

Génomomes de bactériophages - Ines Staes

Diapositive 2 :

Il existe une grande variété de phages et cette diversité se retrouve également au niveau du génome. En raison de leur taille relativement petite et de leur simplicité d'isolement, les bactériophages ont été les premiers génomes complets à être séquencés. Plus de 2000 génomes de bactériophages ont été séquencés jusqu'à aujourd'hui.

Diapositive 3 :

La grande majorité des bactériophages sont des virus à double brin d'ADN (dsDNA) bien que les autres types existent également. Les génomes de phage sont généralement linéaires lorsqu'ils sont empaquetés dans la capsid. Certains ADN de phages ont des extrémités définies. Cela signifie que lorsque vous regardez l'ADN linéaire dans différentes capsides, ils auront toujours les mêmes extrémités gauches et droites. Pour d'autres ce n'est pas le cas, pour ces virus, la composition du génome est la même dans chaque capsid, mais les extrémités ne sont pas toujours identiques, l'ADN semble avoir été linéarisé en ouvrant des cercles identiques sur différents sites. C'est ce qu'on appelle « circulairement permuté ». En outre, le début et la fin sont souvent des séquences identiques et cela porte le nom de redondance des extrémités.

Diapositive 4 :

Les tailles des génomes de phages varient énormément. Les plus petits (tels que les virus ssRNA d'*E. coli*) possèdent un peu plus de 3000 nucléotides tandis que les virus les plus importants (tels que le phage de *Bacillus Megaterium G*) en possèdent près de 500 kbp. En général, l'ADN du phage est densément empaqueté à l'intérieur de la capsid à des densités plus ou moins similaires. La taille de la capsid varie donc selon la taille du génome. Beaucoup de virus, en particulier les bactériophages dsDNA, présentent des couches systématiques d'ADN empaqueté tandis que d'autres, surtout les filamenteux, forment leur génome en hélices entourées de protéines. Vu le peu d'espace dans la capsid, la densité de gènes des génomes du phage est très élevée, ce qui signifie qu'il y a très peu de régions non codantes. Les régions régulatrices sont souvent compactes et, parfois, les régions codantes se chevauchent. Les génomes de bactériophages possèdent potentiellement la plus grande nouveauté génétique du monde biologique, dans la mesure où la plupart de leurs gènes codés (peut-être jusqu'à 80 %) ne sont pas liés à des protéines connues et sont de fonction inconnue.

Diapositive 5 :

La classification taxonomique des virus des archées et des bactéries est très importante afin de développer une compréhension prédictive des écosystèmes microbiens, mais cela s'avère être un travail difficile. Récemment, une énorme quantité de nouveaux génomes viraux et des fragments de génome ont été identifiés ce qui exige une approche qui s'écarte des méthodes de classification traditionnelles et de l'utilisation des génomes pour la taxonomie. La taxonomie des organismes cellulaires, basée sur l'analyse de l'alignement des séquences de gènes homologues universellement conservés, tels que les gènes des ARN ribosomiaux (ARNr), est possible car tous les organismes partagent ces gènes homologues. Cependant, il n'y a pas de gène unique partagé par tous les bactériophages et les virus en général. Cela rend impossible de déterminer les relations évolutives entre tous les bactériophages en se basant sur l'analyse de l'alignement des séquences d'un seul gène.

Rohwer et Edwards ont proposé une façon de classer taxonomiquement les phages en se basant sur leurs protéomes en construisant un « Phage Proteomic Tree ». Il s'agit d'un dendrogramme qui révèle une similarité génomique globale entre des centaines ou des milliers de virus sur la base d'un

algorithme qui utilise le génome pour classer les phages par rapport à leurs voisins les plus proches et à tous les autres phages. Cette méthode n'exige pas la visualisation de virions libres, d'informations sur les virions ou sur leurs modes de vie et peut diviser les phages en taxons qui prédisent plusieurs aspects de la biologie du phage. En outre, l'approche de l'arbre protéomique est efficace pour étudier les génomes des virus nouvellement séquencés et attribuer des fragments de *reads* aux bons génomes. Cependant, avec un grand nombre de phages, cette méthode devient complexe et lente.

Diapositive 6 :

La méthode suivante omet la comparaison des séquences de protéines et considère simplement la présence ou l'absence de gènes. Cette approche basée sur le réseau a été utilisée pour organiser la séquence du génome d'un virus double brin et fonctionne par prédiction des gènes viraux dans tout le génome qui sont ensuite traduits en protéines. Ces dernières sont à leur tour organisées en grappes de protéines. L'évaluation du nombre de grappes de protéines partagées sont comparées avec l'ensemble des données pour établir un profil protéique, la représentation de cette information sous forme de graphique, avec des nœuds représentant les génomes viraux et des flèches représentant le score de similarité de leur contenu protéique partagé vous donnera un réseau de gènes partagés.

Diapositive 7 :

Malgré le fait que leurs génomes ont des ordres de grandeur inférieurs à ceux des bactéries et d'autres organismes, le séquençage des génomes de phage pose plusieurs défis : (1) l'obtention d'un matériel génomique pur, comme les phages ne sont pas auto-réplicatifs, l'isolement du matériel génétique du phage implique plusieurs étapes de purification; (2) L'amplification par PCR peut être biaisée car certains génomes de phage sont particulièrement riches en GC. De telles extrêmes peuvent poser problème pour la PCR et le séquençage; (3) Les génomes de phage sont également connus pour contenir des structures génomiques complexes telles que des répétitions directes ou inversées extrêmement longues et des redondances terminales qui sont problématiques lors de l'assemblage de la séquence complète du génome à partir des *reads* et (4), dans la génomique bactérienne et humaine, le *mapping* des *reads* sur un génome de référence fini peut être un outil très utile, cependant, en génomique des phages, ceci est très rarement possible en raison de l'absence de génome de référence pour un phage donné. Une solution serait d'avoir une approche combinée d'une technologie donnant de longs *reads* combinée à une seconde technologie donnant un grand nombre de courts *reads* afin d'obtenir un séquençage efficace de l'ADN des génomes de bactériophages.