

## TD 2 de Phylogénie - Correction

Arthur Oulès

Björn Wissel



**Un Xenomorphe de la saga Alien nous faisant de jolis sourires ([Xenopedia](#))**

Connaissances travaillées :

- Le maximum de vraisemblance
- Les datations en phylogénie
- Estimer l'évolution d'un caractère morphologique sur une phylogénie

Compétences travaillées :

- Description et interprétation de figures originales tirées d'article
- Analyse d'article sur une application de la phylogénie, vers l'écologie et la biologie fonctionnelle.

## 1 La pharyngognathie chez les Téléostéens

Les poissons Téléostéens possèdent, en plus de leurs mâchoires orales, une deuxième paire de mâchoires appelées mâchoires pharyngiennes (Figure 1).

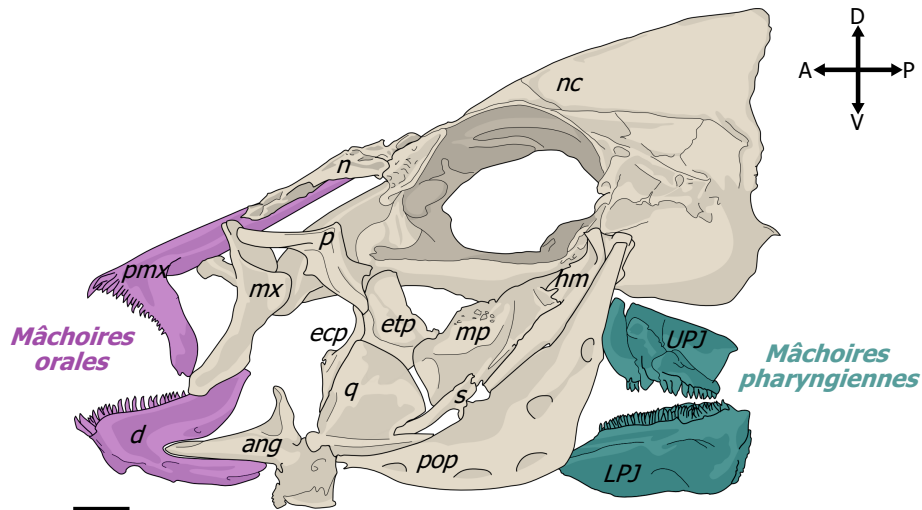


Figure 1: **Diagramme des os du crâne d'un poisson Téléostéen (ici le Cichlidé *Herichthys minckleyi*).**  
Les os de la mâchoire orale sont colorés en violet et ceux de la mâchoire pharyngienne en bleu. La barre d'échelle représente 5 mm. (Adapté d'A. Oulès)

Q1 : Quel peut-être l'intérêt fonctionnel de cette seconde paire de mâchoires ?



### Réponse 1

Chez les poissons chirurgiens (Acanthuridae), les mâchoires pharyngiennes sont réduites à des plaques dentigères dont l'action dans la mastication est très réduite.

A l'inverse, chez les poissons rouges (Cyprinidés), les mâchoires orales sont édentées et seules les mâchoires pharyngiennes portent des dents. Ainsi la mâchoire orale permet la capture de proies par succion et les mâchoires pharyngiennes permettent la mastication.

La présence de mâchoires pharyngiennes permet un découplage fonctionnel des mâchoires orales et pharyngiennes et le relâchement d'une partie des pressions de sélection exercées sur les mâchoires orales.

