

IZABELA MAKALOWSKA, MICHAŁ KABZA, JOANNA CIOMBOROWSKA

*Pracownia Bioinformatyki
Wydział Biologii
Uniwersytet Adama Mickiewicza w Poznaniu
Umultowska 89, 61-614 Poznań
E-mail: izabel@amu.edu.pl*

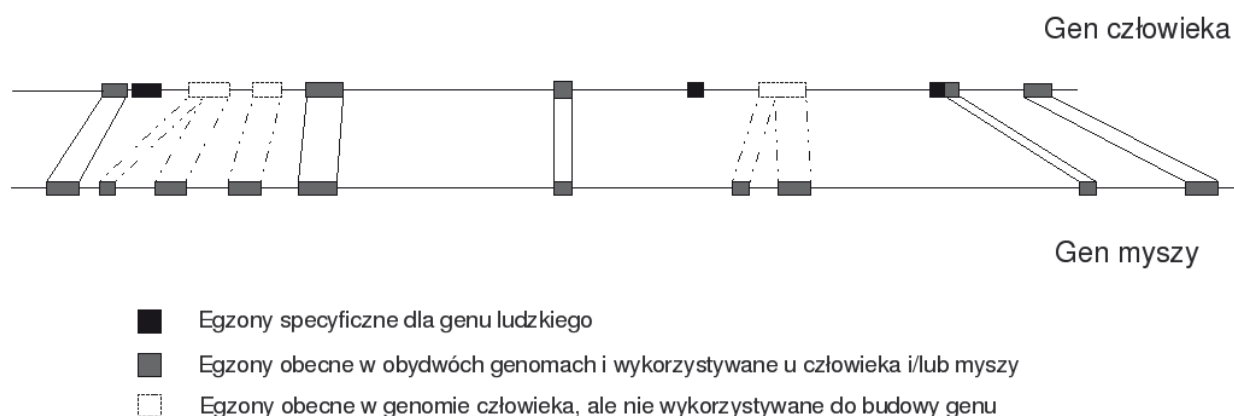
EWOLUCJA STRUKTURY GENÓW

WSTĘP

Genom człowieka (LANDER i współaut. 2001) stanowi katalog cech genetycznych, które, w wyniku interakcji ze środowiskiem, określają naszą biologię, psychologię i odporność na choroby. Ukończenie sekwencjonowania genomów szympansa (CONSORTIUM 2005), makaka (GIBBS i współaut. 2007), myszy (WATERSTON i współaut. 2002), oposa (MIKKELSEN i współaut. 2007), dziobaka (WARREN i współaut. 2008) czy kury (CONSORTIUM 2004) daje nam możliwość przeprowadzenia porównań na skalę genomową w celu zidentyfikowania genów odróżniających człowieka od szympansa, gryzonia czy też innych ssaków lub kręgowców. Jak wykazały analizy porównawcze, zestaw genów kodujących białka jest zadziwiająco podobny u wszystkich kręgowców. Sugeruje to, że do obserwowanej zmienności osobniczej w ogromnym stopniu przyczyniać się może nie tyle powstawanie nowych, specyficznych dla danej linii genów, ile utrata genów, ekspansja rodzin genowych oraz zróżnicowanie struktury genów. Genom blisko spokrewnionego z człowiekiem szympansa nie zawiera ani jednego genu, który nie miałby swojego homologu w genomie człowieka. Typowy gen szympansa różni się od swojego ortologa u człowieka przeciętnie dwoma nukleotydami (CONSORTIUM 2005). Podobnie, ponad 99% genów myszy posiada homologu w genomie człowieka, a pozostałe w większości posiadają homologu w innych genomach ssaków (WATERSTON i współaut. 2002). Do tej pory znaleziono jedynie kilka genów specyficznych wyłącznie dla gryzoni,

nie ma ich w żadnym innym zsekwencjonowanym genomie (WATERSTON i współaut. 2002, GIBBS i współaut. 2007). Liczba tych genów może jednak zmniejszyć się wraz z poznaniem kolejnych genomów.

Potencjał kodujący proteom (liczba białek, jakie genom może kodować) dla danego gatunku jest rozbudowywany poprzez tworzenie nowych genów lub też przez zmiany w strukturze istniejącego już genu. Dotychczasowe badania wskazują, że nowe geny powstają zazwyczaj w wyniku duplikacji lub retropozycji (procesu odwrotnej transkrypcji transkryptu z istniejącego genu i wstawienia tak powstałego odcinka DNA w nowe miejsce w genomie jako nowego genu). Są to więc rezultaty ekspansji istniejących już genów. Duplikacja i retropozycja odgrywają największą rolę w ewolucji transkryptomów i proteomów, otwierają one nowe możliwości ewolucyjne. Często jedna z kopii genu jest tracona, jednakże kiedy oba geny pozostają czynne, jeden z nich może utrzymywać swoją funkcję, podczas gdy drugi ma swobodę ewoluowania bez narażenia na zachwianie funkcjonowania organizmu. Gen może przyjmując nową funkcję również w wyniku ewolucji sekwencji samego genu, jego organizacji lub jego sekwencji regulatorowej (OHNO 1970). Proces ten nazywamy „neofunkcjonalizacją” (RASTOGI i LIBERLES 2005, ZHOU i WANG 2008). Struktura genu może zmienić się w wyniku wbudowania nowego egzonu lub powstawania nowych form splicingowych. Ewolucja genów nie zawsze jest związana z



Ryc. 1. Genomowa struktura ludzkiego genu *LOC12666* i jego ortologa w genomie myszy.

Gen człowieka ma dwa egzony specyficzne jedynie dla małp naczelnych, dodatkowo egzon 6 jest u człowieka dłuższy. Z drugiej strony z 10 egzonów kodujących u myszy jedynie 5 jest wykorzystywanych także w genie człowieka. Pozostałe 5 egzonów znajduje się w rejonie tego genu ale zawierają one mutacje prowadzące do zmiany ramki odczytu lub wprowadzające kodony stop, dodatkowo doszło także do utraty dwóch intronów (introny 2 i 7 w genie myszy).

powstaniem zmienionej formy białka. Często obserwowanym zjawiskiem jest na przykład utrata lub powstanie nowego intronu. Międzgatunkowe różnice w strukturze genu

mogą być dość znaczne i wynikać z działania wielorakich mechanizmów ewolucyjnych (Ryc.1).

SPLICING ALTERNATYWNY

Splicing, to mechanizm za pośrednictwem którego formowane jest dojrzałe mRNA, poprzez usunięcie sekwencji intronów (BLACK 2003). Splicing alternatywny natomiast jest procesem, w wyniku którego z tego samego pre-mRNA produkowanych jest więcej niż jedno mRNA (GRAVELEY 2001). Analizy genomowe ostatnich lat sugerują, że 40–60% genów człowieka posiada formy alternatywne (MIRONOV i współaut. 1999, LANDER i współaut. 2001, MODREK i współaut. 2001). Formy te można, w pewnym sensie, uznać za „wewnętrzne paralogi” (pary genów pochodzących od wspólnego przodka) mogące, podobnie jak zduplikowane geny, w miarę swobodnie ewoluować nie narażając oryginalnej funkcji. Powstawanie alternatywnych form splicingowych jest ściśle powiązane z powstawaniem i/lub wypadaniem egzonów. Większość badań ewolucji struktury genów przeprowadzano właśnie w aspekcie alternatywnego splicingu (MODREK i LEE 2003, ZHANG i CHASIN 2006, ALEKSEYENKO i współaut. 2007, SOREK 2007).

