

WIESŁAW BABIK

*Instytut Nauk o Środowisku Uniwersytetu Jagiellońskiego
Gronostajowa 7, 30-387 Kraków
E-mail: wieslaw.babik@uj.edu.pl*

EWOLUCJA GENOMÓW I POWSTAWANIE NOWYCH GENÓW

W tym artykule chciałbym zająć się dwoma zagadnieniami: najpierw dokonam krótkiego przeglądu wielkości, organizacji oraz głównych trendów ewolucji genomów organizmów komórkowych, następnie przedstawię najważniejsze procesy i mechanizmy ewolucyjne prowadzące do powstawania nowych genów.

Organizmy o budowie komórkowej zaliczamy do trzech wielkich domen życia: bakterii, archeowców i eukariotów. Jednak bardziej tradycyjny podział na organizmy prokariotyczne i eukariotyczne dobrze oddaje zróżnicowanie charakteru komórek i genomów organizmów żywych (KOONIN i WOLF 2008). Bakterie i archeowce, razem określane mianem prokariotów, oddzieliły się od siebie bardzo dawno, na pewno ponad dwa, a prawdopodobnie ponad trzy miliardy lat temu (www.timetree.org). Mają one proste komórki i stosunkowo niewielkie genomy, odmienne od komórek eukariotycznych. Wielkość genomów tradycyjnie mierzy się w

pikogramach ($1 \text{ pg} = 10^{-12} \text{ g}$); 1 pg odpowiada 978 mln par zasad (pz) DNA (978 Mb). Liczbę par zasad określamy wywodzącymi się z języka angielskiego skrótami: $1 \text{ kb} = 1 \text{ tys}$ (10^3) pz, $1 \text{ Mb} = 1 \text{ mln}$ (10^6) pz oraz $1 \text{ Gb} = 1 \text{ mld}$ (10^9) pz. Zakres rozmiarów genomów prokariotycznych obejmuje dwa rzędy wielkości, przy czym zarówno najmniejsze (0,16 Mb), jak i największe (13 Mb) genomy występują u bakterii, zróżnicowanie wielkości genomów archeowców jest jeszcze mniejsze (od ok. 0,5 do 5 Mb). Trzeba tutaj zaznaczyć, iż najmniejsze genomy bakteryjne spotykamy wyłącznie u pasożytów wewnątrzkomórkowych, które wykorzystują wiele procesów metabolicznych komórek-gospodarzy. Najmniejsze genomy wolnożyjących bakterii mają około 1,3 Mb. Wielkość genomów eukariotycznych różni się natomiast o pięć rzędów wielkości, ponad dwieście tys. razy (ich rozmiary wahają się od ok. 2,5 Mb do ok. 700 000 Mb)! (<http://www.genomesize.com/>).

DROGI EWOLUCJI GENOMÓW BAKTERII I ARCHEOWCÓW

Sekwencjonowanie genomów prokariotycznych praktykuje się od początku lat 90., obecnie znane są sekwencje genomów ponad dwu tysięcy bakterii i archeowców (<http://www.ncbi.nlm.nih.gov/genomes/static/gpstat.html>). Dzięki postępowi technicznemu, metodom sekwencjonowania DNA nowej generacji oraz rozwojowi narzędzi bioinformatycznych, sekwencję genomu bakteryjnego może obecnie uzyskać i przeanalizować jedna oso-

ba w ciągu kilku dni kosztem paru tysięcy euro. Poza niewielkimi rozmiarami genomu, czynnikiem ogromnie ułatwiającym sekwencjonowanie genomów prokariotycznych jest minimalna ilość powtarzalnego DNA, to znaczy długich, liczących nawet wiele tysięcy par zasad bloków składających się z licznych kopii identycznych lub prawie identycznych sekwencji. Powtarzalny DNA, występujący powszechnie w genomach eukariotycznych,

utrudnia nie tyle samo sekwencjonowanie, co późniejsze składanie zsekwencjonowanych fragmentów w pełną sekwencję, gdyż trudno ustalić, ile razy dane powtórzenie występuje w genomie. Ponieważ sekwencjonowanie genomów prokariotycznych jest tak łatwe, nagromadziła się ogromna ilość danych porównawczych, co pozwala na szczegółową analizę trendów ewolucyjnych w genomach tych organizmów (KOONIN i WOLF 2008).

Genomy bakterii i archeowców zawierają przede wszystkim kodujący DNA, to znaczy DNA kodujący białka, funkcjonalne RNA, takie jak rybosomalne (rRNA) czy transferowe (tRNA) oraz niewielką ilość DNA niekodującego w ścisłym tego słowa znaczeniu lecz zaangażowanego w regulację replikacji czy transkrypcji, jak np. sekwencje promotorne. Ilość sekwencji niefunkcjonalnych jest minimalna, pseudogeny (niefunkcjonalne kopie genów, np. inaktywowane przez mutacje) występują bardzo rzadko, a gdy się pojawiają, są szybko z genomów usuwane, niewiele jest ruchomych elementów genetycznych, introny są niezwykle rzadkie i odmienne od intronów spotykanych powszechnie u organizmów eukariotycznych. Wiele genów, szczególnie takich, których produkty stanowią elementy jednego szlaku metabolicznego, występuje w postaci operonów, czyli ciągów genów ułożonych jeden za drugim, podlegających wspólnej regulacji.

Porównanie pasożytniczych gatunków bakterii z blisko spokrewnionymi wolnożyjącymi formami wykazało brak wielu genów w genomach pasożytów. Jest to wynikiem szybkiej utraty takich genów, które przestają być niezbędne, gdyż ich produkty spełniają funkcje niepotrzebne w związku z pasożytniczym trybem życia, lub też takich, których funkcje spełniają białka gospodarza (KOONIN i WOLF 2008).

