

Ewolucja chromosomów eukariotycznych – od baz danych do modeli komputerowych i z powrotem

Dorota Mackiewicz, Wojciech Waga, Stanisław Cebrat
Zakład Genomiki, Wydział Biotechnologii, Uniwersytet Wrocławski,
Wrocław

Evolution of eukaryotic chromosomes – from databases to computer modeling

Summary

Computer simulations of chromosomes' and genomes' evolution suggest that the genes located on relatively large and densely packed chromosomes should be grouped in clusters. Clusters located on homologous chromosomes may complement their defects or they may co-operate providing selective advantage to their hosts. Since recombination inside clusters is harmful, selection leads to the uneven distribution of recombination events along chromosomes – relatively high recombination in the subtelomeric regions and low recombination in the central regions of chromosomes. Uneven distribution of recombinations enables sympatric speciation which can not be predicted by the mean field theories of evolution. Further studies of chromosome evolution require more precise data (the best – full sequences) of many closely related genomes belonging to the same species.

Key words:

recombination, Monte Carlo simulation, chromosome, genome evolution.

Adres do korespondencji

Stanisław Cebrat,
Zakład Genomiki,
Wydział Biotechnologii,
Uniwersytet Wrocławski,
ul. Przybyszewskiego 63/77,
51-148 Wrocław;
e-mail:
cebrat@smorfland.uni.wroc.
pl

1. Wstęp

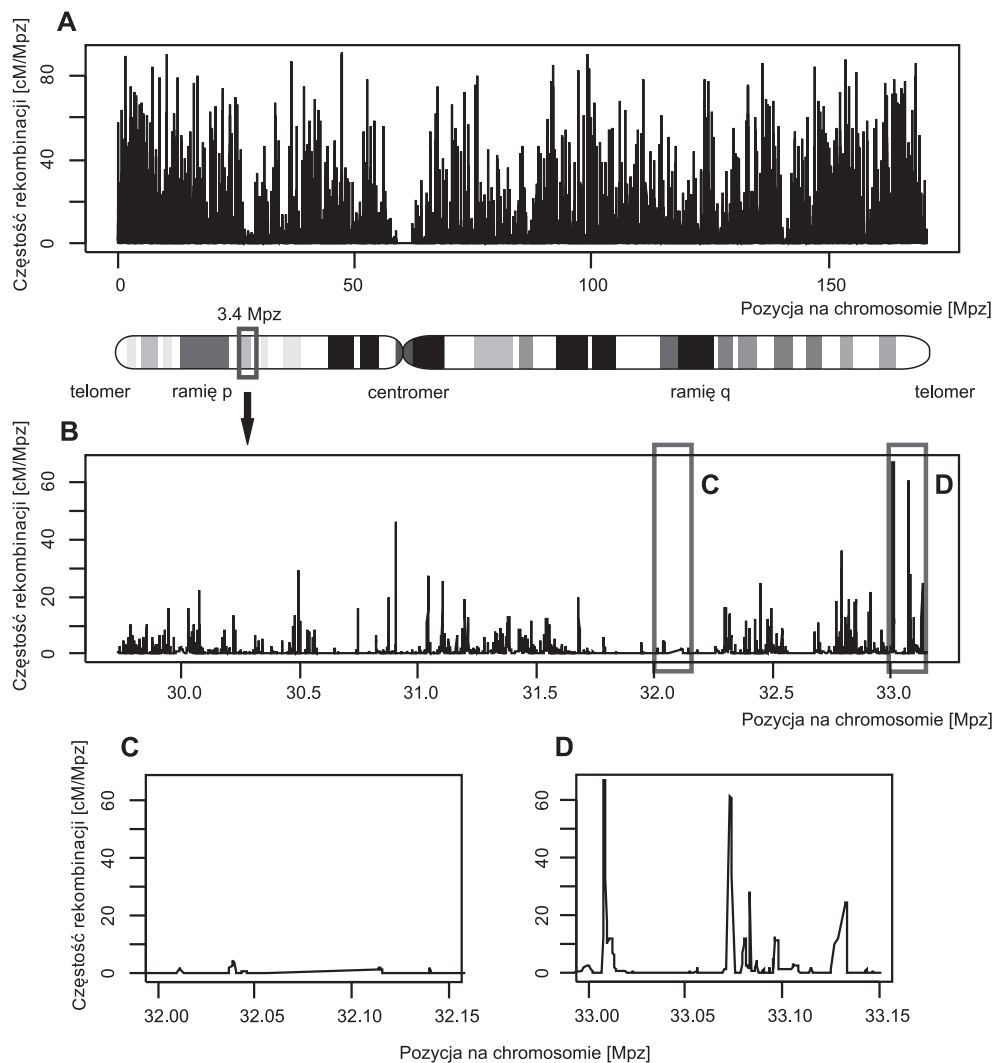
Trudno znaleźć teraz autora pomysłu podziału historii biologii na erę pregenomiczną i postgenomiczną. Terminy te sugerują, że okres sekwencjonowania genomów mamy za sobą, a teraz nastął okres kontemplowania informacji o zebranych sekwencjach genomów przedstawicieli tysięcy gatunków. W wielu

dziedzinach wiedzy znajomość sekwencji pojedynczego genomu z określonego gatunku jest wystarczająca dla wyciągania naukowo istotnych wniosków i dlatego wśród pierwszych poznanych genomów nie było genomów różnych organizmów zaliczanych do tego samego gatunku. Bardzo szybko jednak stało się oczywiste, że nowe pytania dotyczące ewolucji genomów i chromosomów można by stawiać bardziej precyzyjnie, gdyby dysponowało się informacją nie tylko na temat różnic między genomami organizmów przynależnych do różnych taksonów, ale również na temat różnic między genomami organizmów bardziej ze sobą spokrewnionych, należących do różnych ras tego samego gatunku, do różnych rodzin (rodzina jest tutaj rozumiana, jako kategoria społeczna, na przykład rodzice i dzieci), a także do przedstawicieli tej samej rodziny. Porównywanie tak blisko spokrewnionych genomów pozwala na zrozumienie niektórych mechanizmów ewolucyjnych, które do tej pory umykały naszej uwadze.