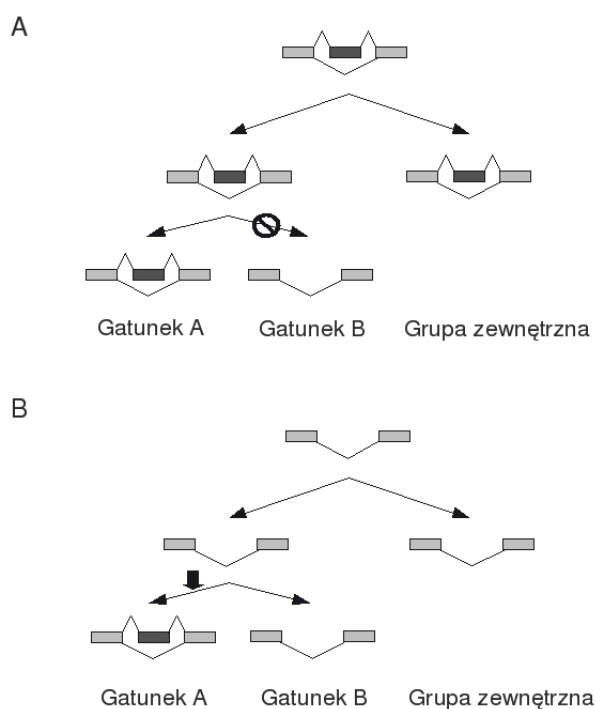
Ewolucja splicingu alternatywnego badana była wielokrotnie za pomocą bioin-

formatycznych analiz na dużą skalę. Wyniki jednej z pierwszych takich prac, opartej na analizie 9434 ortologicznych genów człowieka i myszy, pokazały, że większość egzonów jest zachowanych pomiędzy tymi gatunkami. Dotyczy to jednak głównie egzonów konstytutywnych, a więc takich, które wchodzi w skład wszystkich form splicingowych. Egzony używane jedynie w niektórych wariantach są najczęściej nie zachowane. Spośród 19804 badanych konstytutywnych egzonów człowieka, aż 19309, a zatem 98%, miało swoje odpowiedniki w ortologicznych genach myszy. Jednakże, w przypadku egzonów alternatywnych, a więc używanych jedynie w niektórych formach splicingowych, tylko 27% było zachowane między badanymi gatunkami (MODREK i LEE 2003).

Podobne wyniki otrzymali WANG i współpracownicy (2005), badając 12419 ortologicznych genów człowieka i myszy. Ich analizy wykazały, że 87% nowych egzonów, specyficznych jedynie dla gryzoni, wykorzystywanych było wyłącznie w formach alternatywnych. Drugim wyciągniętym z tych badań wnioskiem było to, że większość nowych eg-

zonów powstała z sekwencji intronowych. Zarówno badania WANGA i współaut. (2005), jak i MODREKA i LEE (2003) pokazują, iż egzony konstytutywne w genomach człowieka i myszy powstały w większości przed rozdzieleniem się tych linii, podczas gdy egzony używane tylko w niektórych wariantach są zazwyczaj nowe i najczęściej nie są zachowane. Inne badania porównawcze transkryptomów człowieka i myszy potwierdzają gatunkową specyficzną nowość egzonów i ich przypuszczalną rolę w kształtowaniu różnic międzygatunkowych (PAN i współaut. 2005, YEO i współaut. 2005).

W analizach ewolucji alternatywnego splicingu ważnym pytaniem jest to, czy brak zachowania większości egzonów używanych w alternatywnych formach jest wynikiem powstawania nowych egzonów czy też może ich utraty. Jeśli egzon jest obecny w genomie organizmu A, ale nie ma go w organizmie B, to może być to konsekwencją zarówno powstania egzonu w organizmie A, jak i utraty w organizmie B. Rozstrzygnięcie tej kwestii możliwe jest jedynie poprzez ustalenie czy egzon ten istniał w genomie przodka. Zrobić to możemy poprzez sprawdzenie, czy dany egzon jest zachowany w bardziej odległym ewolucyjnie organizmie C, określanym jako grupa zewnętrzna. Jeśli egzon znajduje się w genomie organizmu z grupy zewnętrznej, to znaczy, iż musiał być obecny w genomie wspólnego przodka organizmów A i B. Sugeruje to utratę egzonu w organizmie B (Ryc. 2). KONDRASHOV i KOONIN (2003) przeprowadzili analizy alternatywnych form genów człowieka i ich ortologów w genomach innych kręgowców. Jako grupy zewnętrznej użyli białek prokariotów i drożdży (organizmy prokariotyczne nie mają intronów w genach kodujących białka, w przypadku drożdży introny zawiera zaledwie około 3% genów kodujących białka). Analizując 73 białka posiadające dłuższą i krótszą formę splicingową sprawdzali, która z dwóch form obecna jest w genomach z grupy zewnętrznej. Wykazali oni, iż w przypadku 48 białek dłuższe z izoform odpowiadały długością białkom w genomach prokariotycznych i drożdży. Świadczy to o tym, że krótsze formy splicingowe obserwowane u człowieka były konsekwencją utraty egzonu. W 25 przypadkach krótsze formy splicingowe odpowiadały ortologicznym genom u gatunków z grupy zewnętrznej, a zatem powstaniu dłuższych form występujących u człowieka musiało towarzyszyć powstanie nowego egzonu.



Ryc. 2. Metoda 'grupy zewnętrznej' oparta na zasadzie maksymalnej oszczędności pozwalająca na odróżnienie utraty egzonu od jego powstania.

(A) Egzon jest obecny w genomie organizmu z grupy zewnętrznej, oznacza to, że egzon został utracony w gatunku B. (B) W genomie organizmu z grupy zewnętrznej nie ma badanego egzonu a zatem musiał on powstać po rozdzieleniu się gatunków A i B.

Większość przytoczonych badań ewolucji alternatywnego splicingu u kręgowców ogranicza się do porównania człowieka i myszy. ALEKSEYENKO i współpracownicy (2007) przeprowadzili analizy aż 17 genomów. Ich wyniki pokazują, iż gryzonie (analiza mysz *versus* szczur; 40 mln lat od momentu dywergencji) różnią się zaledwie jednym procentem egzonów konstytutywnych, podczas gdy 33% alternatywnie wykorzystywanych egzonów powstało po rozdzieleniu się tych dwóch gatunków. Stopień utraty egzonów nie zależał od tego, czy egzon był wykorzystywany we wszystkich formach czy tylko w niektórych i wynosił 2-4%. U naczelnych (człowiek *versus* szympan; 4 mln lat od momentu dywergencji) tempo powstawania egzonów było wolniejsze, ale tendencja była ta sama, a więc najwięcej nowych egzonów było wśród tych najrzadziej wykorzystywanych w formach alternatywnych (14,5%). Tempo utraty

egzonów było takie jak u gryzoni i wynosiło 2–4%. Taki sam trend zaobserwowano w genomach krowy i psa. Dzięki porównaniu genomów ssaków z genomami innych kręgowców, ALEKSEYENKO i współpracownicy (2007) pokazali także, iż powstawanie egzonów jest

silnie związane z alternatywnym splicingiem nie tylko u ssaków, ale u wszystkich kręgowców. Podobne wyniki uzyskano w innych analizach bioinformatycznych opartych na porównaniu wielu genomów (ZHANG i CHASIN 2006).

MECHANIZM POWSTAWANIA EGZONÓW I ROLA RETROSEKWENCJI W EWOLUCJI STRUKTURY GENÓW

Rozważając problematykę ewolucji genów nie można pominąć mechanizmów powstawania nowych egzonów. Jak dotąd, zaproponowano trzy mechanizmy odgrywające rolę w tworzeniu nowych egzonów: duplikację egzonów (KONDRASHOV i KOONIN 2001), eksaptację elementów powtarzalnych (eksaptacja, to proces ewolucyjny, w którym pewne fragmenty genomu, które pełniły specyficzną funkcję lub też nie pełniły żadnej funkcji, zostały zrekrutowane do nowej funkcji, np. egzonizacja transpozonu, czyli tworzenie egzonu z transpozonu, który jest mobilnym elementem genetycznym) (BROSIOUS i GOULD 1992, MAKALOWSKI i współaut. 1994) i egzonizację sekwencji intronowych (GILBERT 1978, KONDRASHOV i KOONIN 2001, WANG i współaut. 2005).