Kolejnym zaskakującym odkryciem było stwierdzenie, w miarę jak gromadzono sekwencje genomów kolejnych gatunków lub szczepów (definicja gatunku bakteryjnego jest nawet bardziej kontrowersyjna niż w przypadku roślin i zwierząt) (ACHTMAN i WAGNER 2008, FRASER i współaut. 2009), iż nawet blisko spokrewnione bakterie, jak szczepy *Escherichia coli*, różnią się między sobą dramatycznie składem genów. Wśród około 6000 genów obecnych w komórkach 7 szczepów *E. coli*, wspólnych dla porównywanych szczepów jest niecałe 3000 (ABBY i DAUBIN 2007). Obserwacja ta doprowadziła do powstania koncepcji pan-genomu bakteryjnego,

który obejmuje obecne we wszystkich szczepach geny tzw. genomu rdzeniowego (ang. core genome), oraz dodatkowe geny genomu opcjonalnego (ang. dispensable genome), obecne tylko w niektórych szczepach (MEDINI i współaut. 2005). Wydawało się, iż porównanie wielu genomów prokariotycznych umożliwi zidentyfikowanie minimalnego zestawu genów niezbędnych do funkcjonowania żywej komórki. W miarę jednak jak liczba sekwencjonowanych genomów bakteryjnych rosła, liczba genów znajdujących we wszystkich dramatycznie spadała, ulegając redukcji do zaledwie kilkudziesięciu (LAWRENCE i HENDRICKSON 2005). Jest to wynikiem faktu, że chociaż większość, nawet ogromna większość zsekwencjonowanych genomów zawiera dany gen, można znaleźć jeden lub kilka genomów tego genu pozbawionych. Dalsza analiza pan-genomu pozwoliła na wyróżnienie trzech klas genów: a) rozszerzonego rdzenia (ang. extended core), których brak jedynie w znikomej części genomów, b) kodujących cechy obecne w wielu genomach (ang. character genes) oraz c) genów puli dodatkowej (ang. accessory pool), obecnych tylko w nielicznych genomach (LAPIERRE i GOGARTEN 2009). Minimalną liczbę genów dla heterotroficznej komórki żyjącej na bogatej pożywce szacuje się na około 250, a najmniejsze znane genomy wolnożyjących bakterii zawierają około 1100 genów (KOONIN i WOLF 2008).

Kolejną obserwacją, jaką poczyniono, porównując kompletne genomy bakteryjne, było to, że wzajemne ułożenie genów w genomie zmienia się bardzo dynamicznie, o wiele szybciej niż ich sekwencje aminokwasowe, co wskazuje, iż nacisk doboru naturalnego na utrzymanie sekwencji aminokwasów kodowanych przez dany gen jest znacznie silniejszy niż na utrzymanie kolejności genów w genomie. Od tej reguły są jednak pewne wyjątki, np. operony lub białka rybosomalne, gdzie układ genów jest zakonserwowany ewolucyjnie. Prawdopodobnie jest to spowodowane wymaganiami regulacji transkrypcji i translacji. W związku z brakiem rozdziału transkrypcji i translacji u prokariotów, regulacja tych procesów może pozostawiać mniej pola manewru niż u eukariotów, u których transkrypcja i translacja zachodzą w różnym czasie i w oddzielnych przedziałach komórkowych.

Procesem, który w ogromnym stopniu decyduje o kształcie genomów prokariotycznych, jest horyzontalny (poziomy) transfer

genów (HTG) (OCHMAN i współaut. 2000, THOMAS i NIELSEN 2005). Mianem tym określamy przekazywanie fragmentów DNA nie poprzez zwyczajne dziedziczenie przodek-potomek, polegające na replikacji materiału genetycznego i przekazywaniu go komórkom potomnym, nazywane również przekazem pionowym, lecz nabywanie DNA pochodzącego od innych organizmów, nawet daleko spokrewnionych. Istnieją trzy podstawowe mechanizmy horyzontalnego przekazu DNA między komórkami.

1. Transformacja polega na pobieraniu przez komórkę prokariotyczną nagiego DNA obecnego w środowisku; pobrany DNA może następnie ulec integracji do genomu gospodarza, lub też, jeżeli jest to np. plazmid, po dostaniu się do komórki może „żyć własnym życiem”.

2. W procesie transdukcji uczestniczy wektor biologiczny, zazwyczaj bakteriofag, pakujący do swojej otoczki nie tylko własne geny, ale też część genomu gospodarza, który następnie może zostać zintegrowany do genomu innej bakterii, zakażanej przez faga.

3. Wreszcie możliwe jest przekazywanie materiału genetycznego między bakteriami w procesie koniugacji, warunkowanym przez plazmidy koniugacyjne.

Poszczególne grupy bakterii różnią się zdolnością do HTG, jednak proces ten jest powszechny u prokariotów jako całości. Okazało się, że nie wszystkie geny są jednakowo „podatne” na poziomy transfer. Geny, które oddziałują z wieloma innymi genami, oraz zaangażowane w translację podlegają HTG rzadziej, a geny, których produkty obecne są na powierzchni komórki, geny odpowiedzialne za procesy metaboliczne lub zaangażowane w patogenność ulegają HTG częściej. Stwierdzono, że w wyniku horyzontalnego transferu mogą być przenoszone znaczne fragmenty genomu, wielkości kilkudziesięciu kb, zawierające wiele genów i tworzące tzw. „wyspy genomowe”, np. wyspy patogenności czy symbiozy (LAWRENCE i HENDRICKSON 2005). Porównanie między dwoma szczepami *E. coli*: patogennym O157:H7 i laboratoryjnym K12, wykazało, że patogenny szczep zawierał 1387 dodatkowych genów rozmieszczonych w kilku grupach – wyspach o różnej wielkości. Horyzontalny transfer genów prowadzący do powstania wysp patogenności wydaje się być związany z procesem transdukcji fagowej.

