

Définition d'un pipeline automatisé pour la détection des virus endogènes dans les génomes eucaryotes

L'analyse des génomes eucaryotes a révélé la présence de nombreux gènes d'origine virale. Par exemple, on retrouve de l'ordre de 6% de retro-éléments dans le génome humain. Leur présence était attendue, étant donné que le cycle naturel de répllication des retro-virus passe par une étape d'insertion dans les chromosomes de l'hôte. Au delà de ces virus, on s'est rendu compte que tout type de virus pouvait occasionnellement se retrouver intégré dans le matériel génétique des eucaryotes (=endogénisé), même si le cycle de répllication des virus n'implique pas de phase d'intégration. Ainsi, on retrouve des traces de virus à ADN, ou à ARN non-retroviraux dans la plupart des génomes eucaryotes [1]. De plus, certains de ces événements d'endogénisation, ont pu contribuer à des innovations génétiques majeures, comme l'évolution du placenta chez les mammifères [2] ou la domestication de machineries virales par les insectes parasitoïdes [3]. Ces dernières années ont également suggéré que ces virus endogènes pouvaient être impliqué dans l'immunité contre les virus dits "cognate" (apparentés et libres) [4].

Néanmoins, actuellement, il n'existe pas de pipeline automatisé et efficace susceptible de détecter de manière systématique l'ensemble des gènes viraux présents dans les génomes (quel que soit la structure génomique des virus donneurs, i.e. ADN, ARNretroviral, ARNnon-retroviral). Une des difficultés tient en effet au fait que de nombreux virus, en particulier ceux présentant de gros génomes (donc à ADN), codent eux-même des gènes acquis de leurs hôtes. Une analyse de la proximité phylogénétique des candidats avec la diversité eucaryote et virale est donc requise. L'objectif du stage sera donc de définir un pipeline automatisé prenant en entrée un fichier fasta d'un génome eucaryote et des bases de données adaptées et retournant une liste de gènes putativement d'origine virale, associée à une analyse phylogénétique de ces derniers. Le pipeline devra être réalisé en utilisant le système de gestion Snakemake.

Références

[1] Feschotte, C., Gilbert, C., 2012. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 13, 283–296. <https://doi.org/10.1038/nrg3199>

[2] Dupressoir, A., Lavalie, C., Heidmann, T., 2012. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. *Placenta* 33, 663–671. <https://doi.org/10.1016/j.placenta.2012.05.005>

[3] Guinet, B., Lepetit, D., Charlat, S., Buhl, P.N., Notton, D.G., Cruaud, A., Rasplus, J.-Y., Stigenberg, J., De Vienne, D.M., Boussau, B., Varaldi, J., 2023. Endoparasitoid lifestyle promotes endogenization and domestication of dsDNA viruses. *eLife* 12, e85993. <https://doi.org/10.7554/eLife.85993>

[4] Suzuki, Y., Baidaliuk, A., Miesen, P., Frangeul, L., Crist, A.B., Merklings, S.H., Fontaine, A., Lequime, S., Moltini-Conclois, I., Blanc, H., van Rij, R.P., Lambrechts, L., Saleh, M.-C., 2020. Non-retroviral Endogenous Viral Element Limits Cognate Virus Replication in *Aedes aegypti* Ovaries. *Current Biology* 30, 3495–3506.e6. <https://doi.org/10.1016/j.cub.2020.06.057>