

LES PANGÉNOMES, LEUR MISE À JOUR ET EXPLOITATION EN AGRICULTURE

La séance a été organisée par Mylène **DURAND-TARDIF** (section 6)
et Claire **ROGEL-GAILLARD** (section 3).

Après le génome, le pangénome

Le génome, c'est une information à l'échelle d'un individu, d'une variété, d'une lignée ou d'une souche. Le pangénome c'est une information à l'échelle d'une espèce ou d'un groupe d'individus. Il s'agit d'un changement de paradigme en génomique, transverse aux domaines du vivant, que nous avons proposé d'explorer pendant cette séance qui leur est dédiée.

Le séquençage des génomes a été un processus initialement très long qui a abouti, par espèce, à un génome dit de référence, estimé lors de sa réalisation suffisant pour représenter une espèce donnée. Pour mémoire, le séquençage du génome de référence pour l'espèce humaine a été considéré comme une prouesse scientifique et technologique : le projet a démarré officiellement en 1989, les premières versions ont été publiées en 2001, et la première séquence considérée comme totalement complète d'un génome humain n'est disponible que depuis mars 2022.

Jusqu'à récemment le reséquençage des génomes avait comme objectifs principaux d'inventorier le répertoire de gènes d'une espèce et de repérer les variations par rapport à la séquence du génome de référence, ciblant principalement les polymorphismes nucléotidiques, ou SNP (*single nucleotide polymorphism*). Ces travaux ont permis, entre autres, d'élaborer des puces de génotypage des SNP à l'échelle des génomes entiers. Parmi les applications remarquables de ces outils : les études d'associations (GWAS pour *genome wide association study*), la mise en place et le déploiement à grande échelle de la sélection génomique.

Or, grâce aux nouvelles performances des techniques de séquençage actuelles, le reséquençage des génomes met en évidence des variations structurales très conséquentes au sein d'une espèce, allant jusqu'à la présence ou l'absence de gènes complets, démontrant ainsi que le génome est divers et complexe entre individus, variétés, lignées ou souches. Ces découvertes ont conduit à faire émerger la notion de pangénome d'une espèce, avec l'existence d'une portion de ce pangénome qui est commune à tous les individus d'une espèce donnée, et d'une portion qui est distribuée entre les individus.

Cette découverte qui était un concept il y a une dizaine d'années est maintenant source de nombreuses interrogations et de promesses : comment ont évolué ces génomes ? Quel a été l'impact de l'adaptation au milieu, de la domestication, de la sélection ? Quels réservoirs de gènes sont à notre disposition pour améliorer les performances des animaux d'élevage et des espèces cultivées ? Les pangénomes, avec leur portion distribuée au sein d'une même

espèce, représentent une variabilité largement méconnue et constituent une véritable ressource de matériel génétique à disposition. Leur exploitation en agriculture est prometteuse. Cette séance n'a pas l'ambition de couvrir la métagénomique, autre front de science qui tend à caractériser les environnements par la multitude de séquences nucléotidiques qu'ils abritent en identifiant la diversité des espèces et fonctions présentes. La métagénomique s'adresse particulièrement aux microorganismes, y compris les microorganismes eucaryotes.

Le pangénome étant un concept applicable à toutes les espèces, il a été proposé pour cette séance une introduction générale sur les pangénomes par Johann Joets (INRAE), suivie de trois interventions sur les pangénomes des microbes (David Vallenet, CEA), des plantes (François Sabot, IRD) et des animaux (Didier Boichard, INRAE).

Des concepts nouveaux qui se consolident et soulèvent de nombreuses questions

Les intervenants ont rappelé la définition de deux fractions distinctes dans les pangénomes : la fraction commune à tous les individus d'une espèce et celle distribuée entre les individus. Il semble y avoir consensus sur le terme à utiliser pour la fraction commune : *core genome* (génom « cœur ») pour le « génome de base ». En revanche, le vocabulaire ne semble pas totalement convergent pour le génome distribué dit aussi variable, qui est référencé en tant que *dispensable*, *accessory* ou *flexible*. Chez les plantes et chez les microorganismes, il est proposé de répartir la partie variable en deux fractions distinctes (*shell* et *cloud*), conduisant ainsi à trois sous-parties dans un pangénome : le *core* (cœur) ou *persistent* (persistant), le *shell* (coquille) qui regroupe les gènes de fréquence intermédiaire, le *cloud* (nuage) pour les gènes de fréquence faible, avec l'existence d'artefacts dans l'assemblage des séquences, qu'il convient de ne pas minimiser. Chez les bactéries, les gènes variables sont principalement issus de transferts horizontaux qui, pour le *shell*, seraient maintenus dans la population car conférant des propriétés d'adaptation environnementale alors que, pour le *cloud*, seraient issus de transferts plus récents ou non sélectionnés.

Ces termes vont se stabiliser progressivement et l'éclairage apporté par les présentations illustre un historique plus récent sur ces questions pour les animaux que pour les microbes et les plantes.