EWOLUCJA GENOMÓW EUKARIOTYCZNYCH

Genomy eukariotyczne różnią się znacznie od prokariotycznych swoją strukturą, obecnością chromosomów zamkniętych w jądrze komórkowym, powszechnym występowaniem intronów, innym sposobem upakowania DNA i wieloma innymi cechami, których omówienie można znaleźć w podręcznikach (np. BROWN 2009). Genomy eukariotyczne są również zazwyczaj większe od prokariotycznych, lecz zakresy wielkości zachodzą na siebie dość znacznie: najmniejszy genom eukariotyczny jest około pięciu razy mniejszy od największego prokariotycznego. Uderzające w porównaniu z prokariotami jest ogromne zróżnicowanie wielkości genomów eukariotycznych, obejmujące pięć rzędów wielkości. Co więcej, już w latach 60. XX w. zauważono, iż ilość DNA w jądrze komórkowym jest tylko w umiarkowanym stopniu skorelowana ze złożonością organizmów. Ogromne genomy o wielkości kilkudziesięciu-kilkuset Gb spotykamy u wielu jednokomórkowych eukariotów o stosunkowo prostej budowie, a także u niektórych skorupiaków, płazów

ogoniastych i ryb dwudysznych. Istnieją natomiast ryby czy ptaki, a więc organizmy o wysokiej w powszechnym pojęciu złożoności, które mają niewielkie genomy o wielkości poniżej 1 Gb. Ten brak wyraźnej korelacji między wielkością genomu a złożonością organizmu nazwano paradoksem wartości C (C określa ilość DNA w jądrze haploidalnej komórki). Mechanistyczne wyjaśnienie znaleziono stosunkowo szybko. Badania przeprowadzone w końcu lat 60. XX w. doprowadziły do stwierdzenia, że paradoks wartości C jest wynikiem zróżnicowania ilości niekodującego DNA, to znaczy takiego, który nie koduje białek lub funkcjonalnych RNA. Do tej klasy DNA zaliczamy zarówno introny, znajdujące się w różnej obfitości w genomach wszystkich eukariotów, jak również międzygenowy DNA, składający się w znacznym stopniu z sekwencji powtarzalnych. Istnieją doniesienia o transkrypcji ponad 60% nawet tak dużego (2,5Gb) genomu jak myszy (CARNINCI i współaut. 2005), a w stosunkowo niewielkim (100 Mb) genomie *Drosophila melanogaster* więk-

szczość niekodującego DNA jest stosunkowo konserwatywna (jego tempo ewolucji jest niższe niż synonimowych pozycji w genach kodujących białka, które uznaje się za ewoluujące w przybliżeniu neutralnie), co sugeruje, że jest pod wpływem doboru naturalnego i ma znaczenie funkcjonalne (ANDOLFATTO 2005). Jednak duże różnice wielkości genomu między blisko spokrewnionymi gatunkami jak również powtarzalna natura niekodującego DNA dużych genomów, przemawiają za tym, że większość niekodującego DNA w dużych genomach eukariotycznych nie ma znaczenia funkcjonalnego. W miarę jak gromadzono informacje o strukturze genomów okazało się, że również liczba genów, choć zmienna i w pewnym stopniu skorelowana ze złożonością organizmów eukariotycznych, waha się w dość szerokich granicach. Zaskoczenie stanowiło również odkrycie, że w genomie człowieka znajduje się jedynie około 20–25 tys. genów, niewiele więcej niż w genomie nicienia *Caenorhabditis elegans* (18 tys) i prawdopodobnie mniej niż w genomie prostej rośliny – rzodkiewnika *Arabidopsis thaliana* (25 tys). W tym kontekście zaskakiwać może stwierdzenie, że największą liczbę genów wśród poznanych organizmów ma jednokomórkowy patogen układu rozrodczego człowieka *Trichomonas* (około 60 tys.), w którego przypadku wysoka liczba genów jest prawdopodobnie wynikiem poliploidyzacji (CARLTON i współaut. 2007).

Poliploidyzacja lub duplikacja całych genomów jest istotnym procesem w ewolucji genomów eukariotycznych. Można sobie łatwo wyobrazić, że kilka rund duplikacji genomu może doprowadzić do szybkiego wzrostu jego wielkości oraz zwiększenia liczby genów. Ocenia się, iż znaczny procent roślin okrytozalążkowych to poliploidy (RIESEBERG i WILLIS 2007). Choć uważa się, iż poliploidyzacja nie zachodzi równie często u zwierząt, to również w ewolucji strunowców doszło do dwu rund duplikacji genomu, które nastąpiły już po oddzieleniu się linii wiodącej do kręgowców od linii wiodących do lancetnika i osłonicy (PUTNAM i współaut. 2008). Oprócz duplikacji całego genomu do szybkiego wzrostu wielkości genomów eukariotycznych przyczyniają się również duplikacje fragmentów chromosomów, zwane duplikacjami segmentowymi.

Kolejnym czynnikiem umożliwiającym szybkie zmiany wielkości genomów eukariotycznych jest występująca w nich duża liczba ruchomych elementów genetycznych. Nie ma

tutaj potrzeby wchodzenia w szczegóły dotyczące klasyfikacji tych elementów (WICKER i współaut. 2007); z punktu widzenia ewolucji genomu istotne jest to, iż w przypadku większości elementów ruchomych transpozycja jest procesem replikatywnym, w nowe miejsce w genomie wprowadzana jest kopia oryginalnego elementu, który pozostaje na swoim miejscu, a więc transpozycja prowadzi wprost do wzrostu wielkości genomu. Doskonałym przykładem jest tutaj kukurydza: 80% jej genomu o wielkości około 2,5 Gb złożone jest z elementów ruchomych, do których ekspansji doszło w ciągu ostatnich 5–6 mln lat (GAUT i współaut. 2000). Z ruchomych elementów genetycznych wywodzi się również prawie połowa genomu człowieka (wielkość nieco ponad 3 Gb).