Duplikacja genów, choć nie przez wszystkich, uważana jest za najczęstszy mechanizm kreowania funkcjonalnych nowości na poziomie genu (OHNO 1970, HUGHES 1999). W związku z tym można przypuszczać, że niektóre z nowych egzonów tworzących warianty splicingowe, mogły także powstać w wyniku duplikacji, w tym wypadku duplikacji wewnątrzgenowej. Kierując się tą hipotezą KONDRASHOV i KOONIN (2001) zbadali 575 białek posiadających alternatywne egzony. Analiza sekwencji egzonów kodujących wykazała, że 50 z nich (9%) powstało w wyniku duplikacji, a alternatywne sekwencje były do siebie podobne w co najmniej 30%. Co ciekawe, duplikacji ulegał egzon konstytutywny, który po zduplikowaniu zamieniał się w egzon alternatywny. Wskazuje na to fakt, iż zduplikowane egzony nigdy nie występowały jednocześnie w tych samych formach splicingowych. Duplikacje istniejących egzonów, jako mechanizmu powstawania nowych egzonów, potwierdziły także późniejsze badania KONDRASHOVA i KOONINA (2003).

Około 70% nowopowstałych egzonów, to elementy powtarzalne (rola elementów powtarzalnych w ewolucji genów jest dysku-

towana poniżej) lub zduplikowane egzony. Znaczna część pozostałych nowych egzonów wykazuje duży stopień zachowania sekwencji sugerujący, iż powstały one w wyniku mutacji tworzących nowe miejsca splicingowe w obrębie sekwencji intronowej (ALEKSEYENKO i współaut. 2007). Z ewolucyjnego punktu widzenia takie egzony powinny być stosunkowo krótkie, w związku z występującymi w sekwencjach niekodujących kodonami stop. Pierwszych dowodów na powstawanie nowych egzonów z sekwencji intronowych dostarczyły analizy KONDRASHOVA i KOONINA (2003). Spośród 25 zidentyfikowanych przez nich nowopowstałych egzonów u ssaków, jedynie dwa powstały w wyniku duplikacji. Pozostałe nie wykazywały jakiegokolwiek podobieństwa do innych białek, za wyjątkiem ich ortologów w innych gatunkach. Obserwacje potwierdzają przypadki czterech genów, w których alternatywne egzony stanowiły całość intronu dzielącego konstytutywne i zachowane egzony (białka MKO3_RAT i ROM_HUMAN) lub też nowy egzon alternatywny przylegał bezpośrednio do egzonu konstytutywnego i stanowił jego przedłużenie (białka ADRO_BOVINE i PPE1_HUMAN).

WANG i współpracownicy (2005) przeprowadzili porównawczą analizę 71039 egzonów zachowanych w genomach myszy i szczura. 3584 z nich nie miało swoich odpowiedników w genomie człowieka, ale znajdowało się w transkryptomie świni, co dowodzi utracenia tych egzonów u człowieka. 2695 egzonów okazało się być unikalnymi dla gryzoni. Wśród nich 1709 mapowało się do intronów w genomie człowieka, świadcząc iż najprawdopodobniej powstały one w wyniku egzonizacji sekwencji intronowej. Przeprowadzone przez ALEKSEYENKO i współpracowników (2007) analizy bioinformatyczne 17 genomów potwierdzają udział egzonizacji intronów w procesie powstawania nowych form splicingowych. Wszystkie powyższe wyniki sugerują, że ewolucja nowych sekwencji kodujących z sekwencji nie-kodujących jest

aktywnym i ciągle zachodzącym procesem u eukariotów.

Źródłem nowych egzonów w genomach kręgowców bardzo często są retrosekwencje (sekwencje będące wynikiem retropozycji). Retrosekwencje stanowią co najmniej 50% genomów ssaków. Wbrew wcześniejszym opiniom, retrosekwencje nie są po prostu „genomowymi śmieciami”, odgrywają one bardzo ważną rolę w ewolucji genomów. Odbywa się to na wiele sposobów, retrosekwencje służą jako gorące miejsca rekombinacji, dostarczają mechanizmów dla przestawiania poszczególnych elementów w obrębie genomów, są także źródłem gotowych do użycia motywów dla nowych elementów regulatorowych, sygnałów poliadenylacji, miejsc splicingowych oraz sekwencji kodujących (MAKALOWSKI i TODA 2007).

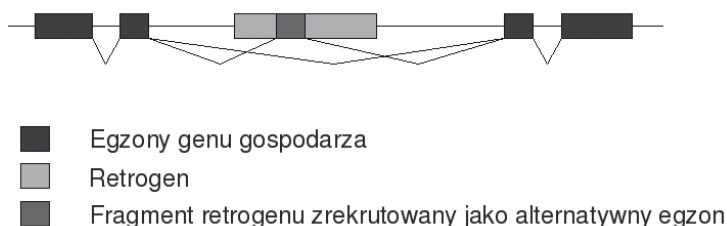
W genomie człowieka najliczniej spośród elementów powtarzalnych reprezentowane są sekwencje *Alu*, jest ich 1,4 miliona, co stanowi około 10% genomu. Ponad 5% alternatywnych egzonów w genomie człowieka powstało w wyniku egzonizacji sekwencji intronowej, przy udziale sekwencji *Alu* (SOREK i współaut. 2002, LEV-MAOR i współaut. 2003, SELA i współaut. 2007). MAKALOWSKI wraz ze współpracownikami (1994) jako pierwsi zwrócili uwagę na istotność egzonizacji *Alu* w ewolucji genów człowieka. Insercja *Alu* w dojrzałym mRNA może przyczyniać się do zwiększenia zmienności i adaptatywności białek. Większość insercji *Alu* do mRNA jest powiązanych z alternatywnym splicingiem, a insercje możliwe są dlatego, iż obie nici *Alu* zawierają sekwencje podobne do sekwencji miejsc splicingowych (MAKALOWSKI i współaut. 1994). Po insercji do intronu, mutacje punktowe w obrębie *Alu* mogą aktywizować miejsca splicingowe, a w konsekwencji *Alu* ulegnie egzonizacji. Teoretycznie, insercja *Alu*, szczególnie jeśli nastąpi ona w obszarze sekwencji kodującej, powinna być szkodliwa dla organizmu. Przykładem może być mutacja sekwencji *Alu* znajdującej się w intronie genu *OAT* (ornitynowa delta-aminotransferaza). Mutacja G→C spowodowała powstanie silnego miejsca donorowego i konstytutywnej insercji nowego egzonu. Nowy egzon zawiera kodon stop, a konsekwencją jest niedobór produktu genu *OAT* (MITCHELL i współaut. 1991, MAKALOWSKI 2000). Mechanizmem zabezpieczającym przed szkodliwymi skutkami tego zjawiska i pozwalającym na tolerowanie *Alu* w sekwencjach kodujących jest alternatywny splicing. Analizy 4151 kon-

tytutywnych i 1176 alternatywnych egzonów w genomie człowieka wykazały, że 61 badanych egzonów „wewnętrznych” zawierało sekwencje *Alu*. We wszystkich przypadkach egzony te były egzonami alternatywnymi (SOREK i współaut. 2002).

Insercje *Alu* mogą mieć też pozytywny wpływ na kształtowanie różnorodności proteomu. Przykładem mogą być geny *DRA* (MAKALOWSKI i współaut. 1994) lub *ADAR2* (GERBER i współaut. 1997). Analizy baz białkowych wykazały, że obok *Alu*, także inne elementy powtarzalne, takie jak MIR (ang. mammalian interspersed repeat) czy CR1 (ang. chicken repeat 1), mogą przyczyniać się do dywersyfikacji białek (LORENC i MAKALOWSKI 2003, GOTEA i MAKALOWSKI 2006)). Ważną funkcją egzonizacji, oprócz dywersyfikacji białek, jest powstawanie form tkankowo specyficznych. Jak wskazują analizy bioinformacyjne, wiele form alternatywnych wykazuje silną regulację tkankową (MODREK i LEE 2003). Wsparciem dla tych analiz są wyniki badań eksperymentalnych, na przykład analiza ekspresji genu dystrofiny, którego wariant zawierający egzon *Alu* wykazuje tkankowo specyficzną ekspresję (FERLINI i współaut. 1998).

Duży udział sekwencji powtarzalnych, a w tym *Alu*, w powstawaniu nowych egzonów, potwierdzają badania przeprowadzone przez grupę Gila Asta (SELA i współaut. 2007). Badając elementy powtarzalne w genomach myszy i człowieka odkryli oni, iż w obu gatunkach 60% tych elementów znajduje się w intronach, a wszystkie rodziny elementów powtarzalnych, wspólnych dla gryzoni i naczelnych, wykazywały podobny stopień egzonizacji. Jednakże egzonizacja *Alu*, specyficznego jedynie dla naczelnych, była znacząco większa niż innych elementów powtarzalnych w genomie człowieka. Efektem zwiększonej aktywności elementów *Alu* jest to, iż proces tworzenia nowych egzonów zachodzi w genomie człowieka na znacznie wyższym poziomie (1824 egzony) niż w genomie myszy (506 egzonów).

