

JAN FILIPSKI

*Institut J. Monod, Uniwersytet Paris VI i Paris VII.  
2 place Jussieu, Tour 43, 75251, PARIS Cedex 05  
e-mail: filipski@ijm.jusiev.fr*

## STRUKTURA I EWOLUCJA SEKWENCJI DNA W CHROMOSOMIE CZŁOWIEKA

### GENEZA CHROMOSOMU

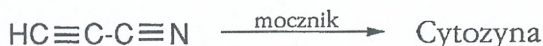
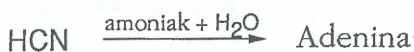
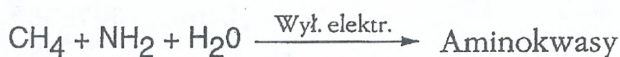
Chromosomy zostały odkryte przez W. Waldeyera już ponad 100 lat temu, w 1888r., jako intensywnie barwiące się (stąd nazwa) organelle, które pojawiają się w komórce Eukaryota tuż przed podziałem mitotycznym i które zanikają w powstających jądrach komórkowych po jego zakończeniu. W 1902 r. W. S. Sutton wysunął hipotezę, że chromosom jest nosicielem mendelowskich genów, co potwierdził rozwój biologii molekularnej. Jednak dopiero obecnie zaczynamy zdawać sobie sprawę z konsekwencji jakie, nawet niepełna wiedza o strukturze chromosomu i technologii wypracowane w trakcie jej zdobywania, pociągają za sobą, jak zaczynają wpływać na nasze życie, a także na sposób w jaki patrzymy na samych siebie i na otaczający nas świat.

Obecnie definiuje się chromosom jako nośnik cech dziedzicznych organizmów żywych oraz wirusów, rodzaj bazy danych zbudowanej z kompleksu kwasów nukleinowych (najczęściej DNA) z białkami. Suma informacji genetycznej zawartej w chromosomach jakiegoś gatunku (wirusa, rośliny, bakterii czy zwierzęcia) lub w DNA mitochondrialnym czy chloroplastycznym, nazywa się genomem i może być zapisana w postaci sekwencji nukleotydowej cząsteczki DNA ale można ją zapisać również na kartce papieru czy w pamięci komputera.

Okoliczności związane z pojawieniem się pierwszego chromosomu są przedmiotem licznych spekulacji między innymi również tej, że został zawleczony na ziemię celowo lub przypadkiem. Nawet jeśli zdarzenie takie miało miejsce, to hipoteza ta nie odpowiada na pytanie skąd się wziął pierwszy chromosom. Wśród specjalistów przeważa opinia, że przodek wszystkich chromosomów, z ludzkimi włącznie, był produktem miejscowym (ORGEL 1998).

Warunki jakie panowały na ziemi przez pierwsze 500 milionów lat po powstaniu układu słonecznego, nieco ponad 4.5 miliarda lat temu, sprzyjały raczej rozkładowi związków organicznych niż ich syntezie. Z drugiej strony, pierwsze geologiczne formacje, których skład chemiczny i izotopowy nie daje się wytłumaczyć inaczej, jak tylko zakładając istnienie organizmów żywych (a więc posiadających chromosomy), datuje się na około 3.85 miliarda lat. Co działo się w okresie kilkuset milionów lat zamkniętych tymi dwoma nawiasami właściwie nie wiadomo. Większość badaczy sądzi, że atmosfera ziemska w tym okresie była beztlenowa i zawierała głównie azot, wodór, metan, dwutlenek węgla, cyjanowodór, amoniak i parę wodną. Doświadczenia S. L. Millera pokazały, że wyładowania elektryczne w mieszaninie zawierającej metan, amoniak, wodór i parę wodną powodują powstanie wielu aminokwasów w tym również takich, które występują w organizmach żywych. Inni, zmieniając nieco warunki, udowodnili powstawanie adeniny, cytozyny i uracylu, a także cukrów (w tym i rybozy) z takich prostych składników, jak para wodna, azot, metan i formaldehyd (Ryc. 1). Jeśli w atmosferze ziemskiej rzeczywiście panowały kiedyś warunki zbliżone do tych, jakie stosowali autorzy wzmiankowanych doświadczeń, to niewątpliwie w kałużach, jeziorach i oceanach młodej podówcza Ziemi rozpuszczone były związki chemiczne stanowiące elementarne cegiełki struktury chromosomu. Hipotetyczny ten roztwór nazywany jest bulionem pierwotnym lub zupą prebiotyczną.

Chyba każdy z nas wykonywał w ramach lekcji fizyki doświadczenie polegające na tym, że nasycony roztwór na przykład sacharozy, w którym zawieszono na nitce kryształek cukru, pozostawiano do powolnego odparowania. Po

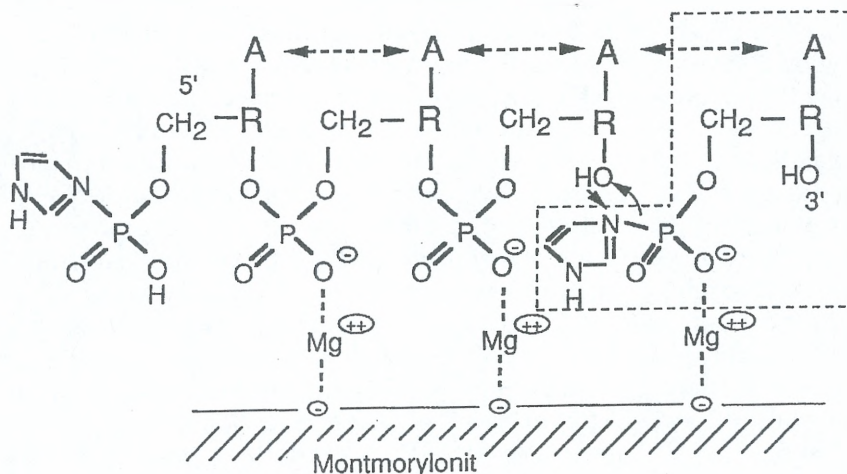


Ryc. 1. Synteza prekursorów biopolimerów. Podobne reakcje mogły zachodzić w „bulionie pierwotnym” (według ORGELA 1998, zmodyfikowane)

kilku dniach stwierdzić można było, że kryształ powiększył się zachowując jednak formę sześciangu. Rosnący kryształ niejako wymusza takie przestrzenne usytuowanie się kolejnych adsorbujących się na jego powierzchni cząsteczek, że geometria sieci zostaje zachowana. Na podobnej zasadzie powstały prawdopodobnie pierwsze cząsteczki polimeru, które można by nazwać proto-chromosomem. Wykazano między innymi, że sieć krystaliczna pospolitego, występującego w glinie minerału montmorylonitu, ułatwia ustawianie się w szereg i polimeryzację aktywowanego rybonukleotydu (5' fosforimidazolidu adenozyne) w wyniku czego powstaje oligomer RNA (Ryc. 2) (FERRIS i ERTEM 1993). Doświadczenia tego typu mają na celu, przynajmniej na razie, nie tyle odtworzenie warunków, w których powstał pierwszy chromosom, ile udowodnienie, że spontaniczna polimeryzacja nukleotydów, która warunkuje powstanie jego prymitywnej formy, tak zwanego replikatora, jest w ogóle możliwa. Replikator jest to hipotetyczna cząsteczka lub kompleks cząsteczkowy, który ma zdolność tworzenia swojej własnej kopii, albo bezpośrednio, bądź też syntetyzując uprzednio matrycę, na której ta kopia powstaje. Ma więc zdolność rozmnażania się. Kolejne kopie replikatora, który kiedyś pojawił się w zupie prebiotycznej różniły się od siebie w drobnych szczegółach struktury, a co za tym idzie w trwałości i we własnościach katalitycznych.

Najtrwalsze z replikatorów trwały z definicji dłużej niż inne a najlepsze autokatalizatory najszybciej się rozmnażały. Powstawały pewnie i takie replikatory, które rozkładały inne replikatory na elementy składowe i z tych elementów budowały własną strukturę. Replikator, który miał skłonności do izolowania się od otoczenia utworzył rodzaj prymitywnej komórki (DAWKINS 1976).

Jeśli chodzi o naturę chemiczną pierwszych replikatorów jesteśmy skazani na domysły. Prawdopodobnie nie były to cząsteczki kwasu dezoksyrybonukleinowego i niekoniecznie były to oligorybonukleotydy. Jakkolwiek by nie było, zarówno replikatory, jak i skład prebiotycznego bulionu ewoluowały. Przypuszcza się, że po pewnym okresie powstało w nim coś co się określa nazwą „świat RNA” (GILBERT 1986). Kompleksy makromolekularne istniejące w tym świecie miały „chromosomy” zbudowane z kwasu rybonukleinowego. Z RNA zbudowane były również katalizatory (rybozimy), które umożliwiały zachodzenie reakcji niezbędnych dla trwania i rozmnażania się tych kompleksów. Nie istniał więc podział na fenotyp i genotyp ani potrzeba translacji. Zsyntetyzowano w laboratorium szereg rybozymów będących w stanie katalizować różne reakcje biochemiczne, co czyni istnienie „świata RNA” prawdopodobnym. Za tym, że „świat RNA” rzeczywiście istniał przemawia również fakt, że niektóre współczesne



Ryc. 2. Tworzenie wiązania fosfodwuestrowego prowadzącego do syntezy 3'-5' oligorybonukleotydów, katalizowane przez adsorbacje aktywowanych monomerów na powierzchni montmorylonitu.

A — reszta adeniny, R — reszta rybozy, w ramce 5' fosforimidazolid adenozyne. Strzałki pomiędzy resztami adeniny symbolizują oddziaływania o charakterze niejonowym (według FERRISA i ERTEMA 1993, zmodyfikowane).

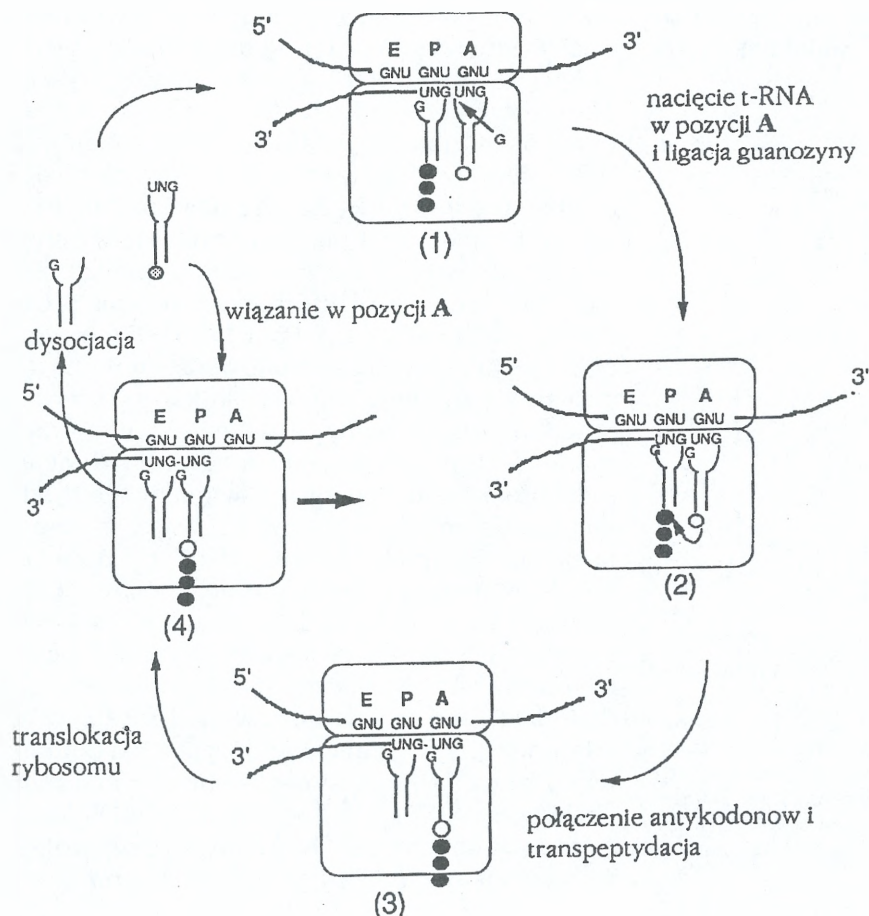
organizmy stosują rybozomy mimo, że białka potrafiłyby z pewnością wykonywać podobne funkcje równie dobrze, jeśli nie lepiej. Sądzi się, że rybozomy stanowią relikty tego zaginionego świata. Co prawda nie zsyntetyzowano jeszcze replikatora, który miałby oczekiwane własności, ale autor tego opracowania sądzi, że jest to tylko kwestia czasu i, że z niewielką pomocą uczonych synteza ta potrwa krócej niż za pierwszym razem.