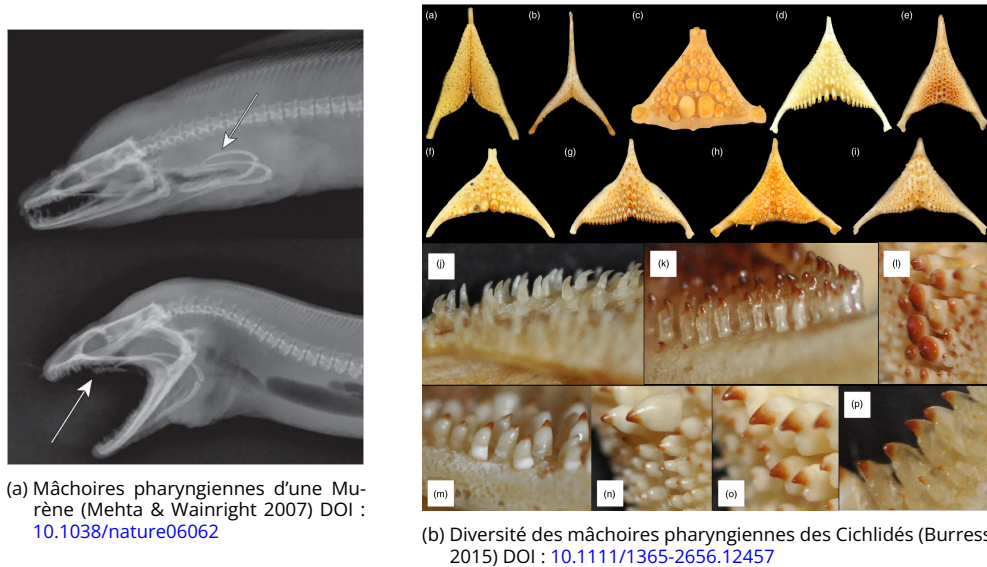


Figure 2: Diversité des mâchoires pharyngiennes chez les Téléostéens

Q2 : Pouvez-vous émettre une hypothèse sur l'origine développementale de ces mâchoires ?

### Réponse 2

Les mâchoires pharyngiennes sont positionnées postérieurement aux opercules, à l'emplacement du cinquième arc branchial chez les poissons sans mâchoires pharyngiennes. Une étude de leur ontogénie montre qu'elles sont issues d'une modification du cinquième arc branchial des Téléostéens.

Chez certains groupes, ce caractère est associé à de nombreuses modifications comme la fusion des demi-mâchoires pharyngiennes inférieures et une modification de la musculature associée (Figure 2). Ces modifications permettent notamment une force de mastication supérieure et l'apparition de régimes alimentaires à base d'aliments coriaces comme les molluscivores ou les corallivores, ce mode d'alimentation à partir de mâchoires pharyngiennes dérivées est appelé pharyngognathie.

i La construction des phylogénies calibrées (ou chronogrammes) est abordée plus en détails dans la section suivante du TD.