W ewolucji genów bardzo często wykorzystywany jest potencjał kodujący retrogenów. Retrogeny są produktami odwrotnej transkrypcji dojrzałego mRNA i insercji powstałego cDNA, do sekwencji genomowej. W ten sposób powstają bezintronowe kopie genów rodzicielskich (BROSIOUS 1991, LONG i współaut. 2003). Powszechnie uważa się, iż proces retropozycji produkuje niefunkcjonalne kopie genów (retropseudogeny), gdyż retropozycja nie obejmuje niezbędnych do



Ryc. 3. Przykład powstania nowej formy splicingowej poprzez wykorzystanie fragmentu retrogeny jako nowego alternatywnego egzonu. Taka eksaptacja retrogeny po jego insercji w intron genu gospodarza nastąpiła na przykład w genach: *BRCA1*, *RORA*, *LIMK2* (BAERTSCH i współaut. 2008).

funkcjonowania genu elementów regulacyjnych (MIGHELL i współaut. 2000). Ostatnie analizy pokazały jednak, iż w genomach ssaków istnieje spora liczba funkcjonalnych retrogenów (BETRAN i współaut. 2002, BETRAN i LONG 2003, BURKI i KAESSMANN 2004, MARQUES i współaut. 2005). Doniesienia na temat funkcjonalnych retrogenów różnią się pod względem oszacowania ich liczby (BALAKIREV i ATALA 2003, SAKAI i współaut. 2007, ZHENG i GERSTEIN 2007), nie ulega jednak wątpliwości, iż retrogeny mają swój udział w formowaniu nowych genów, jak i w ewolucji struktury genów już istniejących (Ryc. 3). Fuzje genów, w których retrogeny uczestniczyły jako „dostarczyciele” nowych domen białkowych dla istniejących genów były odkryte zarówno u zwierząt (LONG i LANGLEY 1993, WANG i współaut. 2002, VICKENBOSCH i współaut. 2006), jak i u roślin (WANG i współaut. 2006b).

Znaczący udział retrogenów w ewolucji struktury genów pokazały niedawne analizy genomu człowieka wykonane przez BAERT-

SCHA i współaut. (2008). Analizując 726 retrokopii genów wykazujących transkrypcję stwierdzili oni, iż 628 z nich stanowiło kompletne retrokopie wbudowane w nowe miejsce, 34 wbudowanych zostało w UTR (ang. untranslated region) istniejących genów, dodając możliwy do wykorzystania potencjał kodujący, a 13 retrokopii zostało wykorzystanych jako źródło nowych egzonów (Ryc. 3). W innych analizach bioinformatycznych VICKENBOSCH i współaut. (2006) pokazali, że 120 retrokopii genów człowieka przekształciło się w funkcjonalne geny będące bezintronowymi, a więc różniącymi się strukturalnie, duplikatami genów rodzicielskich. W 36 przypadkach zaobserwowano fuzję pomiędzy retrogenem i genem sąsiadującym, a w 17 z nich włączenie fragmentu retrogeny w sekwencję kodującą w postaci nowego egzonu. Pozostałe 19 przypadków, to fuzje w obrębie sekwencji niekodujących. Co ciekawe, zidentyfikowano także 27 retrogenów, które zostały rozbudowane poprzez przyłączenie nowopowstałych egzonów niekodujących.

POWSTAWANIE I UTRATA INTRONÓW

Odkrycie intronów w latach 70. było nie małym szokiem dla społeczności naukowej. Od tego czasu introny i historia ich powstania pozostają jednymi z największych tajemnic współczesnej biologii. Funkcja intronów w zasadzie nie jest znana, ale obecność setek tysięcy intronów w genomach zwierząt, roślin, grzybów czy Protista (EICHINGER i współaut. 2005, GALAGAN i współaut. 2005, LOFTUS i współaut. 2005) dostarcza jasnego dowodu na to, że odgrywają one istotną rolę w ewolucji eukariotów. Ewolucja struktury intron-egzon i metody jej badania wzbudzają wiele kontrowersji i są przedmiotem ciągłych „starć” pomiędzy zwolennikami hipotezy wczesnego pojawienia się intronów, jeszcze przed pojawieniem się eukariotów (GILBERT 1987), a zwolennikami koncepcji pojawienia

się intronów dopiero w genomach eukariotycznych (CAVALIER-SMITH 1985, STOLTZFUS i współaut. 1997, LOGSDON 1998). Analizy genomów nie rozwiązały tego problemu i dostarczają argumentów wspierających obie hipotezy. W 2006 r. KOONIN (2006) zaproponował model ewolucji intronów łączący obie hipotezy. W myśl tego modelu prekursorami intronów spliceosomalnych były samowycinające się introny typu II, którym początek dały pasożytnicze retroelementy w komórkach pierwotnych bakterii. Symbioza pomiędzy alfa-proteobakteriami, z których powstały mitochondria, i bliżej nieokreślonymi jednokomórkowymi gospodarzami, dała początek eukariogenezie (procesowi powstawania eukariotów). We wczesnych fazach procesu eukariogenezy nastąpiło masowe przemieszcza-

nie się intronów typu II z mitochondrialnego endosymbiontu do chromosomów gospodarza. Ta inwazja intronów zapoczątkowała formowanie się komórek eukariotycznych i spliceosomu.

Badania porównawcze genomów wskazują na istotne zachowanie pozycji intronów nawet pomiędzy bardzo odległymi ewolucyjnie grupami eukariotów (PENNY i POOLE 1999, FEDOROV i współaut. 2002, ROGOZIN i współaut. 2003). Na przykład, 10–25% (w zależności od zastosowanych metod analitycznych) intronów obecnych w genach ssaków i mających swoje ortologi w genomach roślinnych jest utrzymanych w tych samych miejscach od czasu rozdzielenia się roślin i zwierząt, a więc około 1,5 miliardów lat (FEDOROV i współaut. 2002, ROGOZIN i współaut. 2003). Z drugiej strony, różnorodność pozycji intronów wśród wielu genów homologicznych wskazuje na ich niedawne powstanie (CARMEL i współaut. 2007, ROY i PENNY 2007c). Co ciekawe, w ewolucji ssaków, jak dotąd, obserwowano jedynie utratę intronów. Zarówno badania 1560 ortologów myszy i człowieka (ROY i współaut. 2003), jak i porównania genomów człowieka, myszy i psa (COULOMBE-HUNTINGTON i MAJEWSKI 2007) wykazały ponad 100 przypadków utraty intronu, ale ani jednego nowego intronu specyficznego dla ssaków. Jednakże niedawne analizy transkryptomu człowieka wskazują, iż powstawanie nowych intronów nie zostało zahamowane u ssaków (Makałowska, dane niepublikowane).

W ewolucji genów eukariotycznych obserwujemy zarówno intensywne utraty intronów, jak i powstawanie nowych. Zakres tych procesów różni się znacznie pomiędzy poszczególnymi grupami taksonomicznymi. Niektóre eukarioty, jak na przykład *Saccharomyces cerevisiae*, utraciły niemal wszystkie introny (ROGOZIN i współaut. 2003), podczas gdy utrata intronów u ssaków była stosunkowo nieznaczna (ROGOZIN i współaut. 2003, COULOMBE-HUNTINGTON i MAJEWSKI 2007). Rosnąca w ciągu ostatnich lat liczba badań ewolucji intronów wskazuje, że głównym mechanizmem zmian struktury genów jest

utrata intronów oraz, że zachodziła ona w znacznie większym stopniu niż powstawanie intronów (ROY i współaut. 2003, FEDOROV i współaut. 2003, ROY i GILBERT 2005a, ROY i PENNY 2007b). Nie wszystkie wyniki są jednak pod tym względem zgodne. BABENKO i współpracownicy (2004) analizowali geny z sześciu linii ewolucyjnych organizmów jądrowych i zaobserwowali znaczną przewagę powstawania intronów nad ich utratą. Dodatkowo otrzymane wyniki sugerują różnice w ewolucji pomiędzy genami paralogicznymi i ortologicznymi.