Etapem pośrednim pomiędzy „światem RNA” a obecnym światem istot żywych był prawdopodobnie „świat”, w którym odegrały rolę cząsteczki typu aminoacylo-tRNA. Jak sugerował niedawno GORDON (1995), na tym etapie ewolucji chemicznej mógł powstać przodek współczesnego rybosomu, który katalizował równocześnie translację i replikację. Ten hipotetyczny proto-rybosom zawierałby rybozym, który syntetyzując łańcuch polipeptydowy równocześnie wycinałby tryplety antykononów z przodków dzisiejszych aminoacylo-tRNA, łącząc je w nić matrycy RNA, komplementarnej do nici kodującej (Ryc. 3). Rybozym, którego istnienie postuluje GORDON miałby więc własności zbliżone do współczesnych intronów z grupy I. Taki intron wycina sam siebie z prekur-

sorowego poliribonukleotydu ligując równocześnie nukleotyd leżący po swojej stronie 5' z nukleotydem leżącym po stronie 3'. Hipoteza GORDONA tłumaczy, jak w świecie RNA pojawić się mogły polipeptydy, które mają zdolność tworzenia bardziej wydajnych i trwalszych katalizatorów niż rybozomy.

Dalszym etapem ewolucji było oddzielenie się fenotypu od genotypu. Organizm, który potrafił magazynować swoją informację genetyczną w cząsteczkach trwałego kwasu dezoksyrybonukleinowego miał przewagę nad organizmami, którymi był zaludniony świat RNA. Ukoronowaniem tego etapu było pojawienie się LUCA (ang. last universal common ancestor) — organizmu, którego istnienia dowodzi biochemiczna jedność świata żywego i, który dał początek zarówno eukaryotom, jak i prokariotom.

Na LUCA (zwanym inaczej progenotą) kończy się prehistoria i zaczyna się „pisana” historia chromosomu. Z okresu prehistorycznego pozostały tylko nieliczne relikty, takie jak rybozomy, natomiast po dawno zaginionym LUCA przetrwały kopie (co prawda niezbyt wierne) zapisu informacji zawartej w jego chromosomach dzisiaj istniejących organizmów, z na-



Ryc. 3. Hipotetyczny mechanizm sprzężenia translacji z replikacją poliribonukleotydu na wczesnym etapie ewolucji.

(1) cykl rozpoczyna się od katalizowanego przez rybozym ataku guanozyny na aminoacylo-t-RNA w pozycji A; (2) przecięcie łańcucha t-RNA i ligacja guanozyny do końca 5' fragmentu 3' tego łańcucha; (3) nowo utworzona grupa 3' OH w antykononie znajdującym się w pozycji A reaguje z antykononem w pozycji P co powoduje jego wycięcie (splicing) i włączenie w rosnący łańcuch poliribonukleotydu. Równocześnie następuje transpeptydacja poprzez atak nukleofilowy grupy aminowej aminokwasu z miejsca A na aktywowane wiązanie estrowe pomiędzy t-RNA w pozycji P a grupą karboksylową aminokwasu z nim związanego; (4) translokacja rybosomu o trzy nukleotydy w stronę 3' łańcucha kodonów. Segmenty rozciętego t-RNA oddysocjują a pozycja A zostaje zajęta przez kolejny aminoacylo-t-RNA. E, P, A, pozycje na rybosomie w których wiążą się cząsteczki aminoacylo-t-RNA. G, guanina, U, uracyl, N, nukleotyd (według GORDONA 1995, zmodyfikowane).

szym własnym włącznie. Można mieć nadzieję, że porównując genomy różnych oddalonych ewolucyjnie organizmów żyjących obecnie, uda

się wydedukować strukturę chromosomu naszego przodka.

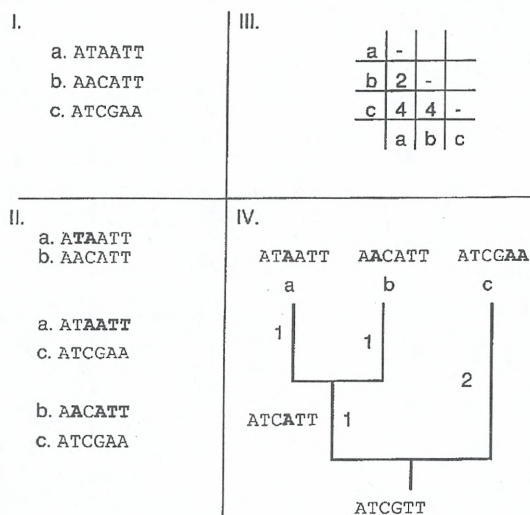
#### KONSTRUKCJA DRZEW GENEALOGICZNYCH NA PODSTAWIE ANALIZY PORÓWNAWCZEJ SEKWENCJI DNA

Problematyką tą zajmuje się gałąź biologii molekularnej zwana teorią ewolucji molekularnej. Podstawowym zajęciem badaczy pracujących w tej dziedzinie jest sporządzanie kladogramów oraz innych bardziej złożonych drzew genealogicznych opisujących ewolucyjną historię sekwencji nukleotydowych, a równocześnie historię organizmów, które są ich nosicielami. (Kladogram jest to takie szczególne drzewo genealogiczne, w którym: (a) nie definiuje się „przodków”, a więc wszystkie rozpatrywane jednostki taksonomiczne są umieszczone na końcach gałęzi oraz (b) drzewo nie zawiera informacji na temat długości gałęzi).

Aby móc narysować drzewo genealogiczne sporządza się wpięrw macierz „odległości” pomiędzy każdymi dwiema z serii badanych sekwencji nukleotydowych. Zapisuje się je w tym celu parami obok siebie i liczy ile wynosi najmniejsza z możliwych ilość zmian, które należy wprowadzić do jednej z sekwencji, aby pokryła się ona z drugą (Ryc. 4). Istnieje szereg modeli matematycznych i programów komputerowych pozwalających wykreślać na podstawie tych

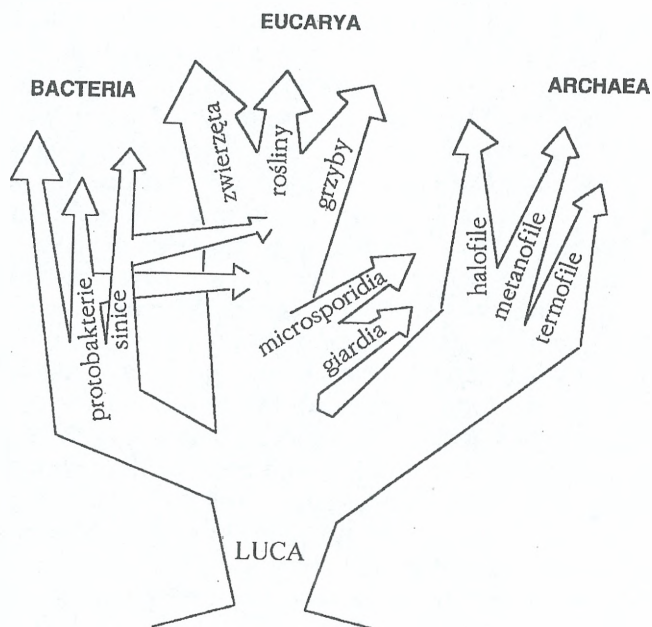
macierzy drzewa genealogiczne. Drzewa genealogiczne można również sporządzać na podstawie macierzy różnic fenotypowych między gatunkami, w szczególności, na podstawie różnic między cechami szkieletów wymarłych zwierząt. Korzysta się z nich dla ustalenia genealogii gatunków kopalnych. Można sporządzać również drzewa genealogiczne na przykład języków stosowanych przez różne grupy etniczne.

W regionach kodujących genomu w czasie ewolucji mutacje gromadzą się z reguły bardzo powoli, ponieważ w zasadzie tylko te zmiany genotypu, które nie powodują ujemnych skutków fenotypowych, są przekazywane z pokolenia na pokolenie. Porównanie sekwencji nukleotydowych genów homologicznych (a więc takich, które u bliskiego wspólnego przodka pełniły tę samą funkcję co u organizmów współczesnych) pozwala badać pokrewieństwo organizmów odległych filogenetycznie, nawet takich, które były izolowane genetycznie przez setki milionów lat. Porównania takiego dokonano dla powoli ewoluujących genów rybosomalnych. Odtworzono w ten sposób drzewo genealogiczne obejmujące główne grupy organizmów żywych, tak zwane „drzewo życia” (Ryc. 5). Konstrukcja takiego drzewa daje nie tylko pojęcie o tym, które z gatunków są blisko z sobą spokrewnione, a które daleko. Zakładając, że ilość mutacji pojawiających się w genomie w jednostce czasu jest stała (hipoteza „zegara molekularnego”, który „tyka” ze stałą prędkością, ZUCKERKANDL i PAULING 1965) można obliczyć, jak dawno nastąpiło oddzielenie się na przykład przodków współczesnych ssaków od przodków współczesnych gadów. Wyniki otrzymane tą drogą nie zawsze są zgodne z danymi paleontologicznymi. Obserwowane niekiedy odstępstwa tłumaczy się tym, że „zegar molekularny” prawdopodobnie nie „tyka” równomiernie. Są okresy, w których mutacje pojawiają się często, i inne, w których są rzadkie. Niektóre organizmy potrafią zwiększać szybkość pojawiania się mutacji, gdy znajdują się w warunkach do których nie są przystosowane. Umożliwia to adaptację przynajmniej niektórym z ich potomków i pozwala na opanowywanie nowych nisz ekologicznych (TADDEI i współaut. 1997). Obok sekwencji kodujących, które mutują stosunkowo wolno istnieją również takie, które gromadzą



Ryc. 4. Konstrukcja drzewa genealogicznego opartej na porównaniu sekwencji nukleotydowych.