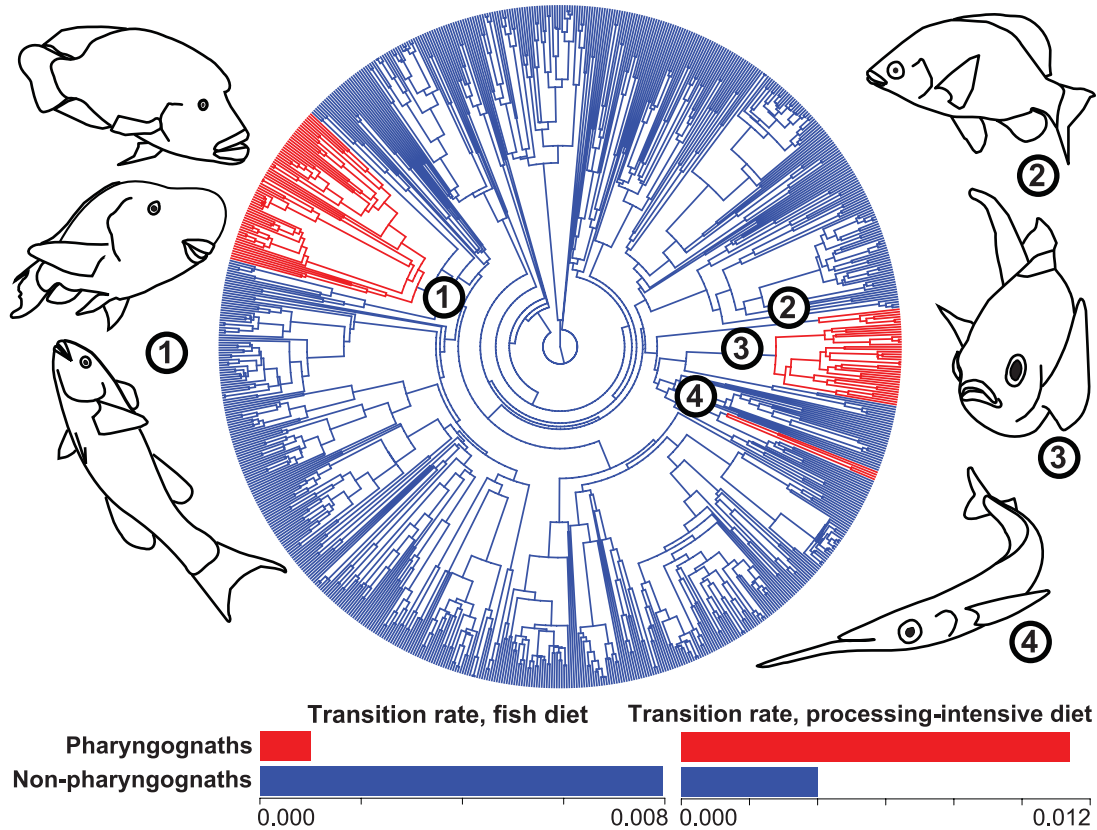
Q3 : À partir du Document 1 A, décrivez l'évolution de la pharyngognathie chez les Actinoptérygiens.

### Réponse 3

C'est un chronogramme, donc un arbre phylogénétique calibré temporellement. Ici, il est représenté circulairement bien qu'il ait été raciné avant sa calibration temporelle. Les feuilles, le long du périmètre du cercle, correspondent au présent et la racine, au centre, est le point le plus ancien temporellement. Cet arbre inclut 851 espèces, c'est ce qu'on appelle une méga phylogénie. Les branches de l'arbre correspondant à des espèces pharyngognathes sont colorées en rouge et à des non-pharyngognathes en bleu.

On observe 4 clades pharyngognathes différents, l'hypothèse la plus parcimonieuse est celle d'une convergence évolutive, les hypothèses alternatives étant moins parcimonieuses (celle qui nécessite le moins de changements évolutifs) car demandant trop de réversions. La pharyngognathie serait donc un caractère polyphylétique.

**Extrait 1 :** The degree of piscivory was defined as the volumetric percentage of non-larval fish in the diet, and the degree of processing-intensive prey was defined as the volumetric percentage of plants, algae, and detritus combined with the percentage of hard-shelled prey like bivalves, shelled gastropods, and hard-bodied echinoderms (*i. e.* no holothurian "sea cucumbers").



Document 1: **Impact de la pharyngognathie sur la transition de régime alimentaire chez les poissons marins.** (McGee et al 2015 DOI : [10.1126/science.aab0800](https://doi.org/10.1126/science.aab0800))

(A) Quatre transitions vers la pharyngognathie sur une phylogénie de 851 Actinoptérygiens calibrée temporellement. (1) Les Labres (Labridae et Scaridae); (2) Les Embiotocidae; (3) Les Demoiselles (Pomacentridae); (4) Les Hemirhamphidae

(B) Comparaison des taux de transitions de régime alimentaire vers la piscivorie (poissons non larvaires) ou vers des proies à mastication intense (détritiques, plantes, proies à coquilles dures), pour des poissons pharyngognathes ou non.

We used the `brownie.lite` function in the `phytools` R package to test whether each diet category was best modeled as a one-rate model across the tree, or a model where rates differed between pharyngognathous and non-pharyngognathous marine fishes. There was a strong preference for a two-rate model for both fish prey ( $\Delta\text{AICc} = -293$ ) and processing-intensive prey ( $\Delta\text{AICc} = -59$ ). For both traits, the difference between pharyngognaths and non-pharyngognaths for the Brownian rate parameter in the two-rate model was always significant ( $p < 0.001$ ).