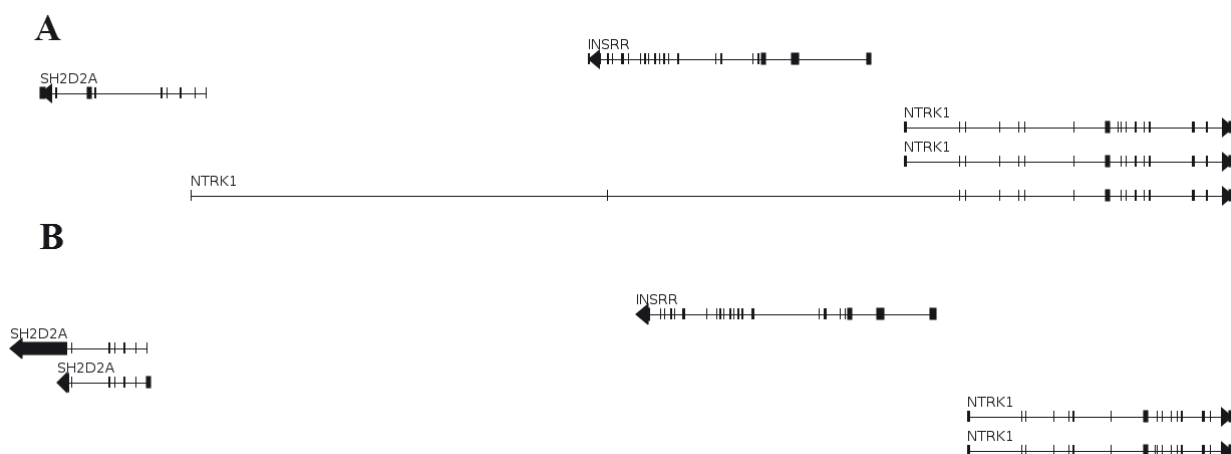
Odmienne rezultaty uzyskane przez BABENKO i współaut. (2004) mogą być konsekwencją zastosowania metody parsymonii Dollo, która nie najlepiej odróżnia powstanie intronów od ich utraty. Te same dane, analizowane przy pomocy innych metod statystycznych, dały odwrotny wynik (ROY i PENNY 2007a). Jak się okazuje, analizy ewolucji intronów są niezwykle czułe na zastosowane metody statystyczne. ROY i GILBERT (2005b) analizując te same 684 klastry ortologicznych genów co ROGOZIN i inni (2003), zastosowali maksymalne prawdopodobieństwo (ML) zamiast parsymonii Dollo i otrzymali wyraźnie mniejszą liczbę na nowo powstałych intronów. Z kolei QIU i współpracownicy (2004) zastosowali statystykę Bayesa analizując 10 rodzin eukariotycznych genów i, podobnie jak BABENKO i współaut. (2004), na podstawie otrzymanych wyników, doszli do wniosku, że większość istniejących obecnie intronów, to introny nowe.

Kontrowersyjnym zagadnieniem w badaniach ewolucji intronów jest także zmiana ich lokalizacji (ang. intron migration) (STOLTZFUS i współaut. 1997). Zwolennicy teorii późnego pojawienia się intronów uważają, iż udział migracji intronów, o ile w ogóle ten proces zachodzi, był w ewolucji nieznaczny. ROGOZIN i jego współpracownicy (2000) przeprowadzili analizę pozycji intronów opartą na teście Monte Carlo. Ich wyniki sugerują, że przesunięcia intronów o 1 nukleotyd występują rzadko, ale są realnym zjawiskiem ewolucyjnym. Nie znaleziono jednak dowodów na większe przesunięcia intronów.

EWOLUCJA NAKŁADAJĄCYCH SIĘ GENÓW

Nakładające się geny są szczególnym przypadkiem, kiedy ewolucja struktury jednego genu wpływa na organizację drugiego, sąsiadującego genu. O nakładających się genach

mówimy wówczas, gdy dwa geny dzielą ten sam fizyczny odcinek DNA i egzon jednego z nich znajduje się w intronie lub egzonie drugiego (Ryc. 4). Na podstawie szeregu badań



Ryc. 4. Przykład zróżnicowanej organizacji genów wynikającej z powstania gatunkowo specyficznego wariantu splicingowego.

(A) Gen *NTRK1* u naczelnych posiada wariant, którego koniec 5' nakłada się z końcem 5' genu *SH2D2A*. Dodatkowo, w wyniku powstania tego wariantu splicingowego gen *INSRR* całkowicie mieści się w genie *NTRK1*. (B) W genomie myszy nie ma nałożenia pomiędzy którymkolwiek z tych trzech genów.

przeprowadzonych w ostatnich latach stwierdzono, że nakładające się geny występują w genomach eukariotów znacznie częściej niż się tego spodziewano. Już odkrycie nakładających się genów u bakteriofaga phiX174 w 1976 r. (BARRELL i współaut. 1976) było nie małą niespodzianką. Tym bardziej zaskoczył fakt obecności nakładających się genów w genomach organizmów jądrowych. Pierwsze doniesienia tego typu genach w eukariotycznym genomie ukazały się w 1986 r., kiedy to SPENCER i współpracownicy (1986) oraz WILLIAMS i FRIED (1986) opublikowali w tym samym numerze *Nature* prace wykazujące obecność takich genów u muszki owocowej i u myszy. Mimo kolejnych raportów o nakładających się genach, dopiero 16 lat później dokonano pierwszych analiz na dużą skalę pokazujących, iż zjawisko to występuje powszechnie u wszystkich eukariotów (IWABE i MIYATA 2001, LEHNER i współaut. 2002, YELIN i współaut. 2003, WANG i współaut. 2006a, MAKALOWSKA i współaut. 2007, NUMATA i współaut. 2007).

Pomimo wielu badań, ciągle nie potrafimy wyjaśnić w jaki sposób dochodzi do powstania nałożeń pomiędzy genami ani też jaki wpływ na ekspresję genu ma taka zmiana organizacji. Geny, których promotory okupują ten sam odcinek DNA, mogą mieć utrudnioną ekspresję, zwłaszcza jeśli leżą one na przeciwnych niciach DNA. W takim przypadku może dojść do kolizji polimeraz poru-

szających się w przeciwnych kierunkach na tym samym odcinku DNA (PRESCOTT i PROUDFOOT 2002). Naturalną konsekwencją będzie utrudniona transkrypcja. Na poziomie posttranskrypcyjnym natomiast sugeruje się kilka mechanizmów wpływających na tworzące się pary mRNA sens-antysens: maskowanie elementów regulatorowych (HASTINGS i współaut. 1997), edytowanie RNA (KIMELMAN i KIRSCHNER 1989), wyciszanie (BORSANI i współaut. 2005).

Pierwszą pracę pokazującą, iż nakładanie się genów jest powszechnie występującą cechą genomów eukariotycznych opublikowano w 2002 r. LEHNER i współpracownicy (2002) przeprowadzili analizę wszystkich dostępnych sekwencji mRNA człowieka i zidentyfikowali 87 par nakładających się genów. Komputerowa analiza sekwencji EST (ang. Expressed Sequence Tag) człowieka i myszy pozwoliła natomiast na zidentyfikowanie 144 i 73 mysich par transkryptów sense-antysense (SHENDURE i CHURCH 2002). Wyniki tych analiz nie oddawały jednak faktycznej częstości występowania nakładających się genów w genomach kręgowców. W 2003 r. YELIN i współpracownicy (2003) zidentyfikowali 2667 par nakładających się genów w genomie człowieka, natomiast KIYOSAWA i współaut. (2003) znaleźli 2481 par takich genów w genomie myszy. Analiza przeprowadzona przez VEERAMACHANENI i współaut. (2004) wykazała natomiast istnienie 774 par nakładają-

cych się genów w genomie człowieka i 578 w genomie myszy. Różnice pomiędzy tymi doniesieniami wynikają przede wszystkim z zestawu analizowanych genów oraz zmian w adnotacjach genomów. Obecnie przyjmuje się, że 40–60% wszystkich genów (kodujących i nie-kodujących) dzieli fragment DNA z innym genem.