(I) Trzy sekwencje nukleotydowe „a”, „b” i „c”, których „pokrewieństwo” chcemy ustalić, (II) porównanie sekwencji, różnice pomiędzy sekwencjami zaznaczone są tłustym drukiem; (III) macierz różnic; (IV) drzewo genealogiczne i wydedukowana prawdopodobna sekwencja ancestralna, liczby obok gałęzi podają ilość mutacji pomiędzy sąsiednimi węzłami w drzewie genealogicznym.



Ryc. 5. Główne gałęzie drzewa genealogicznego organizmów żywych wydedukowane przez porównanie sekwencji rybosomalnych.

Gałęzie, które łączą protobakterie i sinice z eukaryotami symbolizują ewolucję mitochondriów i chloroplastów (według DOOLITTLE i BROWNA 1994, zmodyfikowane)

mutacje szybko niezależnie od warunków. Porównywanie tych sekwencji pozwala badać ewolucję blisko spokrewnionych gatunków oraz ewolucję różnych populacji w ramach jednego gatunku. Poniżej omówię pokrótce różne typy sekwencji, które można napotkać w ludzkich chromosomach oraz przedstawię problemy na

które napotkano starając się zrozumieć najwyższy szczebel organizacji genomu, ten, który przejawia się w postaci prążków na chromosomach. Na zakończenie wspomnę jeszcze o kilku interesujących wynikach z dziedziny prehistorii człowieka, jakie przyniosła analiza różnego typu sekwencji obecnych w chromosomach.

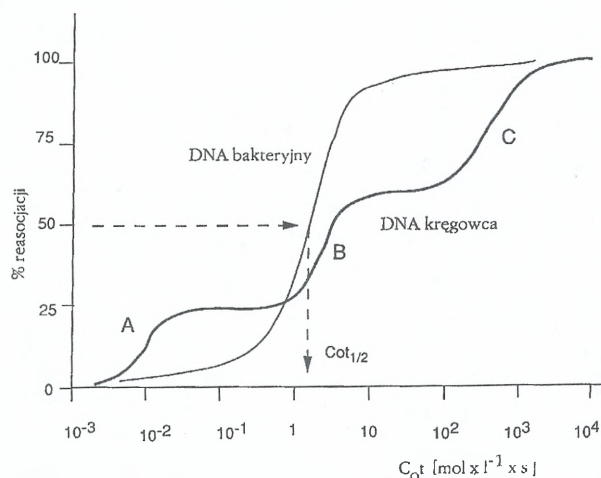
## CHARAKTERYSTYKA SEKWENCJI W DNA CHROMOSOMALNYM

### ZŁOŻONOŚĆ KINETYCZNA

Badając skład DNA bardzo wcześnie stwierdzono, że w chromosomach można znaleźć „elementy” (ciągi nukleotydów) unikalne oraz takie, które powtarzają się wielokrotnie. Badania powtarzalności sekwencji prowadzono oznaczając szybkość renaturacji DNA, to znaczy reakcji polegającej na odtwarzaniu podwójnej spirali z jednoniciowych cząsteczek otrzymanych z pociętego na krótkie segmenty, zdenaturowanego DNA. Renaturacja zachodzi tym szybciej im stężenie cząsteczek o sekwencjach komplementarnych w roztworze jest większe. Mają one wtedy większą szansę spotkania się. W roztworze DNA pociętego na drobne fragmenty stężenie tych z pośród nich, które niosą „elementy” obecne w chromosomie w wielu kopiach jest większe niż stężenie fragmentów o unikalnej sekwencji nukleotydowej. W wyniku tego krzywa przedstawiająca postęp renaturacji ludzkiego DNA w czasie (kinetykę renaturacji) jest multimodalna, to znaczy zawiera szereg składowych kinetycznych, z których każda odpowiada

renaturacji pewnej frakcji genomu o określonej stałej reakcji renaturacji (Ryc. 6). Stała ta jest wyrażana zwykle w postaci  $Cot_{1/2}$ . Równa się ona iloczynowi czasu „ $t_{1/2}$ ” [s] potrzebnego do zrenaturowania połowy cząsteczek danej frakcji DNA, w standardowych warunkach temperatury, stężenia soli i pH oraz stężenia początkowego DNA „Co” [mol/l]. Połowiczny czas renaturacji  $t_{1/2}$  może wynosić od kilku sekund (dla sekwencji obecnych w genomie w wielu kopiach) do kilku dni (dla sekwencji unikalnych.) Istnieje szereg różnych typów sekwencji powtarzających się. W genomach zwierząt i roślin stanowią one niekiedy główną masę chromosomu, bakterie zawierają niewielką ilość (kilka procent) takich sekwencji, a jednokomórkowe Eukaryota, takie jak drożdże *S. cerevisiae*, zajmują pozycję pośrednią. Analiza kinetyki renaturacji ma dziś głównie wartość historyczną, ale pozwoliła ona w swoim czasie zidentyfikować główne składniki genomu.

Sekwencje powtarzające się mogą mieć różną długość jednostki powtarzalności: od jednej pary zasad (pz) do tysiący (ang. kilobase, kb), a nawet milionów (ang. megabase, Mb) par zasad.



Ryc. 6. Wyidealizowany przykład krzywych reasocjacji DNA kręgowca i bakterii.

Stopnie „A” i „B” na krzywej kręgowca reprezentują składowe kinetyczne sekwencji nukleotydowych o wysokim i średnim poziomie powtarzalności. „C” odpowiada sekwencjom unikalnym.

Różnią się również częstością powtórek, organizacją (sekwencje rozproszone lub powtarzające się typu tandem), genezą i funkcją.

#### Sekwencje unikalne

Podział na sekwencje powtarzające się i unikalne jest w pewnym stopniu arbitralny. Podczas ewolucji pewne segmenty chromosomu ulegały duplikacji i łączyły się z innymi segmentami. Różne kopie tego samego segmentu mutowały niezależnie od siebie, na skutek czego w obecnie żyjących organizmach z trudem można poznać, że są one spokrewnione. W szczególności, geny kodujące białka, których jest w genomie ludzkim kilkadziesiąt tysięcy uważa się (z pewnymi wyjątkami) za sekwencje unikalne, mimo, że powstały one wszystkie z kilkuset zaledwie egzonów, które ulegały duplikacjom, mutacjom punktowym i „tasowaniu”, łącząc się podczas ewolucji w różne kombinacje (DE SOUZA i współaut. 1996)

#### Amplikony

Proces duplikacji czy multiplikacji segmentów genomu zdarza się dość często (patrz np. HYRIEN i współaut. 1987). W komórkach organizmów wyższych hodowanych przez kilka pokoleń w obecności związków toksycznych stwierdza się niekiedy na chromosomach obecność jednorodnie barwiących się segmentów (ang. homogenously staining regions, HSR) nieobecnych w komórkach wyjściowych. Pojawiają się również w nich acentryczne mini chromosomy DM (ang. double minute). HSR i DM zawierają wielką ilość kopii dużych segmentów DNA o długości 1Mb lub dłuższych, w obrębie któ-

rych znajduje się zazwyczaj gen, którego ekspresja pozwala na zneutralizowanie efektu czynnika toksycznego obecnego w medium. Te struktury zwane są amplikonami. W komórkach traktowanych lekami antynowotworowymi pojawiają się amplikony zawierające gen kodujący MRP (ang. multidrug resistance-associated protein), co pociąga za sobą syntezę wielu kopii cząsteczek białka wywołującego oporność na te leki. W komórkach nowotworowych obserwuje się czasem regiony amplifikowane zawierające onkogeny, co przyczynia się do ich niekontrolowanego rozmnażania się. Amplifikacja zachodzi albo poprzez wielokrotną inicjację replikacji w jednym lub kilku sąsiadujących replikonach podczas jednego cyklu komórkowego, bądź przez niewzajemny crossing-over. Można sądzić, że podobny proces był również źródłem powstania w chromosomach regionów zawierających wielką ilość kopii genów kodujących rybosomalny RNA oraz histony.

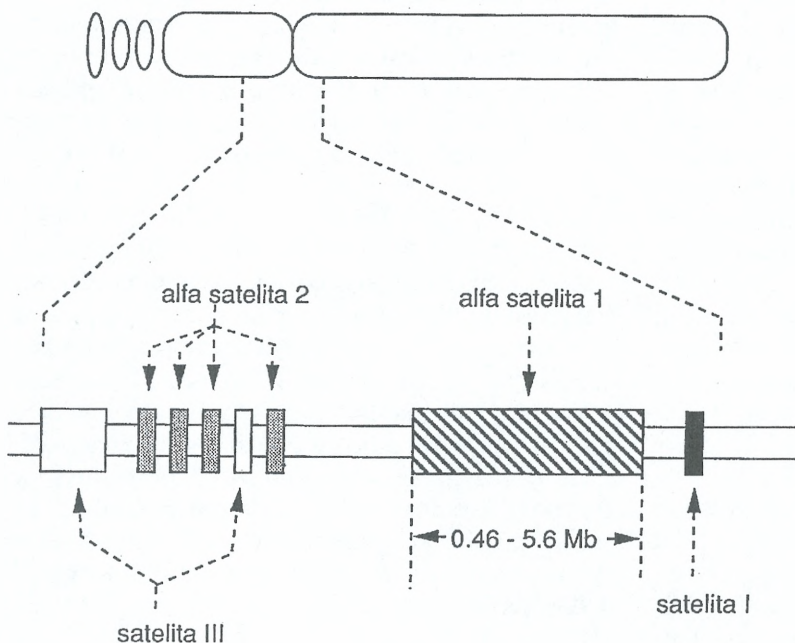
#### Satellity

Z pośród wielu klas sekwencji powtarzających się pierwsze zostały opisane sekwencje satelitarne (termin modny w latach sześćdziesiątych ze względu na podbój kosmosu, który się równocześnie odbywał). Obserwowano je w postaci drobnych frakcji towarzyszących głównej frakcji genomu podczas wirowania w roztworach soli cezu. Satellity występują praktycznie u wszystkich organizmów wyższych stanowiąc niekiedy nawet do połowy DNA chromosomalnego. Sekwencje nukleotydowe satelitów są zbudowane z wielokrotnie powtarzających się krótkich jednostek ułożonych jedna obok drugiej. Satelita kraba *Cancer borealis* ma sekwencję typu poli-d(AT), *Drosophila virilis* ma trzy różne satellity o siedmionukleotydowych jednostkach powtarzalności. U człowieka najlepiej poznane są cztery główne klasy satelitów zwane I, II i III i alfa, których jednostki powtarzalności są trudniejsze do zdefiniowania ponieważ u kręgowców, w odróżnieniu od stawonogów, obserwuje się hierarchiczną organizację elementów powtarzających się. Tak więc można niekiedy wyróżnić w DNA satelitarnym krótkie, kilkunukleotydowe sekwencje, które zostały w trakcie ewolucji powtórzone kilkakrotnie tworząc blok, który stawał się następną podjednostką o długości kilkudziesięciu nukleotydów. Ta podjednostka drugiego rzędu, po nagromadzeniu pewnej ilości mutacji punktowych, ulegała kolejnemu namnożeniu i kolejnym mutacjom itd. Satelitarne DNA ulegały wielu seriom takich replikacji i rekombinacji typu niewzajemny crossing-over, co spowodowało ich ekspansję w chromosomie.