Porównywanie genomów blisko spokrewnionych wymaga zupełnie innych narzędzi, a nawet założeń teoretycznych niż te, które były dotychczas używane w genetyce lub ewolucji populacji. Przyjmowane w tych dziedzinach za paradygmaty założenia prawa Hardy'ego-Weinberga (H-W) (1,2) lub Fishera-Wrighta (F-W) (3-5) nie mają większego zastosowania, lub ich stosowanie jest bardzo ograniczone. Stało się jasne, że w miejsce dotychczasowych założeń średniopoloowych teorii ewolucji należy uwzględnić przestrzenne relacje zarówno między genami na chromosomach, jak i strukturę i zależności w sieci, w jakiej rozmieszczone są poszczególne osobniki populacji (6-9). Te relacje wymagają dodatkowych informacji na temat częstości mutacji, mechanizmów ich eliminowania z puli genetycznej oraz częstości i rozkładu rekombinacji w genomach (8). Komputerowe modelowanie procesów ewolucyjnych w populacjach niemendlowskich (małych, o dużym współczynniku inbrodu, z ograniczoną częstością rekombinacji, w której nie obowiązuje prawo niezależnego dziedziczenia alleli) pozwala na sugerowanie istnienia mechanizmów ewolucyjnych, które do tej pory nie były rozważane lub uważane były za procesy nieistotne z punktu widzenia ewolucji całej biosfery. Do takich procesów można zaliczyć specjację sympatryczną, która, za sprawą Mayra (10), przez pięćdziesiąt lat była uważana za niemożliwą, a później za „pomijalną”(11). Sprawdzanie nowych hipotez wymaga danych genetycznych, które można uzyskać z sekwencjonowania wielu genomów tego samego gatunku bądź porównywania rozkładu polimorfizmów pojedynczych nukleotydów w bardzo blisko spokrewnionych genomach (12).

2. Rekombinacje genetyczne

Znaczenie rekombinacji genetycznych w ewolucji genomów uległo istotnemu przewartościowaniu od czasu, gdy Morgan tworzył chromosomową teorię dziedziczenia (13) i od czasu publikowania wspomnianych praw H-W i F-W (1-5). Według klasycznej neodarwinowskiej teorii ewolucji rekombinacje (obok mutacji) są proce-



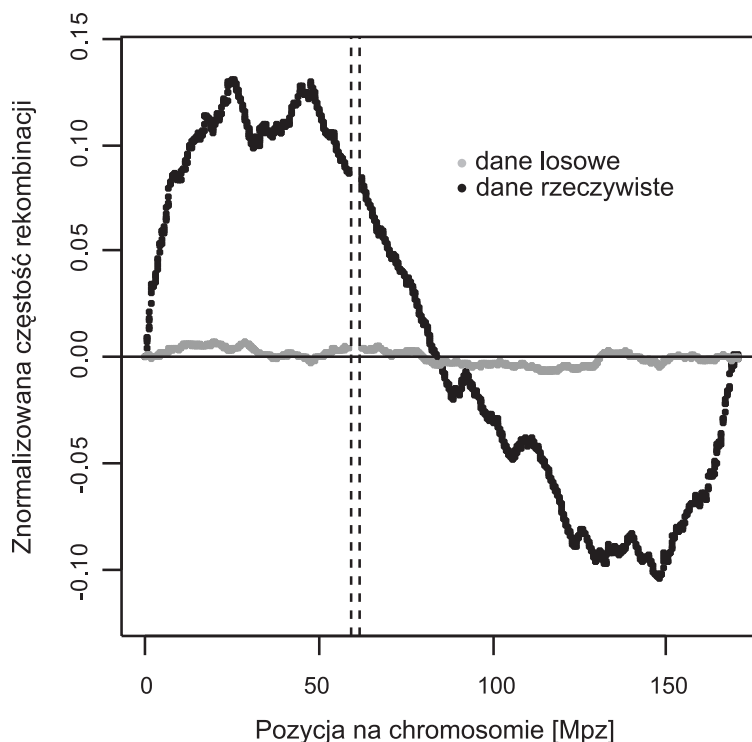
Rys. 1. Rozkład rekombinacji na całym chromosomie 6 człowieka (A) oraz na krótkich jego fragmentach zaznaczonych prostokątami.

sami całkowicie losowymi, a jedyna siła kierunkowa – presja selekcyjna – działa dopiero na produkty mutacji i rekombinacji (14). Obecnie jest raczej bezsporne, że zarówno mutacje, jak i rekombinacje nie są procesami całkowicie losowymi, a specyficzny charakter tych procesów obserwowany we współczesnych genomach jest wynikiem selekcji (6,15).

Przyjrzyjmy się rozkładowi rekombinacji wzajemnych (ang. *crossing over*) w chromosomach człowieka. Na rysunku 1A przedstawiono rozkład rekombinacji na całym

chromosomie 6 człowieka (dane pobrane z <http://www.hapmap.org/>) (12). Na osi x odłożono koordynaty pozycji markerów na chromosomie w milionach par zasad (Mpz), na osi y obserwowaną częstość rekombinacji na odcinku pomiędzy markerami. Na rysunku 1B przedstawiono rozkład rekombinacji w okienku wskazanym na diagramie chromosomu. Jeżeli dokonamy jeszcze większego zbliżenia, jak w panelach C i D (krótsze okienka), to da się zauważyć, że rozkład punktów rekombinacyjnych, jak się zdaje, jest nielosowy. Obserwujemy regiony, w których punkty rekombinacyjne występują znacznie częściej (rys. 1D) jak i rejony, w których są one istotnie niedoreprezentowane (rys. 1C). Nielosowe rozmieszczenie rekombinacji jest cechą charakterystyczną wielu genomów (16-19). Dla rozróżnienia tych obszarów wyróżnia się gorące i zimne miejsca rekombinacji (ang. *hot spots*, *recombination deserts*). U człowieka intensywność rekombinacji w gorących miejscach rekombinacji może wahać się w zakresie od 4×10^{-4} cM do 0,14 cM (20). Natomiast analiza całego ludzkiego genomu pozwoliła na identyfikację ponad 25 000 takich miejsc i oszacować, że odległość pomiędzy nimi wynosi średnio 50 kpz, a około 80% rekombinacji występuje w sekwencjach, o łącznej długości od 10 do 20% genomu (21). Sekwencje, w których skupiają się procesy wymiany pomiędzy niesiostrzanymi chromatydami często nie przekraczają długości od 1 do 2 kpz. Potwierdzono to w analizie nie tylko ludzkiego genomu, ale również genomów grzybów (22,23), myszy (17,20) oraz kukurydzy i *Arabidopsis thaliana* (24).