W przypadku eukariotów powszechnie przyjmuje się, że horyzontalny transfer genów – choć ważny – nie jest tak istotny jak u prokariotów (KEELING i PALMER 2008). Wkrótce po opublikowaniu szkicu sekwencji ludzkiego genomu pojawiły się doniesienia o istnieniu w nim znacznej liczby genów bakteryjnych, co sugerowało, że poziomy transfer genów był dość częsty w linii prowadzącej do człowieka. Późniejsze badania nie potwierdziły jednak tych sugestii, co mogło spowodować niechęć badaczy do zajmowania się zjawiskiem HTG u eukariotów. Tym niemniej HTG ma pewne znaczenie również u eukariotów, choć poszczególne grupy filogenetyczne różnią się bardzo w tym zakresie (KEELING i PALMER 2008). Wydaje się, że HTG od prokariotów ma większe znaczenie u jednokomórkowych eukariotów. Zasadniczą rolę przypisuje się tutaj okazji: szczególnie dużo HTG widzimy u organizmów żyjących w środowisku pełnym bakterii i żywiących się nimi. Jak dotychczas najwięcej genów będących efektem HTG od bakterii stwierdzono u orzęsków żyjących w żwaczu przeżuwaczy i żywiących się bakteriami. Najczęściej przez HTG przekazywane są geny związane z metabolizmem, np. z metabolizmem beztlenowym, co stwierdzono u żyjących w środowisku beztlenowym pasożytniczych eukariotów, takich jak: *Giardia*, *Entamoeba*, *Trichomonas*.

Czynnikiem ograniczającym HTG jest prawdopodobnie wczesne wyodrębnianie się w cyklu życiowym linii płciowej, co może tłumaczyć, dlaczego HTG zachodzi stosunkowo rzadko u zwierząt. Poziomy transfer genów zdarza się także między eukariotami, jest jednak stosunkowo trudny do wykrycia ze względu na techniczno-metodologiczne.

Mimo to stwierdzono, że jest częsty np. u grzybów. Choć znane są pojedyncze przypadki poziomego przekazu genów eukariotycznych do bakterii, uważa się, że taki transfer jest niezwykle rzadki. Sugerowano, że spowodowane jest to występowaniem intronów i/lub złożonej regulacji ekspresji genów eukariotycznych; możliwe jednak, że eukarioty nie mają zbyt wiele do „zaoferowania” prokariotom, biorąc pod uwagę ogromną różnorodność pan-genomu prokariotycznego (KEELING i PALMER 2008).

Szczególne formy horyzontalnego przepływu genów odegrała niemożliwą do przecenienia rolę w historii eukariotów (KEELING i PALMER 2008). Chodzi tutaj oczywiście o przekaz genów prokariotycznych do eukariotów podczas endosymbiozy, związanej z powstaniem organelli. Uważa się, że powstanie mitochondriów z alfa-proteobakterii nastąpiło tylko raz, we wczesnych stadiach ewolucji eukariotów – wszystkie współczesne organizmy eukariotyczne mają mitochondria lub też wykazują oznaki ich wtórnej utraty (zobacz artykuł GOLIKA w tym zeszycie KOSMOSU). Również powstanie plastydów miało miejsce tylko raz, na drodze symbiozy przodka grupy obejmującej rośliny, krasnorosty i glaukofity z sinicą. W wyniku symbiozy w komórkach pierwotnych eukariotów znalazł się niezależny genom, z którego większość genów została przeniesiona do genomu jądrowego. Geny te kodują obecnie białka, które transportowane są z powrotem do organelli za pomocą wyspecjalizowanych mechanizmów. Jedynie stosunkowo nieliczne geny pozostały w organellach, np. niemal wszystkie zwierzęce mitochondria zawierają tylko 13 genów kodujących białka i 24 kodujące funkcjonalne RNA. Proces eksportu genów z organelli do jądra można zaobserwować również współcześnie (ADAMS i współaut. 2000), przy czym często przeniesione kopie są нефunkcjonalne (BENSASSON i współaut. 2001). Jeszcze bardziej złożonymi przykładami HTG są wtórne symbiozy, gdy posiadający plastydy eukariot znajduje się w komórce innego eukariota – geny jądrowe symbionta kodujące białka plastydowe są wtedy przenoszone do jądra komórki gospodarza, a jądro symbionta może zaniknąć zupełnie. Dzięki wtórnej symbiozie z zielenicą plastydy nabyły eugleny, a dzięki symbiozie z krasnorostem – kryptomonady. U bruzdnic znane są nawet symbiozy trzeciorzędowe, polegające na symbiozie z innym eukariotem, który nabył plastyd już wcześniej, w

wyniku wtórnej symbiozy z innym eukariotem (KEELING i PALMER 2008).

