

## Genomy bakteriofagów

Ines Staes

### Slajd 2

Generalnie, istnieje ogromna różnorodność fagów. Ponadto ta różnorodność jest również reprezentowana na poziomie genomu. Ze względu na stosunkowo niewielkie rozmiary i prostotę izolacji, genomy bakteriofagów były pierwszymi, które całkowicie zsekwencjonowano. Obecnie ponad 2000 poszczególnych genomów bakteriofagów zostało zsekwencjonowanych.

### Slajd 3

Większość bakteriofagów zbudowana jest z dwuniciowego DNA (dsDNA), chociaż istnieją również inne typy. Genomy fagowe są z reguły liniowe, gdy są upakowane w kapsydach. Niektóre DNA fagowe mają zdefiniowane końce. Oznacza to, że gdy spojrzymy na liniowy DNA w różnych kapsydach, zawsze będą miały ten sam lewy i ten sam prawy koniec. W przypadkach innych wirusów może być inaczej. Gdzie ogólny skład genomu będzie taki sam w każdym kapsydie, ale końce nie zawsze będą takie same. DNA wydaje się przybierać formę liniową przez otwieranie się form cyklicznych w różnych miejscach. Określane jest to cykliczną permutacją. Dodatkowo, początek i koniec mają często identyczne sekwencje i nazywa się to redundancją końcową.

### Slajd 4

Wielkości genomów fagowych są bardzo różne. Najmniejsze są poniżej 3000 nukleotydów, takie jak wirusy ssRNA *E.coli*, podczas gdy największe wirusy mają prawie 500 kbp, jak w przypadku faga G, którego gospodarzem jest *Bacillus megaterium*. Ogólnie rzecz biorąc, DNA faga jest zazwyczaj mniej lub bardziej upakowany wewnątrz kapsydu, dlatego też rozmiar kapsydu zmienia się w zależności od rozmiaru genomu. Wiele wirusów, zwłaszcza bakteriofagów dsDNA, wykazuje warstwowe upakowanie DNA, podczas gdy inne, głównie nitkowate (filamentarne), wirusy pakują swój genom helikalnie. Ponieważ w kapsydie nie ma dużo miejsca, gęstość genów fagowych w genomie jest bardzo wysoka, co oznacza, że jest niewiele regionów niekodujących. Regiony regulacyjne są często kompaktowe, a czasami nakładają się na regiony kodujące. Genome fagowe posiadają największą pulę nowości genetycznej w świecie biologicznym, dlatego też większość genów (może nawet 80%) nie jest związana ze znanymi białkami i ich funkcja również nie jest znana.

### Slajd 5

Klasyfikacja taksonomiczna wirusów archaea i bakteryjnych jest bardzo ważna dla zrozumienia ekosystemów bakteryjnych, ale okazuje się, że jest to bardzo trudne zadanie. Ostatnio, ogromna ilość nowych genomów wirusowych i fragmentów genomów została zidentyfikowana, co wymaga odrębnego podejścia od tradycyjnych metod klasyfikacji, w celu wykorzystania genomów do taksonomii. Taksonomia organizmów komórkowych, oparta na analizie porównawczej sekwencji genów homologicznych, takich jak geny rybosomalnego RNA (rRNA), jest możliwa, ponieważ wszystkie organizmy posiadają te homologie. Jednakże, nie ma pojedynczego genu, który jest wspólny dla wszystkich bakteriofagów i wirusów. Uniemożliwia to ustalenie zależności ewolucyjnych między wszystkimi bakteriofagami w oparciu o analizę porównawczą sekwencji pojedynczego genu.

Rohwer i Edwards zaproponowali sposób klasyfikacji taksonomicznej fagów w oparciu o ich proteomy, tworząc "Proteomiczne Drzewo Fagów". Jest to program do tworzenia drzew filogenetycznych, który ujawnia globalne podobieństwo genomowe pomiędzy setkami lub tysiącami wirusów w oparciu o algorytm, używający genomów w celu umieszczenia fagów w stosunku do ich najbliższych sąsiadów i

wszystkich innych fagów. Ta metoda nie wymaga wizualizacji wolnych wirionów ani informacji o stylu życia i może podzielić fagi na grupy taksonomiczne, które przewidują kilka aspektów ich biologii. Ponadto, analiza drzewa proteomicznego jest skuteczna w celu badania genomów nowo zsekwencjonowanych wirusów i przyporządkowania fragmentów sekwencji do odpowiednich genomów. Jednak przy dużej liczbie fagów ta metoda staje się złożona i czasochłonna.

#### Slajd 6

Następna metoda pomija porównanie sekwencji białkowych i po prostu rozważa obecność lub brak genów. Jest to podejście oparte na sieciach, które zostało wykorzystane do zorganizowania sekwencji dwuniciowego genomu wirusa i realizuje się przez przewidywanie genów wirusa we wszystkich genomach, które następnie przepisywane są na poszczególne białka. Te ostatnie są z kolei organizowane w klastry białkowe. Ocena liczby wspólnych klastrów białkowych w całym zestawie danych w celu ustalenia białkowego profilu, reprezentowana jest w postaci sieci, z węzłami reprezentującymi genomy wirusa i skorygowana wspólną zawartością białek tworząc sieć uwspólniania genów.

#### Slajd 7

Pomimo faktu, że ich genomy są o rzędy wielkości mniejsze w porównaniu do bakterii i innych organizmów, sekwencjonowanie genów fagowych stwarza kilka wyzwań: (1) uzyskiwanie czystego materiału genomowego faga (fagi replikują się same, izolacja materiału genomowego obejmuje kilka etapów oczyszczania) (2) amplifikacja PCR może być utrudniona, ponieważ niektóre genomy są bardzo bogate w pary GC. Taki skład może stanowić problem dla PCR i sekwencjonowania (3) genomy fagowe są również znane z posiadania skomplikowanych struktur genomowych, takich jak wyjątkowo długie bezpośrednio lub odwrócone sekwencje powtórzone oraz końcowe sekwencje powtórzone, które stwarzają problemy przy składaniu pełnej sekwencji genomu z otrzymanych fragmentów (4) w genomice bakterii czy ludzi, mapowanie nowych genomów odnoszone jest do genomu referencyjnego, przez co może być potężnym narzędziem analitycznym. W genomice fagów jest to bardzo rzadka możliwość ze względu na brak genomu referencyjnego dla danego faga. Możliwym rozwiązaniem jest połączenie technologii sekwencjonowania długich fragmentów z technologią kojarzenia dużej liczby krótkich odcinków DNA w celu skutecznego sekwencjonowania genomów bakteriofagowych.