Wyciąganie statystycznie uprawnionych wniosków z rozkładów przedstawionych na rysunku 1 A-D jest, jak się wydaje, trudne. Spójrzmy jednak na przebieg krzywej na rysunku 2. Jest to tzw. krzywa kumulacyjna pozbawiona trendu wykonana w następujący sposób. Oblicza się całkowitą częstość rekombinacji dla całego chromosomu. Następnie, dla każdego odcinka reprezentowanego na wykresie oblicza się spodziewaną wartość rekombinacji, przy założeniu, że rekombinacje rozłożone są na całym chromosomie równomiernie i wartość tę odejmuje się od rzeczywistej wartości rekombinacji stwierdzanej na tym odcinku. Różnice uzyskane dla kolejnych odcinków dodaje się (kumuluje) i normalizuje przez liczbę analizowanych markerów. Tak zatem współrzędna y każdego punktu na wykresie pokazuje, jaka jest różnica między rzeczywistą a oczekiwaną częstością rekombinacji na całym odcinku chromosomu mierząc od jego początku do punktu określonego współrzędną x . Bardziej interesujące są jednak informacje o występowaniu lokalnych serii odcinków, na których rekombinacje zdarzają się częściej lub rzadziej niż spodziewana średnia. Jeżeli różnica między rzeczywistą częstością rekombinacji a średnią ma wartość dodatnią, to kolejny punkt na wykresie jest „wyżej” od poprzedniego. Jeżeli obserwowana częstość jest niższa od spodziewanej, kolejny punkt znajduje się „niżej”. Ostatni postawiony punkt krzywej kumulacyjnej pozbawionej trendu z definicji powinien mieć wartość $y = 0$. Na rysunku 2 widać (punkty czarne), że istnieją rejony badanego chromosomu, w którym krzywa wznosi się na dłuższym odcinku (rejony położone blisko telomerów) oraz jeden długi rejon, w którym krzywa opada (wnętrze chromosomu), gdzie częstość rekombinacji jest mniejsza od średniej dla tego chromosomu. Ten ob-



Rys. 2. Znormalizowane krzywe kumulacyjne pozbawione trendu sporządzone dla rozkładu częstości rekombinacji na chromosomie 6 człowieka (dane rzeczywiste) oraz na wirtualnym chromosomie (dane losowe), w którym częstości rekombinacji z chromosomu 6 zostały rozłożone losowo wzdłuż chromosomu. Dokładny opis krzywych w tekście. Liniami przerywanymi oznaczono obszar centromeru.

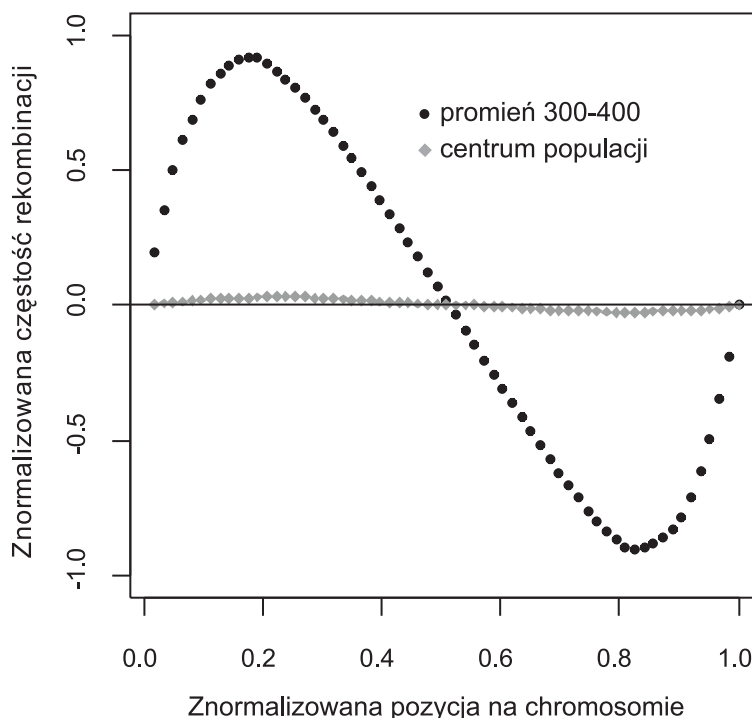
raz rozkładu częstości rekombinacji w chromosomach jest uniwersalnym dla wszystkich autosomów człowieka i chromosomu X (chromosom Y nie ma „sprawnego” partnera do rekombinacji) (25-29). Podobny rozkład obserwuje się również w genomach innych ssaków (28), w genomach roślinnych (30), w genomie *Caenorhabditis elegans* (31) i w genomie drożdży (32).

W wielu prostych testach statystycznych wykazuje się, że rozkład ten jest nielosowy. Najprostszym sposobem wykazania, że rozkład częstości rekombinacji na chromosomach jest nielosowy, jest sporządzenie takiej samej krzywej dla wirtualnego chromosomu, w którym rozkład częstości jest losowy (szare punkty na rys. 2). W tym przypadku, dla każdego odcinka na osi x przypisywano losowo częstość rekombinacji ze zbioru częstości rzeczywistego chromosomu. Z porównania tych krzywych należy wyciągnąć dwa wnioski; – rejonów chromosomów o wyraźnie wyższej częstości rekombinacji są o wiele dłuższe, niż można by się tego spodziewać przy rozkładach losowych i; – występują one głównie w obszarach subtelerowych. Z analizy porównawczej krzywych skumulowanych dla różnych chromoso-

mów (wyników nie pokazano tutaj) można wyciągnąć jeszcze inny wniosek, że środek odcinka o relatywnie małej częstości rekombinacji znajduje się raczej w środku chromosomu, a nie w rejonie centromeru, co wydawało się bardziej intuicyjne. Potwierdzono to również w innych badaniach (28).

