### Examen Octobre 2020 Sujet de Michel Vervoort

Les Tenrecidés forment une famille de petits mammifères insectivores, la plupart vivant exclusivement à Madagascar. Ils appartiennent à une des 4 grandes lignées de Mammifères, appelée Afrotheria. La figure 1 montre un exemple d'un Tenrecidé, *Echinops telfairi*, appelé en anglais « lesser hedgehog tenrec », ainsi qu'une phylogénie très simplifiée des mammifères (d'après Cornelis et al., 2014).

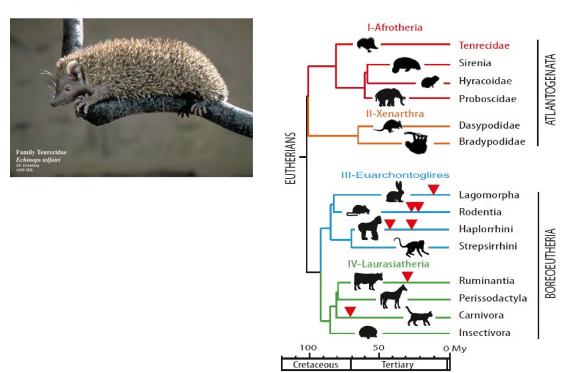


Figure 1 : Photo d'Echinops telfairi et phylogénie simplifiée des mammifères.

Des chercheurs (Cornelis et al., 2014) ont cherché à identifier si *Echinops telfairi* contenait un ou plusieurs gènes de la famille des syncytines, gènes qui codent des protéines importantes pour la formation du placenta (organe reliant l'embryon ou le foetus à la mère, entre lesquels il permet les échanges; le placenta contient des grosses cellules à plusieurs noyaux, ce qu'on appelle un syncytium) chez les mammifères. Ils ont identifié 9 séquences d'*Echinops telfairi* présentant une forte similitude de séquence avec les syncytines des autres mammifères. Les chercheurs ont ensuite caractérisé la transcription de ces 9 séquences et ont observé que seule l'une d'entre elles (appelée *Ten-env1*) est transcrite dans le placenta de *Echinops telfairi*. Les autres séquences sont soit non transcrites soit transcrites dans d'autres tissus. Les chercheurs se sont concentrés par la suite sur *Ten-env1*. Ils ont cloné cette séquence dans un vecteur d'expression (plasmide permettant la synthèse d'ARNm et donc de protéines) et ont introduit ce plasmide dans des fibroblastes (cellules venant de la peau et pouvant être facilement mises en culture *in vitro*). Après introduction du plasmide et production de la protéine codée par *Ten-env1*, ils ont observé que ces cellules fusionnent les unes avec les autres et forment des cellules avec plusieurs noyaux, ce que les cellules contrôles (ne produisant pas la protéine codée par *Ten-env1*) ne font jamais.

#### Question 1 : Que pouvez-vous conclure de cet ensemble de résultats ?

9 séquences d'*Echinops telfairi* qui ont une similitude significative de séquence avec les syncytines des autres mammifères, permettant de déduire que ces 9 séquences identifiées chez *Echinops telfairi* sont donc probablement des homologues (orthologues) des gènes codant les Syncytines des autres mammifères. Parmi les 9 séquences identifiées chez *Echinops telfairi*, seule une seule (*Ten-env1*) est transcrite dans le placenta et donc seule cette séquence peut potentiellement produire une protéine pouvant jouer le rôle de Syncytine. Il a en effet été vu en cours que les syncytines jouent un rôle dans

la formation de syncytium au niveau du placenta chez les mammifères placentaires. Les chercheurs observent qu'introduction et expression de *Ten-env1* dans des fibroblastes, ceux-ci ont tendance à former un syncytium, confirmant que *Ten-env1* code très probablement une Syncytine.

Les chercheurs ont ensuite montré que *Ten-env1* est localisé chez *Echinops telfairi* dans une région génomique définie par la présence de deux gènes (*Fam213A* et *DYDC2*) et très conservée chez les mammifères (voir figure page suivante ; en haut). Cette région a été analysée chez un certain nombre d'autres espèces de mammifères et des séquences similaires à *Ten-env1* n'ont été retrouvées dans aucune de ces espèces. En outre, chez *Echinops telfairi*, des séquences remarquables sont retrouvées à proximité de *Ten-env1* : dans la figure 2 ci-dessous, en jaune, deux séquences présentant une similitude significative de séquence avec des séquences dites « Pol » et, en noir, une courte séquence similaire à des séquences dites « LTR », retrouvées dans les bases de données.