Genomowa organizacja nakładających się genów jest bardzo słabo zachowana nawet pomiędzy blisko spokrewnionymi gatunkami. Większość nałożeń jest gatunkowo specyficznych (VEERAMACHANENI i współaut. 2004, MAKALOWSKA i współaut. 2007). Porównanie położenia i struktury ortologów ludzkich nakładających się genów pokazało, że wraz z dystansem ewolucyjnym maleje liczba genów nakładających się u obydwu porównywanych gatunków. O ile u człowieka i myszy znaleziono wspólne 274 ortologiczne pary genów nakładających się, o tyle pomiędzy człowiekiem i kurczakiem były już tylko 64 pary takich genów, a człowiekiem i *Danio rerio* jedynie 13 (MAKALOWSKA i współaut. 2007). Do odmiennych konkluzji doszli DAHARY i współpracownicy (2005), którzy badając nakładające się geny człowieka i nadymki stwierdzili, iż większość obecnie obserwowanych nałożeń pochodzi od wspólnego przodka. Inne genomowe analizy porównawcze potwierdzają jednak słabe zachowanie nałożeń pomiędzy gatunkami (NUMATA i współaut. 2007, SANNA i współaut. 2008, SOLDA i współaut. 2008).

Istnieją dwie wczesne hipotezy dotyczące mechanizmów powstawania nakładają-

cych się genów. KEESE i GIBBS (1992) zasugerowali, że nakładające się geny powstają w wyniku procesu zwanego nadpisaniem (ang. overprinting), w którym to procesie z istniejącego materiału genetycznego powstają nowe, gatunkowo specyficzne geny. W myśl tej hipotezy, w parze nakładających się genów mamy zawsze jeden gen ewolucyjnie zachowany i drugi, młody, gatunkowo specyficzny i odpowiadający za „styl życia” danego gatunku. SHINTANI i współpracownicy (1999) sugerują natomiast, że nakładanie się pomiędzy dwoma badanymi przez nich genami, *ACAT2* i *TCP1*, powstało w rezultacie translokacji jednego z genów, a następnie adopcji utraconych w procesie translokacji sygnałów regulacyjnych.

Nakładanie się pomiędzy genami może być konsekwencją wielu, często wieloetapowych, procesów ewolucyjnych. Istotny udział zarówno translokacji, jak i powstawania genów *de novo* w wyniku „overprintingu” potwierdzają inne analizy (DAN i współaut. 2002, MAKALOWSKA i współaut. 2007, MAKALOWSKA 2008). Różnice strukturalne pomiędzy genami nakładającymi się w jednym gatunku i nie nakładającymi się w innym, nie dotyczą jedynie wydłużonego, na przykład w wyniku adaptacji nowego sygnału poliadenyacji, końca 3'. W powstawaniu nałożeń pomiędzy genami istotną rolę odgrywają także eksaptacja elementów powtarzalnych, np. *Alu*, czy też powstanie nowych, gatunkowo specyficznych wariantów splicingowych (MAKALOWSKA i współaut. 2007).

PODSUMOWANIE

Pomimo ogromnych osiągnięć w badaniach ewolucji genomów wiele ważnych pytań pozostaje ciągle bez odpowiedzi. Jedną z kwestii spornych jest to, który z mechanizmów odgrywa największą rolę w powstawaniu nowych egzonów. Podczas gdy jedni wskazują na eksaptacje elementów powtarzalnych, inni twierdzą, iż egzonizacja sekwencji intronowych jest głównym napędem powstawania nowych form splicingowych i ewolucji struktury genów. Głównym argumentem w tym drugim przypadku jest to, iż nowe egzony mogą dużo łatwiej powstać w wyniku mutacji punktowych, prowadzących do powstania nowych miejsc splicingowych, niż w wyniku skomplikowanego procesu insercji zewnętrznych elementów. Nierozstrzygnięta jest także kwestia pochodzenia intro-

nów. Co prawda zarówno zwolennicy obu hipotez, zarówno wczesnego pojawienia się intronów, jak i powstania intronów dopiero u eukariotów, złagodnieli w swoich opiniach, ale dostarczane przez analizy dowody wskazują zarówno na niedawne powstanie, masową utratę, jak i zachowanie intronów. Badania genomów dostarczają nam także ciągłych niespodzianek. Elementy powtarzalne uważane wcześniej za genomowe śmieci odgrywają ogromną rolę w ewolucji, są źródłem wielu nowych egzonów lub sygnałów. Przekonaliśmy się też niedawno, że to co przez długie lata przypisywaliśmy jedynie genomom prokariotycznym jest szeroko rozpropagowane także w genomach eukariotycznych, czego przykładem mogą być nakładające się geny. Dynamika sekwencionowania genomów oraz

rozwijające się techniki analiz bioinformatycznych dają zatem nadzieję nie tylko na rozwiązanie istniejących zagadek, ale przede

wszystkim na odkrycie zupełnie nowych tajemnic genomów roślin i zwierząt.

GENE STRUCTURE EVOLUTION

Summary

With the growing number of sequenced genomes comparative studies of lineage specific genomic features become both very rewarding and challenging. Large scale multiple genomes analyses allow to decipher many genomic features. They show that main differences between related species concern not as much the number of genes or the presence of species specific genes as the differences in the gene structure organization. Although much has been learned about gene structure evolution many problems remain unsolved. Alternative splicing is one of the main mechanisms leading to the proteome diversification. The raise of new splice variants is strictly connected with the exon and intron loss and gain. Main mechanisms of how the new exons originate are known, but question which of them, if any,

plays the main role remains open. Another unsolved mystery is the intron origination. The dispute between "intro-early" and "intron-late" hypotheses supporters leads us to many interesting findings but the problem remains unsolved. One of the most fascinating discoveries in the genome studies is the role of so called 'junk DNA' in the evolution of human and other vertebrates. Repetitive elements and retrogenes are one of the most important elements in the gene structure evolution. They provide signals, motifs and coding sequences for new exons, splice sites or regulatory elements. Another phenomenon discovered in the process of whole genomes analyses is the common presence of overlapping genes and, at the same time, their low conservation level.

LITERATURA

- ALEKSEYENKO A. V., KIM N., LEE C. J., 2007. *Global analysis of exon creation versus loss and the role of alternative splicing in 17 vertebrate genomes*. RNA 13, 661-670.
- BABENKO V. N., ROGOZIN I. B., MEKHEDOV S. L., KONIN E. V., 2004. *Prevalence of intron gain over intron loss in the evolution of paralogous gene families*. Nucleic Acids Res. 32, 3724-3733.
- BAERTSCH R., DIEKHANS M., KENT W. J., HAUSSLER D., BROSIUS J., 2008. *Retrocopy contributions to the evolution of the human genome*. BMC Genomics 9, 466.
- BALAKIREV E. S., AYALA F. J., 2003. *Pseudogenes: are they "junk" or functional DNA?* Annu. Rev. Genet. 37, 123-151.
- BARRELL B. G., AIR G. M., HUTCHISON C. A., 3rd, 1976. *Overlapping genes in bacteriophage phiX174*. Nature 264, 34-41.
- BETRAN E., LONG M., 2003. *Dntf-2r, a young Drosophila retroposed gene with specific male expression under positive Darwinian selection*. Genetics 164, 977-988.
- BETRAN E., WANG W., JIN L., LONG M., 2002. *Evolution of the phosphoglycerate mutase processed gene in human and chimpanzee revealing the origin of a new primate gene*. Mol. Biol. Evol. 19, 654-663.
- BLACK D. L., 2003. *Mechanisms of alternative pre-messenger RNA splicing*. Annu. Rev. Biochem. 72, 291-336.
- BORSANI O., ZHU J., VERSLUES P. E., SUNKAR R., ZHU J. K., 2005. *Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis*. Cell 123, 1279-1291.
- BROSIUS J., 1991. *Retroposons—seeds of evolution*. Science 251, 753.
- BROSIUS J., GOULD S. J., 1992. *On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA"*. Proc. Natl. Acad. Sci. USA 89, 10706-10710.
- BURKI F., KAESSMANN H., 2004. *Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux*. Nat. Genet. 36, 1061-1063.
- CARMEL L., WOLF Y. I., ROGOZIN I. B., KOONIN E. V., 2007. *Three distinct modes of intron dynamics in the evolution of eukaryotes*. Genome Res. 17, 1034-1044.
- CAVALIER-SMITH T., 1985. *Selfish DNA and the origin of introns*. Nature 315, 283-284.
- CONSORTIUM. I. C. G. S., 2004. *Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution*. Nature 432, 695-716.
- CONSORTIUM T. C. S. A. A., 2005. *Initial sequence of the chimpanzee genome and comparison with the human genome*. Nature 437, 69-87.
- COULOMBE-HUNTINGTON J., MAJEWSKI J., 2007. *Characterization of intron loss events in mammals*. Genome Res. 17, 23-32.
- DAHARY D., ELROY-STEIN O., SOREK R., 2005. *Naturally occurring antisense: transcriptional leakage or real overlap?* Genome Res. 15, 364-368.
- DAN I., WATANABE N. M., KAJIKAWA E., ISHIDA T., PANDEY A., KUSUMI A., 2002. *Overlapping of MINK and CHRNE gene loci in the course of mammalian evolution*. Nucleic Acids Res. 30, 2906-2910.
- EICHINGER L., PACHEBAT J. A., GLOCKNER G., RAJANDRE-AM M. A., SUGGANG R. i współaut., 2005. *The genome of the social amoeba Dictyostelium discoideum*. Nature 435, 43-57.
- FEDOROV A., MERICAN A. F., GILBERT W., 2002. *Large-scale comparison of intron positions among animal, plant, and fungal genes*. Proc. Natl. Acad. Sci. USA 99, 16128-16133.
- FEDOROV A., ROY S., FEDOROVA L., GILBERT W., 2003. *Mystery of intron gain*. Genome Res. 13, 2236-2241.
- FERLINI A., GALIE N., MERLINI L., SEWRY C., BRANZI A., MUNTONI F., 1998. *A novel Alu-like element rear-*