3. Komputerowe modelowanie ewolucji genomów i chromosomów

Aby zrozumieć znaczenie nielosowego rozkładu rekombinacji, należałoby poznać mechanizmy, które mogą go generować. W tym celu przeprowadzono modelowanie komputerowe ewolucji chromosomów i genomów. W modelu komputerowym, wirtualne osobniki reprezentowane są przez ich diploidalne genomy. Genom złożony jest z określonej liczby par chromosomów homologicznych. Każdy chromosom jest ciągiem bitowym, w którym zero reprezentuje gen funkcjonalny, a 1 gen zmutowany. Dla uproszczenia można przyjąć, że geny zmutowane są recesywne, a rewersje nie zdarzają się. Częstość mutacji najczęściej określa się na 1 mutację w losowo wybranym locus na genom na generację, co odpowiada częstościom spotykanym w naturze. Najbardziej krytycznym parametrem symulacji jest częstość rekombinacji. We wszystkich opisywanych symulacjach przyjmowano, że *crossing over* występuje tylko przy produkcji gamet i jego rozkład jest równomierny na całym chromosomie. Okazało się, że w tak prowadzonych symulacjach ewolucji populacji można zaobserwować pojawianie się specyficznej struktury chromosomów, w zależności od wielkości chromosomów (liczby genów na chromosomie), częstości rekombinacji oraz wielkości populacji (6,15). Odległości między osobnikami nabierają znaczenia, jeżeli populacja ewoluje w ograniczonej przestrzeni (6). W takiej populacji odległość, w granicach której partnerzy mogą się spotkać decyduje o stopniu pokrewieństwa między nimi (współczynnika inbrodu) nawet jeżeli nie ma żadnych barier fizycznych, geograficznych ani innych preferencji w wyborze partnera do rozrodu. Przestrzeń dostępną dla populacji opisuje się siatką kwadratową o odpowiednio dużych rozmiarach, na przykład 1000x1000. W jednym oczku siatki może „żyć” tylko jeden wirtualny osobnik. Jeżeli symulacje zacznie się w centrum siatki od jednej pary doskonałych osobników, bez jakichkolwiek defektów genetycznych (Adama i Ewy), to łatwo sobie wyobrazić ekspansję populacji w czasie. Ze względu na ograniczoną możliwość wędrowania po siatce, populacja zajmuje coraz większy teren, w przybliżeniu kołysty. Sytuacja osobników na tym terenie jest jednak różna. W centrum obserwuje się „przeludnienie” – nowe osobniki mogą się rodzić tylko wtedy, jeżeli starsze umrą. Natomiast na krawędziach populacji, młode osobniki mają miejsce na jeszcze nie skolonizowanych terenach. Te osobniki mogą jednak znajdować partnerów tylko od strony centrum populacji, wobec czego współczynnik inbrodu jest tam większy. Można to porównać z problemami osadników na „Dzikim Zachodzie”; sytuacja pionierów była zupełnie inna niż ich przodków w miastach, z których oni wyemigrowali. Porównanie puli genetycznej tych dwóch klas subpopulacji –



Rys. 3. Znormalizowane krzywe kumulacyjne pozbawione trendu sporządzone dla rozkładu częstości rekombinacji akceptowanych w wirtualnych chromosomach populacji ewoluującej na siatce kwadratowej. Wykresy sporządzono na podstawie częstości rekombinacji w chromosomach osobników znajdujących się w centrum populacji (szare romby) i chromosomów osobników na obrzeżach populacji (czarne kółka). Na obszarach ekspandującej populacji (promień 300-400), gdzie współczynnik inbrodu jest wysoki, rozkład częstości rekombinacji odpowiada rozkładowi obserwowanemu w chromosomach naturalnych (por. z rys. 2).

centrum ekspandującej populacji i „pionierów” pokazuje zasadniczą różnicę między nimi – rozkład częstości akceptowanych rekombinacji. Akceptowane rekombinacje, to te *crossing over* między homologicznymi chromosomami, których produktami są zygoty, dające w rezultacie przeżywające osobniki „żyjące” na siatce. Okazuje się, że rozkład częstości rekombinacji wśród pionierów jest analogiczny do tego, obserwowanego w naturalnych chromosomach, podczas gdy rozkład rekombinacji w centrum populacji jest równomierny, taki jak przyjęto w parametrach modelu (rys. 3) (6). Różnica między tymi subpopulacjami jest tylko we współczynniku inbrodu. Model pozwala, nam zatem na postawienie pytania – dlaczego pokrewieństwo między rodzicami wpływa na rozkład akceptowanych rekombinacji między homologicznymi chromosomami? Aby odpowiedzieć na to pytanie należy zdać sobie sprawę z tego, jaki jest związek między częstością *crossing over* a współczynnikiem inbrodu. Pozwólmy sobie na nieco uproszczony opis tego zjawiska. Rodzice, produ-

kując gamety, przeprowadzają bardzo ograniczone mieszanie informacji genetycznej uzyskanej od swoich rodziców. Na przykład na chromosomie 19, zawierającym niemal 1500 genów zachodzi tylko jeden *crossing over* (33). Zakładając, że rodzeństwo może się stać partnerami, należy przyjąć, że wnuki uzyskają bardzo duże rejonny genomów, które będą identyczne, pochodzące z tego samego chromosomu prarodzica i wtedy należy się spodziewać ewentualnie spotkania tych samych defektywnych alleli, ale można się również spodziewać, że u wnuka spotkają się układy bardzo długich fragmentów genomów lub nawet całych chromosomów w takim samym układzie jak u prarodzica. Ponieważ taki układ został już sprawdzony przez selekcję, jest on zatem do zaakceptowania. Im większa jest odległość genetyczna między rodzicami, tym większe prawdopodobieństwo, że doszło do rekombinacji między dowolnymi genami. Prawdopodobieństwo wystąpienia rekombinacji na danym odcinku jest zatem funkcją odległości genetycznej mierzonej w liczbie mejoz między dwoma osobnikami i prawdopodobieństwa wystąpienia rekombinacji w pojedynczej mejozie. Takie rozważanie wyjaśnia wyniki badania dzietności w rodzinach islandzkich (34). Okazało się, że najwięcej dzieci rodzi się w rodzinach, w których stopień pokrewieństwa między małżonkami jest najwyższy dopuszczalny prawem. Dzietność maleje monotonicznie wraz z odległością genetyczną między rodzicami. Sukces ewolucyjny, mierzony liczbą wnuków jest natomiast najwyższy dla małżeństw w trzecim stopniu kuzynostwa. Ten wynik badań jest do tego stopnia nieintuicyjny, że nie dawano wiary wcześniejszym publikacjom wskazującym na takie zależności dzietności od stopnia pokrewieństwa między rodzicami.

