

Łukasz Dziewit*¹, Dariusz Bartosik¹

Zakład Genetyki Bakterii, Instytut Mikrobiologii, Wydział Biologii Uniwersytetu Warszawskiego
ul. Miecznikowa 1, 02-096, Warszawa, Polska

Wpłynęło w styczniu 2011 r.

1. Wstęp. 2. Wielkość genomów prokariotycznych. 3. Organizacja strukturalna genomów prokariotycznych. 3.1. Genomy zawierające plazmidy. 3.2. Genomy wielochromosomowe. 4. Informacja genetyczna w genomach prokariotycznych. 4.1. Sekwencje kodujące białka. 4.2. Sekwencje kodujące RNA. 4.3. Pseudogeny. 4.4. Sekwencje repetytywne. 4.5. Ruchome elementy genetyczne zintegrowane z genomowym DNA. 5. Minimalny genom i pangenomy. 6. Ewolucja genomów prokariotycznych. 7. Podsumowanie

Prokaryotic genomes in the light of genomic analyses

Abstract: Genomic analyses revealed that prokaryotic genomes are very diverse. They differ in size, architecture and genetic information which they encode. The plasticity of the genomes is the result of several factors, including (i) horizontal gene transfer (responsible for the acquisition of exogenous DNA), (ii) genetic rearrangements (resulting from diverse recombinational events), and (iii) genome reduction, which leads to the loss of genetic information of a low adaptative value. The insertions and deletions of DNA permanently alter the genomes and result in adaptation of the microorganisms to their specific environments by genome optimization. For this reason we observe a straight correlation between metabolic versatility of prokaryotes and the size of their genomes.

1. Introduction. 2. Size of prokaryotic genomes. 3. Structural organization of prokaryotic genomes. 3.1. Genomes containing plasmids. 3.2. Multichromosomal genomes. 4. Genetic information within prokaryotic genomes. 4.1. Protein encoding sequences. 4.2. RNA encoding sequences. 4.3. Pseudogenes. 4.4. Repeated sequences. 4.5. Mobile genetic elements integrated in genomic DNA. 5. Minimal genome and pangenomes. 6. Evolution of prokaryotic genomes. 7. Summary

Słowa kluczowe: genom, horyzontalny transfer genów, pangenom, plazmid

Key words: genome, horizontal gene transfer, pangenome, plasmid

1. Wstęp

Odczytanie sekwencji nukleotydowej genomu bakterii *Haemophilus influenzae* Rd KW20, zapoczątkowało w 1995 roku [14] erę genomiki prokariotów. Od tego czasu obserwuje się coraz bardziej dynamiczny rozwój tej dziedziny badań, szczególnie zintensyfikowany w ostatnich latach dzięki opracowaniu i wprowadzaniu nowych technik sekwencjonowania DNA. Według informacji zawartych w bazie National Center for Biotechnology Information (NCBI), odczytano dotychczas kompletne sekwencje nukleotydowe genomów 1279 szczepów bakterii oraz 93 przedstawicieli *Archaea*. Co więcej, kolejnych 3 531 projektów genomicznych bakterii i 83 archeonów jest obecnie w trakcie realizacji (dane z 13.12.2010). Najwięcej danych zgromadzono dotąd o genomach szczepów bakterii zaliczanych do typów: *Proteobacteria* (607 zsekwenconowanych genomów), *Firmicutes* (317), *Actinobacteria* (123) oraz *Cyanobacteria* (41).

Oprócz kompleksowych badań kompletnych genomów prokariotycznych realizowane są również projekty, których celem jest poznanie sekwencji jedynie mniejszych replikonów, tj. bakteriofagów i plazmidów. Do momentu zakończenia przygotowywania manuskryptu

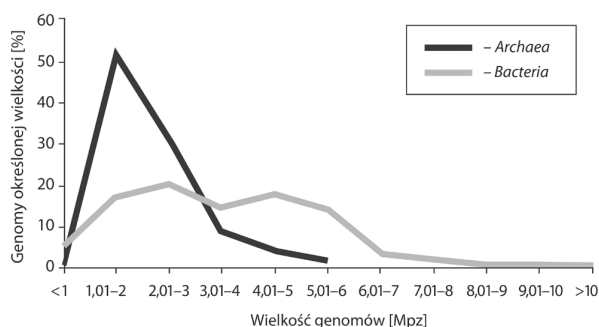
tej pracy poznano kompletne sekwencje 2 155 plazmidów bakteryjnych, 100 plazmidów archeonów oraz 595 bakteriofagów (NCBI).

Analiza tak olbrzymiej puli danych uświadomiła badaczom, że genomy są bardzo zróżnicowane, zarówno pod względem (i) wielkości, (ii) organizacji strukturalnej, jak i (iii) rodzaju niesionej informacji genetycznej. Zmienność ta jest wypadkową wielu procesów, m.in. wprowadzania do komórki nowej, egzogennej informacji genetycznej w wyniku horyzontalnego transferu genów (HGT – horizontal gene transfer), jak i różnego typu rekombinacji, które mogą powodować rozległe zmiany w strukturze DNA (np. delecje, translokacje, inwersje bądź duplikacje).

2. Wielkość genomów prokariotycznych

Wielkość genomów prokariotycznych jest bardzo zróżnicowana. Genomy archeonów mają dość kompaktową strukturę. Ich rozmiary zawierają się w przedziale od 0,49 Mbp do 5,75 Mbp, przy czym ponad połowa z nich liczy około 2 Mbp. Rozpiętość wielkości genomów bakteryjnych jest znacznie większa. Znamy bakterie,

* **Autor korespondencyjny:** Zakład Genetyki Bakterii, Instytut Mikrobiologii, Wydział Biologii Uniwersytetu Warszawskiego, ul. Miecznikowa 1, 02-096, Warszawa; tel: (22) 554 13 44; fax: (22) 554 14 02; e-mail: ldziewit@biol.uw.edu.pl



Rys. 1. Zróżnicowanie wielkości genomów *Archaea* i *Bacteria*, wg [21], (zmodyfikowany)

Powyższe zestawienie wykonano w oparciu o dane zamieszczone do dnia 1 grudnia 2010 w bazie NCBI (<http://www.ncbi.nlm.nih.gov/>).

których genomy nie przekraczają 0,2 Mpz, jak również prawdziwego „giganta” świata bakteryjnego, którego genom (13 Mpz) jest ponad 65 razy większy. Jednak wielkość genomów zdecydowanej większości bakterii (70%) plasuje się w zakresie od 2 do 5 Mpz (rys. 1).

Najmniejsze genomy wśród mikroorganizmów prokariotycznych mają obligatoryjne symbionty i pasożyty. Przyczyną tego jest stopniowa redukcja ich genomów, która jest wynikiem: (i) generowania większej liczby delecji niż insercji (co jest spowodowane m.in. ograniczeniem zakresu wprowadzania egzogennej DNA w wyniku horyzontalnego transferu genów) oraz (ii) utraty genów warunkujących biosyntezę związków, które są dostarczone przez gospodarza. W konsekwencji, z genomów symbiontów i pasożytów eliminowane są geny umożliwiające im przeżycie poza organizmem gospodarza, co jednocześnie czyni stan symbiozy lub pasożytnictwa, obligatoryjnym [29].

