Bonsoir Julian.

Le 12 janvier 2021

Ton texte est remarquablement bien écrit. Voici quelques notes un peu plus longues que ce qui peut entrer confortablement dans la marge du manuscrit.

1. Figure 1 – La Fig. 1A n’est pas claire.

• Pourquoi y a-t-il deux rectangles bleus l’un au dessus de l’autre, ainsi que 2 rectangles verts, dans les matrices de données “1st sampling” et “2nd sampling” ? Les deux rectangles bleus représentent-ils les deux loci avec 2 allèles chacun?

Les 2 loci sont-ils les deux rectangles superposés ou les deux valeurs côte à côte dans chaque rectangle?

• Expliquer les valeurs qu’on peut lire dans ces matrices: a/a, c/a, etc. Je n’ai aucune idée à quoi ces lettres (valeurs) réfèrent.

• Je suggère de présenter les données à plat, comme dans les matrices de données soumises à l’analyse par dist.genpop. Donc, une ligne par site. Exemple : la matrice microsatt$tab (18 x 112) dans les exemples de la fonciton dist.genpop().

Je suggère également de présenter dans le petit exemple un locus à 2 allèles et un autre possédant 3 ou 4 allèles afin de dissiper l’impression que la méthode ne serait applicable qu’à des loci possédant 2 allèles seulement.

2. Fonction TGI.R – Je suggère d’être plus explicite dans les lignes de documentation au-dessus de la fonction :

# method : a number between 1 and 5. Five genetic distances are available in function dist.genpop of the adegenet package. They are : (1) Nei’s D, (2) Edwards’ angular D, (3) Reynolds’ coancestry coefficient, (4) Rogers’ D, (5) Prevosti’s absolute genetic D. Methods 2, 3 and 4 produce Euclidean distances, whereas methods 1 and 5 produce non-Euclidean distances, which produce negative eigenvalues and complex eigenvectors in principal coordinate analysis.

Indiquer aussi dans la section *Software* de l’article que la fonction TGI donne directement accès à ces cinq distances via l’argument "method" pour le calcul des indices TGI.

3. Remarque sur la distance de Rogers. – Celle-ci est en fait la distance entre profiles d’abondances relatives (D18 dans *Numerical ecology* (2018) calculée pour chaque locus. Les valeurs des différents loci sont ensuite sommées et la somme est divisée par le nombre de loci.

Il manque certaines propriétés intéressantes à la distance entre profils d’abondances relatives (et donc aussi à la distance de Rogers), soit les propriétés P4, P6 et P7 de Legendre & De Caceres (2013) ; voir le tableau 2 de cette publication, p. 957. Voici les propriétés que cette distance ne possède pas :

P5 : Sites without species in common have the largest dissimilarity. – Cela n’est pas vrai pour la distance entre les profils d’abondances relatives, donc pas vrai pour la distance de Rogers.

P6 – Dissimilarity does not decrease in series of nested species assemblages. Idem.

P7 – Species replication invariance. Idem.

Je suggère de conserver cette idée pour un prochain article. Ne mélangeons pas les messages dans cet article-ci. Il faudra cependant alerter, un jour, les généticiens des populations quant aux propriétés des différentes distances qu’ils utilisent.

Un tel article pourrait fournir une occasion pour suggérer à Thibaut Jonbart d’ajouter la distance de corde de Cavalli-Sforza & Edwards (1967) à la fonction dist.genpop. [Si mes souvenirs sont bons, la distance de corde des généticiens est la distance de Hellinger des écologistes, c’est-à-dire la distance de corde calculée à partir des racines carrées des fréquences.]

4. Exemple des gobies (Kinziger et al., 2015) – Je suggère de décrire ces résultats un peu plus longuement. Bien des lecteurs liront d’abord cette section en détail afin de comprendre le type d’information nouvelle qu’ils peuvent obtenir de la méthode TGI.

Tu pourrais par exemple :

• Ajouter une carte montrant les populations significatives. – Je vois que tu avais prévu de présenter cette carte dans le fichier « Supp Mat ». Je la placerais plutôt dans l’article principal si possible.

• Dans une annexe, montrer les résultats de ta fonction R pour les données des gobies.

Cet exemple m’amène à la réflexion suivante : pour les données de communautés, la méthode TBI permet de quantifier les pertes et les gains d’espèces (pour les données de présence-absence) ou les pertes et les gains d’individus dans les différentes espèces (pour les données de fréquences). Avec TBI, peux-tu extraire des statistiques de ce type ? Pour les données des gobies, sur quoi te bases-tu pour affirmer (ligne 459) : « … there was a loss of genetic diversity in the Elk River population » ? La diversité génétique est-elle calculée dans ta fonction TGI ? Faudrait-il ajouter quelques lignes de code pour fournir cette information à la sortie de la fonction, ou encore montrer comment l’utilisateur peut faire appel à une autre fonction R pour calculer l’un ou l’autre des indices de diversité génétique ?

Pierre

========