**i** AIC = Akaike Information Criterion (Critère d'information d'Akaike) est une mesure de la qualité relative d'un modèle statistique sachant un ensemble de données. Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un paramètre. Le critère d'information d'Akaike prend cela en compte en pénalisant les modèles en fonction du nombre de paramètres. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible. Le  $\Delta\text{AICc}$  inclut une correction liée à la taille du jeu de données.

Q4 : Qu'est-ce qu'un mouvement Brownien ? Comment peut-il être utilisé pour modéliser l'évolution d'un trait continu ?

#### 💡 Réponse 4

La description historique du mouvement Brownien est attribuée à Robert Brown en 1827 alors qu'il observait des grains de pollen dans une goutte d'eau au microscope. Leurs déplacements stochastiques l'ont intrigué mais la modélisation mathématique de ce phénomène n'a eu lieu qu'à la fin du siècle.

Statistiquement, la position d'une particule suivant un mouvement Brownien  $\bar{z}(t)$  est décrite par une marche aléatoire avec les caractéristiques suivantes :

1.  $E[\bar{z}(t)] = \bar{z}(0)$ , le mouvement moyen est nul autour de la position de départ. Le mouvement n'a pas de tendance particulière
2. Les différents "pas" infinitésimaux de la marche sont indépendants
3.  $\bar{z}(t) \sim \mathcal{N}(\bar{z}(0), \sigma^2 t)$  Le mouvement de la particule à un instant  $t$  suit une loi normale de moyenne  $\bar{z}(0)$  et de variance  $\sigma^2 t$

Dans le cas de notre grain de pollen, si l'on considère que ses mouvements sont influencés par la somme des chocs imposés par les molécules d'eau environnantes, peu importe les lois décrivant les mouvements de ces molécules, leur somme tend vers une loi normale par le théorème central limite.

Ainsi, c'est le paramètre  $\sigma^2$  qui est déterminant dans la description de cette marche.

Revenons à la description de l'évolution de traits continus. Prenons la masse corporelle, on peut légitimement supposer que ce trait est contrôlé par une multitude de gènes subissant des mutations aléatoires. Ainsi, la somme de petits écarts de la valeur de ce trait dues à des mutations de ces gènes peut être vu comme une marche aléatoire, de la même manière que les molécules d'eau décalant le grain de pollen par collision.

On peut donc appliquer des hypothèses de travail similaire et utiliser le même modèle mathématique. Ici, on cherchera à retrouver les paramètres  $\bar{z}(0)$ , l'état ancestral et  $\sigma^2$  le taux d'évolution.

La modélisation et les hypothèses de travail sont grandement simplifiées dans ces explications, pour approfondir ce concept, je vous redirige vers le chapitre 3 cet excellent livre de Luke J. Harmon : <https://lukejharmon.github.io/pcm/chapters/>.

Q5 : Rappeler le principe du maximum de vraisemblance. La fonction *brownie.lite* inclut ce principe, à votre avis, de quelle façon ?

### Réponse 5

Commençons par un exemple simple et concret, une pièce de monnaie que l'on lance douze fois avec le tirage suivant :

$$P - F - F - P - F - P - P - F - F - F - P - F$$

On veut à présent vérifier si la pièce est truquée ou non, on va donc dans un premier temps faire l'hypothèse qu'elle ne l'est pas et donc que les chances d'obtenir pile ou face sont de 50% chacune, on a alors la probabilité d'obtenir ce tirage sachant que la pièce est équilibrée qui vaut :

$$P(\text{Tirage}|\text{Pièce équilibrée}) = (0,5)^7 \times (0,5)^5 = 2,4 \cdot 10^{-4}$$

Maintenant, avec l'hypothèse d'une pièce biaisée avec 70% d'obtenir face, on obtient :

$$P(\text{Tirage}|\text{Pièce truquée}) = (0,7)^7 \times (0,3)^5 = 2,0 \cdot 10^{-4}$$

On a donc  $P(\text{Tirage}|\text{Pièce équilibrée}) > P(\text{Tirage}|\text{Pièce truquée})$ , on dit qu'il est plus vraisemblable que la pièce soit équilibrée.

