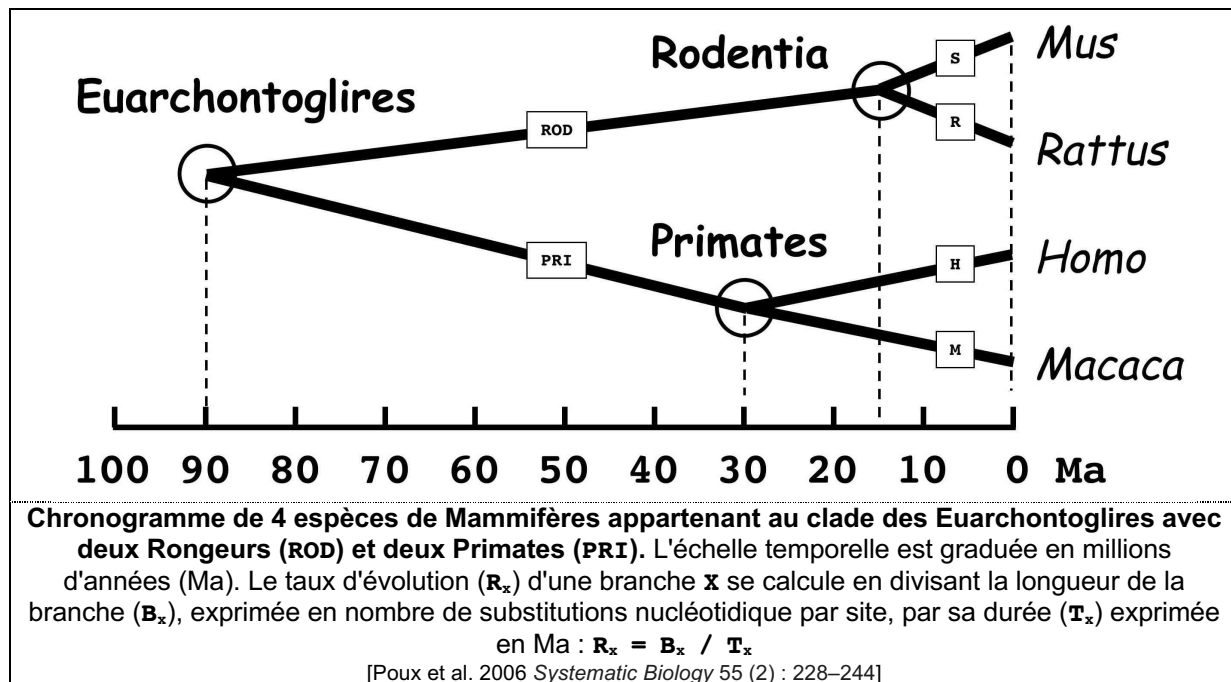




ÉVOLUTION MOLÉCULAIRE COMPARÉE DES EXONS NUCLÉAIRES ORTHOLOGUES DE PRIMATES ET DE RONGEURS



Dans ce projet, une facette de l'évolution moléculaire des séquences d'ADN codantes sera abordée chez les Mammifères. Il s'agit de mesurer et comparer les vitesses d'évolution de séquences d'exons orthologues chez deux Primates — l'homme (*Homo sapiens* : **H**) et le macaque (*Macaca fascicularis* : **M**) — et chez deux Rongeurs — la souris (*Mus musculus*, : **S**) et le rat (*Rattus norvegicus* : **R**).



0. Exploitation de la base de données OrthoMaM (Orthologous Mammalian Markers).

- Interroger OrthoMaM v10b : <https://orthomam.mbb.cnrs.fr> → Query
- Se focaliser uniquement sur les **exons** (et non sur les CDS) ;
- Paramètres par défaut :
 - Relative evolutionary rate = de 0 à 2 ;
 - %G+C₃ = de 23 à 45% ;
 - α shape = de 0 à 2 ;
 - Alignment length = de 0 à 40 000 sites ;
 - Human sequence length = de 0 à 40 000 sites.
- Genes present in **ALL** selected species and/or groups ;
- Sélectionner 4 espèces qui devront toujours être présentes :
 - *Homo sapiens* (homme) ;
 - *Macaca fascicularis* (macaque) ;
 - *Mus musculus* (souris) ;
 - *Rattus norvegicus* (rat).
- Décompte préliminaire → **463** exons orthologues sont disponibles.
- **Étendre à son maximum la gamme de chaque paramètre → 5 334 exons.**
- Extraire les alignements d'exons (Nucleotide (NT) alignments) et les arbres (Maximum likelihood trees) correspondant à une gamme donnée de composition nucléotidique (%G+C₃) :

Catégorie/Gamme	01	02	03	04	05	06	07	08	09	10
%G+C ₃ minimum	22.0	35.0	39.0	42.0	44.0	45.5	47.0	48.5	50.0	51.5
%G+C ₃ maximum	34.9	38.9	41.9	43.9	45.4	46.9	48.4	49.9	51.4	52.9
Nombre d'exons	98	106	113	100	104	99	121	91	138	106