- ranged in the dystrophin gene causes a splicing mutation in a family with X-linked dilated cardiomyopathy. *Am. J. Hum. Genet.* 63, 436-446.
- GALAGAN J. E., CALVO S. E., CUOMO C., MA L. J., WORTMAN J. R., BATZOGLOU S. i współaut., 2005. *Sequencing of Aspergillus nidulans and comparative analysis with A. fumigatus and A. oryzae.* *Nature* 438, 1105-1115.
- GERBER A., O'CONNELL M. A., KELLER W., 1997. *Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette.* *RNA* 3, 453-463.
- GIBBS R. A., ROGERS J., KATZE M. G., BUMGARNER R., WEINSTOCK G. M., MARDIS E. R. i współaut., 2007. *Evolutionary and biomedical insights from the rhesus macaque genome.* *Science* 316, 222-234.
- GILBERT W., 1978. *Why genes in pieces?* *Nature* 271, 501.
- GILBERT W., 1987. *The exon theory of genes.* *Cold Spring Harb. Symp. Quant. Biol.* 52, 901-905.
- GOTEA V., MAKALOWSKI W., 2006. *Do transposable elements really contribute to proteomes?* *Trends Genet.* 22, 260-267.
- GRAVELEY B. R., 2001. *Alternative splicing: increasing diversity in the proteomic world.* *Trends Genet.* 17, 100-107.
- HASTINGS M. L., MILCAREK C., MARTINCIC K., PETERSON M. L., MUNROE S. H., 1997. *Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels.* *Nucleic Acids Res.* 25, 4296-4300.
- HUGHES A. L., 1999. *Adaptive evolution of Genes and Genomes.* Oxford University press, New York, Oxford.
- IWABE N., MIYATA T., 2001. *Overlapping genes in parasitic protist Giardia lamblia.* *Gene* 280, 163-167.
- KEESE P. K., GIBBS A., 1992. *Origins of genes: "big bang" or continuous creation?* *Proc. Natl. Acad. Sci. USA* 89, 9489-9493.
- KIMELMAN D., KIRSCHNER M. W., 1989. *An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes.* *Cell* 59, 687-696.
- KIYOSAWA H., YAMANAKA I., OSATO N., KONDO S., HAYASHIZAKI Y., 2003. *Antisense transcripts with FANTOM2 clone set and their implications for gene regulation.* *Genome Res.* 13, 1324-1334.
- KONDRASHOV F. A., KOONIN E. V., 2001. *Origin of alternative splicing by tandem exon duplication.* *Hum. Mol. Genet.* 10, 2661-2669.
- KONDRASHOV F. A., KOONIN E. V., 2003. *Evolution of alternative splicing: deletions, insertions and origin of functional parts of proteins from intron sequences.* *Trends Genet.* 19, 115-119.
- KOONIN E. V., 2006. *The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?* *Biology Direct* 1, 22.
- LANDER E. S., LINTON L. M., BIRREN B., NUSBAUM C., ZODY M. C., BALDWIN J. i współaut., 2001. *Initial sequencing and analysis of the human genome.* *Nature* 409, 860-921.
- LEHNER B., WILLIAMS G., CAMPBELL R. D., SANDERSON C. M., 2002. *Antisense transcripts in the human genome.* *Trends Genet.* 18, 63-65.
- LEV-MAOR G., SOREK R., SHOMRON N., AST G., 2003. *The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons.* *Science* 300, 1288-1291.
- LOFTUS B. J., FUNG E., RONCAGLIA P., ROWLEY D., AMEDEO P., BRUNO D. i współaut., 2005. *The genome of the basidiomycetous yeast and human pathogen Cryptococcus neoformans.* *Science* 307, 1321-1324.
- LOGSDON J. M., Jr., 1998. *The recent origins of split-ceosomal introns revisited.* *Curr. Opin. Genet. Dev.* 8, 637-648.
- LONG M., LANGLEY C. H., 1993. *Natural selection and the origin of jingwei, a chimeric processed functional gene in Drosophila.* *Science* 260, 91-95.
- LONG M., BETRAN E., THORNTON K., WANG W., 2003. *The origin of new genes: glimpses from the young and old.* *Nat. Rev. Genet.* 4, 865-875.
- LORENC A., MAKALOWSKI W., 2003. *Transposable elements and vertebrate protein diversity.* *Genetica* 118, 183-191.
- MAKALOWSKA I., 2008. *Comparative analysis of an unusual gene arrangement in the human chromosome 1.* *Gene* 423, 172-179.
- MAKALOWSKA I., LIN C. F., HERNANDEZ K., 2007. *Birth and death of gene overlaps in vertebrates.* *BMC Evol. Biol.* 7, 193.
- MAKALOWSKI W., 2000. *Genomic scrap yard: how genomes utilize all that junk.* *Gene* 259, 61-67.
- MAKALOWSKI W., TODA Y., 2007. *Modulation of host genes by mammalian transposable elements.* *Genome Dyn.* 3, 163-174.
- MAKALOWSKI W., MITCHELL G. A., LABUDA D., 1994. *Alu sequences in the coding regions of mRNA: a source of protein variability.* *Trends Genet.* 10, 188-193.
- MARQUES A. C., DUPANLOUP I., VINCKENBOSCH N., REYMOND A., KAESSMANN H., 2005. *Emergence of young human genes after a burst of retroposition in primates.* *PLoS Biol.* 3, e357.
- MIGHELL A. J., SMITH N. R., ROBINSON P. A., MARKHAM A. F., 2000. *Vertebrate pseudogenes.* *FEBS Lett.* 468, 109-114.
- MIKKELSEN T. S., WAKEFIELD M. J., AKEN B., AMEMIYA C. T., CHANG J. L., DUKE S. i współaut., 2007. *Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences.* *Nature* 447, 167-177.
- MIRONOV A. A., FICKETT J. W., GELFAND M. S., 1999. *Frequent alternative splicing of human genes.* *Genome Res.* 9, 1288-1293.
- MITCHELL G. A., LABUDA D., FONTAINE G., SAUDUBRAY J. M., BONNEFONT J. P., LYONNET S., BRODY L. C., STEEL G., OBIE C., VALLE D., 1991. *Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation.* *Proc. Natl. Acad. Sci. USA* 88, 815-819.
- MODREK B., LEE C. J., 2003. *Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss.* *Nat. Genet.* 34, 177-180.
- MODREK B., RESCH A., GRASSO C., LEE C., 2001. *Genome-wide detection of alternative splicing in expressed sequences of human genes.* *Nucleic Acids Res.* 29, 2850-2859.
- NUMATA K., OKADA Y., SAITO R., KIYOSAWA H., KANAI A., TOMITA M., 2007. *Comparative analysis of cis-encoded antisense RNAs in eukaryotes.* *Gene* 392, 134-141.
- OHNO S., 1970. *Evolution by Gene Duplication.* Springer, Berlin, Alemania.
- PAN Q., BAKOWSKI M. A., MORRIS Q., ZHANG W., FREY B. J., HUGHES T. R., BLENCOWE B. J., 2005. *Alternative splicing of conserved exons is frequently species-specific in human and mouse.* *Trends Genet.* 21, 73-77.
- PENNY D., POOLE A., 1999. *The nature of the last universal common ancestor.* *Curr. Opin. Genet. Dev.* 9, 672-677.
- PRESCOTT E. M., PROUDFOOT N. J., 2002. *Transcriptional collision between convergent genes in bud-*