Najmniejszy genom wśród archeonów ma *Nanoarchaeum equitans* Kin4-M, obligatoryjny symbiont innego archeona – *Ignicoccus hospitalis* KIN4/I, który jest hipertermofilnym chemolitoautotrofem [33]. Pojedynczy, kolisty chromosom *N. equitans*, o wielkości 490 885 pz, koduje jedynie 540 białek. Archeon ten nie posiada większości genów umożliwiających syntezę *de novo*: aminokwasów, lipidów, kofaktorów i nukleotydów (które są dostarczane przez gospodarza), może on jednak przeprowadzać proteolizę białek i deaminację aminokwasów z uwolnieniem amoniaku. Stwierdzono, że *N. equitans*, po osiągnięciu odpowiedniej liczebności hamuje wzrost *I. hospitalis*, co może sugerować, że archeon ten nie jest symbiontem, lecz pasożytem swego gospodarza [45].

Najmniejszy z poznanych do tej pory genomów bakteryjnych (143 795 pz) występuje u „*Candidatus* Hodgkinia cicadicola” Dsem (*Alphaproteobacteria*). Bakteria ta jest obligatoryjnym endosymbiontem cykadki *Diceroprocta semicincta*. Jej genom koduje zaledwie 169 białek oraz 19 strukturalnych RNA, a w przeciwieństwie do innych symbiontów prokariotycznych,

charakteryzuje się dość dużą zawartością par GC w sekwencji nukleotydowej (58,4%).

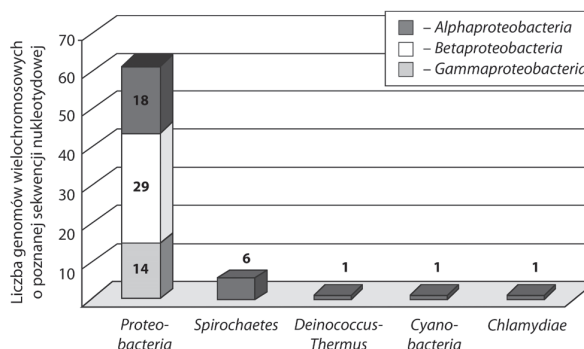
Z kolei, najmniejsze genomy wśród wolno żyjących mikroorganizmów prokariotycznych mają odpowiednio: (i) wspomniany wyżej archeon – *Ignicoccus hospitalis* KIN4/I – 1 297 538 pz [33] oraz (ii) bakteria należąca do klasy *Alphaproteobacteria* – „*Candidatus* Pelagibacter ubique” HTCC 1062 – 1 308 759 pz [19]. Na drugim końcu skali znajduje się bakteria glebowa *Sorangium cellulosum* ‘So ce 56’ (*Deltaproteobacteria*), której genom (13 033 799 pz) jest o około 10% większy od chromosomu *Saccharomyces cerevisiae* S288c (12 Mpz). Genom tej bakterii składa się z pojedynczego, kolistego chromosomu, w obrębie którego zidentyfikowano 9 381 genów kodujących m.in. białka zaangażowane w syntezę wielu metabolitów wtórnych, np. barwników karotenoidowych oraz różnorodnych substancji cytotoksycznych [36].

3. Organizacja strukturalna genomów prokariotycznych

Ponieważ pierwsze poznane genomy bakteryjne (*H. influenzae*, *E. coli*) zawierały pojedynczy, kolisty chromosom, dość długo błędnie uważano, że jest to typowe dla chromosomów wszystkich bakterii. Okazało się jednak, że w wielu przypadkach genomy mają strukturę wieloreplikonową, składają się bowiem z dwóch, bądź trzech różnych chromosomów, wśród których spotyka się też replikony o strukturze liniowej (rys. 2). Ważną częścią wielu genomów są również plazmidy, których wielkość może niekiedy przekraczać rozmiary pewnych chromosomów.

3.1. Genomy zawierające plazmidy

Plazmidy są to pozachromosomowe, kolisty lub liniowe cząsteczki DNA zdolne do autonomicznej replikacji. Występują one powszechnie w bardzo wielu



Rys. 2. Występowanie genomów wielochromosomowych u bakterii. Powyższe zestawienie wykonano w oparciu o dane zamieszczone do dnia 1 grudnia 2010 w bazie NCBI.

organizmach prokariotycznych, a także u niektórych eukariotów. Analiza poznanych dotąd genomów wskazuje, że ok. 20% archeonów i 36% bakterii niesie co najmniej jeden plazmid.

Plazmidy, choć zwykle nie są niezbędne dla swoich gospodarzy (nie noszą kluczowych genów metabolizmu podstawowego – house keeping genes), odgrywają ważną rolę w ewolucji prokariotów, umożliwiając bowiem wprowadzanie do komórki puli egzogennej DNA, w drodze horyzontalnego transferu genów. Wiele z nich niesie informację genetyczną warunkującą różnorodne cechy fenotypowe (np. oporność na antybiotyki bądź jony metali ciężkich, czy też zdolność do wykorzystywania trudno przyswajalnych źródeł węgla), które w określonych warunkach środowiska mogą zapewnić przewagę gospodarzowi w konkurencji z innymi mikroorganizmami zasiedlającymi tę samą niszę ekologiczną.

Wśród archeonów najwięcej plazmidów, bo aż siedem, zidentyfikowano w genomie *Haloarcula marismortui* ATCC 43049 [1], zaliczanej do rodziny *Halobacteriaceae*. W halofilnych archeonach z tej rodziny występuje niemal połowa wszystkich poznanych dotąd plazmidów *Archaea*. Wykryto tam również najmniejszy oraz największy zidentyfikowany dotąd plazmid archeonów, odpowiednio: (i) pZMX201 (1668 pz) z *Natri-nema* sp. CX2021, który zawiera tylko jeden gen oraz (ii) pHLAC01 (413 338 bp) z *Halorubrum lacusprofundi* ATCC 49239 niosący aż 402 geny.

