux taxons séparés par "=" ont le même âge ; si deux taxons sont séparés par "<", l'âge du deuxième est supérieur à celui du premier. L'absence de séparateur équivaut à "=".
- D'une ligne à la suivante, l'âge augmente, sauf si la ligne commence par "=", auquel cas l'âge des taxons de la ligne (au moins le premier) est le même que celui du dernier taxon de la ligne précédente. L'opérateur "<" en début de ligne est facultatif.

Autrement dit, on peut aussi bien utiliser une syntaxe sans opérateur, en plaçant sur une même ligne les taxons de même âge et en passant à la ligne pour incrémenter l'âge, qu'une syntaxe avec opérateur. Ainsi, les trois listes suivantes sont équivalentes :

```
HiFi                            HiFi                            HiFi
A B                            A = B < C < D = E < F ;            A B < C < D E < F ;
C
D E
F ;
```

### Format du fichier en sortie.

Taxons ordonnés :

`3iahifi` affiche la liste ordonnée des taxons telle qu'il la comprise suite à l'analyse de la section `HiFi`.

Cladogramme XX :

Le cladogramme de base fourni par la section `Descriptions`.

Elimination des polytomies :

Si, et seulement si, le cladogramme possède des polytomies, le cladogramme résultant de leur suppression est affiché.

H+ :

L'horogramme H+ calculé par réordonnement des taxons selon leur âge.

Après fusion des polytomies :

Si, et seulement si, l'horogramme subit une modification par fusion des polytomies, l'horogramme H+ résultant est affiché.

H- :

L'horogramme H- calculé par réordonnement des taxons selon leur âge.

Après fusion des polytomies :

Si, et seulement si, l'horogramme subit une modification par fusion des polytomies, l'horogramme H- résultant est affiché.

HiFi =  $n / d = x.xxxx$  ; PF HiFi =  $y.yyyy$

$n$  et  $d$  sont les nombres de 3is constituant le rapport HiFi ( $n = B$ ,  $d = A+B+C$  dans la terminologie du schéma du § 4).  $x.xxxx$  est le rapport des nombres de 3is,  $y.yyyy$  le rapport des sommes des PF des mêmes 3is.

ortho / para =  $no / np$

$no$  et  $np$  sont les nombres de nœuds orthologues et paralogues respectivement dans le cladogramme de base.

### Détail des informations affichées sous l'option -v.

Avec l'option -v sont affichés en plus les informations suivantes.

#### 1/ Modifications unitaires de la structure de l'arbre.

Chaque modification unitaire d'un arbre est affichée. Les différents cas possibles sont les suivants.

## suppression du taxon *XX*

Suppression d'un taxon lors de la réduction des polytomies du cladogramme de base.

## permutation *XX* <-> *YY*

Permutation de deux taxons lors du calcul d'un horogramme.

## permutation *XX* <-> (*AA* ((*BB* *CC*) *DD*))

Permutation entre un taxon et un nœud paralogue, décrit par son sous-arbre, lors du calcul d'un horogramme.

## regroupement des taxons *XX* et *YY*

Fusion en une polytomie de deux nœuds orthologues successifs en fin de traitement d'un horogramme.

#### 2/ Liste des 3is.

La liste des 3is du système est détaillée. Pour chaque 3is sont affichées :

- Sa présence (+) ou son absence (-) dans le cladogramme de base, dans l'horogramme H+ et dans l'horogramme H-.
- Sa prise en compte dans le calcul du rapport HiFi :
  - « N » au numérateur et au dénominateur (B dans le schéma du § 4).
  - « D » au dénominateur seul (A ou C dans le schéma du § 4).
- Sa pondération fractionnaire s'il a été généré par le cladogramme de base.

### Historique.

version 1.3 : décembre 2009

Changement de nom → 3iahifi.

version 1.2 : novembre 2006

Ajout de l'affichage des nombres de nœuds orthologues et paralogues.

version 1.1 : octobre 2006

Les pondérations fractionnaires sont traitées sur 64 bits.

version 1 : juillet 2006.