DNA typu satelitarnego ewoluuje szybko, tak, że nawet zbliżone gatunki mają niekiedy zupełnie różne satelity. Znajdują się one w centromerach oraz w regionach subtelerowych chromosomów. Przez pewien czas istniały wątpliwości co do tego, czy satelity pełnią jakąś funkcję, zważywszy na to, że tak bardzo różnią się sekwencjami nukleotydowymi oraz długościami. Na przykład długość bloku sekwencji satelity alfa 1 obecnego w centromerze chromosomu 21 człowieka wykazuje zmienność osobniczą w granicach od 420 do 5600 kb, nawet u osób należących do tej samej rodziny (Ryc. 7). Obecnie uważa się że tą funkcją jest tworzenie heterochromatyny, a więc regionów, w których włókno chromatynowe jest silnie skondensowane i które są niezbędne dla poprawnego funkcjonowania chromosomów. Szczególną rolę w centromerach chromosomów ludzkich odgrywają satelity typu alfa. Posiadają one szesnastonukleotydową sekwencję typu konsensus zwaną „CENP-B box”, która jest specyficznie rozpoznawana przez typowe dla centromeru białko CENP-B (ang. centromeric protein B). Kompleks jaki tworzy to białko z „CENP-B box” ma prawdopodobnie udział w budowie centromeru (YODA i współaut. 1998). Sekwencja spokrewniona z „CENP-B box” znaleziona została w satelicie mniejszym (minor) myszy i w satelitarniej sekwencji obecnej w centromerach chromosomów ryżu. Z drugiej strony nie stwierdzono, aby genetycznie modyfikowane myszy typu „knock-out” z genomu, których usunięto gen kodujący białko CENP-B, różniły się czymkolwiek od myszy normalnych. Prawdopodobnie więc białko to nie jest niezastąpione.

#### Sekwencje telomerowe

Zadaniem telomerów jest ochrona końców chromosomu przed atakiem enzymów nukleolitycznych i przed systemem naprawiającym dwuniciowe przecięcia, który ma tendencję do łączenia wolnych końców DNA (BLACKBURN 1994). Dodatkowo, telomery zapobiegają utracie informacji genetycznej podczas replikacji (patrz niżej). Sekwencja nukleotydowa DNA w telomerach przypomina sekwencję satelitów znajdujących w centromerach z tą różnicą, że kopie sekwencji znajdowanej na końcach telomerów są u wszystkich Eukaryota podobne w tym, że zawierają ciągi guanin. Sekwencje telomerowe są identyczne we wszystkich chromosomach danego gatunku. W chromosomach kręgowców (również człowieka) jest to sekwencja TTAGGG. W pewnej odległości od końca chromosomu obserwuje się jej warianty typu TTGGGG czy TTTAGGG. Obecność takich wariantów dowodzi, że procesy, które biorą udział w generowaniu zmienności i w ekspansji satelitów, a więc mutacje punktowe, rekombinacja typu niewzajemny crossing-over i konwersja genów, wpływają na strukturę sekwencji telomerowych w pewnej odległości od końca chromosomu. Pochodzenie sekwencji nukleotydów na samym końcu chromosomu jest jednak inne niż pochodzenie większości satelitów. Telomery DNA jest syntetyzowany przez enzym telomerazę, który jest polimerazą DNA typu odwrotnej transkryptazy. Posiada on podjednostkę będącą cząsteczką RNA, która dla telomerazy pełni rolę matrycy. Na tej matrycy telomeraza syntetyzuje wielką ilość kopii DNA telomerowego.



Ryc. 7. Struktura regionu centromeru na chromosomie 21. Pomiedzy blokami DNA satelitarnego znajdują się segmenty DNA, których natura i długość jest nieznana (według MARCAISA i współaut. 1991, zmodyfikowane).

Telomeraza obecna jest tylko w komórkach linii zarodkowej. Sądzi się, że dzięki temu właśnie komórki linii zarodkowej nie starzeją się. W komórkach somatycznych ekspresja genu telomerazy jest wygaszona. Na skutek tego telomery w tych komórkach ulegają skracaniu podczas kolejnych podziałów. I tak na przykład telomerowy DNA w chromosomach spermatozoidów człowieka ma długość około 20 kb natomiast w limfocytach jego długość wynosi 5 do 8 kb w zależności od wieku dawcy (im bardziej zaawansowany wiek, tym krótsze telomery). Efekt skracania telomerów jest wynikiem specyficznego mechanizmu działania polimeraz DNA, które podczas replikacji są niezdolne do rozpoczęcia „od zera” syntezy nici DNA. Najczęściej, w pierwszym etapie replikacji syntetyzowany jest odcinek RNA („starter”) przez polimerazę RNA, która potrafi syntetyzować komplementarny oligonukleotyd startując od zera (relikt „świata RNA”). Dopiero w dalszym etapie replikacji polimeraza DNA zaczyna dobudowywać dezoksyrybonukleotydy do startera. Starter obecny na początku zreplikowanej nici zostaje następnie strawiony i tak część informacji obecna na matrycy przepada. Skrócenie telomeru powoduje to, że geny w pobliżu końca chromosomu przestają ulegać ekspresji prawdopodobnie w wyniku „inwazji” chromosomu przez heterochromatynę telomeru. W pobliżu końców chromosomów w prążkach typu „T” (patrz poniżej) zlokalizowane są często geny metabolizmu podstawowego, których ekspresja jest niezbędna do życia komórki. Można sądzić, że to wygaszanie ekspresji jest przyczyną tego, że komórki somatyczne w kulturze przestają się rozmnażać po około 40 podziałach. Efekt starzenia się komórek w kulturach zaobserwowano już dawno i zjawisko to uważano za wynik działania „zegara mitotycznego”, dzięki któremu komórka „wie” ile razy uległa mitozie od czasu, gdy linia komórek do niej prowadząca zaczęła się różnicować. Ponieważ komórki, w których przez manipulacje typu inżynierii genetycznej spowodowano ekspresję telomerazy nie starzeją się, można sądzić, że skracanie się telomerów jest tym od dawna poszukiwanym mechanizmem „zegara”. Starzenie się komórek somatycznych nie jest bez związku ze starzeniem się organizmu jako całości (zob. JARUGA 1999).

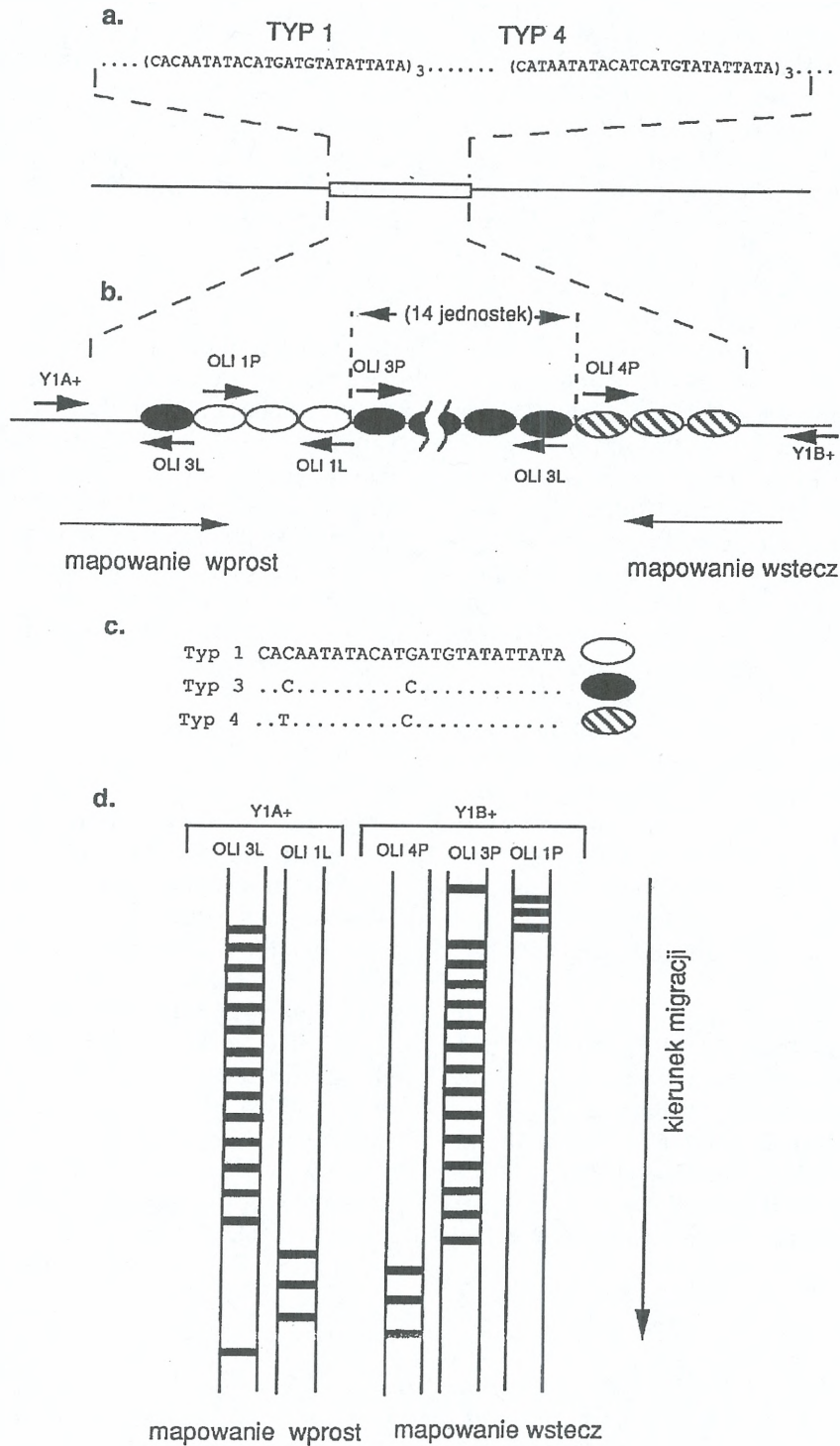
#### Minisatelity

Oddzielną klasę sekwencji powtarzających się stanowią minisatelity czasem zwane VNTR (ang. variable number tandem repeats) (JEFFREYS i współaut. 1985). Ich jednostki powtarzalności o długości od 10 do 30 pz. tworzą bloki

$10^3$  do  $10^4$  pz. rozproszone w różnych miejscach genomu. Charakteryzują się wyjątkowym polimorfizmem jeśli chodzi o długość bloków, co powoduje, że każdy osobnik z wyjątkiem bliźniąt monozygotycznych posiada inny, łatwy do zidentyfikowania wzór minisatelitów. Najbardziej znane minisatelity są bogate w guaninę i należą do rodziny, która wykazuje pewne podobieństwo do sekwencji bakteryjnych „chi” odgrywających rolę w rekombinacji. Minisatelity w genomie ludzkim są homologiczne do minisatelitów znalezionych w genomach niektórych innych gatunków ssaków. Wyjątkowo duży polimorfizm minisatelitów znalazł zastosowanie w identyfikacji materiału biologicznego w analizach sądowych oraz w ustalaniu ojcostwa ludzi i rasowych psów. Początkowo porównywano tylko długość bloków minisatelitów u różnych osobników. Nowa wersja tej techniki korzysta z tego, że niektóre z minisatelitów zbudowane są z bloków różniących się nieznacznie sekwencją. Analiza typu MVR-PCR (ang. minisatellite variant repeat-polymerase chain reaction) (Ryc. 8) pozwala na oznaczenie kolejności tych wariantów w bloku i na przedstawienie wyników w postaci „kodu kreskowego” (JOBLING i współaut. 1998). Analiza DNA kilkuset osób z różnych kontynentów pokazała istnienie olbrzymiej ilości alleli takich bloków. Ponadto wykazano, że mutacje akumulujące się w blokach w czasie ewolucji mają charakter kierunkowy: strona arbitralnie nazwana 3' mutuje znacznie szybciej niż strona 5'. Pozwala to na grupowanie wyników analiz w klasy i dostarcza informacji o pokrewieństwie pomiędzy osobami, grupami osób, i populacjami. Ostatnio znaleziono blok minisatelitów (MSY1) w chromosomie Y człowieka. Blok ten jest haploidalny i nie zachodzi w nim rekombinacja podczas mejozy często obserwowana w autosomalnych minisatelitach. Jest to najbardziej zmienny marker chromosomalny i jedyny, który został przebadany dla znacznej liczby osób.