W przypadku bakterii, prawdziwym rekordzistą pod względem liczby niesionych plazmidów, jest znany patogen *Borrelia burgdorferi* B31, wywołujący boreliozę. W bakterii tej występuje aż dwadzieścia jeden plazmidów, z których dwanaście ma strukturę liniową. Łączna wielkość tych replikonów wynosi około 533 kpz, podczas gdy liniowy chromosom szczepu B31 ma zaledwie 911 kpz. Plazmidy te noszą niemal 50% wszystkich genów *B. burgdorferi* B31 (rys. 3). Przypuszcza

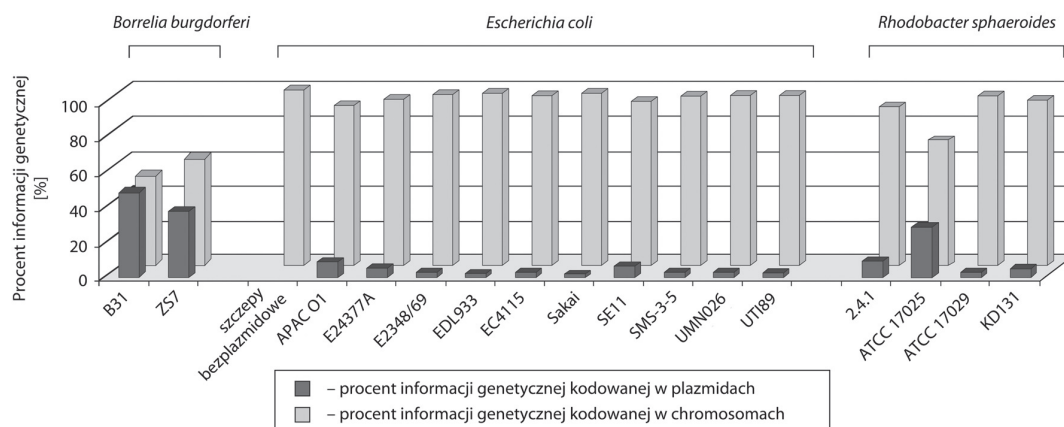
się, że wiele z tych genów może kodować białka odpowiedzialne za wirulencję, w tym za zmienność antygenów powierzchniowych [15]. Co ciekawe, w jednym z plazmidów tego szczepu (cp26; 26,5 kpz) zidentyfikowano gen kodujący resolwazę telomerów – enzym zaangażowany w proces replikacji liniowego chromosomu [6]. Plazmidy tego typu, zawierające pojedyncze, niezbędne dla bakterii geny metabolizmu podstawowego pochodzenia chromosomowego, określa się jako *essential genetic elements*, a więc można je nazwać „niezbędnikami”.

Największym poznany dotąd plazmidem jest pNGR234b (2 430 033 pz) pochodzący z *Rhizobium* sp. NGR234 (*Alphaproteobacteria*). Chociaż plazmid ten niesie ponad 35% genów szczepu NGR234, to jednak nie jest on „niezbędnikiem”, gdyż nie zawiera unikalnych genów metabolizmu podstawowego. Niemniej jednak koduje on wiele systemów transportu (ponad 50% transporterów tego szczepu, głównie typu ABC) oraz potencjalne elementy transpozycyjne (ok. 40% całkowitej puli transpozonów zidentyfikowanych w szczepie NGR234), co sugeruje zarówno obecność genów o charakterze adaptacyjnym, jak i duży potencjał rekombinogeny tego replikonu [35].

U bakterii przeważają jednak znacznie mniejsze plazmidy. Najmniejsze z nich to pRKU1 i pRQ7 (oba o wielkości 846 pz), które zidentyfikowano, odpowiednio, w *Thermotoga petrophila* RKU-1 oraz *Thermotoga* sp. RQ7. Są one niemal identyczne i zawierają tylko jeden gen – kodujący białko replikacyjne.

3.2. Genomy wielochromosomowe

Jak wynika z danych zamieszczonych w bazie NCBI, poznano dotąd 70 genomów bakteryjnych, w skład których wchodzi dwa lub trzy chromosomy. W większości (ponad 80%) genomy takie występują u przedstawicieli



Rys. 3. Dystrybucja sekwencji kodujących białka w plazmidach i chromosomach szczepów wybranych gatunków bakterii. Powyższe zestawienie wykonano w oparciu o dane zamieszczone do dnia 1 grudnia 2010 w bazie NCBI.

typu *Proteobacteria*, głównie z klasy *Betaproteobacteria* (rys. 2). Z kolei wśród *Archaea*, jedynie *Haloarcula marismortui* ATCC 43049 (halofilny archeon wyizolowany z Morza Martwego) oraz (ii) *Halorubrum lacusprofundi* ATCC 49239 (psychrofilny halofil wyizolowany na Antarktydzie) posiadają dwa chromosomy [1].

Analiza informacji genetycznej zawartej w genomach wielochromosomowych wskazuje, że chromosomy nie są równocenne, można bowiem wyróżnić wśród nich: (i) chromosom główny (oznaczany jako chromosom 1), który zawiera zasadniczą część genów metabolizmu podstawowego, rozmieszczonych równomiernie w całym replikonie oraz (ii) chromosom (bądź chromosomy) dodatkowy (secondary chromosomes), niosący zazwyczaj mniejszą pulę genów metabolizmu podstawowego, skupionych w obrębie stosunkowo niewielkich segmentów DNA. W chromosomach tych spotyka się charakterystyczne dla plazmidów systemy replikacyjne (np. replikony typu *repABC* występujące często w bakteriach z klasy *Alphaproteobacteria*) oraz geny o charakterze adaptacyjnym, które zazwyczaj sprzężone są z plazmidami. Obserwacje te sugerują, że drugi i kolejne chromosomy powstają prawdopodobnie z protoplasty plazmidowego, na skutek wewnątrzkomórkowego transferu genów z pierwszego chromosomu [38]. Wyniki analiz bioinformatycznych sekwencji genomów bakterii należących do rzędu *Rhizobiales* potwierdzają tę hipotezę, wskazując jednocześnie, że plazmidy (protoplasti chromosomów) ulegały wielu zmianom w wyniku kolejnych duplikacji, delecji, a także insercji segmentów DNA wprowadzanych w horyzontalnym transferze genów [3].

Istnieją także inne możliwości pojawiania się dodatkowych chromosomów, takie jak (i) defektywny rozdział chromosomów po replikacji, (ii) defektywny rozdział kointegratów chromosomów utworzonych w wyniku różnego typu rekombinacji – homologicznej, specyficznej wobec miejsca bądź transpozycji oraz (iii) kumulacji mutacji punktowych w chromosomie występującym w komórce w wielu kopiach.

Sugeruje się, że wspólny przodek wielu współczesnych bakterii posiadał zapewne pojedynczy chromosom, zaś stopniowe zmiany, zachodzące na zasadzie przypadkowych rekombinacji, doprowadziły do obserwowanej obecnie różnorodności w architekturze genomów bakteryjnych [42].

Na podkreślenie zasługuje fakt, że bakterie posiadające złożone genomy charakteryzuje bardzo plastyczny metabolizm. Doskonale radzą sobie one w środowiskach otwartych, a także mogą wchodzić w interakcje z innymi organizmami. Przykładowo, bakterie z rodzajów *Agrobacterium* oraz *Brucella* są patogenami (odpowiednio roślin i zwierząt), a z rodzaju *Rhizobium* – symbiontami roślin motylkowych. Występowanie złożonego genomu może być korzystne dla organizmów żyjących

w zmiennych warunkach środowiska, bowiem genom taki szybciej ulega replikacji (co skraca czas generacji), ponadto łatwiej ulega różnym rearanżacjom strukturalnym, co z kolei może pociągać za sobą pojawienie się nowych cech o charakterze adaptacyjnym [13]. Przyjmuje się również, że kolejne chromosomy stają się miejscem akumulacji nowej informacji genetycznej wówczas, gdy pojemność kodująca pierwszego chromosomu osiągnęła maksimum [38].

