

## M1 Bionformatique / Biotechnologie

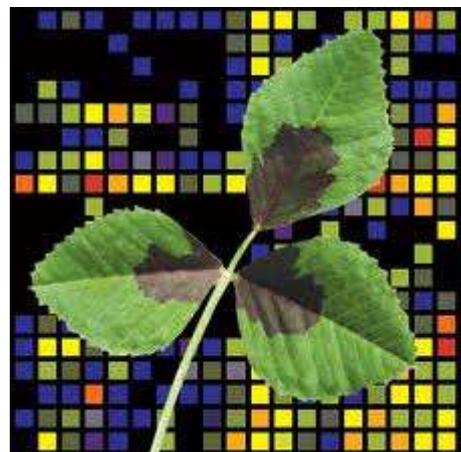
### TP Evolution Moléculaire : Détection moléculaire de l'adaptation

Bien que le concept Darwinien d'adaptation ait été formalisé il y a un siècle par les généticiens (Fisher, Wright, Haldane), il a été difficile de démontrer rigoureusement que des différences en acide aminés entre protéines homologues avaient une signification adaptative. Depuis plus de 30 ans, les généticiens des populations et les théoriciens de l'évolution ont développé des méthodes pour détecter la signature de la sélection naturelle à l'échelle des séquences d'ADN (codant mais aussi non codant). Grâce à l'avènement des technologies de séquençage à haut-débit, un grand nombre de données de séquences peuvent maintenant être explorées pour évaluer l'influence de la sélection naturelle dans l'évolution des populations et des espèces.

Ce TP vise à explorer 3 types d'approches permettant de détecter la signature de la sélection naturelle. Ces méthodes dépendent du type d'échantillonnage des séquences d'ADN (alignement intra/interspécifiques) et du type de séquence (codante, non codantes). Nous utiliserons 2 logiciels : DNAsp (Librado and Rozas, 2009) et Mega (Kumar and Tamura et al., 2011).

#### Partie 1

L'espèce *Medicago truncatula* est une plante légumineuse modèle très étudiée en laboratoire, notamment dans le domaine des interactions plantes-microorganismes. En effet, cette espèce est capable d'établir une symbiose avec des bactéries fixatrices d'azote (*Rhizobium*) et des champignons endomycorhiziens à arbuscules (*Glomus intraradices*). Cette espèce est aussi la cible de microorganismes pathogènes. (ex : *Aphanomyces euteiches*). ***Medicago truncatula* est donc un bon modèle pour rechercher des gènes impliqués dans l'adaptation de la plante à différents microorganismes.** Différentes accessions (lignées) de *Medicago truncatula* ont été séquencées par des méthodes haut-débit (séquençage 454). L'alignement des lectures (reads) sur le génome de référence (lignée A17) a permis de détecter des polymorphismes de séquences à l'échelle d'une base (Single Nucleotide Polymorphisms= SNPs).



Le travail consiste à effectuer un « genome scan » pour repérer les régions du génome soumises à des pressions de sélection. Nous travaillons ici sur une petite région du génome séquencée chez plusieurs accessions (individus) de l'espèce *Medicago truncatula*.

- Ouvrir le fichier « *Medicago\_SNPs\_sequences.fasta* » dans DNAsp
  - Quelle est la longueur de la séquence ? combien de sites polymorphes étudions-nous ?
- Effectuez une analyse du polymorphisme ADN (globale et par fenêtres glissantes. Utiliser plusieurs paramètres de fenêtres glissantes) :
  - Que signifient **S**, **Theta (per sequence) from S**, **k**, **Theta (per site) from S**, et **Nucleotide diversity  $\pi$  (Pi)**, et sur quoi vous renseignent-ils ?
  - Existe-t-il une variation de ces paramètres le long de cette région du génome ?
- Avec quel test mettriez-vous en évidence un potentiel effet de la sélection ?
- Effectuez le test (global, et par fenêtres glissantes), et interprétez.

D'après l'annotation du génome de référence, la région génomique que nous étudions est plutôt pauvre en gènes, mais elle comporte un petit cluster de 3 gènes connus pour être impliqués dans la reconnaissance d'effecteurs microbiens (gène de type NBS-LRR).

- Quel lien pouvez-vous faire entre la nature des gènes localisés dans cette région et vos résultats ? Quelle interprétation évolutive pouvez-vous donner ?

## Partie 2

La mouche du vinaigre, du genre *Drosophila*, présente un grand nombre d'espèces qui se sont en partie diversifiées en raison de leur régime alimentaire, ces mouches se nourrissant d'une grande variété de fruits sucrés. Les fruits sucrés contiennent des proportions variables de sucres, et donc d'alcool, au fur et à mesure de leur maturation, jusqu'à leur pourrissement. **L'enzyme Adh (alcool dehydrogenase)** participe à la dégradation de l'alcool qui est toxique pour l'organisme à certaines concentrations. Nous étudions la séquence codante de cette enzyme chez différents individus de 2 espèces de drosophiles : *Drosophila simulans* et *Drosophila yakuba*. Est-ce que cette enzyme a subi une évolution adaptative chez ces espèces ?