Z omówienia i porównania genomów eukariotycznych i prokariotycznych wynika pytanie o kluczowym znaczeniu: Co odpowiada za obserwowane zróżnicowanie wielkości oraz wzorców ewolucji tych genomów? Postawiono wiele hipotez, które krótko omówię poniżej. Ostatnio za najbardziej przekonującą uważa się hipotezę sformułowaną przez Michaęla Lyncha i współpracowników (LYNCH i CONERY 2003, LYNCH 2007), stwierdzającą, iż wzrost wielkości i złożoności genomu nie jest przejawem ewolucji adaptacyjnej, a więc odbywającej się pod wpływem doboru naturalnego, lecz przeciwnie – efektem słabego działania doboru oczyszczającego w niewielkich populacjach (patrz artykuł KORONY w tym zeszycie KOSMOSU). Teoria genetyki populacji mówi, iż mutacje o niewielkiej szkodliwości będą w małych populacjach zachowywać się neutralnie, co oznacza, że mogą utrwalić się w wyniku działania procesów losowych – dryfu genetycznego. Graniczny współczynnik doboru jest równy odwrotności czterokrotności efektywnej wielkości populacji. Prokarioty mają gigantyczne efektywne wielkości populacji, rzędu 10^8 (setki milionów). Wiele jednokomórkowych eukariotów ma również duże populacje, rzędu 10^7 , podczas gdy oszacowania efektywnej wielkości populacji u organizmów wielokomórkowych są rzędu 10^4 – 10^6 . Okazuje się, iż wstawienie do genomu „zbędnych” fragmentów DNA, takich jak introny czy elementy ruchome, będzie najczęściej szkodliwe. Szkodliwość dodatkowego DNA wynika z tego, iż mogą zajść w nim mutacje powodujące powstanie „fałszywych” sygnałów regulujących ekspresję genów, w przypadku intronów mutacje części sekwencji kluczowych dla ich wycinania mogą zaburzyć proces składania transkryptu, a wstawienie elementu ruchomego w sekwencję kodującą genu najczęściej spowoduje inaktywację genu (LYNCH 2007). Współczynnik doboru przeciw temu nadmiarowemu DNA szacuje się na 10^{-8} – 10^{-6} . Oznacza to, że w gigantycznych populacjach prokariotycznych dobór oczyszczający będzie efektywnie usuwał nadmiarowy DNA, podczas gdy ten DNA będzie efektywnie neutralny w populacjach organizmów wielokomórkowych, a więc będzie gromadzić się w wyniku działania dryfu genetycznego, prowadząc do wzrostu wielkości genomu. Nie wyklucza to oczywiście faktu, iż dodatkowy DNA, kiedy już znalazł się w komórkach, mógł zostać wy-

korzystany w procesach adaptacyjnych. Teoria Lyncha, aczkolwiek nadal kontrowersyjna, znalazła liczne grono zwolenników (KOONIN 2009). Na jej korzyść przemawia fakt, iż oparta jest na znanych od dawna i niekontrowersyjnych podstawach genetyki populacji – po prostu, jeżeli oszacowania współczynników doboru i efektywnych wielkości populacji są poprawne, to procesy postulowane przez Lyncha będą zachodzić.

Istnieją również konkurencyjne teorie dotyczące przyczyn zróżnicowania ilości DNA w jądrze komórkowym. Hipoteza samolubnego DNA (ang. selfish DNA hypothesis) sugeruje, że elementy ruchome będą zwiększały swoją liczbę w genomie aż do punktu, w którym dobór naturalny powstrzyma ich ekspansję; teoria ta nie tłumaczy jednak zadowalająco wzrostu zawartości intronów i powtarzalnych sekwencji DNA nie mających charakteru elementów ruchomych. Hipoteza wypełniającego DNA (ang. bulk DNA hypo-

thesis), na której korzyść mógłby przemawiać obserwowany silny związek między ilością DNA w jądrze a wielkością komórki mówi, iż większość niekodującego DNA odgrywa rolę strukturalną, wypełniacza zapewniającego utrzymanie odpowiedniego stosunku objętości jądra komórkowego do cytoplazmy, co może mieć znaczenie dla efektywności transportu białek i RNA między cytoplazmą a jądrem komórkowym; wielkość komórek, a zarazem ilość DNA w jądrze komórkowym jest negatywnie skorelowana z tempem metabolizmu (SZARSKI 1983, KOZŁOWSKI i współaut. 2003). Zwrócono również uwagę, iż metaboliczny i/lub czasowy koszt replikacji nadmiarowego DNA może prowadzić do usuwania go przez dobór naturalny z populacji szybko dzielących się komórek prokariotycznych, oraz iż kierunkowa presja mutacyjna – przewaga mutacji typu insercji może prowadzić do wzrostu wielkości genomu.

POWSTAWANIE NOWYCH GENÓW

Zagadnienie ewolucji genów jest bardzo obszerne i obejmuje wiele aspektów, których nie sposób omówić czy nawet zasygnalizować w krótkim, przekrojowym przeglądzie. Dlatego też skupię się tutaj jedynie na mechanizmach, jakie prowadzą do powstawania nowych genów.

Nowe geny powstają najczęściej z genów już istniejących lub ich fragmentów. Oczywistym mechanizmem prowadzącym do ich powstania jest poziomy przekaz (HTG), omówiony powyżej. W wyniku tego procesu organizm otrzymuje geny już „gotowe”, spełniające konkretną funkcję, czasem wraz z sekwencjami regulatorowymi. Znaczenie poziomego transferu w uzyskiwaniu nowych genów przez bakterie i archeowce znajduje odzwierciedlenie we wspomnianej wcześniej koncepcji pan-genomu prokariotycznego. Poziomy przekaz może być również źródłem nowych genów u eukariotów, w ich przypadku wydaje się jednak, iż najważniejszym źródłem nowych genów są zachodzące w obrębie genomu procesy duplikacji (TAYLOR i RAES 2004). Duplikacja obejmować może fragmenty wielkości kilku kb do części chromosomu obejmujących wiele Mb, mówimy w tym przypadkach o duplikacjach segmentowych. Może również dotyczyć całego genomu, jak to omówiono powyżej. Duplikacja może być ponadto efektem działalności elementów