Catégorie/Gamme	11	12	13	14	15	16	17	18	19	20
%G+C ₃ minimum	53.0	54.5	56.0	57.0	58.0	59.0	60.0	61.0	62.0	63.0
%G+C ₃ maximum	54.4	55.9	56.9	57.9	58.9	59.9	60.9	61.9	62.9	63.9
Nombre d'exons	139	102	97	108	101	112	129	124	139	126

Catégorie/Gamme	21	22	23	24	25	26	27	28	29	30
%G+C ₃ minimum	64.0	65.0	66.0	67.0	68.0	69.0	70.0	71.0	72.0	73.0
%G+C ₃ maximum	64.9	65.9	66.9	67.9	68.9	69.9	70.9	71.9	72.9	73.9
Nombre d'exons	135	151	153	158	162	164	150	179	183	177

Catégorie/Gamme	31	32	33	34	35	36	37	38	39	40
%G+C ₃ minimum	74.0	75.0	73.0	77.0	78.0	79.0	80.0	81.0	82.0	83.0
%G+C ₃ maximum	74.9	75.9	73.9	77.9	78.9	79.9	80.9	81.9	82.9	83.9
Nombre d'exons	176	170	141	127	116	143	100	82	95	76

Catégorie/Gamme	41	42	43
%G+C ₃ minimum	84.0	85.5	88.0
%G+C ₃ maximum	85.4	87.9	97.0
Nombre d'exons	96	80	69

Dans cette partie, chaque **arbre phylogénétique** d'OrthoMaM — inféré à partir de chaque alignement d'exons — sera analysé en langage **R**.

1.A. Estimer le **taux d'évolution** de chacune des 6 branches, à savoir celui de la souris (R_s), du rat (R_r), de l'homme (R_h), du macaque (R_m), de l'ancêtre des Rongeurs (R_{rod}), et de l'ancêtre des Primates (R_{pri}).

Paquet R	APE	Analyses of Phylogenetics and Evolution [Paradis 2012]
Fonctions R	read.tree	Reads a file which contains one or several trees in parenthetic format.
	getMRCA	Returns the MRCA of two or more tips.
	dist.nodes	Computes pairwise distances between pairs of (internal and terminal) nodes from a phylogenetic tree using its branch lengths.

1.B. En déduire la **médiane** et les **quantiles à 5% et 95%** de chacun des 6 taux d'évolution pour l'ensemble des exons de la catégorie considérée.

1.C. Représenter sous la forme d'un **unique histogramme** la distribution des taux d'évolution à la fois pour la souris (R_s) et pour l'homme (R_h), ceci pour l'ensemble des exons de la catégorie analysée. N'oubliez pas d'étiqueter les axes.

1.D. Comparez les résultats obtenus. Vous émettrez une brève **hypothèse** pour expliquer vos observations.

1.E. Synthèse et Perspectives — Regroupez sous forme de tableau pour l'ensemble de la promotion du M1 BioInformatique les résultats concernant la **médiane** et les **quantiles à 5% et 95%** des taux d'évolution pour la souris (R_s) et pour l'homme (R_h), en fonction du pourcentage de G+C₃ de chacune des catégories d'exons. Commentez le tableau récapitulatif des résultats. Proposez alors une ou plusieurs perspectives à donner à ce projet.

TABLEAU RÉCAPITULATIF DES RÉSULTATS POUR DIFFÉRENTS POURCENTAGES DE G+C ₃														
Catégorie/Gamme	01	04	07	10	13	16	19	22	25	28	31	34	37	etc ...
%G+C ₃ minimum														
Nombre d'exons														
Taux R _S de la souris : médiane														
Taux R _S de la souris : quantile 5%														
Taux R _S de la souris : quantile 95%														
Taux R _H de l'homme : médiane														
Taux R _H de l'homme : quantile 5%														
Taux R _H de l'homme : quantile 95%														

2. Compétences à acquérir lors de ce projet.

- **Théoriques** — comprendre les paramètres d'évolution moléculaire du génome :

- taux relatif d'évolution entre gènes ;
- hétérogénéité du taux de substitution entre sites ;
- composition en G+C en position 3 du codon.

- **Informatiques** — prendre en main un nouvel outil bioinformatique (ici, une base de données génomiques) ; automatiser la manipulation des données sous R grâce à des scripts commentés et expliqués ; gérer une centaine d'alignements de séquences ; naviguer dans les arbres en repérant et en extrayant les feuilles, les nœuds internes, et les longueurs de branches.

- **Analytiques** — calculer des taux d'évolution ; produire et comparer des statistiques descriptives (médiane, quantiles) ; générer les représentations graphiques des distributions de paramètres.

3. Cahier des charges.

- Rédaction d'un rapport de **5 pages maximum** au format **PDF** (Nom1_Nom2-HAU703I_OrthoMaM~2022-23.pdf).

- Format : marges minimales de 2,5 cm ; police 12 ; interligne 1,5.

- Structure : rédiger une **courte introduction** présentant le contexte général du projet, puis répondre aux questions **1.A.** jusqu'à **1.E.** du TP (cf. ci-dessus).

- Produire les codes **R**, commentés et expliqués (compris dans les 5 pages).

- Date-butoir pour le rendu : Jeudi **15 décembre 2022**, avant **18h00** (avec 1 point de pénalité par heure de retard).