4. Informacja genetyczna w genomach prokariotycznych

Podstawowy etap projektów genomowych zakłada przeprowadzenie szczegółowych analiz sekwencji nukleotydowych w celu zdefiniowania występującej w nich informacji genetycznej. Konieczna jest zatem (i) identyfikacja i charakterystyka sekwencji kodujących białka i cząsteczki RNA, (ii) wyróżnienie pseudogenów oraz (iii) sekwencji repetytywnych, a także (iv) zdefiniowanie innych składowych, w tym ruchomych elementów genetycznych, z których wiele występuje w formie zintegrowanej z genomowym DNA.

4.1. Sekwencje kodujące białka

Genomy bakterii i archeonów mają kompaktową strukturę, bowiem kodują maksymalną ilość informacji genetycznej w obrębie jak najmniejszego regionu DNA. Zagęszczenie sekwencji kodujących (w przeliczeniu na 1 kpb genomowego DNA) wynosi od 0,8 do 1,2 genu, przy średniej długości pojedynczego genu wynoszącej około 1 kpb [21].

Na podstawie analizy sekwencji nukleotydowych 573 szczepów bakterii stwierdzono, że przeciętny genom składa się z 3 053 genów, wśród których można wyróżnić: (i) geny stanowiące konserwowany rdzeń genomów, (ii) geny nadające cechy charakterystyczne danemu gatunkowi oraz (iii) geny dodatkowe. Geny konserwowanego rdzenia należą do prawie 250 rodzin genów i stanowią około 8% każdego genomu bakteryjnego, niezależnie od gatunku. Kodują one białka zaangażowane w translację, replikację i zachowanie równowagi energetycznej – są to zatem geny metabolizmu podstawowego. Znajdują się one stale pod wpływem presji selekcyjnej, bowiem ich inaktywacja wywołuje efekt letalny. Geny nadające cechy charakterystyczne danemu gatunkowi stanowią około 64% przeciętnego genomu bakteryjnego i zalicza się je do prawie 7 900 rodzin. Z kolei geny zaliczane do ostatniej kategorii (zajmują około 28% genomu) stanowią niezwykle różnorodną pulę, w której spotyka się reprezentantów aż około 140 000 rodzin. Postuluje się, że stanowią one swoisty rezerwuar sekwencji kodujących białka o nieznaney

funkcji, z których wiele może pełnić ważną rolę w specyficznych warunkach środowiska [2, 22].

Kompleksowe analizy z zakresu genomiki porównawczej pozwoliły wyróżnić grupy genów ortologicznych (COG – cluster of ortholog genes), pochodzących od wspólnego przodka [20]. Obecność ortologów w genomach różnych gatunków prokariotów nie jest jednak jednoznaczna z filogenetycznym pokrewieństwem tych organizmów, geny te mogą być bowiem wprowadzane do różnych gospodarzy w drodze horyzontalnego transferu genów. Okazało się, że przeciętnie około 80% genów w danym genomie należy do znanych grup ortologów, a tym samym wykazuje pokrewieństwo z innymi genami [21].

W obrębie genomów bakterii i archeonów zidentyfikowano również wiele potencjalnych genów, które kodują białka niewykazujące znaczącego podobieństwa sekwencji aminokwasowej do poznanych dotąd białek. Białka te określane są w bazach danych jako „hipotetyczne białka o nieznannej funkcji”, a ich potencjalne geny noszą angielskie nazwy ORFans lub ELFs (evil little fellows). Te „hipotetyczne geny” stanowią nawet do 15% całkowitej puli genowej organizmów prokariotycznych i zaliczane są do puli genów rzadko spotykanych, które nie zostały dotąd zaklasyfikowane do żadnej z grup ortologów. Niekiedy postuluje się ich fagowe pochodzenie, co w świetle wciąż dość słabo rozwiniętej genomiki bakteriofagów, mogłoby tłumaczyć brak zidentyfikowanych homologów [21]. Nie można też wykluczyć, że przynajmniej część z nich jest wynikiem błędów popełnianych podczas adnotacji sekwencji nukleotydowych [31].

4.2. Sekwencje kodujące RNA

Analizując długość regionów międzygenowych u *Archaea* i *Bacteria*, uzyskujemy wartości bimodalne, osiagające maksima dla: (i) około 0 pz (głównie w przypadku operonów, gdzie poszczególne geny zachodzą na siebie) oraz (ii) około 100 pz (dla regionów położonych między operonami) [21]. Duże regiony międzygenowe, o wielkości przekraczającej 1 000 pz, kodują tzw. funkcjonalny RNA lub mogą zawierać pseudogeny.

Do funkcjonalnego RNA zaliczamy: (i) transportujący RNA (tRNA), który odpowiada za przyłączanie aminokwasów i ich transport do rybosomów (gdzie zostają one włączane do powstającego, w trakcie translacji, łańcucha polipeptydowego) oraz (ii) mały, regulatorowy RNA (sRNA – small RNA), nazywany także niekodującym RNA (ncRNA – non-coding RNA). Bakteryjne ncRNA mają zwykle od 50 do 400 nukleotydów. Są one kluczowymi regulatorami ekspresji genów zaangażowanych w różne procesy komórkowe, m.in. w adaptację do warunków stresowych, wirulencję, transdukcję sygnałów komórkowych oraz szeroko pojęty metabolizm.

Działają one na różnych etapach ekspresji genów, m.in. poprzez kontrolę procesu: (i) transkrypcji, (ii) stabilizacji i procesowania RNA, (iii) translacji oraz (iv) translacji i degradacji białek [8, 25].

Sekwencje, które są transkrybowane w niekodujące, regulatorowe RNA stanowią średnio 10–15% genomu [46]. W ciągu ostatnich 10 lat, dzięki rozwojowi genomiki porównawczej oraz technik bioinformatycznych, udało się ponad dziesięciokrotnie zwiększyć liczbę identyfikowanych ncRNA. Są to jednak, w większości, cząsteczki wyróżnione jedynie *in silico*, których obecność nie została dotąd zweryfikowana eksperymentalnie. Nie wiadomo również, jaką specyficzną rolę mogą one odgrywać w komórkach. Dobry przykład stanowią genomy szczepów *E. coli* gdzie, stosując metody bioinformatyczne, wykryto ponad 100 cząsteczek ncRNA, z których jedynie dwudziestu udało się przypisać funkcje biologiczne [4].

4.3. Pseudogeny

Pseudogeny to niekodujące regiony DNA powstałe z defektywnych genów, które utraciły swą pierwotną funkcję. Początkowo uważano, że występują one sporadycznie u prokariotów, jednak kompleksowa analiza 64 wybranych genomów bakteryjnych, wykazała obecność ponad 7 000 pseudogenów. W niektórych szczepach mogą one stanowić nawet do 5% wszystkich sekwencji o charakterze genów [26].