ruchomych lub procesów warunkowanych działalnością elementów ruchomych – jak retrotranspozycja. Jeżeli zduplikowany zostanie cały gen, wraz z sekwencjami regulującymi jego transkrypcję, może on potencjalnie zachować swoją funkcję. Geny spokrewnione ze sobą w wyniku duplikacji nazywamy paralogami (genami paralogicznymi), podczas gdy geny zajmujące to samo miejsce w chromosomie, homologiczne między różnymi organizmami, to ortologi lub geny ortologiczne. Geny paralogiczne ewoluują niezależnie od momentu duplikacji, który można wyznaczyć na podstawie pomiaru liczby różnic, jakie nagromadziły się między sekwencjami paralogów (ich dywergencji). Uważa się, że kolejne duplikacje prowadzą do powstawania rodzin genów – grup genów wywodzących się w drodze duplikacji od wspólnego przodka oraz spełniających zazwyczaj zbliżone, lecz nie identyczne funkcje. Klasycznym przykładem rodziny genów są globiny kręgowców. Liczne geny zgrupowane są w rodziny o zróżnicowanej liczbie członków. Zarówno u człowieka jak i u drożdży najczęstsze są rodziny genów liczące po dwa paralogi, lecz zdarzają się rodziny daleko liczniejsze, liczące ponad tysiąc paralogicznych genów, czego przykładem są geny receptorów węchowych ssaków. Rodziny genów paralogicznych mogą być również wynikiem duplikacji całych ge-

nomów. Obserwuje się bardzo różne stopnie dywergencji paralogów w rodzinach genów, co może sugerować ich różny wiek. W tym kontekście należy wspomnieć o mechanizmie, nazwanym ewolucją zespołową (ang. concerted evolution), który powoduje, że stopień dywergencji między paralogami może być minimalny mimo dawnej duplikacji (NEI i ROONEY 2005). Klasycznym przykładem rodziny ewoluującej na drodze ewolucji zespołowej są eukariotyczne geny rybosomalnego RNA. Zazwyczaj geny te obecne są w genomie w kilkudziesięciu-kilkuset kopiach, a ich sekwencje są praktycznie identyczne. Mechanizmem molekularnym odpowiedzialnym za homogenizację sekwencji między paralogami jest tutaj konwersja genów – szczególny proces rekombinacji powodujący zastąpienie jednej sekwencji DNA drugą. Geny rRNA są zduplikowane prawdopodobnie dlatego, że ogromne ilości rRNA potrzebne są do szybkiego wytwarzania dużej liczby rybosomów w komórce. Natomiast ich ewolucja zespołowa ma znaczenie adaptacyjne zapewniając, iż poszczególne cząsteczki rRNA będą identyczne. Ewolucja zespołowa jest niezbyt często obserwowanym procesem.

Jakie mogą być losy zduplikowanych genów? Oczywiście często po duplikacji dochodzi do utraty funkcji genu – pseudogenizacji. Jeżeli duplikacja jest niepełna, zduplikowana kopia pozbawiona jest ważnych sekwencji regulatorowych lub też jeżeli w wyniku retrotranspozycji zostanie wstawiona w nieodpowiednie środowisko genomowe, kopia taka będzie niefunkcyjna i od momentu swojego powstania będzie pseudogenem (ang. dead-on-arrival pseudogene). Jeżeli nawet początkowo zduplikowana kopia będzie funkcjonalna, to szkodliwe mutacje, które pojawiają się w jednej z kopii zduplikowanego genu, doprowadzą do utraty funkcji (ang. nonfunctionalization), nieszkodliwej dla organizmu, gdyż druga kopia, paralogiczna, będzie nadal funkcjonalna. Tak powstały pseudogen może utrwalić się w wyniku działania dryfu genetycznego w populacji. Wydaje się, że pseudogenizacja w wyniku jednego z omówionych wyżej procesów jest najczęstszym losem duplikatów. W wyniku duplikacji może jednak również dojść do dwu innych procesów, skutkujących zachowaniem obu zduplikowanych genów oraz powodujących powstanie genów o nowej funkcji. Pierwszym z tych procesów jest neofunkcjonalizacja (ang. neofunctionalization), mająca miejsce, gdy jedna ze zduplikowanych kopii nabywa, zanim

zostanie dezaktywowana przez mutacje, korzystnych mutacji warunkujących nową funkcję, co może doprowadzić do jej utrwalenia się w wyniku działania doboru naturalnego. Najczęściej przyjmuje się, iż ta nowa funkcja upośledzałaby oryginalną funkcję genu, dlatego też mutacje takie nie mogłyby się utrwalić w genie oryginalnym. Warunki genetyczno-populacyjne, w jakich dochodzi do neofunkcjonalizacji są dość restrykcyjne, co oznacza, że powinna występować stosunkowo rzadko (LYNCH 2007). Innym i jak się obecnie uważa, częstszym mechanizmem zachowania obu zduplikowanych kopii genu jest subfunkcjonalizacja (ang. subfunctionalization), zachodząca wg mechanizmu DDC (duplikacja-degeneracja-komplementacja). Odbywa się to w ten sposób, że w jednej kopii zduplikowanego genu zachodzi mutacja, powodująca utratę jednej z funkcji oryginalnego białka, co zapewnia zachowanie w stanie funkcjonalnym drugiej kopii, w której dochodzi do utrwalenia się innej mutacji. To z kolei powoduje utratę funkcji, którą spełnia kopia pierwsza, co zapewnia zachowanie w stanie funkcjonalnym tejże. W ten sposób obie zduplikowane kopie stają się niezbędne, co zapewnia ich zachowanie i umożliwia ewolucję, która może doprowadzić do powstania bardziej wyspecjalizowanych form białek spełniających odmienne nieco funkcje, co obserwujemy w wielu rodzinach białek. Trzeba zwrócić uwagę, iż w gruncie rzeczy początkowe etapy subfunkcjonalizacji wymagają zajścia mutacji upośledzających funkcje białka, a więc zjawisk, które powinny być częste. Wiele danych przemawia za tym, że subfunkcjonalizacja jest dominującym procesem powodującym zachowanie zduplikowanych paralogów w stanie funkcjonalnym (LYNCH 2007).