#### Mikrosatelity

Mikrosatelitami nazwano powtarzające się sekwencje o długości jednostki powtarzalności od jednego do kilku nukleotydów i o liczbie powtórzeń, która nie przekracza naogół kilkadziesięciu nukleotydów (LEVINSON i GUTMAN 1987). Za mikrosatelity uważane są sekwencje poli A, które często towarzyszą pseudogenom, między innymi sekwencjom „Alu”, powstałym w wyniku retrotranskrypcji RNA (patrz niżej). Najczęściej spotykanymi mikrosatelitami są stosunkowo regularnie rozmieszczone sekwencje  $(CA)_n/(GT)_n$  znajdujące przeciętnie co 25–100 kb. Rzadziej spotyka się mikrosatelity typu



Ryc. 8. Mapowanie wariantów sekwencji powtarzającej się w minisatelicie MSY1.

(a) Częściowa sekwencja nukleotydowa segmentu MSY1; (b) rozmieszczenie wariantów oraz przykładowe pozycje starterów (OLI 1P, OLI 1L, OLI 3P, OLI 3L, OLI 4P, OLI 4L, Y1B+ i Y1A+) w pięciu reakcjach PCR wykonanych w celu analizy wariantów w minisatelicie, (c) sekwencje nukleotydowe trzech wariantów elementu powtarzającego się, (d) wynik elektroforezy czasteczek powstałych w reakcji PCR z użyciem par starterów pokazanych nad kolejnymi ścieżkami (według JOBLINGA i współaut. 1998, zmodyfikowane).

(TTA)<sub>n</sub>/(AAT)<sub>n</sub> oraz (AGC)<sub>n</sub>/(TCG)<sub>n</sub>. Spotykane są również motywy czteronukleotydowe. Są one stosowane jako markery polimorfizmu, ponieważ wykazują wielką zmienność jeśli chodzi o

długość. Polimorfizm długości sekwencji mikro-satelitarnych wykazuje się w ten sposób, że wykonuje się reakcje PCR regionu w którym się one znajdują używając do reakcji starterów

komplementarnych do sekwencji sąsiadujących z mikrosatelitą z lewej i prawej strony. Elektroforeza produktów namnażania pozwala wykazać różnice nawet dwóch nukleotydów pomiędzy allelami. Analizę polimorfizmu mikrosatelitów stosuje się do identyfikacji materiału biologicznego w sprawach sądowych, a także do diagnozowania chorób genetycznych, ponieważ pozwala odróżnić allel zmutowany obecny u członków badanej rodziny od alleli niezmutowanych. Poślizg replikacyjny (Ryc. 9), który jest odpowiedzialny za niestabilność sekwencji mikrosatelitarnych i za ich duży polimorfizm, może spowodować tak zwaną mutację dynamiczną, która może być przyczyną choroby genetycznej, jeśli sekwencja, która uległa poślizgowi, znajduje się w obrębie genu. I tak na przykład dystrofia Steinera jest powodowana przez ekspansję mikrosatelitarnego DNA towarzyszącego genowi kinazy miotoninowej. Tryplet CTG jest w genie osób zdrowych powtórzony 5 do 27 razy. W przypadku osób cierpiących na dystrofię Steinera liczba powtórzeń wynosi 52 do 1000. Liczba powtórzeń zmienia się z pokolenia na pokolenie i nie dziedziczy się w sposób przewidywalny przez klasyczną genetykę mendelowską. Mutacje dynamiczne przechodzą niekiedy stadium pre-mutacji, w którym liczba powtórzeń przekracza ilość standardową, ale nie daje efektów patologicznych. Niekiedy mutacje cofają się w wyniku „kontrakcji” mikrosatelity.

#### Pseudogeny

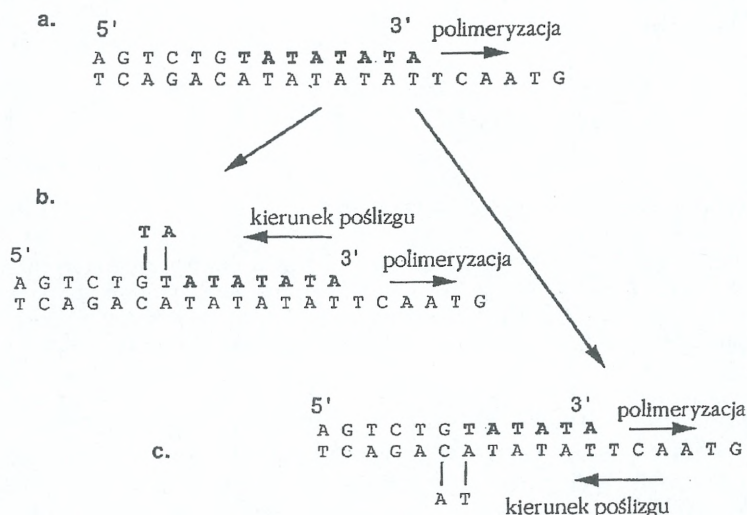
W genomie ludzkim znajduje się duża ilość tak zwanych „pseudogenów”, a więc sekwencji nukleotydowych przypominających geny lecz transkrypcyjnie nieaktywnych. Wyróżnia się dwa rodzaje pseudogenów. Pierwszy z nich powstaje w wyniku duplikacji odcinka DNA w procesie na przykład niewzajemnego crossing-

over. Niekiedy kopie zduplikowanego genu „dzieli się” funkcjami (taka jest np. geneza wariantów genu beta globiny, z których niektóre są aktywne we wczesnych, a inne w późniejszych stadiach rozwojowych ssaków). W innych przypadkach następuje wygaszenie ekspresji jednej z kopii poprzez mutacje. Tego typu pseudogeny zachowują strukturę egzonów i intronów genów funkcjonalnych, od których pochodzą.

Drugi rodzaj pseudogenów powstaje w wyniku integracji do chromosomu segmentu DNA utworzonego przez retrotranskrypcję cząsteczki RNA z udziałem enzymu rewertazy (odwrotnej transkryptazy). Takie pseudogeny nie mają intronów ani promotora i towarzyszy im sekwencja poli A na końcu 3'. Niekóre segmenty DNA generują w ten sposób ogromną ilość kopii. W komórkach człowieka funkcjonuje, obok systemów zawleczonych przez retrowirusy, endogeny system retrotranskrypcyjny, który generuje sekwencje typu LINE.

#### Sekwencje typu LINE

Skrótem tym oznacza się „długie, rozproszone, powtarzające się sekwencje” (ang. long interspersed repetitive element, patrz SKOWRONSKI i SINGER 1985) znajdujące się w genomach ssaków. Kompletny element LINE ma długość 6-7 kb, ale u człowieka większość (>90%) kopii to sekwencje niekompletne. Wszystkie kopie są zakończone sekwencją poliA. Ponadto, w miejscu na chromosomie, w które zostaje wstawiona kopia LINE, kilka nukleotydów ulega duplikacji. Stwierdza się ich obecność na końcach sekwencji LINE. Kompletnie sekwencje LINE mają dwie otwarte ramki odczytu (ORF 1 i 2). Jedna z nich (ORF 2) koduje białko o cechach odwrotnej transkryptazy. Od retrowirusów (patrz niżej) sekwencje LINE różnią się brakiem sekwencji



Ryc. 9. Mechanizm poślizgu replikacyjnego.

(a) sekwencja ancestralna, (b) poślizg nici rosnącej w kierunku 3'-5' powoduje na niej wypętlenie dinukleotydu TA, w wyniku replikacji następuje dodanie dinukleotydu TA do elementu powtarzającego się (ekspansja mikrosatelity). (c) Poślizg nici rosnącej następuje w kierunku 5'-3' pętla TA tworzy się na matrycy. Syntetyzowana sekwencja powtarzająca posiada o jeden dinukleotyd TA mniej niż matryca. (kontrakcja mikrosatelity) (według LEVINSONA i GUTMANA 1987, zmodyfikowane).

LTR (ang. long terminal repeat). Nazywa się je retrotranspozonomami klasy II.

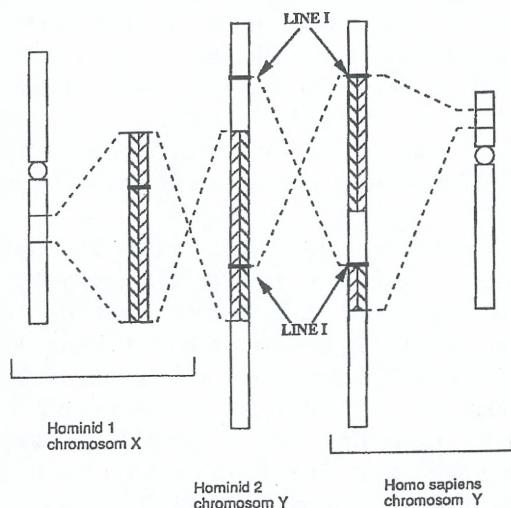
Proces retrotranspozycji rozpoczyna się od transkrypcji jednej z kompletnych kopii elementu LINE obecnych w genomie. Transkrypcja jest ściśle kontrolowana i zachodzi tylko w komórkach linii zarodkowej, w tkankach wczesnych embrionów i w niektórych nowotworach. Transkrypt jest poliadenylowany i transportowany do cytoplazmy, gdzie następuje translacja RNA. Produkty translacji ORF 1 i ORF 2 tworzą kompleks rybonukleoproteinowy z RNA LINE. Kompleksy takie zostały wyizolowane z komórek nowotworowych. Stanowią one najprawdopodobniej źródło retrotranskryptu, który po przetransportowaniu do jądra zostaje zintegrowany z chromosomem. Uważa się, że podczas ewolucji proces transpozycji ma okresy „eksplozji”. W takim okresie genom zostaje „zaśmiecany” przez „rodzinę” transpozonów. Po pewnym czasie następuje wyciszenie źródła RNA retrotranspozonu na przykład przez mutację w promotorze. Następuje wówczas okres, w którym transpozony akumulują w sposób bierny mutacje, z rzadka tylko ulegając transpozycji, do momentu aż któraś z kopii wbuduje się w region chromosomu obok silnego promotora, co powoduje „eksplozję” następnej „rodziny” LINE. Obecność sekwencji typu LINE w genomie pociąga za sobą szereg konsekwencji. Po pierwsze, retrotranspozon może ulec integracji wewnątrz sekwencji kodującej gen. Znane są przypadki chorób genetycznych, które mają takie źródło. Zidentyfikowano na przykład przypadek hemofilii, który jest spowodowany włączeniem się LINE w egzon 14 genu kodującego czynnik VIII krzepliwości krwi. Drugą konsekwencją integracji sekwencji LINE jest zwiększona plastycz-