Formellement, la vraisemblance d'un modèle sachant un jeu de donnée est la probabilité conditionnelle d'obtenir un jeu de données sachant un modèle.

Ainsi, on peut définir une fonction qui à un ensemble de modèles va attribuer une valeur de vraisemblance : la fonction de vraisemblance. C'est cette fonction que l'on cherche à maximiser afin de trouver le modèle qui, sachant nos données, est le "meilleur" d'après nos critères.

Dans notre exemple simplifié, on peut écrire que pour toute chance  $p$  de tomber sur Face qui décrit un modèle on a :

$$\mathcal{L}(\text{modle}(p)|\text{Tirage}) = P(\text{Tirage}|\text{modle}(p)) = (p)^7 \cdot (1-p)^5$$

Où  $\mathcal{L}$  est la fonction de vraisemblance du modèle sachant le tirage qu'on a eu. Il est donc simple de trouver une solution analytique dans ce cas !

Cependant, quand on travaille avec des phylogénie ou des modèles complexes de manière générale, on a bien plus de paramètres et pas toujours de solution analytique et il faut passer par des estimations de cette fonction qui sont gourmandes en puissance et temps de calcul.

Dans le cas de la fonction `brownie_lite`, l'évolution du trait le long d'un arbre est modélisée par un mouvement Brownien de paramètres  $\bar{z}(0)$  et  $\sigma^2$  comme vu en réponse à la question précédente et le maximum de vraisemblance permet de trouver le jeu de paramètres (le taux d'évolution du degré de piscivorie et le caractère ancestral) le plus vraisemblable sachant notre jeu de donnée (la topologie d'arbre et les caractères actuels).

Q6 : Quel est le résultat obtenu à l'issue de ces tests ?

### Réponse 6

Les chercheurs ont testé si le taux d'évolution des régimes alimentaires testés étaient différents en fonction du caractère pharyngognathe ou non-pharyngognathe.

Dans un premier temps, ils étudient la transition vers la piscivorie.

Pour cela, ils ont construit deux modèles dont l'un est la simplification de l'autre (on parle de modèles emboîtés).

Dans le modèle M0 le plus simple, le paramètre estimé (transition vers la piscivorie) a la même valeur dans tout l'arbre. Dans le modèle M1 plus complexe, le taux de transition vers la piscivorie peut avoir deux valeurs différentes dans les branches rouges et bleues. Ils calculent le maximum de vraisemblance pour M0 et M1 et comparent ces vraisemblances pour choisir le meilleur modèle, c'est-à-dire le modèle le plus simple possible qui explique au mieux les données (grâce à l'AIC).

Ils choisissent M1 et concluent que les taux de transitions sont significativement différents dans les deux cas (pharyngognathes vs. non-pharyngognathes) et pour les deux régimes alimentaires testés (piscivorie et mastication intense).

**Extrait 2 :** Our Brownian rate analysis treated the percentage of either processing-intensive or fish prey as a continuous character. However, if we define each species as a prey specialist if it eats over a certain percentage of that prey in its diet, we can also analyze diet as a categorical variable using stochastic character mapping.

We also used stochastic character mapping in the `phytools` package to examine transitions between dietary strategies. Fish were scored as processing-intensive if they had a mean value of processing-intensive prey over 25% of their diet, and as piscivorous if the mean value of non-larval fish in their diet was higher than 25%. Each fish was then scored as non-pharyngognath non-piscivore, non-pharyngognath piscivore, pharyngognath non-piscivore, and pharyngognath



piscivore.