Poza duplikacją istnieją jeszcze inne mechanizmy formowania nowych genów (LONG i współaut. 2003). Geny praktycznie wszystkich eukariotów zawierają introny, choć ich liczba dramatycznie różni się między grupami taksonomicznymi. Eksonowo-intronowa budowa genów eukariotycznych pozwala na mechanizm powstawania nowych genów, zwany tasowaniem eksonów. W wyniku rekombinacji zachodzącej w intronach całe eksony mogą być przenoszone między białkami. Ponieważ często granice eksonów odpowiadają granicom domen białkowych – funkcjonalnych części budulcowych białek – mechanizm tasowania eksonów może prowadzić do wymiany całych fragmentów warunkujących określone funkcje. Ocenia się, iż

19% eksonów w genach eukariotycznych jest wynikiem tasowania eksonów (LONG i współaut. 2003).

Procesem, który prowadzi do wzrostu różnorodności białek bez duplikacji genów, jest alternatywne składanie (ang. alternative splicing) transkryptu, polegające na tworzeniu więcej niż jednego mRNA z sekwencji genu poprzez łączenie eksonów w różnych kombinacjach (MAKAŁOWSKA i współaut. 2009). Alternatywne składanie jest częste u organizmów wielokomórkowych; ocenia się, że około 75% ludzkich genów ma co najmniej dwie formy będące jego wynikiem. Co ciekawe, gdy gen ulegnie duplikacji, alternatywnie składane formy mogą być utrwalone jako paralogi (TAYLOR i RAES 2004).

Nowe eksony genów mogą również powstawać w wyniku wstawienia ruchomych elementów genetycznych w introny; sekwencje elementów ruchomych ulegają następnie mutacjom skutkującym utratą zdolności do transpozycji oraz wytworzeniem odpowiednich sygnałów składania, które umożliwiają funkcjonowanie sekwencji wywodzącej się z elementu ruchomego jako nowego eksonu. Ocenia się, że około 4% nowopowstałych eksonów w ludzkich genach wywodzi się z elementów ruchomych.

Połączenie dwu genów w jeden lub rozszczepienie jednego genu w dwa to również procesy mogące doprowadzić do powstania nowych genów, szczególnie częste u prokariotów. Mogły one być zaangażowane w tworzenie 0.5% genów prokariotycznych (LONG i współaut. 2003).

Seqwencje kodujące mogą wreszcie powstawać *de novo*, np. z sekwencji intronowych, w wyniku uzyskania przez nie odpowiednich sygnałów zapewniających właściwe składanie nowopowstałych eksonów. Powstawanie sekwencji kodujących *de novo* dotyczy raczej nowych eksonów, a nie całych genów.

Należy wreszcie wspomnieć o kombinowanych mechanizmach powstawania nowych genów, jak *jingwei* u *Drosophila*, gdzie prześledzenie tych mechanizmów okazało się możliwe i stwierdzono, iż w powstaniu tego nowego genu (wiek określa się na 2 mln lat) brały udział duplikacje segmentalne, retrotranspozycja oraz tasowanie eksonów (WANG i współaut. 2000).

Szczególny mechanizm powstawania nowych genów opisano niedawno u wrotków z grupy Bdelloidea (POUCHKINA-STANTCHEVA i współaut. 2007). Jest to największa znana grupa zwierząt rozmnażająca się bezpłciowo od bardzo dawna (kilkadziesiąt mln lat). Konsekwencją rozmnażania bezpłciowego jest całkowity lub prawie całkowity brak rekombinacji, czego efektem jest bardzo wysoka dywergencja sekwencji między allelami w tym samym locus, znana jako efekt Meselsona. Okazało się, że w jednym przypadku, allele tego samego genu nie tylko wykazują wysoką dywergencję sekwencji, lecz również funkcjonalne zróżnicowanie. Wrotki te są zdolne do przechodzenia w stan anabiozy, całkowitego wyschnięcia a następnie powrotu do życia. Białko będące produktem jednego z alleli zapobiega podczas wysychania organizmu tworzeniu agregatów (złogów) przez wrażliwe na wysychanie enzymy. Produkt drugiego allelu nie ma natomiast takiej zdolności; wiąże się on z dwuwartową lipidową i prawdopodobnie zaangażowany jest w zachowanie integralności błon biologicznych.

Przedstawione przykłady pokazują iż powstawanie nowych genów odbywać się poprzez działanie wielu mechanizmów; niektóre z nich poznano już stosunkowo dobrze, lecz w związku z błyskawicznym postępem genomiki porównawczej można spodziewać się wykrycia nowych, a także jeszcze pełniejszego zrozumienia już znanych mechanizmów.

EVOLUTION OF GENOMES AND THE ORIGIN OF NEW GENES

Summary

Genomes of Bacteria and Archaea are extremely compact, almost devoid of noncoding DNA. Sizes of these "prokaryotic" genomes span only two orders of magnitude and their evolution is characterized by: strong pressure for the removal of non-functional DNA, frequent structural rearrangements resulting in randomization of gene order, profound differences in gene content between related forms and ubiquitous horizontal gene transfer (HGT). Genome sizes in Eukaryotes vary enormously, span-

ning five orders of magnitude. A relatively weak correlation between the genome size and organismal complexity in Eukaryotes, known as the C-value paradox, results from interspecific differences in the amount of noncoding DNA, composed of introns, repetitive sequences and mobile elements. The plausible explanation for the disparities between prokaryotic and eukaryotic genomes are the differences of the effective population sizes between organisms, which affect efficiency of natural

selection. The accumulation of “extra” DNA is weakly deleterious and it is efficiently removed by selection in huge populations of Bacteria and Archaea. In smaller populations of eukaryotes, particularly multicellular organisms, drift overcomes selection, rendering this “extra” DNA effectively neutral, enabling its accumulation and consequently increase of genome size. New genes may emerge through multiple mechanisms. In bacteria and Archaea HGT

is very important in this respect. In Eukaryotes duplications, both whole genome and segmental, are of utmost importance. One copy of a duplicated gene most often accumulates deleterious mutations and becomes a pseudogene. However, sometimes both duplicated copies are retained – one of them evolves a new function in the process of neofunctionalization or each copy undergoes specialization in the process of subfunctionalization.

