Projet :

Etude des propriétés de liaison à l’ADN de 529 facteurs de transcription chez *Arabidopsis thaliana*

Puisque chaque cellule d'un même organisme possède le même ADN, la grande variabilité de types cellulaires dans cet organisme nécessite une expression différentielle des gènes pour chaque type cellulaire.

~~La plupart des processus biologiques comme les réponses à l’environnement ou les processus de développement impliquent des régulations de l’expression des gènes.~~ Les facteurs de transcription (TF) sont une classe importante de protéines impliquées dans la régulation des gènes. Ces facteurs se fixent sur des régions spécifiques du génome et recrutent des co-répresseurs ou des co-activateurs pour réguler l’expression des gènes à proximité. L’ADN d’un génome contient toutes les informations pour recruter les facteurs adéquats au bon moment.Tous les processus de régulation sont très robustes, ce qui indique qu’il existe des règles, en particulier au niveau de la fixation des TF à l’ADN. Cependant, la découverte de ces règles reste un enjeu majeur en biologie.EPar conséquent, il est très difficile de mettre en relation les chaque gène et les gènes qu'il régule en se basant sur des séquences génomiques. Cette mise en relation, c'est la construction d'un réseau de facteur de transcription.

L'émergence des méthodes de séquençage haut débit a permis de faire des avancées considérables dans le domaine de la régulation en fournissant des listes exhaustives de régions liées par un facteur donné ou des listes de gènes régulés en réponse à une stimulation ou lors d’un processus développemental.

Différentes techniques donnent accès aux régions liées par un TF donné. Les expériences de ChIP-Seq (Immunoprécipitation de chromatine) permettent l'identification de ces séquences in vivo . Cependant, dans le contexte d'une cellule, de nombreux paramètres viennent moduler la fixation des facteurs. Une autre technique in vitro consister en une expérience de ChIP-Seq mais utilisant de l’ADN génomique nu, c'est le DAPSeq.

Ces données peuvent alors être exploitées pour tenter de comprendre les règles qui régissent la fixation d’un facteur à l’ADN : quelles sont les séquences reconnues (TFBS), comment ces séquences se positionnent-elles dans une région liée ?

On peut alors construire des modèles prédictifs de liaison (appelé PWM) qui permettront ensuite de scanner n’importe quelle séquence d’ADN pour prédire l’affinité de liaison d’un facteur avec cette séquence. Confronter les modèles prédictifs à la réalité permet ainsi de comprendre mieux les paramètres impliqués dans la liaison : on peut palier les différences observées par introduction de ces paramètres dans le modèle.

Mon projet s’étendra selon 2 axes.

Axe 1 : Alimenter la base Jaspar avec des modèles issus du DAP-Seq. Il existe 529 facteurs d’*Arabidopsis thaliana* pour lesquels les expériences de DAP-Seq ont été publiées. Ces modèles seront validés et insérés dans la base Jaspar la base de données de référence qui recense les PWM de centaines de facteurs de transcription dans de nombreux organismes. L'accès gratuit, les normes de qualités exigées pour la construction des PFM et l'absence de doublon pour un même facteur de transcription font de JASPAR un outil unique et incontournable pour l'étude des séquences cis-régulés. La base est actualisée tous les deux ans et un grand nombre y est à chaque fois incorporé. Mon équipe au laboratoire LPCV a contribué aux deux dernières mises à jour.

Axe 2 : Etudier la disposition des TFBS dans les régions liées. Les TF ne se lient pas souvent à un site unique dans une région. Certains ont des préférences en terme de disposition (orientation, inter-distan ces). Ma thèse porte sur les facteurs de transcription LEAFY et ARF qui contrôlent le développement des fleurs. J’ai déjà réalisé de nombreuses analyses sur ces deux facteurs et il serait intéressant de les généraliser à l’ensemble des données disponibles.

Intérêt de l’échange :

Le travail proposé est d'une grande ampleur. Étant étudiant en thèse de bioinformatique dans l'équipe Régulation du Développement Floral au Laboratoire de Physiologie Cellulaire et Végétale (LPCV), les question relatives aux sites de liaison me concernent directement. Le LPCV ne dispose pas réellement d'expertise bioinformatique sur des traitements de grands nombres de jeux de données. L’équipe d’accueil (Mathelier à Oslo) est une référence dans le domaine des TFBS. Anthony Mathelier a été le principal acteur dans la complétion de la base JASPAR en 2014 et en 2016. Il est spécialiste de modèles innovants qui permettent en particulier d’intégrer dans les modèles les paramètres de forme de l’ADN. Ce séjour serait une occasion unique de me former à ces modèles pour pouvoir les implémenter ensuite dans mon laboratoire grenoblois.