

Étude de jeux de séquences de populations naturelles de drosophiles (*Drosophila simulans*).

Directives :

À rédiger en Times taille 12 et interligne 1.5. Merci.

Suivre la structure d'un article scientifique : introduction (1 page et demi à deux pages), matériel et méthodes (décrire le type de données, les outils de mesure de la variabilité du polymorphisme moléculaire : estimateurs et tests = formules et description), résultats et discussion-conclusion.

À remettre par courriel le vendredi 01 mars 2024. (points retirés si retard : 5%/jour).

Soignez la grammaire et l'orthographe, 5% des points y sont alloués.

Barème :

expression = 5%

introduction + matériel & méthodes = 15%

c) Estimateurs : 15%

d) *Fst* : 10%

e-g) Tests : 25%

Discussion-conclusion : 30%

Bonne chance !

a) Pour chaque jeu de séquences de *Drosophila simulans* : les ouvrir séparément (*Nrg*, *sxl* et *vermillion*) dans le logiciel *DNAAsp* (Rozas *et al.* 2017).

b) i- Partitionner en quatre populations (Madagascar, Mayotte, La Réunion et Kenya) : dans menu « *data* », sélectionner « *define sequence sets* ».

ii- Dans menu « *analysis* », sélectionner « *DNA polymorphism* » et « *Fu & Li test* » : pour chaque population, noter les différentes informations sur ces jeux de séquences (*n*, longueur séquence, *k*, *S*, *Hd*, η_s , π ...).

Interpréter les valeurs des différents estimateurs pour chacune des pops, puis entre elles. Traduisent-ils des évènements sélectifs, démographiques ?

- c) Calculer les F_{ST} entre chaque paire de populations et pour les trois gènes : dans menu « *analysis* », sélectionner « *gene flow, genetic differentiation* ».

Interpréter les valeurs des F_{ST} obtenues (flux génique).

- d) Tests D_T de Tajima et $D_{Fu \& Li}$ sans recombinaison.

Interpréter les valeurs des différents tests aux trois gènes pour chacune des populations, puis entre elles. Traduisent-ils des évènements de sélection, démographiques ?

- e) Tests par simulation de coalescent, D_T de Tajima, et $D_{Fu \& Li}$ Hd et h avec et sans recombinaison

($R = 0 ; 6$ et 60). 10 000 itérations pour $R = 0$ et 1 000 itérations pour $R = 6$ et 60 .

Interpréter les valeurs obtenues à partir des différents taux de recombinaison pour chacune des pops, puis entre elles.

Sachant que le taux de recombinaison moyen chez la drosophile est de 10 par gène et par génération pour une taille efficace (Ne) de 1 000 000 individus, quels sont les taux de recombinaison conservatifs pour chaque type de tests ? Dans quels cas ?

- f) Pour les populations de **Madagascar** et de **La Réunion**, construire les trois combinaisons de deux gènes pour le test *HKA* pour les lignées « SR » seulement (*Nrg-sxl*, *Nrg-vermillion*, *sxl-vermillion*). Note, pour les gènes *sxl* et *vermillion*, prendre un nombre de séquences (*SR* et *non-SR*) équivalent à celui des séquences *SR* de *nrg*.

Utilisez les jeux de données avec la séquence correspondante de *Drosophila melanogaster* pour les 3 gènes *Nrg*, *Sxl* et *Vermilion* : c'est le groupe externe pour calculer le *HKA*. Ils portent le nom *GENE-wOUTGROUP.fas*.

Ensuite, il faut construire 3 combinaisons de séquences en concaténant les séquences de 2 fichiers ensemble : *Nrg + Vermilion* ; *Nrg + Sxl* et *Vermilion + Sxl*.

Ensuite, partitionner les jeux de séquences concaténées en 2 populations + séquence du groupe externe (*D. Melanogaster*) : *Madagascar + D. melanogaster* et *La Réunion + D. melanogaster*.

Effectuer les *HKA* pour les deux populations (Madagascar et La Réunion) et les 3 combinaisons de séquences : *Nrg + Vermilion* ; *Nrg + Sxl* et *Vermilion + Sxl*. (= 6 tests).

Interpréter les valeurs de chaque paire de gènes pour chacune des pops, puis entre elles. Sélection ? Démographie ?

Faire un scénario général synthétisant toutes les informations que vous avez obtenues à partir de ce jeu de données.