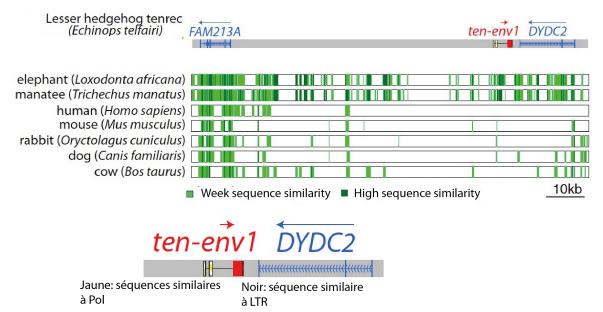


Figure 2 : Région génomique où se trouve Ten-env1. La séquence similaire à LTR correspond au fin trait noir immédiatement adjacent à Ten-env1 (carré rouge).

# Question 2 : A l'aide des résultats obtenus et de vos connaissances, en justifiant votre réponse, proposez une hypothèse quant à l'origine évolutive de *Ten-env1*.

Il a été vu en cours que les syncytines des mammifères comme l'Homme et la souris dérivent de la « domestication » de rétrovirus ou de rétrotransposons à LTR (les deux sont apparentés et très proches). Plus particulièrement, ces gènes dérivent de la phase ouverte de lecture *env* de ces virus/rétrotransposons à LTR. On peut faire l'hypothèse qu'il en est de même pour *Ten-env1* d'*Echinops telfairi*. L'analyse de la région génomique de la région où se trouve *Ten-env1* conforte cette hypothèse. En effet, à proximité de *Ten-env1* se trouvent, du côté 3', une séquence similaire à un LTR (typique des rétrotransposons à LTR et des rétrovirus) et, du côté 5', une séquence similaire à la région *pol* (elle aussi retrouvée dans les rétrotransposons à LTR et les rétrovirus). De plus si on examine la région plus globale où se trouve *Ten-env1*, on constate que cette séquence n'existe que chez *Echinops telfairi* et pas chez une série d'autres espèces appartenant à différentes lignées de mammifères. L'hypothèse plus précise serait donc qu'un rétrotransposon à LTR ou un rétrovirus se serait inséré dans cette région génomique chez un ancêtre de *Echinops telfairi*. Ce rétrotransposon à LTR /rétrovirus aurait dégénéré (perte d'un LTR et peut-être de la région Gag) et la phase ouverte de lecture *env* serait devenu *Ten-env1*, la syncytine d'*Echinops telfairi*.

Grâce à des amorces spécifiques de la région où se trouve *Ten-env1*, les chercheurs ont recherché par PCR la présence de séquence similaire chez d'autres espèces de Tenrécidés et de quelques autres

espèces. Les fragments d'ADN obtenus par PCR ont été séquencés et leur séquence comparée à celle de *Ten-env1*. Dans les espèces où une séquence avec une forte similitude de séquence a été trouvée, cette séquence a été clonée et, par l'expérience mentionnée à la page précédente, la capacité de ces séquences à induire la fusion de fibroblastes et la formation de cellules à plusieurs noyaux a été testée. Ces résultats sont résumés dans la figure 3 ci-dessous :

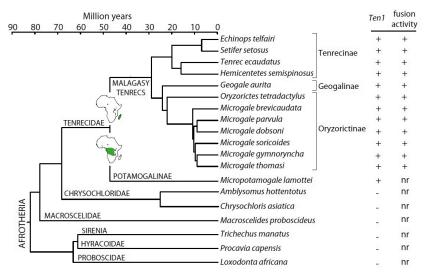
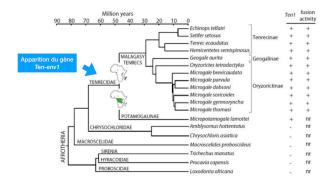


Figure 3 : Arbre phylogénétique d'espèces de Tenrécidés (Tenrecidae en anglais) et de groupes proches. Dans cette figure, + indique la présence d'une séquence similaire à Ten-envl et l'activité de fusion de cette séquence, - indique l'absence de séquence similaire, nr = not relevant. Dans le cas de Micropotamogale lamottei, seule une séquence partielle a été amplifiée par PCR et l'activité de fusion de cette séquence n'a donc pas pu être testée.