Pseudogeny identyfikuje się zarówno w chromosomach, jak i w plazmidach bakterii oraz archeonów. Analizując ich dystrybucję w genomach, nie dostrzega się żadnych prawidłowości. Sytuacja ta wynika z losowego generowania pseudogenów, które mogą powstawać w wyniku (i) powolnej i stopniowej akumulacji mutacji w obrębie poszczególnych genów bądź (ii) szybkiej inaktywacji genów, spowodowanej insercją egzogenego DNA, np. w wyniku transpozycji elementu transpozycyjnego [16].

Szczególne dużo pseudogenów zawierają genomy niektórych wewnątrzkomórkowych patogenów, głównie z rodzajów *Mycobacterium* i *Rickettsia*. Na przykład w *Mycobacterium leprae* TN zidentyfikowano aż 1 115 pseudogenów, stanowiących prawie 30% całego genomu. Jest wielce prawdopodobne, że nagromadzenie pseudogenów w genomach patogenów wewnątrzkomórkowych wynika ze specyficznego środowiska życia tych mikroorganizmów. Bogactwo substancji odżywczych wewnątrz komórek gospodarza oraz ograniczona presja selekcyjna środowiska zewnętrznego sprzyjają akumulacji mutacji w obrębie genów kodujących zbędną, w tych warunkach, informację genetyczną. Zmiany te są szybko utrwalane w toku ewolucji, a z czasem DNA zawierający zmutowane geny zostaje usunięty, co prowadzi do redukcji genomu [16, 26].

4.4. Sekwencje repetytywne

W genomach prokariotycznych występuje wiele sekwencji repetytywnych. Analizy bioinformatyczne wykazały, że stanowią one średnio 6,9% genomu, niekiedy jednak przekraczają nawet 40%. Rekordzistą (42,2%) jest bakteria *Orientia tsutsugamushi* str. Ikeda, w genomie której występuje około 1 000 kopii sekwencji o wielkości 770 pz [44].

Większość powtórzeń w DNA to tzw. sekwencje rozproszone (interspaced repeats), które występują powszechnie. Uważa się, że mogą odgrywać ważną rolę w organizacji i ewolucji genomów prokariotycznych. Wykazano, że w obrębie takich sekwencji dochodzi dość często do różnego typu zdarzeń rekombinacyjnych, które prowadzą do mniej bądź bardziej rozległych zmian w strukturze DNA [44].

Sekwencje repetytywne są zwykle zlokalizowane w regionach międzygenowych, dzięki czemu mogą być transkrybowane wraz z przyległymi genami. Ich obecność niejednokrotnie umożliwia wytworzenie struktur drugorzędowych mRNA, które determinują stabilność powstałych transkryptów. Sekwencje te mogą zatem pełnić również ważne funkcje regulacyjne [9, 32].

Jedną z najlepiej scharakteryzowanych klas sekwencji repetytywnych są MITE (miniature inverted-repeat transposable element), które zaliczane są do grupy nieautonomicznych elementów transpozycyjnych. Ich wielkość zawiera się w przedziale od 70 do 300 pz. MITE to prawdopodobnie zredukowane formy sekwencji insercyjnych, których mobilność jest determinowana przez dostarczone *in trans* transpozazy, kodowane przez inne, funkcjonalne elementy transpozycyjne. Występują one powszechnie u wielu prokariotów, w liczbie od kilkudziesięciu do kilkuset kopii, zaś w niektórych przypadkach mogą stanowić nawet około 2% genomu gospodarza [10]. Elementy typu MITE oprócz pełnienia funkcji regulacyjnych mogą także odgrywać istotną rolę w ewolucji białek oraz regulacyjnego RNA [11].

Powtórzone sekwencje są także elementem składowym systemów CRISPR (clustered, regularly interspaced short palindromic repeats), zawierających regularnie rozmieszczone, krótkie sekwencje palindromowe. Systemy te to największe zgrupowania sekwencji repetytywnych, jakie poznano dotąd w genomach prokariotycznych. Występują one aż w około 90% genomów *Archaea* oraz w ponad 40% genomów bakteryjnych. Regiony te mogą zawierać od 2 do 249 powtórzeń prostych, o długości w zakresie od 24 do 47 pz, które są rozdzielone regionami łącznikowymi (26 pz – 72 pz). Wraz z przyległymi obszarami kodującymi CAS (CRISPR associated sequences) oraz tzw. sekwencjami liderowymi, tworzą one systemy CRISPR-CAS, które pełnią funkcję analogiczną do systemu interferencyjnego RNA (siRNA) *Eucaryota* [27, 39].

Za sekwencje repetytywne można także uznać ruchome elementy genetyczne oraz genomy profagów, które występują w większości genomów prokariotycznych, niekiedy w wielu kopiach.

4.5. Ruchome elementy genetyczne zintegrowane z genomowym DNA

Ruchome elementy genetyczne (MGE – mobile genetic elements) są ważnym składnikiem genomów prokariotycznych. Występują one powszechnie zarówno u bakterii jak i archeonów. Niekiedy mogą one stanowić nawet 40% ogólnej puli informacji genetycznej zawartej w danym szczepie.

Do grona ruchomych elementów genetycznych zalicza się, oprócz plazmidów i genomów bakteriofagów: (i) elementy transpozycyjne (sekwencje insercyjne i ranspozony), (ii) elementy integrujące z DNA (w tym elementy koniugacyjne i mobilizowalne, odpowiednio, ICE – integrative and conjugative elements i IME – integrative and mobilizable elements) oraz (iii) introny i introny. Elementy te (z wyjątkiem plazmidów) występują w formie zintegrowanej z genomowym DNA.

Niektóre z nich (plazmidy, ICE, IME) mogą być przekazywane drogą koniugacji nawet do odległych filogenetycznie gatunków – są to zatem naturalne wektory genetyczne odgrywające zasadniczą rolę w horyzontalnym transferze genów. Należy również podkreślić, że niemal wszystkie typy MGE kodują różnego typu rekombinazy, których aktywność może prowadzić do licznych zmian w strukturze materiału genetycznego gospodarza. Z tej przyczyny uważane są one za jeden z najbardziej rekombinogennych czynników, który decyduje, w dużej mierze, o architekturze genomu oraz odgrywa rolę istotnego czynnika zwiększającego zmienność, a tym samym tempo ewolucji swoich gospodarzy [37].

5. Minimalny genom i pangenu

Wyniki dogłębnych analiz genomowych skłaniają do przypuszczeń, że minimalny genom bakteryjny może składać się, teoretycznie, z zaledwie 206 genów, których produkty umożliwiają zachowanie podstawowych funkcji życiowych mikroorganizmu, takich jak: (i) replikacja DNA, (ii) naprawa DNA, (iii) transkrypcja i translacja, (iv) obróbka potranslacyjna, (v) sekrecja i procesowanie białek, (vi) podział komórki, (vii) transport substratów, (viii) pozyskiwanie energii, (ix) odpowiedź na stresowe warunki środowiska oraz (x) synteza aminokwasów, lipidów, nukleotydów i niektórych kofaktorów [18].

