# Données de spectrométrie

Le dossier “csv” contient les données de spectrométrie de masse.

Chaque ligne représente un pic. Pour chaque pic, il y a deux informations (une par colonne):

* le rapport m/z (masse / charge) du pic
* l’intensité du pic

Toutes les espèces sont des espèces actuelles. Leur provenance n’est pas connue.

s1, s2, s3… : réplicats.

XXX : espèce inconnue

# Tables de features

Les tables de feature ont été obtenues via MetaboAnalyst, avec les paramètres suivants :

* mass tolerance (mzwid) 0.019 : width of overlapping m/z slices to use for creating peak density chromatograms and grouping peaks across samples
* minfrac 0 : minimum fraction of samples necessary in at least one of the sample groups for it to be a valid group

Le [tableau suivant](https://drive.google.com/file/d/1n-8wio_rIPmZ6Y-YGWM6AevxkPP6_N4w/view?usp=sharing) contient les intensités brutes.

Dans le [tableau suivant](https://drive.google.com/file/d/10oTU-qi7ri2ebTo3E570VhQgZ3tS1pCo/view?usp=sharing), toutes les valeurs non nulles d’intensités ont été remplacées par 1.

Dans le [tableau suivant](https://drive.google.com/file/d/1Bz3mkmXUWKE96v7ZJKPJooZxoPC7CxPD/view?usp=sharing), chaque valeur d’intensité a été divisée par la valeur de l’intensité maximale de l’échantillon, de sorte que pour chaque échantillon, le pic majoritaire a toujours une valeur de 1.

Le [même tableau](https://docs.google.com/spreadsheets/d/1GkpkI0ZANdG9C95LKr02XplFk_Lfr1jYKB8ByTq16WM/edit?usp=sharing) avec une mise en forme conditionnelle sur les valeurs d’intensité permet de visualiser facilement les features communes ou au contraire différentes entre espèces.

# Phylogénie

Le jeu de données contient des données de spectrométrie de 29 espèces actuelles (voir fichier [espèces actuelles](https://docs.google.com/spreadsheets/d/1e4qObRFHSEUhJybHW6xrQaBCaurulapQ/edit?usp=sharing&ouid=114062525516244985806&rtpof=true&sd=true) pour le détail).

## COL1A1

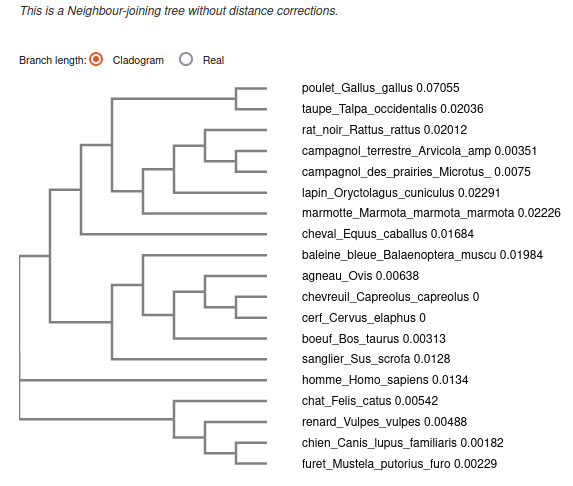
Parmi les 29 espèces, nous disposons de la séquence de [COL1A1](https://drive.google.com/file/d/1g8cob_dSu8rVlAmSIErHM_ODCOWRcKoh/view?usp=sharing) pour 21 espèces (ou d’une espèce proche ou représentative).

Nous ne possédons aucune séquence de poisson pour COL1A1.

Voir onglet “NCBI\_data” du fichier [espèces actuelles](https://docs.google.com/spreadsheets/d/1e4qObRFHSEUhJybHW6xrQaBCaurulapQ/edit?usp=sharing&ouid=114062525516244985806&rtpof=true&sd=true) pour le détail.

Un alignement multiple de ces séquences a été réalisé avec Muscle ainsi que l’arbre phylogénétique correspondant (algorithme Neighbour Joining), en supprimant les séquences de l’autruche et du pigeon, qui ne semblent pas de bonne qualité.

Voir [ici](https://drive.google.com/file/d/19P8W-DqPGBWBxztNYiQYPK7tOg5IPV_R/view?usp=sharing) l’alignement multiple entre les séquences disponibles de collagène.



## 

## COL1A2

Parmi les 29 espèces, nous disposons de la séquence de [COL1A2](https://drive.google.com/file/d/1QU9v3yq5lbBkMv6BtMTnzwiTO1TwM5g_/view?usp=sharing) pour 19 espèces (ou d’une espèce proche ou représentative).

A défaut de séquences chez le thon, nous avons inclus la séquence de Danio rerio pour les poissons.

Voir onglet “NCBI\_data” du fichier [espèces actuelles](https://docs.google.com/spreadsheets/d/1e4qObRFHSEUhJybHW6xrQaBCaurulapQ/edit?usp=sharing&ouid=114062525516244985806&rtpof=true&sd=true) pour le détail.

Un alignement multiple de ces séquences a été réalisé avec Muscle ainsi que l’arbre phylogénétique correspondant (algorithme Neighbour Joining).

Voir [ici](https://drive.google.com/file/d/11YBeR2452gAm4NzLT48VxZuEzhV7v_65/view?usp=sharing) l’alignement multiple entre les séquences disponibles de collagène.