ność genomu. Stwierdzono na przykład, że chromosom ludzki Y zawiera oprócz regionu pseudoautosomalnego, a więc tego, który umożliwia parowanie z chromosomem X podczas mejozy, również i inne segmenty, homologiczne do X. Jeden z tych segmentów zawiera dwa bloki markerów genetycznych, identycznych z markerami genetycznymi obecnymi na chromosomie X z tym, że na chromosomie X tworzą one jeden blok (SCHWARTZ i współaut. 1987). Ponadto kolejność markerów w mniejszym bloku jest odwrócona (Ryc. 10). Porównanie chromosomu Y ludzi i szympanów (których materiał genetyczny jest w 98% identyczny z ludzkim) nie wykazało obecności tego bloku markerów na chromosomie Y małp. Sugeruje to, że u naszego przodka nastąpiła transpozycja całego bloku sekwencji z chromosomu X na chromosom Y, już po oddzieleniu się linii małp człekokształtnych od linii prowadzącej do naszego gatunku. W dalszej kolejności, u innego, bliższego przodka, nastąpiła inwersja segmentu chromosomu Y zawierającego blok markerów z chromosomu X oraz część markerów chromosomu Y. Otóż ta ostatnia rekombinacja nastąpiła z udziałem sekwencji LINE obecnych po obydwu stronach segmentu, który uległ inwersji. Ten przykład pokazuje, że obecność sekwencji powtarzających ułatwia wycinanie całych odcinków genomu i łączenie ich w inne kombinacje. Zwykle nie jest to korzystne dla osobników, których genom uległ takiemu przekształceniu, ale stwarza możliwość wypróbowania nowej kombinacji genów, która może okazać się korzystna.

#### Sekwencje typu SINE

Duży procent genomu człowieka jest zbudowany z sekwencji typu SINE (ang. short interspersed repetitive element, JELINEK i SCHMID 1985) zwanych sekwencjami Alu (od enzymu restrykcyjnego Alu I, który rozpoznaje tetranukleotydy AGCT znajdujący się w większości kopii tego pseudogenu). Sekwencja Alu jest blisko spokrewniona (80% homologii) z genem kodującym RNA 7SL, który bierze udział w transporcie białek przez błony. Sekwencje Alu zostały znalezione wyłącznie w genomach prymatów. Sekwencje typu SINE obecne w genomach innych ssaków są spokrewnione z t-RNA.

Ewolucja sekwencji Alu została prześledzona dość dokładnie między innymi dzięki pracom grupy Jerzego Jurki (patrz np. JURKA i SMITH 1988). We wczesnym okresie ewolucji prymatów pojawiła się w genomie sekwencja FAM (ang. fossil Alu monomer) powstała w wyniku delekcji bloku nukleotydów w genie 7SL RNA. Następnie (bądź równocześnie) w wyniku kolejnych delekcji powstały FRA i FLA (ang. free right Alu i free left



Ryc. 10. Transpozycja bloku sekwencji z chromosomu Xq21 do Yp (według SCHWARTZA i współaut. 1987, zmodyfikowane).

Alu), monomeryczne formy Alu obecne w dużej ilości kopii w genomie człowieka. Ostatecznie elementy te utworzyły dimer przy pomocy bogatego w adeninę łącznika. Zarówno formy monomeryczne, jak i dimer Alu dobrze zaadaptowały się do któregoś systemu retropozycyjnego. Wykorzystując ten system dokonały one kilku „inwazji” genomu naczelnych w różnych okresach czasu. Elementy Alu, pochodzące z różnych inwazji, można zidentyfikować podobnie jak sekwencje LINE. Pochodzą one bowiem od różnych, wcześniej zintegrowanych kopii, które przez długie okresy trwały „uśpione” w chromosomie akumulując przypadkowe mutacje. W wyniku inwazji powstawała więc „rodzina”, której wszyscy członkowie noszą mutacje sekwencji „matki”. Najwcześniejsza z inwazji miała miejsce, jak się sądzi, około 65 milionów lat temu. Te „stare” Alu znaleziono u wszystkich prymatów z lemurami włącznie. Ostatnia z inwazji miała miejsce już po oddzieleniu się linii prowadzącej do małp człekokształtnych od linii prowadzącej do naszego gatunku. Od tego czasu sekwencje Alu włączają się do genomu człowieka sporadycznie powodując niekiedy inaktywację genów (stwierdzono np. przypadek neurofibromatozy spowodowany wbudowaniem się Alu do genu NF1 itp.).

#### HERV

U ludzi podobnie jak u innych ssaków znaleziono kilkadziesiąt kopii genomu retrowirusa HERV (ang. human endogenous retrovirus) na stałe wbudowanych do chromosomów. Zbudowane są one tak jak wszystkie genomy retrowirusów, tzn. mają sekwencje LTR (ang. long terminal repeat) biorące udział w regulacji transkrypcji wirusowego DNA, kodują białka, między którymi znajduje się odwrotna transkryptaza dokonująca retrotranskrypcji RNA na DNA oraz integraza, która wbudowuje DNA wirusa do chromosomu gospodarza. Endogenne retrowirusy są reliktem zakażeń, którym ulegli nasi przodkowie wiele milionów lat temu i które spowodowały wbudowanie się kopii genomów wirusa do chromosomów w komórkach linii zarodkowej. Znajdują się one w genomie w formie uśpionej, mogą jednak niekiedy w pewnych typach komórek ulegać aktywacji. Aktywację endogennego retrowirusa stwierdzono w nowotworach wywodzących się z komórek linii zarodkowej. Nie jest jednak jasne czy ta aktywacja jest przyczyną czy skutkiem transformacji nowotworowej. Niewiele wiemy dotychczas na temat znaczenia obecności sekwencji retrowiralnych dla funkcjonowania genomu. Niektó-

rzy sądzą, iż sam fakt, że znajdują się w genomie od dziesiątków milionów lat dowodzi, że pełnią jakąś funkcję pożyteczną dla organizmu (np. że biorą udział w cyklu rozwojowym), w przeciwnym bowiem przypadku zostałyby wyeliminowane. Podejrzewa się również, że z ich obecnością mogą być związane choroby autoimmunologiczne. Aktywacja wirusów endogennych stanowi potencjalne niebezpieczeństwo towarzyszące ksenotransplantacji (transplantacja tkanek z obcego gatunku) oraz terapii genowej z zastosowaniem wektorów retrowiralnych. Rekombinacja wektora z endogennym retrowirusem może dać nowy potencjalnie niebezpieczny gatunek retrowirusa.

#### ZŁOŻONOŚĆ KOMPOZYCYJNA

Wiadomo, że tylko mały procent (1–2%) DNA to sekwencje kodujące. Z drugiej strony wiadomo, że DNA ewoluował ulegając przypadkowym mutacjom. Wydawałoby się więc, że sekwencja nukleotydów w genomie powinna być, przynajmniej w regionach niekodujących, zbliżona do przypadkowej. To, że tak nie jest, wynika z znacznym stopniu z tego, że mutacje w genomie przy całej przypadkowości, charakteryzuje pewna „tendencyjność” różna w różnych regionach genomu. W genomach kręgowców zjawisko to związane jest z pojawianiem się prążków na chromosomach w komórkach mitotycznych traktowanych różnymi barwnikami. Wzór prążków na chromosomach człowieka jest dobrze poznany, skatalogowany i stosowany do identyfikacji i mapowania chromosomów.

Różne techniki barwienia dają w zasadzie jeden z czterech zasadniczych typów prążków: C, G, R i T. Prążki typu „C” pojawiają się podczas barwienia barwnikiem Giemzy chromosomów traktowanych uprzednio wodoro-tlenkiem baru. Obserwuje się je głównie w okolicach centromerów, gdzie znajduje się szczególnie gęsto upakowany kompleks białek z DNA tak zwana heterochromatyna konstytutywna. Kompleks taki jest tworzony przeważnie przez DNA satelitarne.

Podczas barwienia chromosomów metodą Giemzy w temperaturze pokojowej otrzymuje się prążki typu „G”, natomiast traktowanie chromosomów tym samym barwnikiem, lecz w podwyższonej temperaturze pozwala otrzymać prążki typu „R” (franc. „revers” — odwrotna strona) nazwane tak, ponieważ powstają w miejscach, które pozostają niezabarwione podczas barwienia metodą Giemzy. Prążki typu „R” stopniowo zanikają w miarę podwyższania tem-

peratury. Te, które zanikają ostatnie nazwane zostały prążkami „T”. Znajduje się je głównie w okolicach telomerów.

Związek pomiędzy prążkami na chromosomach a budową DNA zaczęto badać, gdy okazało się, że DNA tak zwanej frakcji głównej genomu, to znaczy tej, która pozostaje po oddzieleniu frakcji satelitarnych, można podzielić na podfrakcje (FILIPSKI i współaut. 1973). Wirowanie w roztworach siarczanu cezu kompleksów DNA z solami rtęci lub srebra daje podfrakcje różniące się ciężarem właściwym, stąd nazwano je L i H (ang. light i heavy) oraz składem zasad. Podfrakcje H są bogate w guaninę i cytozyny, a podfrakcje L w adeninę i tyminę. Prążki chromosomalne „G” zawierają głównie DNA odnajdywane we frakcjach L, prążki „T” zawierają DNA z frakcji H, natomiast prążki „R” zawierają DNA o składzie pośrednim, który odnajduje się zarówno we frakcjach L, jak i H. Frakcje genomu różniące się składem zasad zawierają klasy genów różniące się budową funkcją i sposobem regulacji aktywności transkrypcyjnej (BERNARDI i współaut. 1985).

W genomach organizmów wyższych rozróżnia się dwie główne klasy genów: geny specyficzne tkankowo i geny metabolizmu podstawowego (ang. housekeeping). Te ostatnie są aktywne we wszystkich komórkach organizmu, gdyż zapewniają ich normalne funkcjonowanie. Są wśród nich na przykład geny kodujące enzymy związane z syntezą aminokwasów czy nukleotydów, polimerazy DNA i RNA i tym podobne. W Tabeli 1 podano charakterystykę tych dwóch klas genów. Należy jednak pamiętać, że istnieje wiele genów tkankowo specyficznych, które charakterystyką kompozycyjną i lokalizacją na chromosomach przypominają geny metabolizmu podstawowego.

Wszystkie bez wyjątku geny metabolizmu podstawowego są zbudowane w sposób charakterystyczny: mianowicie segment DNA po stronie 5' (początek sekwencji kodującej) ma charakter „wyspy CpG” (patrz np. artykuł przeglądowy IRA 1997). Są to krótkie (0.5–2 Kb) segmenty chromosomu, w których obserwuje się dużą częstotliwość występowania dinukleotydów CpG, skądinąd rzadkich w genomie (CROSS i BIRD 1995). Większość z nich, z wyjątkiem tych właśnie, które znajdują się w wyspach, jest w genomie kręgowców zmodyfikowana przez podstawienie grupy metylowej w pozycji 5 cytydyny w obydwu niciach (dinukleotyd CpG jest palindromem, a więc jest komplementarny do samego siebie).