W 1999 roku stworzono komputerowy model komórki (projekt E-CELL; <http://www.e-cell.org/ecell/>), której minimalny genom zawierał jedynie 127 genów. Teoretycznie jest to wystarczająca ilość informacji

genetycznej, która powinna zapewnić takiej wirtualnej komórce zdolność do samodzielnego funkcjonowania. Chociaż komórka taka, jak zakładano, mogłaby osiągnąć metaboliczną homeostazę, nie byłaby ona jednak zdolna do reprodukcji i ewolucji. Większość genów wykozystanych w tym projekcie pochodziła z *Mycoplasma genitalium*, której chromosom, o wielkości 580 kbp, koduje 482 białka [43].

Dla porównania – w obrębie jednego z najmniejszych, poznanych dotychczas genomów, należącego do bakterii „*Candidatus Carsonella ruddii*” PV, kodowane są zaledwie 182 białka, a całkowita liczba genów wynosi 213. Należy jednak pamiętać, że bakteria ta jest obligatoryjnym symbiontem, który nie jest zdolny do przeżycia poza organizmem swojego gospodarza – wykorzystuje zatem jego informację genetyczną, kompensując własne „braki” [30]. Z kolei „*Candidatus Pelagibacter ubique*” HTCC 1062, który posiada najmniejszy genom wśród bakterii swobodnie żyjących (1389 genów), koduje aż 1354 białka. Zestawienie tych danych uświadamia, że chociaż minimalny, uniwersalny rdzeń genomu prokariotycznego może być rzeczywiście niewielki, to jednak mikroorganizmy swobodnie żyjące, ze względu na zmienne warunki środowiska oraz konieczność przeprowadzania złożonych procesów metabolicznych muszą kodować, co najmniej, kilkaset białek [7].

W ostatnich latach rozpoczęto również bardziej kompleksowe analizy, których celem było zdefiniowanie pangenu (nazywanego również supragenomem), stanowiącego całkowitą pulę genów obecnych w genomach bakterii zaliczanych do danej grupy taksonomicznej (zwykle gatunku) [22, 40]. Analizy te wykazały, że pangeny większości analizowanych bakterii są „otwarte”. Oznacza to, że kolejne poznawane szczepy niosą zestawy unikatowych genów, które wzbogacają pangenu danego gatunku [2]. Sugeruje się, że pangeny nielicznych bakterii są „zamknięte”, a więc zidentyfikowano ich całkowitą pulę genową. Wykazano, na przykład, że dla zdefiniowania pełnego pangenu *Bacillus anthracis* wystarczy wiedza o informacji genetycznej niesionej przez zaledwie cztery szczepy tej bakterii. Przypuszcza się jednak, że *B. anthracis* stanowi odmianę klonalną *Bacillus cereus*. Gdyby to stwierdzenie było prawdziwe, pula genów tego gatunku zostałaby znacznie zwiększona. Z „zamkniętymi” pangenomami prawdopodobnie mamy również do czynienia w przypadku: *Staphylococcus aureus*, *Streptococcus pyogenes* oraz *Ureoplasma ureolyticum* [28, 41]. Wprowadzenie pojęcia „zamkniętego” pangenu budzi jednak pewne wątpliwości, zakłada bowiem występowanie w genomach ograniczonego zestawu genów, co nie jest zgodne z obserwacjami wskazującymi na powszechność i zakres horyzontalnego transferu genów.

Analizy pangenu umożliwiły również zdefiniowanie zestawu genów stanowiących konserwowany

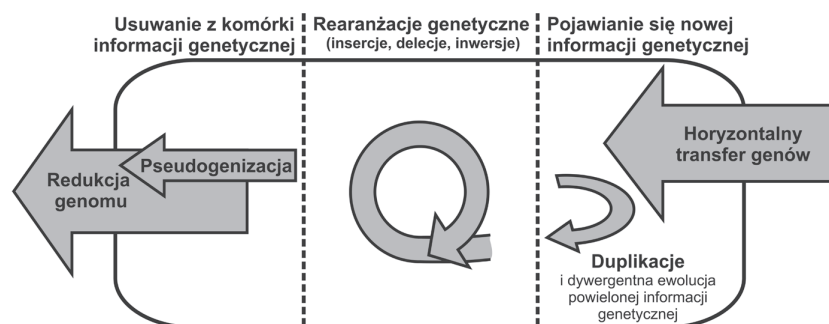
rdzeń genomu, który jest specyficzny dla danego gatunku. Wykazano, na przykład, że dla ośmiu analizowanych szczepów *Streptococcus agalactiae* rdzeń taki składa się z 1806 genów, co stanowi około 80% każdego z badanych genomów [40]. Analizując z kolei genomy 17 szczepów *E. coli* stwierdzono obecność 2344 genów wspólnych, co stanowi średnio mniej niż 50% pojedynczego genomu i, co ciekawe, jedynie około 17% pangenu tego gatunku [34]. Przytoczone dane dowodzą ogromnej zmienności genetycznej szczepów w obrębie jednego gatunku bakterii i wskazują na rolę HGT w tym procesie.

6. Ewolucja genomów prokariotycznych

Przytoczone w tej pracy przykłady wskazują, że genomy prokariotyczne nie są statycznymi, monolitycznymi strukturami, lecz podlegają licznym rearranżacjom genetycznym. Plastyczność i ewolucja genomów jest przede wszystkim wypadkową trzech procesów: (i) horyzontalnego transferu genów, (ii) rekombinacji prowadzących do różnego typu rearranżacji genetycznych oraz (iii) redukcji, która prowadzi do zmniejszenia rozmiaru genomów.

Powszechnie przyjmuje się, że horyzontalny transfer genów jest siłą napędową ewolucji bakterii. Kompensuje on bakteriom brak możliwości rozmnażania płciowego, które u innych organizmów warunkuje zmienność genetyczną. Istnieje wiele dróg wprowadzania egzogennej DNA do komórek bakteryjnych, takich jak koniugacja, transdukcja, transformacja, naturalna elektroporacja (indukowana wyładowaniami atmosferycznymi) czy też transfer za pośrednictwem cząstek GTA (gene transfer agents) [12, 24]. W wyniku tych procesów, teoretycznie dowolny segment genomowego DNA może zostać przeniesiony nawet do odległych filogenetycznie gatunków bakterii, co pozwala zdefiniować informację genetyczną tych mikroorganizmów jako ogromną, wspólną pulę różnorodnych genów. Wprowadzony do komórki DNA zawiera niekiedy kompletne operony bądź większe zgrupowania genów chromosomowych o wyspecjalizowanych funkcjach. Często są to geny metabolizmu podstawowego, bądź geny o charakterze adaptacyjnym, których obecność może zmienić profil metaboliczny nowego gospodarza. Tego typu egzogenną informację określa się często jako wyspy genomowe (GI lub GEI – genomic islands).

