

Identyfikacja i adnotacja funkcjonalna taksonomicznie-specyficzných genów bakterii

Geny tworzą rodziny — to znaczy trwające w czasie linie rodowe, które zapoczątkował gen założycielski, który mógł ulegać kopiowaniu i zmianom w każdym kolejnym pokoleniu. Przez ostatnie 40 lat naukowcy uważali, że taki jest właśnie główny mechanizm powstawania nowych genów - gen po prostu powstaje z kopii istniejących już starych genów (*duplikacja*), które muszą nieustannie działać, wykonując niezbędne dla organizmu funkcje (*presja selekcyjna*). Natomiast nowa kopia genu - w odpowiedzi na zmieniające się warunki środowiskowe - może się swobodnie różnicować, aż wyspecjalizuje się na tyle, aby spełniać nową, ważną funkcję biologiczną (*neofunkcjonalizacja*).

Ostatnie doniesienia naukowe wskazują jednak, że narodziny pewnej części genów nie odbywają się w tego typu 'rodzinných kręgach'. Geny takie nie wykazują bowiem pokrewieństwa (*homologii*) z żadnym genem obecnym u innych znanych organizmów, a ich powstawanie określane jest w literaturze naukowej jako *tajemnicze*, *enigmatyczne* i *nieznane*. Stąd geny te nazywane są często "genami sierocymi" lub bardziej precyzyjnie, genami taksonomicznie-specyficznymi (TRG, *Taxonomically-Restricted Genes*). Mimo, że zdecydowana większość genów TRG nie została w pełni scharakteryzowana, wyniki najnowszych badań wskazują, że zazwyczaj odpowiadają one za specyficzne dla danego taksonu funkcje, na przykład podczas regeneracji kończyn u salamandry, określania ról społecznych pszczół czy hamowania rozwoju nowotworów u człowieka.

Prowadzone do tej pory badania mające na celu identyfikację na szeroką skalę genów TRG najczęściej sztucznie zawyżały ich liczbę, ponieważ zwykle obejmowały niekompletny zbiór sekwencji i/lub wykorzystywały nieprecyzyjne metody komputerowe do identyfikacji sekwencji homologicznych. W rezultacie tysiące genów, początkowo zaproponowanych jako 'osierocone', znajdowało swoich dalekich 'krewnych' i traciło status taksonomicznej unikalności. Dlatego w proponowanym projekcie zamierzamy wykorzystać cały arsenał bioinformatycznych narzędzi (w tym także własne rozwiązania) w celu kompleksowej identyfikacji i pełnej charakterystyki genów TRG w reprezentatywnej grupie organizmów, obejmującej ponad 40 tysięcy szczepów bakteryjnych (tj. około 140 milionów genów).

Zasadniczym celem proponowanego przez nas projektu jest znalezienie odpowiedzi na fundamentalne dla biologii molekularnej pytania dotyczące natury genów TRG u bakterii, m.in.: - *Jak powszechnie jest występowanie genów TRG u różnych rodzajów, gatunków i szczepów bakteryjnych?* - *Czy możliwe jest określenie czynników biologicznych (np. specyficzne warunki środowiskowe), które preferencyjnie wpływają na powstawanie genów TRG?* - *W jakim stopniu zróżnicowanie składu genów TRG poszczególnych bakteryjnych grup taksonomicznych odzwierciedla ich ekologiczne potrzeby i powstawanie nowych cech (np. oporność na antybiotyki)?* - *Czy geny TRG pojawiają się częściej wśród określonych grup bakteryjnych (np. bakterii chorobotwórczych)?* - *Jakie mechanizmy molekularne zaangażowane są w powstawanie genów TRG?* - *Jaką funkcję pełnią niescharakteryzowane geny TRG?*

Nasze pilotażowe badania obejmujące ponad 6 tys. szczepów bakteryjnych wskazują, że każdy z tych organizmów posiada średnio 3% specyficznych genów TRG. Co ciekawe, najwięcej genów tego typu posiadają bakterie chorobotwórcze. Rekordzistą jest szczep *Prevotella copri* związany z reumatoidalnym zapaleniem stawów - wszystko wskazuje na to, że w genomie tej bakterii funkcjonuje ponad 40% genów (2025 z 4835) niespotykanych w żadnym innym organizmie, nawet w szczepach należących do tego samego gatunku!

Proponowana przez nas identyfikacja genów występujących wyłącznie w genomach bakterii chorobotwórczych, nie mających swoich odpowiedników w innych bakteriach oraz w genomach nosicieli, pozwoli na opracowanie nowych leków 'wyłączających' specyficzne dla bakterii białka kluczowe dla rozwoju choroby. Obiecujące wyniki uzyskaliśmy w badaniach szczepu gronkowca złocistego opornego na niemal wszystkie dostępne antybiotyki (*S. aureus MRSA*), który jest odpowiedzialny za jedne z największych fal infekcji w ostatnich 5 latach. W genomie tej bakterii zidentyfikowaliśmy 39 specyficznych genów, które produkują białka bezpośrednio zaangażowane w procesy chorobotwórcze (np. toksyny, czynniki osłabiające odpowiedź immunologiczną człowieka, białka zmniejszające poziom hemu w hemoglobinie).

Końcowym produktem tego projektu będzie stworzenie publicznie dostępnego portalu internetowego zawierającego pełen zasób informacji na temat zidentyfikowanych przez nas genów TRG u bakterii. Unikalny i kompleksowy charakter tych danych powinien być referencją i inspiracją dla biologów molekularnych badających funkcje genów TRG w różnych szczepach, gatunkach i grupach taksonomicznych bakterii. Poza informacjami czysto poznawczymi, zawarte w tym portalu dane umożliwią m.in. identyfikację sekwencji, które mogą posłużyć jako markery molekularne, np. w metagenomice środowiskowej lub badaniach medycznych.