Wykład I: molekularne podstawy ewolucji

Anna Gambin
3.10.2002

1 Filogenetyka molekularna

Filogenetyka molekularna to nauka badająca genetyczne podobieństwa i zależności ewolucyjne pomiędzy organizmami (najczęściej należącymi do różnych gatunków) lub pojedynczymi genami (białkami). Zależności te prezentowane są najczęściej przy pomocy drzewa filogenetycznego. Większość rozważań opiera się na naturalnym założeniu że gatunki (lub białka/geny) o podobnych cechach są genetycznie bliskie. Początkowo, w klasycznym podejściu do filogenetyki, porównywano różne cechy morfologiczne takie jak budowa ciała, u barwienie, etc. Nowoczesna filogenetyka molekularna, której rozwój rozpoczął się we wczesnych latach 60-tych wraz z pierwszymi metodami sekwencjonowania, skupia się do badaniu materiału genetycznego [1]. Ogromny postęp biotechnologii, który umożliwił badanie przyrastających w coraz większym tempie danych molekularnych, pozwolił również na o wiele szersze i ogólniejsze niż wcześniej podejście do zagadnień ewolucji i systematyki.

Sekwencje kodujące białka i kwasy nukleinowe o wiele lepiej niż dane morfologiczne nadają sie do badań nad ewolucją gatunków z następujących względów:

- każdy żywy organizm posiada swój unikalny materiał genetyczny (zazwyczaj DNA, lub RNA w przypadku niektórych wirusów), co pozwala porównywać między sobą nawet bardzo odlegle ewolucyjnie gatunki;
- niektóre mniej złożone organizmy takie jak bakterie, grzyby lub pierwotniaki wymykały sie klasycznym klasyfikacjom, właśnie z powodu dość ubogiego zestawu cech morfologicznych;
- ewolucja DNA podlega mniej lub bardziej regularnym wzorcom, co pozwala na sformułowanie modeli matematycznych opisujących różnice w sekwencjach genetycznych jakie obserwujemy u różnych organizmów.

2 Mechanizmy ewolucyjne

Podstawowym mechanizmem ewolucji jest mutacja genu. Taki gen-mutant (kodowany przez sekwencję DNA, w której nastąpiła substytucja, insercja, delecja, lub inna mutacja) może się rozpowszechnić w populacji dzięki zjawisku określanemu jako *dryf genetyczny*. Proces ten może zostać zintensyfikowany poprzez *naturalną selekcję*, o ile gen-mutant spowodował powstanie jakiejś nowej cechy organizmu mającej wpływ na jego zdolność rozmnażania.

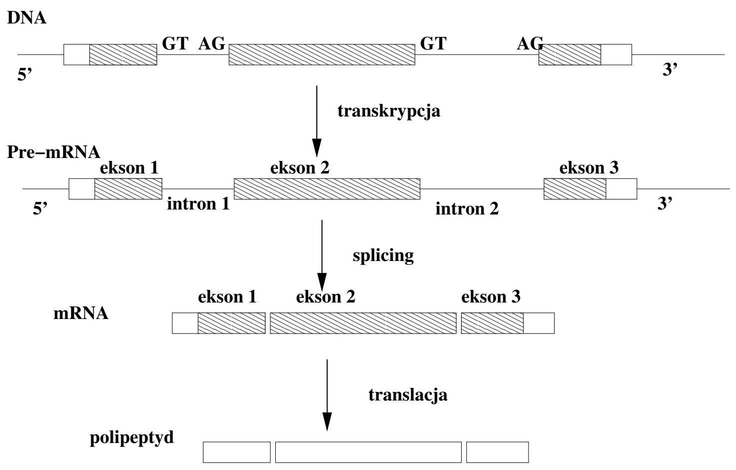
2.1 Struktura i funkcja genów

Nie będziemy na tym wykładzie zajmować się zagadnieniami biologii molekularnej genu; podamy tylko kilka bardzo wstępnych informacji potrzebnych do zrozumienia omawianych zagadnień.

Biorąc pod uwagę funkcje jakie pełnia, możemy zaklasyfikować geny do dwóch grup: **geny kodujące białka** oraz **geny kodujące RNA**. Pierwsza grupa genów w procesie *transkrypcji* zostaje przekształcona w **mRNA — ang. messenger RNA**, a nastepnie przetłumaczona w sekwencję aminokwasów tworzących białka. Z drugiej grupy genów powstają inne rodzaje RNA takie jak **tRNA — ang. transfer RNA** niezbędne w procesach syntezy białek, **rRNA — RNA rybosomalne**, które jest głownym składnikiem rybosomów oraz **snRNA — ang. small nuclear RNA** zaangażowane m.in. w proces wycinania (ang. splicing) *intronów*.