*Drosophila simulans* (vigne)

*Drosophila yakuba*



- Ouvrir le fichier « Drosophila\_Adh.phy » dans DNAsp. Le format Phylip est un format d'alignement de séquences que l'on peut générer par exemple avec un logiciel d'alignement tel que Clustal X, et peut être lu par différents logiciels d'analyse évolutive des polymorphismes. Sous DNAsp, assignez ces séquences comme une région codante.
- A quels types de mutations s'intéresse-t-on ? où se situent-elles sur un codon donné ?
- Quel test basé sur la comparaison du polymorphisme et de la divergence entre espèces pourriez-vous utiliser afin d'identifier une éventuelle évolution adaptative de l'Adh ?
- Que pouvez-vous conclure ?
- Les auteurs de cette étude (Mc Donald and Kreitman, 1991) ont montré que la séquence codante de l'Adh chez l'ancêtre de *Drosophila simulans* et *Drosophila yakuba* est très proche de la séquence consensus de *Drosophila simulans*. Chez laquelle de ces 2 espèces y-a-t-il eu évolution adaptative suite à la sélection positive ?

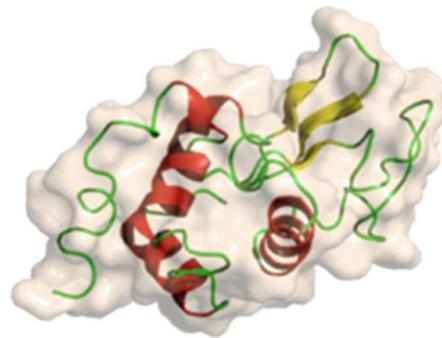
### Partie 3

Le lysozyme est une enzyme bactériolytique présente chez la plupart des mammifères. Chez les primates elle est excrétée dans différentes parties du tube digestif selon le régime alimentaire de l'espèce. Nous étudions une partie de la séquence codante de cette enzyme (les exons 1 et 2) chez **19 espèces représentant les grandes familles de primates (Singes du nouveau monde, Colobinae, Cercopithecinae et Hominoïdes)**. Nous cherchons à savoir si cette enzyme a subi une évolution adaptative, et chez quels primates (d'après l'étude de Messier and Stewart, Nature, 1997).

Gang de Primates



Lysosyme



- Ouvrir MEGA (v5), et convertissez le fichier « Primates\_lysozymes.fasta » en fichier MEGA (.meg). Puis ouvrez ce nouveau fichier (Open a file).
- Construisez un arbre phylogénétique de ces séquences avec la méthode de Neighbor-Joining avec la distance Kimura-2p. Combien de « familles » (ou groupes) observez-vous ?
- A quoi correspondent les longueurs de branches ?
- A quels types de mutations s'intéresse-t-on ?
- Quelle approche basée sur l'étude de la divergence entre espèces pourriez-vous utiliser afin d'identifier une éventuelle évolution adaptative du lysozyme ? Que concluez-vous ?

Supplémentaire si assez de temps :

- Pour détecter une adaptation à l'intérieur de chaque groupe (une espèce ou un groupe d'espèce): comparer le  $d_N - d_S$  moyen au sein des 4 groupes
- Pour détecter une adaptation d'un groupe par rapport à un autre (chez l'ancêtre commun de l'un ou l'autre des groupes) : comparer le  $d_N - d_S$  moyen pour des groupes composites tels que « colo/cerco », « colo/hom ».

Pour cela vous devez créer de nouveaux groupes dans l'ongle « Data => Select Taxa and Groups »)

- En quoi l'arbre phylogénétique peut vous donner une indication sur le groupe de primates dans lequel le lysosyme a subi une évolution adaptative ?
- Les colobinae (« leaf-eating monkeys ») ont un estomac antérieur complexe dans lequel les bactéries fermentent le matériel végétal, suivi d'un vrai estomac qui exprime des niveaux importants de lysosyme. Quelle interprétation biologique pouvez-vous faire de vos résultats ?

Liste des espèces

**Hom : Hsa, Ggo, Ppy, Ptr, Ppa**

(*Homo sapiens*, *Gorilla gorilla*, *Pongo pygmaeus*, *Pan troglodytes*, *Pan paniscus*)

**New: Ssc, Soe**

(*Saimiri sciurus*, *Saguinus oedipus*)

**Colo: Pen, Tve, Tob, Tfr, Can, Cgu**

(*Semnopithecus entellus*, *Semnopithecus vetulus*, *Trachypithecus obscurus*, *Trachypithecus francoisi*, *Colobus angolensis*, *Colobus guereza*)

Cerco: Cae, Cto, Epa, Ani, Pha, Mmu

(*Cercopithecus aethiops*, *Cercocebus torquatus*, *Erythrocebus patas*, *Allenopithecus nigroviridis*, *Papio hamadryas*, *Macaca mulatta*)