

Animation scientifique bioinformatique : « Pangénomique végétale »

Lundi 1^{er} juillet 2019, 147 rue de l'université, Paris VII.

Introduction

Avec l'appui des deux animateurs Johann Joets (INRA) et Jeroen Wilmer (Biogemma), le GIS BV a organisé la première concertation en France entre chercheurs issus du secteur public et du secteur privé. Les objectifs de la journée étaient d'identifier les liens entre recherche publique et recherche privée, leurs forces et leurs lacunes. Des priorités de recherche communes aux deux secteurs ont été dégagées. 71 participants ont été accueillis, dont 20 issus du secteur privé.

Déroulé

Sept orateurs ont effectué les présentations listées ci-dessous :

- **J. Joets** – (GQE Gif-sur-Yvette) Les pangénomes de plantes
- **C. Donnadieu** – (Plateforme GeT-PlaGe, Toulouse) Développements et innovations dans le domaine du séquençage pour le projet SeqOcln
- **D. Vallenet** – (Génomique métabolique du Génoscope, Evry) Un graphe de pangénome partitionné pour capturer la diversité microbienne
- **E. Rivals** – (LIRMM, Montpellier) Enjeux computationnels de la pangénomique
- **S. Nicolas** - (GQE Gif-sur-Yvette) Contribution des insertions/délétions à la variation des caractères, l'adaptation et aux performances hybrides au moyen d'une puce de génotypage à haut débit
- **F. Sabot** – (DIADE, Montpellier) Pangénomes chez le riz et impact de la domestication
- **F. Choulet** – (GDEC, Clermont-Ferrand) Variations structurales et dynamique des éléments transposables du génome du blé

M. Conte (Syngenta), **H. Le** (Vilmorin & Cie), **C. Donnadieu** (GeT-PlaGe) & **E. Paux** (GDEC) ont ensuite participé à une table ronde et à des discussions avec la salle sur la thématique : De la recherche à l'application : demandes et contraintes.

Synthèse globale des discussions

La pangénomique est une thématique de recherche très récente puisque les premières publications marquantes datent des années 2000. Dans son acceptation large la pangénomique consiste en l'identification de blocs d'haplotypes présents dans un ou quelques génotypes seulement, par opposition au *core genome* ou génome de base, présent dans tous les génotypes de l'espèce. Chez les microorganismes, le nombre de souches séquencées autorise la distinction entre le *core genome*, le *shell genome* ou génome distribué et le *cloud genome* ou génome rare. Pangénomique et pantranscriptomique offrent des potentialités vastes pour l'amélioration des plantes puisque les variations de type présence/Absence (PAV) ou de nombre de copies (CNV) seraient impliquées dans l'hétérosis, l'adaptation à l'environnement, y compris à divers stress biotiques et abiotiques, la domestication et autres caractères d'importance agronomique. La recherche de liens de causalité entre génomes distribués et phénotypes est un objectif affiché en amélioration des plantes. Les variations de structure peuvent aussi modifier le taux de recombinaison le long des chromosomes.

Les participants sont engagés dans des études diverses du pangénome et développent des outils spécifiques. À ce stade une concertation est nécessaire pour optimiser les efforts de chacun.



Principales pistes de recherche identifiées entre secteurs public & privé

- Le séquençage ne semble plus être limitant, même si des efforts doivent encore être fait pour maîtriser les coûts tout en obtenant la qualité nécessaire pour 1/ découvrir de manière fiable les variations structurales et 2/ être capable de les localiser. Une concertation entre acteurs publics et privés est nécessaire pour définir quels et combien de génotypes seraient nécessaires à l'exploration des variations en PAV et CNV au sein d'une espèce. En général les pangénomes de plante sont fermés, c'est-à-dire qu'ils sont composés d'un nombre fini de gènes. Mais le nombre d'individus à séquencer pour bien couvrir le pangénome peut aller jusqu'à plusieurs milliers si l'espèce est très variable.
- Le matériel stratégique des semenciers ne peut pas être intégré à cet effort collectif, mais un panel de lignées fondatrices des programmes de sélection serait un matériel très informatif. Ce matériel doit être assorti de variétés ancestrales, *landraces* et/ou de ressources génétiques publiques qui se montrent souvent plus riches en gènes que le matériel élite utilisé en sélection.
- Aujourd'hui, plusieurs variétés des espèces de rente majeures sont séquencées pour représenter le pangénome. Il faut généraliser cette démarche et viser l'établissement d'un pangénome de référence, qui puisse être partagé et utilisé par tous et non plus, une variété de référence qui *de facto* est lacunaire.
- L'intégration des données acquises en masse et leurs analyses doivent progresser conjointement. De même, il faut travailler des solutions pour compresser les données tout en conservant des possibilités de requêtes.
- Des solutions doivent être mises en place pour l'intégration des PAV et des CNV dans les analyses d'associations génotype-phénotype de type GWAS. Le génotypage des haplotypes tenant compte des variations de structure est une voie possible. Il n'existe pas de solution à haut débit et peu coûteuse pour génotyper les variations structurales, en dehors de l'utilisation du déséquilibre de liaison lorsque cela est possible. Des approches sont en développement (puces, cartographie de *reads* sur graphe de variations...).
- Les approches de représentation et d'analyse des pangénomes par graphe de variation semblent prometteuses. Elles permettent d'analyser simultanément l'ensemble de la variabilité génétique connue, mais sont pour l'instant difficilement applicables à de grands génomes et pour de nombreux individus. Les algorithmes doivent donc évoluer et dans cet objectif, des compétences informatiques sont nécessaires.
- Il est aussi nécessaire de définir les besoins en matière de visualisation des données et d'adapter les outils existants à la complexité des pangénomes. Une piste pourrait être le réemploi des outils de génomique comparative.
- Il faut prendre garde à ne pas se limiter aux variations structurales qui incluent des séquences codantes : les régions régulatrices sont aussi potentiellement déterminantes pour l'expression des phénotypes, de même que les épivariations.