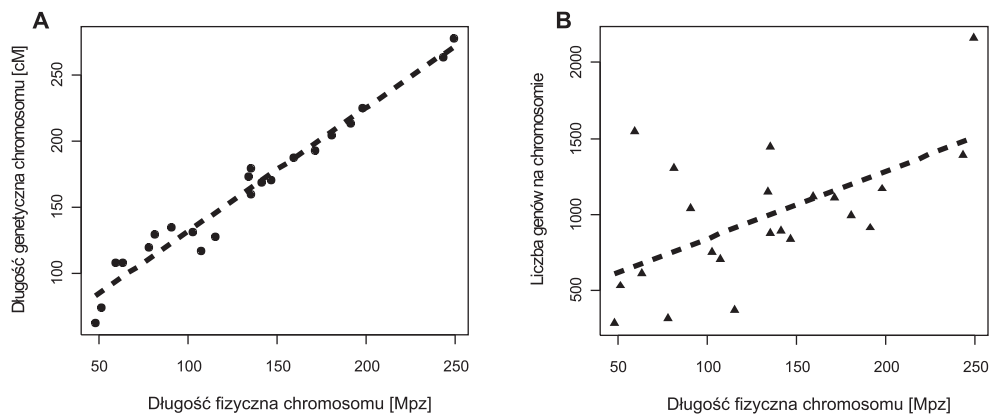
4. Czy selekcja może faworyzować nierównomierny rozkład rekombinacji?

Wyniki wykazujące nielosowy rozkład rekombinacji wzdłuż chromosomów, przedstawiono na rysunkach 2 oraz 3 i pokazują one jedynie, że gamety, których fuzyja dała przeżywające zarodki zostały utworzone po rekombinacjach nierównomiernie rozłożonych na chromosomach, czyli są to wyniki po selekcji. Bezpośrednio możemy to obserwować w modelu komputerowym, gdzie implementowany rozkład rekombinacji jest losowy, natomiast przeżywające osobniki posiadają chromosomy tworzone przez rekombinacje rozłożone nielosowo (6). Z wyników symulacji wyraźnie widać, że to selekcja nie dopuszcza do przeżycia osobników, albo istotnie zmniejsza szansę przeżycia osobników, które dziedziczą chromosomy po *crossing over* w centrum. Patrząc na te wyniki z punktu widzenia kosztów ewolucyjnych, wydaje się oczywiste, że bardziej ekonomicznie byłoby wprowadzić taki mechanizm rekombinacyjny, który bardziej odpowiadałby „oczekiwaniom” selekcji. Kowalski i wsp. (8) opracowali model komputerowy, w którym częstość rekombinacji pomiędzy poszczególnymi genami była dziedziczona, ale mogła ulegać powolnej ewolucji. Na podstawie wyników symulacji wykazano, że rozkład częstości rekombinacji ewoluował w ten sposób, że po pewnym czasie odpowiadał on faktycznemu rozkładowi

rekombinacji akceptowanych. Selekcja doprowadziła zatem do rozkładu minimalizującego koszty ewolucyjne. W innych symulacjach (35) potwierdzono te wyniki i wykazano istnienie ujemnego sprzężenia zwrotnego w tym procesie ewolucji; rozsuniecie częstości rekombinacji w kierunku telomerów nie jest nieograniczone, zbytne zwiększenie częstości rekombinacji w regionach subtelerowych jest niekorzystne i również kontrolowane przez selekcję.

5. Znaczenie nierównomiernego rozkładu częstości rekombinacji

Jeżeli selekcja preferuje nierównomierny rozkład rekombinacji na chromosomach, to muszą istnieć różnice w przeżywalności osobników zależne od położenia punktu rekombinacji i ich częstości, co wykazano w symulacjach komputerowych (36). Istotnym problemem jest jednak, w jaki sposób nielosowy rozkład rekombinacji może wpływać na inne elementy struktury chromosomów powodując zmianę przystosowania i przeżywalności organizmów. W symulacjach komputerowych wykazano, że populacje mogą przyjmować dwie różne strategie ewolucji w zależności od częstości rekombinacji, a ściślej w zależności od liczby genów przypadających na jednostkę rekombinacyjną (15,37). Jeżeli ta liczba jest wysoka, to całe serie genów dziedziczą się razem (są sprzężone) i istnieje tendencja do komplementowania defektów w takich blokach genów (klastrach). W genomach diploidalnych pojawiają się klastry homologiczne, które wzajemnie się uzupełniają – komplementują. Rekombinacja wewnątrz nich jest niedozwolona, ponieważ prowadzi do pojawienia się nowego typu klastra, który nie pasuje do już istniejących w puli genetycznej populacji. Jeżeli gęstość kodowania jest niska i na jednostkę rekombinacyjną przypada mała liczba genów, wtedy klastry się nie tworzą, a darwinowska selekcja oczyszczająca bardzo wydajnie eliminuje defektywne allele z puli genetycznej populacji. Przejście pomiędzy tymi dwoma strategiami ewolucyjnymi ma charakter przejścia fazowego, jest bardzo wrażliwe na gęstość kodowania i wielkość populacji efektywnej (36). W warunkach symulacji komputerowych, średnia różnica w częstości rekombinacji wynosząca zaledwie 0,1 cM na chromosom może decydować o wyborze jednej ze strategii (9,36). Przejście pomiędzy tymi strategiami jest niekorzystne dla potencjału rozrodczego populacji. W tym samym genomie mogą jednak istnieć pary chromosomów wykorzystujące różne strategie – komplementowania lub selekcji oczyszczającej – zależy to od gęstości kodowania chromosomu, całkowitej liczby zawartych w nim genów i częstości rekombinacji. Porównanie chromosomów człowieka pokazuje, że różnią się one bardzo istotnie pod względem tych parametrów (tab.). Korelacja między fizyczną wielkością chromosomów a ich długością mierzoną w jednostkach rekombinacyjnych jest bardzo wysoka, natomiast korelacja między wielkością chromosomu a liczbą genów jest niska (rys. 4). Na przykład chromosomy 18 i 19 są bardzo podobne pod względem wielkości fizycznej i częstości rekombinacji, natomiast różnią się wielokrotnie pojemnością kodowania (33,38-40). W bada-