#### Question 3 : Que pouvez-vous conclure de ces résultats ?

Des homologues de *Ten-env1* sont retrouvés chez tous les Tenrecidés et, chez les tous Tenrecidés de Madagascar, il a été montré que ces homologues codent des protéines capables d'induire la formation de syncytium, donc potentiellement de jouer un rôle de Syncytine. Dans le cas de l'espèce d'Afrique centrale (*Micropotamogale lamottei*), l'activité n'a pas pu être testée. Chez les autres mammifères représentés dans la figure, y compris les plus proches parents des Tenrecidés que sont les Chrysochloridés, il n'ya pas d'homologue de *Ten-env1*. La parcimonie nous suggère donc que le gène serait apparu chez un ancêtre des Tenrecidés après la séparation avec les Chrysochloridés (éventuellement représenter une version schématique de l'arbre de la figure et placer l'événement sur cet arbre). En se basant sur l'hypothèse faite précédemment, on peut aussi suggérer (mais sans certitude) que le rétrotransposon à LTR/rétrovirus à l'origine de *Ten-env1* se serait inséré chez cet ancêtre des Tenrecidés.



Les séquences obtenues par séquençage des fragments de PCR amplifiés dans différentes espèces de Tenrécidés ont été utilisées pour construire un arbre phylogénétique par la méthode du Maximum de Vraisemblance (figure 4 ci-dessous). Il s'agit d'un arbre raciné — l'enracinement s'est fait par la méthode dite du « mid-point rooting ».

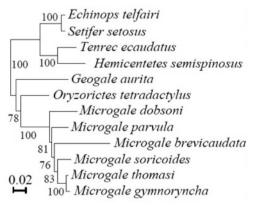


Figure 4 : Arbre phylogénétique des séquences Ten-env1 retrouvées dans les espèces indiquées dans l'arbre. Les valeurs à côté des nœuds de l'arbre sont les valeurs (pourcentages) de bootstrap. Seules les valeurs supérieures à 70% sont indiquées.

# Question 4a : Expliquez quelles sont les différentes étapes ayant permis la construction de cet arbre phylogénétique.

- (0) (optionnel à mentionner) Obtention des données de séquences.
- (1) Alignement multiple des séquences qui va permettre de mettre en regard les positions homologues des différentes séquences comparées et donc de comparer, pour chacune de ces positions (correspondant à un caractère), le type de nucléotides ou d'acides aminés présents, ce qu'on appelle l'état du caractère. C'est cette comparaison qui sera utilisée pour reconstruire ensuite l'arbre phylogénétique.
- (2) Construction de l'arbre phylogénétique par la méthode du maximum de vraisemblance détermination à l'aide d'une approche heuristique (vu le nombre de séquences) en calculant la vraisemblance des données en fonction du modèle d'évolution des séquences choisi et de l'arbre testé. On retient l'arbre pour lequel on a une vraisemblance des données maximale. Il s'agit d'un seul arbre, dit de maximum de vraisemblance, non raciné.
- (3) Enracinement de l'arbre par la méthode du mid-point rooting. Dans cette méthode, la racine est placée au point de l'arbre équidistant de toutes les OTUs et coupe l'arbre en deux parties d'égale longueur. La racine représente l'ancêtre commun à toutes les OTUs de l'arbre et le fait de placer une racine permet de définir le chemin unique qui conduit à tous les autres noeuds. La direction entre la racine et les différentes OTUs représente le temps évolutif.
- (4) Détermination de la robustesse des nœuds- cfr. question 4b.

# Question 4b : Précisez quelle information nous apportent les valeurs de bootstrap indiquées à côté des nœuds de l'arbre. Il n'est pas nécessaire d'expliquer la méthode des bootstrap.

Ces valeurs donnent une estimation de la robustesse du nœud en question, c'est-à-dire si le nœud en question est fortement soutenu par les données. Autrement dit si ce nœud sera retrouvé même si on modifie les données (ré-échantillonnage aléatoire dans le cas des bootstraps).

## Question 4c : Quelle autre méthode aurait pu être utilisée pour faire l'enracinement de cet arbre ? Expliquez brièvement cette méthode.

L'enracinement aurait pu se faire grâce à l'utilisation d'un groupe externe. Il s'agit inclure dans l'analyse une ou un groupe de séquences homologues aux séquences analysées mais dont on sait a priori qu'elles sont extérieures aux séquences analysées. Il pourrait s'agir ici par exemple de séquences *env* de rétrotransposons à LTR/rétrovirus connus dans d'autres espèces de mammifères. La racine est définie alors par le nœud reliant le groupe extérieur aux séquences étudiées.