IDENTIFICATION SANS *A PRIORI* DE NOUVEAUX VIRUS ARN PHYTOPATHOGÈNES : APPLICATION A L'ÉTUDE DE L'ÉTIOLOGIE DE MALADIES DES PLANTES ET À CELLE DU MÉTAGÉNOME VIRAL

par Thierry **Candresse**[1]

Analysés collectivement les virus représentent le groupe d'agents biologiques présentant la plus grande diversité, les virus phytopathogènes ne faisant pas exception à cette règle. Encodant leur génome sur différents types d'acides nucléiques (ADN ou ARN, simple ou double brin, circulaire ou linéaire…), ayant des organisations génétiques ou des stratégies de réplication différentes, il n'existe de fait aucune séquence nucléotidique ou protéique qui soit commune à tous les phytovirus. A cette diversité « globale » correspond également une très grande diversité aux niveaux taxonomiques moins élevés des taux de divergence nucléotidique intra-espèce allant jusqu'à 30% étant tolérés dans certains genres. Cette diversité énorme et l'absence de marqueur généraliste qu'elle entraîne ont jusqu'à présent fortement limité notre capacité à identifier ou à caractériser sans *a priori* les phytovirus présents dans une plante hôte ou, plus largement, dans les populations végétales.

Le développement des nouvelles technologies de séquençage à très haut débit (NGS : *next generation sequencing*) modifie fortement cette situation. L'augmentation énorme des débits de séquences et l'abaissement des coûts correspondants permettent en effet aujourd'hui de séquencer de façon très large les acides nucléiques présents dans une plante. Il est alors possible de rechercher parmi ceux-ci d'éventuelles séquences d'origine virale, que ce soit par des recherches d'homologie avec des virus connus ou par une analyse différentielle avec les séquences provenant d'un hôte de statut sanitaire connu. De telles démarches sont conceptuellement à positionner dans le grand domaine de la métagénomique, qui connaît actuellement un développement très important.

Les phytovirus n'ayant très généralement pas de phase libre facilement accessible hors de leur hôte ou de leur vecteur, diverses stratégies correspondant à différents types d'acides nucléiques et à la mise en œuvre d'éventuelles étapes d'enrichissement sont envisageables : séquençage avec ou sans sélection des ARN messagers, d'ARN bicaténaires produits durant la réplication des virus ARN, de petits ARN issus du *silencing* (siRNAs) ou d'acides nucléiques circulaires. Chacune de ces approches présente ses avantages et ses inconvénients, en particulier au niveau du spectre viral potentiellement détectable, mais aussi au niveau des coûts, de la sensibilité ou des difficultés de mise en œuvre.

Le premier niveau d'analyse pour lequel ces nouvelles approches présentent un très fort potentiel est celui de la plante individuelle, pour des démarches d'étiologie ou de diagnostic. En effet, si la recherche ciblée d'un agent viral connu est aujourd'hui généralement directe, l'indexage global d'une plante pour la présence de n'importe quel agent viral, connu ou inconnu, reste un challenge du fait de la très large diversité des phytovirus. De plus, de nombreuses plantes induisent des difficultés supplémentaires, liées à la présence de substances d'origine végétale interférentes plus ou moins bien caractérisées (latex, polyphénols et polysaccharides…). Ceci explique que les agents viraux responsables de maladies fréquentes ou connues depuis de nombreuses années soient parfois encore aujourd'hui non caractérisés.

Une analyse systématique des banques d'EST (*expressed sequence tags*, obtenues par séquençage systématiques de banques ADNc) de plante pour la présence de séquence de phytovirus nous a permis

[1] [1]INRA, UMR 1332 BFP, BP81, 33883 Villenave d'Ornon cedex, France.
[2] Université de Bordeaux, UMR 1332 BFP, BP81, 33883 Villenave d'Ornon cedex, France.