Rys. 4A. – Zależność między fizycznymi rozmiarami chromosomów człowieka mierzonych w milionach par zasad a ich rozmiarami genetycznymi, mierzonymi w centyMorganach. B – Zależność między fizycznymi rozmiarami chromosomów człowieka a liczbą kodowanych przez nie genów.

niach komputerowych wykazano, że wiele z chromosomów genomu człowieka znajdowałoby się w warunkach bliskich przejściu fazowemu, gdyby człowiek ewoluował w populacjach efektywnych o wielkości rzędu kilkuset osobników (41). Wydaje się, że ta wielkość populacji jest do przyjęcia. Wskazują na to też oszacowania oparte na analizach teoretycznych i danych genetycznych pochodzących z 52 populacji ludzkich (42). Tak jak wspomniano, przejście z jednej strategii ewolucyjnej w drugą jest bardzo kosztowne ewolucyjnie, co związane jest z drastycznym spadkiem potencjału rozrodczego populacji (36). Należy przypuszczać, że struktura chromosomów i całych genomów populacji, po dłuższej ewolucji w warunkach określonego inbrodu i częstości rekombinacji, dostosowuje się do tych warunków, a ich zmiana może negatywnie wpływać na potencjał rozrodczy populacji. Wyniki badań dzietności i pokrewieństwa w małżeństwach islandzkich (34) mogą odzwierciedlać właśnie tę sytuację.

Tabela

Charakterystyka ludzkich autosomów

Nr	Fizyczny rozmiar chromosomu w Mpz*	Długość genetyczna chromosomu w cM**	Liczba genów*
1	2	3	4
1	249,3	278,1	2163
2	243,2	263,4	1390
3	198,0	224,6	1168
4	191,2	213,2	911

1	2	3	4
5	180,9	204,0	993
6	171,1	193,0	1110
7	159,1	187,0	1115
8	146,4	170,2	839
9	141,2	168,3	889
10	135,5	179,4	876
11	135,0	159,5	1445
12	133,9	173,0	1148
13	115,2	127,2	369
14	107,3	117,1	707
15	102,5	131,4	753
16	90,4	135,0	1042
17	81,2	129,5	1308
18	78,1	119,6	317
19	59,1	107,9	1544
20	63,0	108,1	608
21	48,1	62,3	280
22	51,3	73,6	529

*Dane pobrano z bazy Ensembl (wersja 56) znajdującej się pod adresem: <http://www.ensembl.org/index.html>. Liczba genów obejmuje zarówno geny znane jak i nowe. **Długości map genetycznych dla autosomów pobrano z bazy HapMap, z adresu: <http://www.hapmap.org/>

6. Struktura rzeczywistych chromosomów

W genomie człowieka geny zajmują około 30% sekwencji, ale same egzony stanowią niewiele ponad 1% (39). Geny nie są jednak upakowane w chromosomach w równym stopniu, niektóre chromosomy są bogate w geny, jak na przykład chromosom 19 u człowieka (33), gdzie średnio 28 genów przypada na 1 Mpz, przy średniej wartości dla genomu rzędu 10 genów na 1 Mpz. Przykładem chromosomów ubogich w geny w genomie człowieka jest chromosom 18 (40) i 13 (43), zawierające odpowiednio 4,4 geny i 6,5 genów na 1 Mpz. W chromosomie 13 jest nawet region, w którym liczba genów spada do 3 na 1 Mpz. Innym przykładem nierównomiernego rozkładu genów są chromosomy 2 i 19 człowieka, różniące się czterokrotnie wielkością, a zawierające zbliżoną liczbę genów. W gęsto upakowanym chromosomie 19 prawie 30% sekwencji kodujących zorganizowanych jest w duże klastry, z których większość ma swoje homologiczne odpowiedniki w syntenicznym położeniu w genomie myszy (44). Właśnie takie struktury są przewidywane w modelowaniu komputerowym. Innymi przykładami klastrów są duże domeny tzw. RIDGEs (ang. *Region of Increased Gene Expression*) złożone głównie z genów *housekeeping*, czyli ge-

nów odpowiedzialnych za utrzymanie głównych funkcji życiowych komórki i ulegających ekspresji w każdej tkance (45,46). Szacuje się, że w genomie człowieka około 65% genów podlegających skoordynowanej ekspresji tworzy klastry, a grupowanie genów jest szczególnie częste na chromosomach charakteryzujących się wysoką gęstością kodowania (46).

Na szczególną uwagę zasługują dwie rodziny genów, występujących w klastrach: geny głównego układu zgodności tkankowej (MHC, ang. *Major Histocompatibility Complex*) kodujące białka odpowiedzialne za prezentację antygenów limfocytom T, oraz geny receptorów węchowych (ang. OR – *Olfactory Receptors genes*), których liczba w genomie człowieka sięga około 1000. Klastry genów OR występują na wszystkich chromosomach człowieka oprócz chromosomu 20 i Y (47). Rola tych genów nie ogranicza się jedynie do produkcji receptorów na powierzchni komórek węchowych, ponieważ stwierdza się ich ekspresję również w czasie spermatogenezy (48-50). Przypuszcza się, że mogą brać udział w rozpoznawaniu gamet oraz mogą współdziałać z genami MHC w doborze płciowym. Przesłanką wskazującą na takie współdziałanie jest sprzężenie jednego z klastrów OR z klastrem MHC (51). Takie sprzężenia klastrów określono terminem superklastrów, a przykładem superklastra jest właśnie rozszerzony klaster MHC (ang. *extended MHC*) (52).