We used the `fitDiscrete` function in the R package `geiger` to compare the transition rate matrix for two models of character evolution. Specifically, we tested a model with equal transition rates between non-pharyngognath non-piscivores and non-pharyngognath piscivores versus between pharyngognath non-piscivores and pharyngognath piscivores. We then compared this equal-rate model to one where the transition rates were allowed to differ. All other rates were allowed to differ. Transitions were always modeled as symmetric, *i. e.* equal rates into and out of a category.

Q7 : Pourquoi recalculer les taux de transition en considérant, cette-fois, le trait comme discret ? Quelle(s) différence(s) avec le modèle précédent ?



#### Réponse 7

Des caractères comme le régime alimentaire ont des composantes discrètes et continues. Comme montré par ces chercheurs, on peut catégoriser les régimes alimentaires en variable discrète, ce qui permet de mieux détecter des événements de convergence par exemple, ou au contraire comme une variable continue. En réalité, le régime alimentaire est rarement discret et, au moins par opportunisme, les organismes peuvent dévier en partie de leur régime alimentaire préféré. Ce degré de liberté dans le régime alimentaire peut être un paramètre pertinent à étudier.

Une autre façon de voir les choses est que le modèle continu se rapproche d'une vision gradualiste de l'évolution et le modèle discret d'une évolution par équilibres ponctuels. Ces deux visions sont complémentaires et décrivent des processus différents, ainsi coupler ces deux approches peut permettre de détecter ces processus. Outre ces considérations, multiplier les approches apporte de la robustesse aux résultats.

**Extrait 3 :** AIC strongly favored the model where non-pharyngognaths had a different transition rate into and out of piscivory as compared to pharyngognaths ( $\Delta AIC = -20.3$ ). We used a similar character mapping scheme and analysis for pharyngognathy and processing-intensive prey, and found a preference for different rates of transition within pharyngognaths as compared to non-pharyngognaths ( $\Delta AIC = -8.6$ ). Rates of transition between piscivory and non-piscivory were an order of magnitude higher in non-pharyngognaths ( $7.96e-03$  vs  $9.95e-04$ , Document 1 B). For processing-intensive prey, rates were an order of magnitude lower in non-pharyngognaths ( $4.03e-03$  vs  $1.15e-02$ , Document 1 B).

Q8 : Ces résultats corroborent-ils ceux obtenus en considérant la variable comme continue ? Quelles différences observe-t-on entre les poissons pharyngognathes et non-pharyngognathes ?