Les exemples présentés sont éloquentes quant à la diversité des pangénomes selon les domaines du vivant et la taille du génome variable peut représenter plus du double du génome cœur, notamment chez les plantes avec des exemples chez le maïs, le riz, le sorgho. Chez le blé tendre, l'étude de 18 cultivars révèle que 12500 gènes sont absents du génome de référence. De manière tout à fait intéressante, la taille du pangénome diminue au cours du processus de domestication, entraînant aussi une réduction du ratio entre génomes cœur et variable. Chez les bovins, avec les données fraîchement disponibles, le génome variable semble ne pas dépasser quelques dizaines de mégabases pour un génome de 3000 mégabases (1 à 2%). Chez les bactéries, le rapport de taille entre génome cœur et génome variable pourrait amener à reconsidérer la définition d'espèce chez ces organismes.

En toile de fond se pose la question des gènes (et néogènes) présents dans chaque portion des pangénomes, de leurs origines et de leurs flux, de leurs rôles en termes d'adaptation et d'évolution. Le génome variable serait-il un tampon de gènes issus du génome cœur ou en devenir dans le génome cœur ? Est-il important de reconsidérer les approches en génomique comparative ? A ces questions s'ajoutent la finitude d'un pangénome d'une espèce (génome ouvert ou fermé) et la collection d'individus (lignées, variétés, souches, etc.) à séquencer, nécessaires pour le caractériser de la manière la plus exhaustive possible.

Des défis méthodologiques

Le séquençage de l'ADN nous a conduit à une représentation linéaire de la molécule, avec la possibilité de visualiser simplement les différences par alignement de séquences. La caractérisation des pangénomes nécessite de visualiser les variations de structures. Cette complexité conduit à passer d'une représentation linéaire à une représentation par graphes. Ces nouvelles visualisations nécessitent le développement d'outils bioinformatiques appropriés, et leur élaboration pour optimiser la lisibilité des graphes est un défi actuellement.

Les études pangénomiques mettent en évidence que la variabilité allélique issue du génotypage des SNP ne sera pas suffisante pour les études de génétique à venir. Ainsi, outre les évolutions de représentation mentionnées, se posera à terme la question de l'exploitation à grande échelle des informations du pangénome.

Vers une caractérisation renouvelée des ressources génétiques

Les ressources génétiques sont considérées comme un réservoir de gènes pour le futur. Des efforts considérables ont été consentis depuis des dizaines d'années pour collecter et préserver ces ressources génétiques aussi bien microbiennes, végétales qu'animales. Leur caractérisation génomique et fonctionnelle reste encore limitée, beaucoup de ressources étant conservées sans beaucoup d'informations associées.

Les travaux sur les pangénomes vont conduire à renouveler les approches à mettre en place pour la caractérisation génomique des ressources génétiques, afin d'explorer aussi bien le génome cœur que le génome variable. En effet, le génome variable contient de nombreux gènes d'intérêt, avec comme exemple emblématique le gène *SUB1A* chez le riz, responsable de la résistance à la submersion. Chez les plantes, il comprend également de très nombreux éléments transposables dont certains peuvent modifier les réseaux de régulation des gènes en s'insérant dans des locus spécifiques. Chez les mammifères, le génome variable actuellement exploré semble contenir des gènes de l'olfaction et de l'immunité, importants pour améliorer des caractères de santé favorables à la durabilité des élevages. Chez les microorganismes, connaître les briques de base indispensables au fonctionnement de la cellule apportera des connaissances pour la biologie synthétique et explorer la diversité contenue dans le génome variable offrira des opportunités de découverte de nouvelles fonctions et molécules d'intérêt thérapeutique ou biotechnologique.

Ces caractérisations vont s'appuyer sur du séquençage à grande échelle, avec l'ambition de rechercher des polymorphismes au-delà des variations alléliques nucléotidiques. Elles devraient apporter des informations expertes pour estimer la rareté et le potentiel des accessions conservées.

Pour conclure : de l'importance de l'accessoire

Les travaux sur les pangénomes sont attendus pour améliorer nos connaissances sur l'évolution des génomes et leurs processus adaptatifs. Ils s'appuient sur des développements méthodologiques de pointe (nouvelles technologies de séquençage, bioinformatique) et donnent accès à un réservoir de variabilité encore inexplorée. Ils renouvellent aussi le rôle majeur des consortiums scientifiques compétents pour générer, interpréter et partager les données.

La recherche sur les pangénomes est un véritable front de science en génomique. Il prend son essor, notamment grâce aux nouvelles capacités de séquençage avec les lectures

longues. Cette séance a eu l'ambition de faire un état des lieux des connaissances sur le sujet chez les microorganismes, les plantes et les animaux.

Les interventions de très grande qualité ont permis de faire un point sur les concepts, les connaissances et les verrous, ouvrant des discussions riches à l'adresse de la biologie fondamentale, des biotechnologies et des enjeux liés à l'adaptation de l'agriculture, la durabilité et l'agroécologie dans un contexte de changement climatique.