W wyniku HGT możliwe jest zatem uzyskanie *en bloc* dużej ilości informacji genetycznej. Takie skokowe zmiany są niewątpliwie ważnym czynnikiem przyspieszającym znacznie tempo ewolucji bakterii. Alternatywnym, lecz bardzo długotrwałym procesem prowadzącym do pojawienia się w komórce nowej informacji genetycznej jest duplikacja segmentów DNA,



Rys. 4. Procesy wpływające na strukturę i ewolucję genomów prokariotycznych, wg [29] (zmodyfikowany)

a następnie ich dywergentna ewolucja, która może doprowadzić do powstania genów, kodujących białka o nowych właściwościach (rys. 4).

Poczynione obserwacje wskazują, że wielkość genomu jest bardzo często skorelowana ze środowiskiem bytowania mikroorganizmu. Duże znaczenie odgrywa tu m.in. dostępność związków odżywczych oraz stabilność warunków w danej niszy ekologicznej. Zasiedlanie zmiennych środowisk, w których bakteria narażona jest na wpływ czynników stresowych i musi współzawodniczyć z innymi mikroorganizmami, promuje utrzymanie dużej różnorodności metabolicznej. Konieczne jest tym samym zachowanie dużej puli informacji genetycznej. W warunkach bardziej stabilnych ta informacja może stać się zbędna – ulega więc redukcji, co pociąga za sobą zmniejszenie wielkości genomu.

Redukcji podlegają genomy wszystkich organizmów prokariotycznych. Według hipotezy zakładającej „optymalizację genomu” (streamlining hypothesis), zjawisko to polega na eliminacji informacji genetycznej o małej wartości adaptacyjnej do danych warunków środowiska, w celu ograniczenia obciążenia metabolicznego wynikającego z utrzymywania i replikacji „mało wartościowego” DNA. Tęgo typu selekcja „oczyszczająca” odgrywa bardzo ważną rolę i jest szczególnie widoczna w przypadku genomów prokariotycznych endosymbiontów i pasożytów wewnątrzkomórkowych. Wzrastająca liczba pseudogenów jest pierwszym symptomem wskazującym na możliwość utraty części informacji genetycznej, co w konsekwencji prowadzi do redukcji genomu.

Należy też pamiętać, że informacja genetyczna utracona w wyniku delecji, potencjalnie może być ponownie odzyskana za sprawą HGT (z tej przyczyny niektórzy przyrównują genomy do palimpsestów – średniowiecznych manuskryptów, w których pozornie usunięte teksty, z czasem pojawiały się ponownie i stawały się w pełni czytelne). Przykładem jest transpozon występujący w obrębie wyspy genomowej PAI II *E. coli* 536, który niesie homologi genów występujących w operonie *pdu* *Salmonella enterica* sv. Typhimurium (operon warunkuje rozkład 1,2-propanodiolu). W świetle

przeprowadzonych analiz wydaje się, że operon ten został utracony przez protoplastę rodzajów *Salmonella* i *Escherichia*, a następnie odzyskany przez bakterie z rodzaju *Salmonella* oraz szczep *E. coli* 536 na drożdże niezależnych zdarzeń HGT [5, 23]. Obserwacje te potwierdzają założenie, że żadna informacja genetyczna nie musi być tracona bezpowrotnie.

7. Podsumowanie

Wyniki analiz genomicznych przyniosły wiele cennych danych świadczących o ogromnej zmienności genomów prokariotycznych. Tempo ewolucji prokariotów jest bardzo wysokie, co wynika, w dużej mierze, z szerokiego zakresu horyzontalnego transferu genów oraz częstych mutacji i reanżacji genetycznych, które prowadzą do dywersyfikacji pierwotnie homogennej informacji genetycznej. Proces kształtowania i ewolucji genomów jest zatem wypadkową ciągłych zmian, spowodowanych kolejnymi insercjami, rearanżacjami i delecjami w obrębie DNA (rys. 4). Poziom tych zmian jest dość zbalansowany, co powoduje, że genomy nie rozrastają się do monstrualnych rozmiarów, a ich wielkość, w obrębie danego gatunku, nie jest zbyt zróżnicowana.

Szczegółowe analizy sekwencji nukleotydowych genomów przynoszą wiele informacji na temat potencjału metabolicznego danego mikroorganizmu, jak również umożliwiają wyciągnięcie bardziej ogólnych wniosków na temat plastyczności genomu, jego ewolucji oraz zakresu zdarzeń horyzontalnego transferu genów. Ze względu na ogromne zróżnicowanie bakterii, nie jesteśmy jednak w stanie stworzyć uniwersalnego modelu typowego genomu prokariotycznego. Możliwe jest jednak określenie, które z jego elementów są absolutnie niezbędne do prawidłowego funkcjonowania komórki. Analizy takie są doskonałym punktem wyjścia dla biologii syntetycznej, której największym osiągnięciem jest, jak dotąd, stworzenie „sztucznej” bakterii – *Mycoplasma mycoides* JCVI-syn1.0, której genom został w całości zsintetyzowany chemicznie w laboratorium [17].

Zmienność genomów jest również ważnym przyczynkiem do dyskusji na temat taksonomii prokariotów, a w szczególności kryteriów stosowanych przy definiowaniu gatunków tych organizmów. Nie ulega wątpliwości, że klasyfikacja taka powinna być oparta wyłącznie na wynikach kompleksowych analiz z zakresu genomiki porównawczej. Dynamiczny rozwój wysoko-przepustowych technik sekwencjonowania DNA, pozwala przypuszczać, że może to nastąpić już w bardzo niedalekiej przyszłości.

Podziękowania

Praca ta powstała w trakcie realizacji dwóch projektów badawczych finansowanych przez MNiSW (IP2010 008670 i N N303 816340), dotyczących analiz genomicznych bakterii z rodzajów *Paracoccus* (*Alphaproteobacteria*) oraz *Psychrobacter* (*Gamma proteobacteria*). Autorzy dziękują dr Jadwidze Baj za krytyczne uwagi na temat manuskryptu.