Ewolucjonowanie chromosomów eukariotycznych i pojawianie się w ich strukturach klastrów genów, w których rekombinacje obniżają szanse przeżycia nosiciela wskazuje na istnienie opisanego zjawiska komplementowania funkcji klastrów w diploidalnym genomie. Taką hipotezę potwierdzono we wcześniej opisanych komputerowo symulacjach ewolucji genomów (6,9,15,36). Przyjęcie tej hipotezy wymaga jednak dodatkowego założenia, że układy par chromosomów w diploidalnym genomie są nielosowe – należy się spodziewać istnienia preselekcji gamet. Preselekcja gamet bardzo istotnie zwiększyłaby potencjał rozrodczy populacji (53), i być może stanowiłaby również rozwiązanie dylematu Haldane'a (54). Dylemat Haldane'a wynika z szacunków wskazujących, że koszty ewolucji są, jak się wydaje, tak wielkie, że sama ewolucja posługująca się znanymi nam dotychczas mechanizmami jest niemożliwa. Sposób dziedziczenia superklastra MHC może, jak się wydaje, potwierdzać istnienie preselekcji gamet. Produkty genów MHC kooperują w procesie rozpoznawania immunologicznego z produktami genów klastra genów KIR (KIR, ang. *Killer Cell Immunoglobulin Like Receptor*) kodujących immunoglobulinopodobne receptory komórek cytotoksycznych (55). Niektóre klastry MHC (zwane haplotypami) preferencyjnie występują wspólnie z pewnymi haplotypami KIR (56-61). Takie sprzężenie nie byłoby niczym szczególnym, gdyby te dwa klastry leżały obok siebie na tym samym chromosomie, ale one są zlokalizowane na różnych chromosomach – szóstym i dziewiętnastym. Na chromosomie 19. zlokalizowano również klaster innych genów, odpowiadających za supresję reakcji immunologicznej matki na rozwijający się w niej płód (62).

7. Wnioski

W przedstawionych danych, będących wynikiem zarówno eksperymentalnych badań genomów, jak i symulacji komputerowych wykazano jednoznacznie, że mechanizmy rekombinacji, które w głównej mierze są odpowiedzialne za strukturę i funkcję genomów same podlegają selekcji i ewoluują. Relacje między prawdopodobieństwem rekombinacji między genami a funkcją tych genów są bardzo skomplikowane i nie można ich poznać analizując pojedyncze genomy, należy przeprowadzać analizy porównawcze wielu blisko spokrewnionych genomów w celu poszukiwania korelacji zarówno funkcjonalnych, jak i strukturalnych. Szczególnie interesujące jest, jak się wydaje, poszukiwanie „stanów zakazanych” – układów, które ze statystycznego punktu widzenia powinny występować, a nie stwierdza się ich. Znajomość takich stanów mogłaby rozwiązać dylemat Haldane’a, ale byłaby nieoceniona również dla biotechnologa konstruującego organizm transgeniczny.

Literatura

1. Hardy G. H., (1908), *Science*, 28, 49-50.
2. Weinberg W., (1908), *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, 64, 368-382.
3. Fisher R. A., (1930), *The genetical theory of natural selection*, Clarendon Press, Oxford.
4. Wright S., (1931), *Genetics*, 16, 97-159.
5. Wright S., (1932), *Proceedings of the 6th International Congress of Genetics*, 1, 356-366.
6. Waga W., Mackiewicz D., Zawierta M., Cebrat S., (2007), *Theory Biosci.*, 126, 53-59.
7. Cebrat S., Mackiewicz D., Waga W., (2009), *Kosmos*, 58, 29-41.
8. Kowalski J., Waga W., Zawierta M., Cebrat S., (2009), *Int. J. Mod. Phys. C*, 20, 1299-1309.
9. Waga W., Zawierta M., Kowalski, Cebrat S., (2009), in: *Series on Advances in Mathematics for Applied Sciences*, vol. 79, Eds. J. Miekisz, M. Lachowicz, 70-102, World Scientific, Singapore.
10. Mayr E., (1942), *Systematics and the Origin of Species*, Columbia Press, New York.
11. Jiggins C. D., (2006), *Curr. Biol.*, 16, R333-R334.
12. The International HapMap Consortium. The International HapMap Project, (2003), *Nature*, 426, 789-796.
13. Morgan T. H., Sturtevant A. H., Muller H. J., Bridges C. B., (1915), *The Mechanism of Mendelian Heredity*, Henry Holt and Company.
14. Ayala F. J., Kiger J. A., (1980), *Modern Genetics*, The Benjamin/Cummings Pub. Comp. Inc., California.
15. Zawierta M., Biecek P., Waga W., Cebrat S., (2007), *Theory Biosci.*, 125, 123-132.
16. Jeffreys A. J., Kauppi L., Neumann R., (2001), *Nat. Genet.*, 29, 217-222.
17. Petes T. D., (2001), *Nat. Rev. Genet.*, 2, 360-369.
18. Yu A., Zhao C., Fan Y., Jang W., Mungall A. J., Deloukas P., Olsen A., Doggett N. A., Ghebranious N., Broman K. W., Weber J. L., (2001), *Nature*, 409, 951-953.
19. Mézard C., (2006), *Biochem. Soc. Trans.*, 34, 531-534.
20. Kauppi L., Jeffreys A. J., Keeney S., (2004), *Nature Rev. Genet.*, 5, 413-424.
21. Myers S., Bottolo L., Freeman C., McVean G., Donnelly P., (2005), *Science*, 310, 321-324.
22. Lichten M., Goldman A. S., (1995), *Annu. Rev. Genet.*, 29, 423-444.
23. de Massy B., (2003), *Trends Genet.*, 19, 514-522.
24. Gaut B. S., Wright S. I., Rizzon C., Anderson L. K., Dvorak J., (2007), *Nature Rev. Genet.*, 8, 77-84.
25. Payseur B. A., Nachman M. W., (2000), *Genetics*, 156, 1285-1298.