Metylacja DNA w chromosomach zachodzi progresywnie w miarę różnicowania komórek

organizmu. Metylowane dinukleotydy CpG wiążą specyficznie białka, które sprzyjają gęstemu upakowaniu DNA w chromatynie. Metylacja jest więc jednym z mechanizmów, który przyczynia się do wyciszenia ekspresji genów „nieużywanych” w danej linii komórkowej, aby „nie przeszkadzały” genom aktywnym. Cena jaką płaci genom za stosowanie tego mechanizmu jest wysoka. Mianowicie 5-metylocytozyna ma skłonność do deaminacji, w wyniku której zmienia się w tyminę a więc w miejsce pary zasad MetC:G powstaje źle sparowana struktura T:G. Podczas replikacji w jednej z nowopowstałych nici DNA powstaje w tym miejscu para G:C a w drugiej para A:T. W konsekwencji mutacje tego typu wyeliminowały podczas ewolucji z genomu kręgowców, 80% dinukleotydów CpG. Stwierdzono, że wiele chorób genetycznych (do 30%) człowieka jest spowodowanych przez te właśnie mutacje.

Wracając do wysp CpG, są to jedyne odcinki genomu, w których nie zachodzi (z pewnymi wyjątkami) metylacja dinukleotydów CpG. Nie podlegają one w związku z tym eliminacji i ich

Tabela 1. Dwie klasy genów w genomach kręgowców

| Geny metabolizmu podstawowego                              | Geny tkankowo specyficzne                                 |
|------------------------------------------------------------|-----------------------------------------------------------|
| bogate w guaninę i cytozyny                                | bogate w adeninę i tyminę                                 |
| zlokalizowane w regionach bogatych w guaninę i cytozyny    | zlokalizowane w regionach bogatych w adeninę i tyminę     |
| obecność wysp CpG                                          | brak wysp CpG                                             |
| replikujące we wczesnej fazie S                            | replikujące w późnej fazie S                              |
| geny krótkie, egzony i introny krótkie i nieliczne         | geny długie, egzony i introny długie i liczne             |
| regiony często rekombinujące                               | regiony rzadko rekombinujące                              |
| zlokalizowane w prążkach T na chromosomach                 | zlokalizowane w prążkach R na chromosomach                |
| zlokalizowane w regionach o dużej gęstości genów           | zlokalizowane w regionach o małej gęstości genów          |
| stosunkowo duża częstotliwość sekwencji powtarzających Alu | stosunkowo duża częstotliwość sekwencji powtarzających L1 |

Tabela przedstawia dwie klasy genów w sposób wyidealizowany, tak aby podkreślić różnice między nimi. W rzeczywistości wiele genów tkankowo specyficznych (około połowy) posiada charakterystykę zbliżoną do genów metabolizmu podstawowego. Klasyfikacja przedstawiona powyżej jest szczególnie zaakcentowana w genomie człowieka.

częstotliwość jest w wyspach kilkakrotnie wyższa niż w sąsiednich odcinkach chromosomu. Efekt ten jest prawdopodobnie związany z fun-

kcją odcinków chromosomu, w których są one zlokalizowane. Znajdują się tam liczne „elementy regulatorowe” odpowiedzialne za regulację transkrypcji genów, którym towarzyszą. Białka regulatorowe, które tworzą kompleksy z elementami regulatorowymi chronią dinukleotydy CpG w „wyspach” przed metylacją. W komórkach linii zarodkowej, a więc w tych komórkach, w których powstają mutacje dziedziczone w następnych pokoleniach, silne kompleksy regulatorowe towarzyszą wszystkiemu genom metabolizmu podstawowego i dlatego wszystkie te geny mają po stronie 5' wyspy CpG. Geny tkankowo specyficzne nie są na ogół aktywne w linii zarodkowej i wyspy CpG towarzyszą tylko niektórym z nich.

Geny, którym towarzyszą wyspy CpG są ponadto bogate w guaninę i cytozynę. Nie jest to wynikiem tego, że kodowanie ich sekwencji aminokwasowych wymaga użycia kodonów bogatych w te nukleotydy. Jest raczej odwrotnie, to znaczy geny te leżą w regionach chromosomów (tzw. izochorach), bogatych w guaninę i cytozynę (znaczną ich część w prążkach „T”) i do budowy swoich kodonów używają takiego budulca jaki mają „pod ręką”. Dlatego tak się dzieje, że pewne regiony w genomie są bogatsze

w guaninę i cytozynę, a inne w adeninę i tyminę nie jest całkiem jasne. Istnieją dwie klasy hipotez na ten temat. Selekcjoniści uważają, że zarówno w części DNA kodującej, jak i w niekodującej, w organizmach o podwyższonej temperaturze ciała (ptaki i ssaki), pary G:C akumulowane są z tego względu, że są bardziej trwałe. Neutraliści natomiast przypisują różnice w składzie zasad temu, że geny metabolizmu podstawowego oraz sekwencje niekodujące im towarzyszące są ekspozowane na działanie innych systemów reperacji uszkodzeń w DNA niż większość genów specyficznych tkankowo. Wiadomo, że różne systemy reperacyjne mają różne preferencje. Na przykład istnieje system, który naprawia źle sparowane zasady G: T produkując zawsze pary G:C co wzbogaca DNA w te zasady. Inny natomiast system reperacji daje w wyniku raz parę G:C a raz A:T. Istnieje również system, który podczas replikacji matrycy częściowo zdepurynowanej (często spotykane uszkodzenie DNA), włącza w nią nowo syntetyzowaną zawsze adeninę itp. Ubocznym skutkiem istnienia różnych systemów reperacyjnych jest to, że szybkość akumulacji mutacji w DNA bywa różna w regionach genomu będących pod ich kontrolą.

#### ANALIZA SEKWENCJI W ZASTOSOWANIU DO BADANIA GENEALOGII GATUNKU LUDZKIEGO, ZWIĄZKÓW POMIĘDZY GRUPAMI ETNICZNYMI ORAZ POWIĄZAŃ RODZINNYCH

Od kilkunastu lat stosuje się analizę sekwencji nukleotydowych w DNA do badania historii ludzkiego gatunku. Analizowane są dwojakiego rodzaju dane. Po pierwsze porównuje się DNA różnych ludzkich populacji, a po drugie bada się sekwencje DNA wyizolowanego z materiału biologicznego pochodzącego od osobników dawno zmarłych. I tak na przykład analiza mitochondrialnego DNA wyizolowanego z kości człowieka neandertalskiego sugeruje, że był on tak różny od człowieka *Homo sapiens*, iż wydaje się wykluczone, aby mogło następować między nimi krzyżowanie się. Neandertalczyki stanowili odrębny od naszego gatunek hominidów. Niektórzy jednak poddają w wątpliwość tę interpretację sugerując, że rekombinacja mogła w ostatnich kilkunastu tysiącach lat bardzo ujednolicić nasz mitochondrialny DNA. W przeszłości mógł on być znacznie bardziej zróżnicowany. Niedawnej domieszki DNA „małpoluda” w naszym genomie nie da się na razie wykluczyć. Sugerowano nawet, że obfite owłosienie mężczyzn, biały kolor skóry i jasne włosy współczesnych mieszkańców Europy są skutkiem takiej domieszki.

DNA mitochondrialny był do niedawna chętnie stosowany do śledzenia ewolucji. Sądzone, że DNA występujący w mitochondriach dziedziczy się ściśle matrylinearnie, przechodząc do komórek embrionu wyłącznie z komórki jajowej matki. Uważano również, że nie ulega on rekombinacji i że, mutacje gromadzą się w nim podczas ewolucji ze stałą prędkością (tzw. „zegar mitochondrialny”). Analiza sekwencji regionu „pętli D” w DNA mitochondrialnym pochodzącym z wielu różnych ludzkich populacji pozwoliła przy tych założeniach wysunąć hipotezę tak zwanej mitochondrialnej Ewy (wspólnego żeńskiego przodka wszystkich ludzi) żyjącej około dwieście tysięcy lat temu w Afryce. W myśl tej teorii człowiek współczesny tam właśnie ewoluował i stamtąd, niedawno, około sto tysięcy lat temu wyemigrował do Azji, Australii, Oceanii i Europy, gdzie całkowicie zastąpił archaiczne populacje hominidów. Jest to teoria zwana „Out of Africa”. Obecnie wiadomo, że przyjęte założenia nie są ściśle. DNA mitochondrialny dziedziczy się rzeczywiście głównie w sposób matrylinearny, ale zdarza się czasami, że mitochondrium plemnika przenika w czasie zapłod-

nienia do komórki jajowej. Stwierdzono również, że DNA mitochondrialny ulega rekombinacji. Ponadto założenie o stałej szybkości akumulacji mutacji również okazało się niesłuszne. Hipoteza mitochondrialnej Ewy musi więc ulec rewizji.

Wykonana niedawno analiza sekwencji genu beta-globiny 349 probantów, pochodzących z różnych ludzkich populacji, daje nieco inny obraz ewolucji człowieka. Sugeruje ona mianowicie, że nasi przodkowie żyli zarówno w Afryce, jak i w Azji jeszcze przed dwustu tysiącami lat, a wielokierunkowy przepływ genów miał miejsce w średnim i późnym pleistocenie w obrębie całego zamieszkałego świata (HARDING i współaut. 1997). Wynik ten jest zgodny z teorią istnienia wielu centrów ewolucji człowieka („teoria multiregionalizmu”), a nie z teorią „Out of Africa”. Teoria multiregionalizmu jest niezbyt lubiana i wręcz bywa uważana za politycznie niesłuszną przez lewicowych ekstremistów, z tego względu, że może być interpretowana przez prawicowych ekstremistów jako dowód na istnienie wśród ludzi ras prymitywnych i ras ewolucyjnie bardziej zaawansowanych. Nie wdając się w te spory należy mieć nadzieję, że wątpliwości dotyczące naszej ewolucji zostaną usunięte, gdy uda się zbadać sekwencje DNA ludzi prehistorycznych.

W ostatnich latach rolę DNA mitochondrialnego w badaniach z dziedziny ewolucji molekularnej zaczyna przejmować chromosom Y, na którym zidentyfikowano szereg szybko ewoluujących mini- i mikrosatelitów. Chromosom Y, podobnie jak nazwisko, przechodzi z ojca na syna i, z wyjątkiem regionu pseudoautosomalnego, nie ulega rekombinacji z chromosomem X. Z tego względu nadaje się szczególnie dobrze do śledzenia ewolucji człowieka. Zbadano na przykład sekwencje DNA markerów na chromosomach Y wielu populacji Indian z Południowej i Północnej Ameryki i porównano je z sekwencjami 60 innych populacji (KARAFET i współaut. 1999). Badacze stwierdzili, że obecnie żyjący Indianie są w większości potomkami dwóch grup, które wyemigrowały z Azji. Jedna z nich była spokrewniona z przodkami ludów zamieszkujących dziś region środkowego Jeniseju, druga natomiast z przodkami ludów żyjących w okolicach jeziora Bajkał. Inne grupy Azjatów również przechodziły cieśninę Beringa, ale ich wkład w genetyczną strukturę populacji Indian jest mniejszy. Interesujące jest to, że niektórzy członkowie jednej z tych grup wyewoluowali również do Europy Zachodniej i potomkowie ich zamieszkują obecnie Wielką Brytanię.