LITERATURA

- ABBY S., DAUBIN V., 2007. *Comparative genomics and the evolution of prokaryotes*. Trends Microbiol. 15, 135–141.
- ACHTMAN M., WAGNER M., 2008. *Microbial diversity and the genetic nature of microbial species*. Nature Rev. Microbiol. 6, 431–440.
- ADAMS K. L., DALEY D. O., QIU Y. L., WHELAN J., PALMER J. D., 2000. *Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants*. Nature 408, 354–357.
- ANDOLFATTO P., 2005. *Adaptive evolution of non-coding DNA in Drosophila*. Nature 437, 1149–1152.
- BENSASSON D., ZHANG D. X., HARTL D. L., HEWITT G. M., 2001. *Mitochondrial pseudogenes: evolution's misplaced witnesses*. Trends Ecol. Evol. 16, 314–321.
- BROWN T. A., 2009. *Genomy*. PWN, Warszawa.
- CARLTON J. M., HIRT R. P., SILVA J. C. i współaut., 2007. *Draft genome sequence of the sexually transmitted pathogen Trichomonas vaginalis*. Science 315, 207–212.
- CARNINCI P., KASUKAWA T., KATAYAMA S. i współaut., 2005. *The transcriptional landscape of the mammalian genome*. Science 309, 1559–1563.
- FRASER C., ALM E. J., POLZ M. F., SPRATT B. G., HANAGE W. P., 2009. *The bacterial species challenge: making sense of genetic and ecological diversity*. Science 323, 741–746.
- GAUT B. S., D'ENNEQUIN M. L., PEEK A. S., SAWKINS M. C., 2000. *Maize as a model for the evolution of plant nuclear genomes*. Proc. Natl. Acad. Sci. USA 97, 7008–7015.
- KEELING P. J., PALMER J. D., 2008. *Horizontal gene transfer in eukaryotic evolution*. Nature Rev. Genet. 9, 605–618.
- KOONIN E. V., 2009. *Darwinian evolution in the light of genomics*. Nuc. Acid. Res. 37, 1011–1034.
- KOONIN E. V., WOLF Y. I., 2008. *Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world*. Nuc. Acid. Res. 36, 6688–6719.
- KOZŁOWSKI J., KONARZEWSKI M., GAWELCZYK A. T., 2003. *Cell size as a link between noncoding DNA and metabolic rate scaling*. Proc. Natl. Acad. Sci. USA 100, 14080–14085.
- LAPIERRE P., GOGARTEN J. P., 2009. *Estimating the size of the bacterial pan-genome*. Trends Genet. 25, 107–110.
- LAWRENCE J. G., HENDRICKSON H., 2005. *Genome evolution in bacteria: order beneath chaos*. Curr. Opin. Microbiol. 8, 572–578.
- LONG M., BETRAN E., THORNTON K., WANG W., 2003. *The origin of new genes: Glimpses from the young and old*. Nature Rev. Genet. 4, 865–875.
- LYNCH M., 2007. *The origins of genome architecture*. Sinauer, Sunderland.
- LYNCH M., CONERY J. S., 2003. *The origins of genome complexity*. Science 302, 1401–1404.
- MEDINI D., DONATI C., TETTELIN H., MASIGNANI V., RAPUOLI R., 2005. *The microbial pan-genome*. Curr. Opin. Genet. Dev. 15, 589–594.
- NEI M., ROONEY A. P., 2005. *Concerted and birth-and-death evolution of multigene families*. Annual Rev. Genet. 39, 121–152.
- OCHMAN H., LAWRENCE J. G., GROISMAN E. A., 2000. *Lateral gene transfer and the nature of bacterial innovation*. Nature 405, 299–304.
- POUCHKINA-STANTCHEVA N. N., MCGEE B. M., BOSCHETTI C. i współaut., 2007. *Functional divergence of former alleles in an ancient asexual invertebrate*. Science 318, 268–271.
- PUTNAM N. H., BUTTS T., FERRIER D. E. K. i współaut., 2008. *The amphioxus genome and the evolution of the chordate karyotype*. Nature 453, 1064–1071.
- RIESEBERG L. H., WILLIS J. H., 2007. *Plant speciation*. Science 317, 910–914.
- SZARSKI H., 1983. *Cell size and the concept of wasteful and frugal evolutionary strategies*. J. Theor. Biol. 105, 201–209.
- TAYLOR J. S., RAES J., 2004. *Duplication and divergence: The evolution of new genes and old ideas*. Annual Rev. Genet. 38, 615–643.
- THOMAS C. M., NIELSEN K. M., 2005. *Mechanisms of, and barriers to, horizontal gene transfer between bacteria*. Nature Rev. Microbiol. 3, 711–721.
- WANG W., ZHANG J. M., ALVAREZ C., LLOPART A., LONG M., 2000. *The origin of the Jingwei gene and the complex modular structure of its parental gene, yellow emperor, in Drosophila melanogaster*. Mol. Biol. Evol. 17, 1294–1301.
- WICKER T., SABOT F., HUA-VAN A. i współaut., 2007. *A unified classification system for eukaryotic transposable elements*. Nature Rev. Genet. 8, 973–982.