Copyright – Académie d'Agriculture de France – 2012. Séance du 16 mai.

d'identifier de nombreux nouveaux agents viraux mais surtout d'obtenir une première vision large de la fréquence des séquences virales et, partant de la profondeur de séquençage requise pour une recherche approchant l'exhaustivité. Ces données indiquent que seulement environ 50% des virus détectés sont représentés dans les banques d'EST avec une fréquence supérieure ou égale à 0.1%, cette valeur montant à 80% des virus pour un seuil abaissé à 0.01%. Une recherche exhaustive de virus par le séquençage d'ARN messagers doit donc nécessairement approcher voire dépasser la profondeur de 100.000 séquences par échantillon. Les autres approches que nous avons commencé à développer et qui semblent aujourd'hui particulièrement prometteuses sont le séquençage des siRNA produits par l'action du *silencing* cellulaire et le séquençage d'ARN bicaténaires (dsRNAs). Cette dernière démarche limite le champ de détection aux virus ARN, qui représentent cependant de l'ordre de 85% des phytovirus connus. Par contre, l'enrichissement en séquences virales obtenu lors de la purification des dsRNAs permet de considérablement réduire la profondeur de séquençage et donc les coûts associés. Ces différentes approches seront présentées sur des exemples concrets illustrant leur puissance et leur intérêt potentiel.

Pour la stratégie d'indexage par les dsRNAs que nous avons optimisée, la préparation des échantillons en amont du séquençage peut être conduite en quelques jours et sans recours à des équipements sophistiqués. En partenariat avec des bioinformaticiens, nous avons également développé une stratégie pour l'analyse rapide et automatisée des gros volumes de données de séquence générés. Grace à une analyse multiplexée des échantillons, le coût est aujourd'hui estimé à 300-400 euros par échantillon en incluant les salaires, ce qui est compétitif par rapport au coût d'un indexage biologique complet tel que conduit actuellement. De telles approches pourraient donc trouver rapidement des applications dans le domaine du phytodiagnostic.

Le deuxième domaine dans lequel ces approches présentent un fort intérêt est celui de l'analyse globale des phytovirus présents dans une population végétale ou dans un écosystème. De telles démarches de métagénomique phytovirale n'étaient jusqu'à présent pas accessibles mais sont aujourd'hui rendues possibles par l'augmentation des débits et la démocratisation des approches NGS. En utilisant l'analyse des ARN bicaténaires et des approches de séquençage multiplex, nous avons ainsi commencé à analyser le métagénome phytoviral dans deux environnements contrastés. L'un, les Iles Kerguelen, peut être considéré comme un environnement modèle non anthropisé et caractérisé par des contraintes biotiques fortes, avec une flore paucispécifique et une absence d'insectes vecteurs de virus jusqu'à un passé très récent. L'autre environnement étudié correspond à un milieu agricole tempéré caractérisé par une anthropisation forte et par la présence d'une diversité d'espèces hôtes cultivées ou adventices. Les premières données d'analyse de ces deux métagénomes contrastés seront présentées, mettant en particulier en évidence un enrichissement en virus tempérés à ARN bicaténaires dans celui des Iles Kerguelen et permettant une première évaluation de la part de la diversité virale connue dans ces deux environnements.

Au-delà de ces premiers résultats démontrant la faisabilité et l'intérêt de ce type de démarche, il est clair que de nouveaux développements tant technologiques que conceptuels sont à attendre pour les années à venir. Au plan technologique, les NGS continuent à progresser, générant des volumes de séquence toujours plus importants pour des prix de plus en plus faibles. Les challenges sont donc aujourd'hui à rechercher dans l'automatisation du traitement des échantillons, dans le multiplexage et, surtout, dans le traitement des données de séquence. De nombreuses questions existent dans ce domaine, nécessitant des développements spécifiques. On peut par exemple citer le problème du comptage des espèces ou des OTU (*Operation Taxonomy Units*) pour parvenir à une quantification objective de la diversité virale ou celui de l'identification des séquences de virus trop divergents pour pouvoir être identifiés par les techniques actuelles basées sur la recherche d'homologies de séquence. Au plan conceptuel, c'est à une nouvelle vision de l'écologie phytovirale que nous appellent ces développements.

Copyright – Académie d'Agriculture de France – 2012. Séance du 16 mai.