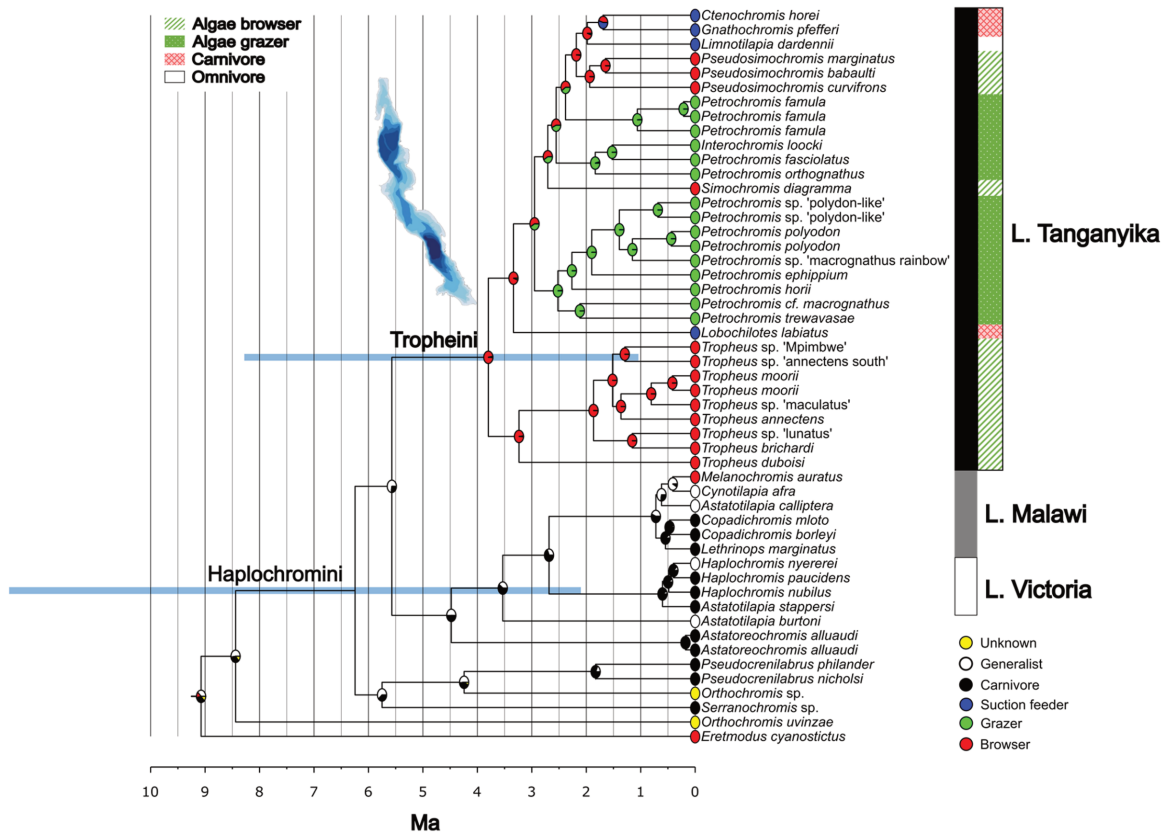


#### Réponse 8

Oui, ces résultats sont en accord avec les résultats précédents. On observe que, pour un régime alimentaire donné, le taux de transition est différent entre pharyngognathes et non-pharyngognathes.

Ce taux est plus important pour les pharyngognathes pour une transition vers la mastication intense et plus important pour les non-pharyngognathes pour une transition vers la piscivorie. Ainsi, la pharyngognathie semble favoriser des transitions vers des régimes alimentaires demandant une mastication intense mais réduire les transitions vers la piscivorie.

## 2 Diversification des régimes alimentaires et radiation évolutive chez les Cichlidés des grands lacs Africains



Document 2: **Phylogénie RAxML des Tropheini calibrée temporellement et faisant apparaître les reconstructions des traits ancestraux pour la spécialisation trophique.** (Singh et al 2022 DOI : [10.1002/ece3.9077](https://doi.org/10.1002/ece3.9077))

L'arbre a été reconstitué par maximum de vraisemblance. Les temps de divergences ont été inférés avec RelTime à partir d'un schéma de calibration vicariance-dépendant incluant des fossiles datés et la fragmentation du Gondwana. Les abscisses représentent le temps en millions d'années avant le présent. Les intervalles de confiance des nœuds clefs sont marqués par des boîtes bleues.

Q9 : Quels sont les deux types d'informations utilisés pour calibrer temporellement cet arbre ? En quoi sont-elles complémentaires ? Sont-elles suffisantes pour calibrer l'entièreté de l'arbre ?



### Réponse 9

Les deux types d'informations utilisés sont les fossiles datés (par datation relatives ou absolues) et certains événements géologiques datés et corrélés à des événements de spéciation ou de divergence. Ces informations permettent de placer des "points d'ancrages" en datant certains nœuds de la phylogénie, avec une incertitude, mais il faut encore associer une date aux autres points.

C'est avec des modèles d'évolution moléculaires que l'on peut dater les autres nœuds de l'arbre. On peut citer le plus simple et connu, le modèle de Jukes-Cantor qui considère des taux de substitutions égaux entre toutes les bases azotées, une indépendance des sites de substitution et des fréquences nucléotidiques égales.

Q10 : Interpréter le Document 2





**Réponse 10**

Pas fait en TD.