Informacje, które można uzyskać badając sekwencje szybko ewoluujące wzbogacają niekiedy dane genealogiczne dotyczące krótszych okresów czasu. Interesujący wynik tego typu dotyczy żydowskich kapłanów (THOMAS i współaut. 1998). Zgodnie z tradycją biblijną pierwszy z nich został mianowany około 3300 lat temu i odtąd jego potomkowie w linii męskiej, stanowiący około 5% populacji dzisiejszych Żydów, przekazywali z ojca na syna zdolność do piastowania tej godności. Porównanie markerów genetycznych, w tym mikrosatelitarnych, wykazało, że istotnie, ze względu na strukturę DNA na chromosomie Y mężczyźni z tej grupy różnią się od innych Żydów, natomiast wewnątrz grupy różnice są nieznaczne, mimo, że pobrano materiał biologiczny od mężczyzn pochodzących z populacji będących w rozproszeniu od setek lat. Inny interesujący wynik otrzymano zbadawszy blok minisatelitów MSY1 na chromosomie Y u męskich potomków rodziny Jeffersona, jednego z pierwszych prezydentów USA, oraz męskich potomków jednej z jego czarnych niewolnic (FOSTER i współaut. 1998). Okazało się, że niektórzy z tych ostatnich noszą ten sam blok MSY1 co biali członkowie rodziny prezydenta. Skądinąd wiadomo, że blok ten nie występuje u potomków mężczyzn pochodzących z Afryki. Mimo, że badanie minisatelitów nie pozwala rozstrzygnąć, który z Jeffersonów przekazał swój chromosom Y dzieciom niewolnicy, jego wynik dowodzi w sposób niepodlegający dyskusji, że czarni i biali Amerykanie są bliższymi krewnymi, niż wielu z nich chciałoby przyznać.

Badanie genomów żyjących obecnie ludzi mówi nam tylko kto jest z kim spokrewniony i jak blisko, ale nic nie mówi o tym kiedy i gdzie żyli nasi przodkowie. Dane archeologiczne natomiast pozwalają sytuować w czasie i w przestrzeni różne populacje ludzkie, ale nie potrafią dać odpowiedzi na pytanie czy istnieje pokrewieństwo pomiędzy członkami dawno zaginionych grup a żyjącymi obecnie ludźmi. Paleogenetyka pozwoli powiązać te dwie grupy danych. Dane archeologiczne sugerują na przykład, że twórcy pierwszej cywilizacji, Sumerowie, byli w Mezopotamii imigrantami. Pytanie, gdzie mieszkali ich przodkowie, nie znalazło do dziś odpowiedzi. Albo inny, bliższy nam problem: istnieje teoria, która mówi że pierwsze struktury państwowe na ziemiach polskich powstały w wyniku podboju rdzennej ludności słowiańskiej przez irański lud Sarmatów, który następnie zmieszał się z autochtonami (SULIMIRSKI 1970). Czy tak istotnie było? Znaki zapyta-

nia tego typu można będzie prawdopodobnie usunąć. Paleogenetyka odpowie nam, przynajmniej w części, na pytanie, które nas zawsze fascynowało: skąd przyszliśmy i kim jesteśmy? Co z tą wiedzą zrobimy zależeć będzie od polity-

ków. Pierwsze chwiejne kroki genetyki posłużyły im niestety dla uzasadnienia podbojów i ludobójstwa. Miejmy nadzieję, że nauka posłuży im tym razem do budowania mostów, a nie do stawiania murów między ludźmi.

## STRUCTURE AND EVOLUTION OF DNA SEQUENCES IN THE HUMAN CHROMOSOME

### Summary

The origin of the first chromosome which appeared on earth about 4 billions years ago is essentially unknown. It is supposed that, in the reducing atmosphere prevailing at that time, the monomer components of the chromosome were synthesized and dissolved in water producing a "prebiotic soup". These monomers subsequently polymerized to give a "replicator molecule" capable of catalyzing its own replication. The "prebiotic soup" evolved into the "RNA world" in which both the storage of genetic information and the catalytic reactions necessary for the continuation and expansion of this early life were performed by RNA molecules. The complexity of the nucleotide sequence of the replicator increased during its evolution by duplication, recombination, retrotransposition and other mutational processes. These processes were probably similar to those which shape the structure of the modern chromosomes. The

biochemical unity of all living organisms suggests that the LUCA (Last Universal Common Ancestor) eventually appeared in that primordial world. The phylogenetic tree of the species, which evolved from this primitive organism, could be reconstructed by comparing the slowly evolving nucleotide sequences present in the chromosomes of the organisms living today. Various classes of sequences present in the chromosomes evolve with unequal speed. The mini- and micro- satellites are evolving so fast that each population and even each individual has its own characteristic set of these genetic markers. Studies of these polymorphic "elements" in the living and/or extinct organisms allow to reconstruct the phylogenetic relationships between species, populations, subpopulations and individuals, and to shed light on the prehistory of the mankind.

### LITERATURA

- BERNARDI G., OLOFSSON B., FILIPSKI J., ZERIAL M., SALINAS J., CUNY G., MEUNIER-ROTHVAL M., RODIER F., 1985. *The mosaic genome of warm-blooded vertebrates*. Science. 228, 953-958.
- BLACKBURN E. H., 1994. *Telomeres: no end in sight*. Cell. 77, 621-623.
- CROSS S., BIRD A. P., 1995. *CpG islands and genes*. Current Opinion in Genet. Develop. 5, 309-314.
- DAWKINS R., 1976. *The selfish gene*. Oxford University Press Inc., New York, Oxford.
- DE SOUZA S. J., LONG M., GILBERT W., 1996. *Introns and gene evolution*. Genes Cells 1, 493-505.
- DOOLITTLE W. F., BROWN J. R., 1994. *Tempo, mode, the progenote, and the universal root*. Proc. Natl. Acad. Sci. USA. 91, 6721-6728.
- FERRIS J. P., ERTEM G., 1993. *Montmorillonite catalysis of RNA oligomer formation in aqueous solution. A model for the prebiotic formation of RNA*. J. Am Chem. Soc. 115, 12270-12275.
- FILIPSKI J., THIERY J. P., BERNARDI G., 1973. *An analysis of bovine genome by Cs<sub>2</sub>SO<sub>4</sub>-Ag<sup>+</sup> density gradient centrifugation*. J. Mol. Biol. 80, 177-197.
- FOSTER E. A., JOBLING M. A., DE KNIJFF P., MIEREMET R., ZERJAL T., TYLER-SMITH C., 1998. *Jefferson fathered slave's last child*. Nature. 396, 27-28.
- GILBERT W., 1986. *The RNA world*. Nature. 319, 618.
- GORDON K. H. J., 1995. *Were RNA replication and translation directly coupled in the RNA (+ protein ?) world ?* J. theor. Biol. 173, 179-193.
- HARDING R., FULLERTON S. M., GRIFFITHS R. C., BOND J., COX M. J., SCHNEIDER J. A., MOULIN D. S., CLEGG J. B., 1997. *Archaic african and asian lineages in the genetic ancestry of modern humans*. Am. J. Hum. Genet. 60, 772-789.
- HYRIEN O., DEBATISSE M., BUTTIN G., DE SAINT VINCENT B. R., 1987. *A hotspot for novel amplification joints in a mosaic of Alu-like repeats and palindromic A+T-rich DNA*. EMBO J. 6, 2401-2408.
- IRA G., 1997. *Wyspy CpG — jedyne niezmetylowane odcinki DNA charakterystyczne dla genomów kręgowców*. Post. Bioch. 43, 189-198.
- JARUGA E., 1999. *Udział telomerów w regulacji podziałów komórkowych i starzeniu*. Kosmos. 48, 255-264.
- JELINEK W. R., SCHMID C. W., 1982. *Repetitive sequences in eukaryotic DNA and their expression*. Ann. Rev. Biochem. 51, 813-844.
- JEFFREYS A. J., WILSON V., THEIN S. L., 1985. *Hypervariable minisatellite regions in human DNA*. Nature. 314, 67-73.
- JURKA J., SMITH T., 1988. *A fundamental division in the Alu family of repeated sequences*. Proc. Natl. Acad. Sci. USA 85, 4775-4778.
- JOBLING M. A., BOUZEKRI N., TAYLOR P. G., 1998. *Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, MSY1 (DYF155S1)*. Hum. Mol. Genet. 7, 643-653.
- KARAFET T. M., ZEGURA S. L., POSUKH O., OSIPOVA L., BERGEN A., LONG J., GOLDMAN D., KLITZ W., HARIHARA S., DE KNIJFF P., WIEBE V., GRIFFITHS R. C., TEMPLETON A. R., HAMMER M. F., 1999. *Ancestral asian sources of New World Y-chromosome founder haplotypes*. Am. J. Hum. Genet. 64, 817-831.
- LEVINSON G., GUTMAN G. A. 1987. *Slipped strand mispairing: A major mechanism for DNA sequence evolution*. Biol. Mol. Evol. 4, 203-221.
- MARCAIS B., BELLIS M., GERARD A., PAGES M., BOUBLIK Y., ROIZES G., 1991. *Structural organization and polymorphism of the alpha satellite DNA sequences of chromosome 13 and 21 as revealed by pulse field gel electrophoresis*. Hum. Genet. 86, 311-316.
- ORGEL L. E., 1998. *The origin of life — a review of facts and speculations*. TIBS. 23, 491-495.
- SCHWARTZ A., CHAN D. C., BROWN L. G., ALAGAPPAN R., PETTAY D., DISTECHE C., MCGILLIVRAY B., DE LA CHAPPELLE A., PAGE D. C., 1998. *Reconstructing hominid Y evolution: X-homologous block created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination*. Hum. Mol. Genet. 7, 1-11.

- SKOWRONSKI J., SINGER M. F., 1985. *Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line*. Proc. Natl. Acad. Sci. USA. 82, 6050-4.
- SULIMIRSKI T., 1970. *Sarmaci*. PIW., Warszawa. Przekład z angielskiego A i T. Baranowscy.
- TADDEI F., RADMAN M., MAYNARD-SMITH J., TOUPANCE B., GOUYON P. H., GODELLE B., 1997. *Role of mutator alleles in adaptive evolution*. Nature. 387, 700-702.
- THOMAS M. G., SKORECKI K., BEN-AMI H., PARFITT T., BRADMAN N., GOLDSTEIN D. B., 1998. *Origin of Old Testament priests*. Nature. 394, 138-139.
- ZUCKERKANDL E., PAULING L., 1965. *Evolutionary divergence and convergence in proteins*. [W:] *Evolving Genes and Proteins*. BRYSON V., VOGEL H. J., (red.) Academic Press London, New York, pp. 97-106.
- YODA K., ANDO S., OKUDA A., KIKUCHI A., OKAZAKI T., 1998. *In vitro assembly of the CENP-B/alpha satellite DNA /core histone complex: CENP-B causes nucleosome positioning*. Genes to Cells. 3, 533-548.