- ding yeast. Proc. Natl. Acad. Sci. USA 99, 8796–8801.
- QIU W. G., SCHISLER N., STOLTZFUS A., 2004. *The evolutionary gain of spliceosomal introns: sequence and phase preferences*. Mol. Biol. Evol. 21, 1252–1263.
- RASTOGI S., LIBERLES D. A., 2005. *Subfunctionalization of duplicated genes as a transition state to neofunctionalization*. BMC Evol. Biol. 5, 28.
- ROGOZIN I. B., LYONS-WEILER J., KOONIN E. V., 2000. *Intron sliding in conserved gene families*. Trends Genet. 16, 430–432.
- ROGOZIN I. B., WOLF Y. I., SOROKIN A. V., MIRKIN B. G., KOONIN E. V., 2003. *Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution*. Curr. Biol. 13, 1512–1517.
- ROY S. W., GILBERT W., 2005a. *Rates of intron loss and gain: implications for early eukaryotic evolution*. Proc. Natl. Acad. Sci. USA 102, 5773–5778.
- ROY S. W., GILBERT W., 2005b. *The pattern of intron loss*. Proc. Natl. Acad. Sci. USA 102, 713–718.
- ROY S. W., PENNY D., 2007a. *On the incidence of intron loss and gain in paralogous gene families*. Mol. Biol. Evol. 24, 1579–1581.
- ROY S. W., PENNY D., 2007b. *Widespread intron loss suggests retrotransposon activity in ancient apicomplexans*. Mol. Biol. Evol. 24, 1926–1933.
- ROY S. W., PENNY D., 2007c. *A very high fraction of unique intron positions in the intron-rich diatom *Thalassiosira pseudonana* indicates widespread intron gain*. Mol. Biol. Evol. 24, 1447–1457.
- ROY S. W., FEDOROV A., GILBERT W., 2003. *Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain*. Proc. Natl. Acad. Sci. USA 100, 7158–7162.
- SAKAI H., KOYANAGI K. O., IMANISHI T., ITOH T., GOJOBORI T., 2007. *Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes*. Gene 389, 196–203.
- SANNA C. R., LI W. H., ZHANG L., 2008. *Overlapping genes in the human and mouse genomes*. BMC Genomics 9, 169.
- SELA N., MERSCH B., GAL-MARK N., LEV-MAOR G., HOTZWAGENBLATT A., AST G., 2007. *Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome*. Genome Biol. 8, R127.
- SHENDURE J., CHURCH G. M., 2002. *Computational discovery of sense-antisense transcription in the human and mouse genomes*. Genome Biol. 3, RESEARCH0044.
- SHINTANI S., O'HUIGIN C., TOYOSAWA S., MICHALOVA V., KLEIN J., 1999. *Origin of gene overlap: the case of *TCP1* and *ACAT2**. Genetics 152, 743–754.
- SOLDA G., SUYAMA M., PELUCCHI P., BOI S., GUFFANTI A., RIZZI E., BORK P., TENCHINI M. L., CICCARELLI F. D., 2008. *Non-random retention of protein-coding overlapping genes in Metazoa*. BMC Genomics 9, 174.
- SOREK R., 2007. *The birth of new exons: mechanisms and evolutionary consequences*. RNA 13, 1603–1608.
- SOREK R., AST G., GRAUR D., 2002. *Alu-containing exons are alternatively spliced*. Genome Res 12, 1060–1067.
- SPENCER C. A., GIETZ R. D., HODGETTS R. B., 1986. *Overlapping transcription units in the dopa decarboxylase region of *Drosophila**. Nature 322, 279–281.
- STOLTZFUS A., LOGSDON J. M., JR., PALMER J. D., DOOLITTLE W. F., 1997. *Intron "sliding" and the diversity of intron positions*. Proc. Natl. Acad. Sci. USA 94, 10739–10744.
- VEERAMACHANENI V., MAKALOWSKI W., GALDZICKI M., SOOD R., MAKALOWSKA I., 2004. *Mammalian overlapping genes: the comparative perspective*. Genome Res. 14, 280–286.
- VINCKENBOSCH N., DUPANLOUP I., KAESSMANN H., 2006. *Evolutionary fate of retroposed gene copies in the human genome*. Proc. Natl. Acad. Sci. USA 103, 3220–3225.
- WANG W., BRUNET F. G., NEVO E., LONG M., 2002. *Origin of sphinx, a young chimeric RNA gene in *Drosophila melanogaster**. Proc. Natl. Acad. Sci. USA 99, 4448–4453.
- WANG W., ZHENG H., YANG S., YU H., LI J., JIANG H., SU J., YANG L., ZHANG J., MCDERMOTT J., SAMUDRALA R., WANG J., YANG H., YU J., KRISTIANSEN K., WONG G. K., WANG J., 2005. *Origin and evolution of new exons in rodents*. Genome Res. 15, 1258–1264.
- WANG H., CHUA N. H., WANG X. J., 2006a. *Prediction of trans-antisense transcripts in *Arabidopsis thaliana**. Genome Biol. 7, R92.
- WANG W., ZHENG H., FAN C., LI J., SHI J., CAI Z., ZHANG G., LIU D., ZHANG J., VANG, S., LU Z., WONG G. K., LONG M., WANG J., 2006b. *High rate of chimeric gene origination by retroposition in plant genomes*. Plant Cell 18, 1791–1802.
- WARREN W. C., HILLIER L. W., MARSHALL GRAVES J. A., BIRNEY E., PONTING C. P. i współpracownicy, 2008. *Genome analysis of the platypus reveals unique signatures of evolution*. Nature 453, 175–183.
- WATERSTON R. H., LINDBLAD-TOH K., BIRNEY E., ROGERS J., ABRIL J. F. i współpracownicy, 2002. *Initial sequencing and comparative analysis of the mouse genome*. Nature 420, 520–562.
- WILLIAMS T., FRIED M., 1986. *A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends*. Nature 322, 275–279.
- YELIN R., DAHARY D., SOREK R., LEVANON E. Y., GOLDSTEIN O., SHOSHAN A., DIBER A., BITON S., TAMIR Y., KHOSRAVI R., NEMZER S., PINNER E., WALACH S., BERNSTEIN J., SAVITSKY K., ROTMAN G., 2003. *Widespread occurrence of antisense transcription in the human genome*. Nat. Biotechnol. 21, 379–386.
- YEO G. W., VAN NOSTRAND E., HOLSTE D., POGGIO T., BURGE C. B., 2005. *Identification and analysis of alternative splicing events conserved in human and mouse*. Proc. Natl. Acad. Sci. USA 102, 2850–2855.
- ZHANG, X. H., CHASIN L. A., 2006. *Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons*. Proc. Natl. Acad. Sci. USA 103, 13427–13432.
- ZHENG D., GERSTEIN M. B., 2007. *The ambiguous boundary between genes and pseudogenes: the dead rise up, or do they?* Trends Genet. 23, 219–224.
- ZHOU Q., WANG W., 2008. *On the origin and evolution of new genes—a genomic and experimental perspective*. J. Genet. Genomics 35, 639–648.