26. Kong A., Gudbjartsson D. F., Sainz J., Jonsson G. M., Gudjonsson S. A., Richardsson B., Sigurdardottir S., Barnard J., Hallbeck B., Masson G., et al., (2002), *Nat. Genet.*, 31, 241-247.
27. Nachman M. W., (2002), *Curr. Opin. Genet. Devel.*, 12, 657-663.
28. Jensen-Seaman M. I., Furey T. S., Payseur B. A., Lu Y., Roskin K. M., Chen C.F., Thomas M. A., Hausler D., Jacob H. J., (2004), *Genome Res.*, 14, 528-538.
29. Cheung V. G., Burdick J. T., Hirschmann D., Morley M., (2007), *Am. J. Hum. Genet.*, 80, 526-530.
30. Tenailon M. I., Sawkins M. C., Anderson L. K., Stack S. M., Doebley J., Gaut B. S., (2002), *Genetics*, 162, 1401-1413.
31. Prachumwat A., DeVincentis L., Palopoli M. F., (2004), *Genetics*, 166, 1585-1590.
32. Barton A. B., Pekosz M. R., Kurvathi R. S., Kaback D. B., (2008), *Genetics*, 179, 1221-1235.
33. Grimwood J., Gordon L. A., Olsen A., Terry A., Schmutz J., Lamerdin J., Hellsten U., Goodstein D., Couronne O., Tran-Gyamfi M., et al., (2004), *Nature*, 428, 529-535.
34. Helgason A., Pálsson S., Guðbjartsson D. F., Kristjánsson Þ., Stefánsson K., (2008), *Science*, 319, 813-816.
35. de Oliveira P. M. C., Sa Martins J. S., Moss de Oliveira S., Stauffer D., Cebrat S., (2009), Self-organization of crossing points for biological recombination, (personal communication).
36. Zawierta M., Waga W., Mackiewicz D., Biecek P., Cebrat S., (2008), *Int. J. Mod. Phys. C*, 19, 917-926.
37. Bońkowska K., Kula M., Cebrat S., Stauffer D., (2007), *Int. J. Mod. Phys. C*, 18, 1329-1338.
38. International Human Genome Sequencing Consortium, (2001), *Nature*, 409, 860-921.
39. Venter J. C., Adams M. D., Myers E. W., Li P. W., Mural R. J., Sutton G. G., Smith H. O., Yandell M., Evans C. A., Holt R. A., et al., (2001), *Science*, 291, 1304-1351.
40. Nusbaum C., Zody M. C., Borowsky M. L., Kamal M., Kodira C. D., Taylor T. D., Whittaker C. A., Chang J. L., Cuomo C. A., Dewar K., et al., (2005), *Nature*, 437, 551-555.
41. Mackiewicz D., Zawierta M., Waga W., Cebrat S. (2010), Genome analyses and modelling the relationships between coding density, recombination rate and chromosome length. arXiv:1001.5419, *J. Theor. Biol.*, 267, 186-192.
42. Liu H., Prugnolle F., Manica A., Balloux F., (2006), *Am. J. Hum. Genet.*, 79, 230-237.
43. Dunham A., Matthews L. H., Burton J., Ashurst J. L., Howe K. L., Ashcroft K. J., Beare D. M., Burford D. C., Hunt S. E., Griffiths-Jones S., et al., (2004), *Nature*, 428, 522-528.
44. Dehal P., Predki P., Olsen A. S., Kobayashi A., Folta P., Lucas S., Land M., Terry A., Ecale Zhou C. L., Rash S., et al., (2001), *Science*, 293, 104-111.
45. Zhang L., Li W. H., (2004), *Mol. Biol. Evol.*, 21, 236-239.
46. Elizondo L. I., Jafar-Nejad P., Clewing J. M., Boerkoel C. F., (2009), *Curr. Genomics*, 10, 64-75.
47. Glusman G., Yanai I., Rubin I., Lancet D., (2001), *Genome Res.*, 11, 685-702.
48. Spehr M., Gisselmann G., Poplawski A., Riffell J. A., Wetzel C. H., Zimmer R. K., Hatt H., (2003), *Science*, 299, 2054-2058.
49. Spehr M., Schwane K., Heilmann S., Gisselmann G., Hummel T., Hatt H., (2004), *Curr Biol.*, 14, R832-R833.
50. Feldmesser E., Olender T., Khen M., Yanai I., Ophir R., Lancet D., (2006), *BMC Genomics*, 7, 121.
51. Ziegler A., Dohr G., Uchanska-Ziegler B., (2002), *Am. J. Reprod. Immunol.*, 48, 34-42.
52. Horton R., Wilming L., Rand V., Lovering R. C., Bruford E. A., Khodiyar V. K., Lush M. J., Povey S., Talbot C. C. Jr, Wright M. W., et al., (2004), *Nat. Rev. Genet.*, 5, 889-899.
53. Cebrat S., Stauffer D., (2008), *Int. J. Mod. Phys. C*, 19, 259-265.
54. Haldane J. B. S., (1957), *J. Genet.*, 55, 511-524.
55. Carrington M., Norman P. J., (2003), *The KIR Gene Cluster*, Vol. 2003, National Library of Medicine (U.S.), National Center for Biotechnology Information, Bethesda, MD, USA.
56. Trowsdale J., (2001), *Immunity* 15, 363-374.
57. Martin M. P., Gao X., Lee J. H., Nelson G. W., Detels R., Goedert J. J., Buchbinder S., Hoots K., Vlahov D., Trowsdale J., et al., (2002), *Nat. Genet.*, 31, 429-434.
58. Hiby S. E., Walker J. J., O'shaughnessy K. M., Redman C. W., Carrington M., Trowsdale J., Moffett A. J., (2004), *Exp. Med.*, 200, 957-965.
59. Kelley J., Trowsdale J., (2005), *Transplant. Immunology*, 14, 129-134.

60. Parham P., (2005), *Nat. Rev. Immunol.*, 5, 201-214.
61. Yawata M., Yawata N., Draghi M., Little A. M., Partheniou F., Parham P., (2006), *J. Exp. Med.*, 203, 633-645.
62. Than N. G., Romero R., Goodman M., Weckle A., Xing J., Dong Z., Xu Y., Tarquini F., Szilagyí A., Gal P., et al., (2009), *Proc. Natl. Acad. Sci. USA.*, 106, 9731-9736.