Literatura

- Baliga N.S., Ng W.V. i wsp.: Genome sequence of *Haloarcula marismortui*: a halophilic archaeon from the Dead Sea. *Genome Res.* **14**, 2221–2234 (2004) (praca jest dziełem 15 autorów)
- Bentley S.: Sequencing the species pan-genome. *Nat. Rev. Microbiol.* **7**, 258–259 (2009)
- Boussau B., Karlberg E.O., Frank A.C., Legault B.A., Andersson S.G.: Computational inference of scenarios for alpha-proteobacterial genome evolution. *Proc. Natl. Acad. Sci. USA*, **101**, 9722–9727 (2004)
- Brantl S.: Bacterial chromosome-encoded small regulatory RNAs. *Future Microbiol.* **4**, 85–103 (2009)
- Brzuszkiewicz E., Dobrindt U. i wsp.: How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic *Escherichia coli* strains. *Proc. Natl. Acad. Sci. USA*, **103**, 12879–84 (2006) (praca jest dziełem 13 autorów)
- Byram R., Stewart P.E., Rosa P.: The essential nature of the ubiquitous 26-kilobase circular replicon of *Borrelia burgdorferi*. *J. Bacteriol.* **186**, 3561–3569 (2004)
- Carbone A.: Computational prediction of genomic functional cores specific to different microbes. *J. Mol. Evol.* **63**, 733–746 (2006)
- Chang T.-H., Wub L.-C., Lin J.-H., Huang H.-D., Liu B.-J., Cheng K.-F., Horng J.-T.: Prediction of small non-coding RNA in bacterial genomes using support vector machines. *Expert Syst. Appl.* **37**, 5549–5557 (2010)
- Cozzuto L., Petrillo M., Silvestro G., Di Nocera P.P., Paoletta G.: Systematic identification of stem-loop containing sequence families in bacterial genomes. *BMC Genomics*, **9**, 20 (2008)
- De Gregorio E., Abrescia C., Carlomagno M.S., Di Nocera P.P.: Ribonuclease III-mediated processing of specific *Neisseria meningitidis* mRNAs. *Biochem. J.* **374**, 799–805 (2003)
- Delihias N.: Small mobile sequences in bacteria display diverse structure/function motifs. *Mol. Microbiol.* **67**, 475–481 (2008)
- Demaneche S., Bertolla F., Buret F., Nalin R., Sailland A., Auriol P., Vogel T.M., Simonet P.: Laboratory-scale evidence for lightning-mediated gene transfer in soil. *Appl. Environ. Microbiol.* **67**, 3440–3444 (2001)
- Egan E.S., Fogel M.A., Waldorf M.K.: Divided genomes: negotiating the cell cycle in prokaryotes with multiple chromosomes. *Mol. Microbiol.* **56**, 1129–1138 (2005)
- Fleischmann R.D., Venter J.C. i wsp.: Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512 (1995) (praca jest dziełem 15 autorów)
- Fraser C.M., Venter J.C. i wsp.: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, **390**, 580–586 (1997) (praca jest dziełem 40 autorów)
- Fuxelius H.H., Darby A.C., Cho N.H., Andersson S.G.: Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction. *Genome Biol.* **9**, R42 (2008)
- Gibson D.G., Venter J.C. i wsp.: Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, **329**, 52–56 (2010) (praca jest dziełem 24 autorów)
- Gil R., Silva F.J., Pereto J., Moya A.: Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**, 518–37 (2004)
- Giovannoni S.J., Mathur E.J. i wsp.: Genome streamlining in a cosmopolitan oceanic bacterium. *Science*, **309**, 1242–1245 (2005) (praca jest dziełem 14 autorów)
- Koonin E.V.: Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **39**, 309–338 (2005)
- Koonin E.V., Wolf Y.I.: Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. *Nucleic Acids Res.* **36**, 6688–6719 (2008)
- Lapierre P., Gogarten J.P.: Estimating the size of the bacterial pan-genome. *Trends Genet.* **25**, 107–110 (2009)
- Lawrence J.G., Roth J.R.: Evolution of coenzyme B12 synthesis among enteric bacteria: evidence for loss and reacquisition of a multigene complex. *Genetics*, **142**, 11–24 (1996)
- Leung M.M., Florizone S.M., Taylor T.A., Lang A.S., Beatty J.T.: The gene transfer agent of *Rhodobacter capsulatus*. *Adv. Exp. Med. Biol.* **675**, 253–264 (2010)
- Liu J.M., Camilli A.: A broadening world of bacterial small RNAs. *Curr. Opin. Microbiol.* **13**, 18–23 (2010)
- Liu Y., Harrison P.M., Kunin V., Gerstein M.: Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* **5**, R64 (2004)
- Marraffini L.A., Sontheimer E.J.: CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science*, **322**, 1843–1845 (2008)
- Medini D., Donati C., Tettelin H., Massignani V., Rappuoli R.: The microbial pan-genome. *Curr. Opin. Genet. Dev.* **15**, 589–594 (2005)
- Mira A., Ochman H., Moran N.A.: Deletional bias and the evolution of bacterial genomes. *Trends Genet.* **17**, 589–596 (2001)
- Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H.E., Moran N.A., Hattori M.: The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science*, **314**, 267 (2006)
- Ochman H.: Distinguishing the ORFs from the ELFs: short bacterial genes and the annotation of genomes. *Trends Genet.* **18**, 335–337 (2002)
- Ogata H., Audic S., Barbe V., Artiguenave F., Fournier P.E., Raoult D., Claverie J.M.: Selfish DNA in protein-coding genes of *Rickettsia*. *Science*, **290**, 347–350 (2000)
- Podar M., Stetter K.O. i wsp.: A genomic analysis of the archaeal system *Ignicoccus hospitalis*-*Nanoarchaeum equitans*. *Genome Biol.* **9**, R158 (2008) (praca jest dziełem 27 autorów)
- Rasko D.A., Ravel J. i wsp.: The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. *J. Bacteriol.* **190**, 6881–6893 (2008) (praca jest dziełem 34 autorów)
- Schmeisser C., Streit W.R. i wsp.: *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Appl.*

- Environ. Microbiol.* **75**, 4035–4045 (2009) (praca jest dziełem 18 autorów)
36. Schneiker S., Muller R. i wsp.: Complete genome sequence of the myxobacterium *Sorangium cellulosum*. *Nat. Biotechnol.* **25**, 1281–1289 (2007) (praca jest dziełem 59 autorów)
 37. Siefert J.L.: Defining the mobilome. *Methods Mol. Biol.* **532**, 13–27 (2009)
 38. Slater S.C., Wood D.W. i wsp.: Genome sequences of three *Agrobacterium* biovars help elucidate the evolution of multichromosome genomes in bacteria. *J. Bacteriol.* **191**, 2501–2511 (2009) (praca jest dziełem 36 autorów)
 39. Sorek R., Kunin V., Hugenholtz P.: CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**, 181–186 (2008)
 40. Tettelin H., Fraser C.M. i wsp.: Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA*, **102**, 13950–13955 (2005) (praca jest dziełem 46 autorów)
 41. Tettelin H., Riley D., Cattuto C., Medini D.: Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477 (2008)
 42. Teyssier C., Marchandin H., Jumas-Bilak E.: The genome of *Alphaproteobacteria*: complexity, reduction, diversity and fluidity. *Can. J. Microbiol.* **50**, 383–396 (2004)
 43. Tomita M., Hutchison C.A. 3rd i wsp.: E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84 (1999) (praca jest dziełem 11 autorów)
 44. Treangen T.J., Abraham A.L., Touchon M., Rocha E.P.: Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol. Rev.* **33**, 539–571 (2009)
 45. Waters E., Noordewier M. i wsp.: The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. USA*, **100**, 12984–12988 (2003) (praca jest dziełem 22 autorów)
 46. Westhof E.: The amazing world of bacterial structured RNAs. *Genome Biol.* **11**, 108 (2010)