Budowa genów kodujących białka jest naszkicowana na rysunku 1. Wyróżniamy region podlegający transkrypcji oraz regiony regulujące proces transkrypcji DNA na RNA znajdujące się na skrajach nici, zwane $5'$ (początek) oraz $3'$ (koniec). W procesie transkrypcji powstaje RNA zawierające odcinki, które w procesie translacji zostają zgodnie z kodem genetycznym przetłumaczone na ciąg aminokwasów (fragmenty kodujące czyli **eksony**) poprzeplatane z fragmentami (zwanymi **intronami**), które nie kodują aminokwasów lecz stanowią informację niezbędną do regulacji procesu produkcji łańcucha aminokwasów. Introny zostają usunięte w procesie *splicingu* poprzedzającym własciwą translację. Geny organizmów prokariotycznych nie posiadaja intronów, natomiast dla eukariontów ich liczba zmienia się w zależności od genu, dochodząc nawet do kilkudziesięciu. Podczas translacji nić RNA czytana jest sekwencyjnie po trzy nukleotydy jednocześnie. Taka trójka zwana kodonem tłumaczona jest na jeden z 20 aminokwasów.

Kod genetyczny jest redundantny: pośród 64 kodonów wyróżniamy 3,



Rysunek 1: Budowa genu kodującego białko u eukariotów

które nie reprezentują żadnego aminokwasu tylko sygnalizują koniec translacji (**kodony stopu**: UAA, UAG, UGA), pozostałe 61 trójek koduje 20 aminokwasów. Niektóre aminokwasy (np. metionina) kodowane są tylko przez jedną trójkę, inne jak leucyna przez 6 kodonów.

Kodon AUG który reprezentuje metioninę spełnia dodatkową rolę sygnalizując początek translacji.

2.2 Zmiany mutacyjne w sekwencjach DNA

Obserwujemy cztery podstawowe typy zmian (mutacji) jakie mogą wystąpić w sekwencji DNA:

- **substytucje** czyli zamiany jednego nukleotydu na inny;
- **delecje** czyli zniknięcie pewnego (spójnego) ciągu nukleotydów;
- **insercje** czyli wstawienie fragmentu nici;
- **inwersje** czyli odwrócenie kierunku pewnego fragmentu nici.

Jeżeli insercja lub delecja ma miejsce we fragmencie genu, który koduje ciąg aminokwasów może ona spowodować drastyczną zmianę w kodowanym ciągu poprzez przesunięcie ramki odczytu podczas procesu translacji. Substytucje nukleotydów dzielimy na dwie grupy:

- **tranzycje** czyli zamiany puryny (adenina lub guanina) na inną purynę, lub zamiany pomiędzy pirymidami (tymina i cytozyna).
- **transwersje** czyli zamiany pomiędzy grupami.

W większości przypadków tranzycje zachodzą łatwiej (częściej) niż transwersje. Zamiana nukleotydu, która przekształci dany kodon na kodon reprezentujący nadal ten sam aminokwas nazywana jest **synonimiczną**, natomiast mutacja, która powoduje przekształcenie aminokwasu w kodon stopu nazywana jest **nonsensowną**.

Insercje i delecje zdarzają się dość często, zwłaszcza w niekodujących fragmentach DNA. Krótkie fragmenty nici mogą zostać wstawione lub pominięte podczas błędnej replikacji DNA. Dluższe spowodowane są głównie przez zjawisko **transpozycji** DNA polegające na przemieszczaniu się całych fragmentów nici z jednego miejsca w chromosomie w inne lub zjawisko **horyzontalnego transferu genów** pomiędzy gatunkami.

3 Drzewa filogenetyczne

Od czasów Karola Darwina wielu biologów marzy o zrekonstruowaniu historii ewolucji wszystkich organizmów żyjących na ziemi i przedstawieniu jej z pomocą drzewa filogenetycznego.

Innym ciekawym projektem badawczym współczesnej filogenetyki molekularnej jest zrekonstruowanie genetycznej historii gatunku ludzkiego. Rywalizują ze sobą dwie teorie: jedna próbuje dowieść, że gatunek ludzki wywodzi się z populacji zamieszkującej Afrykę, druga natomiast zakłada, że przodkowie wspóczesnego człowieka zamieszkiwali wiele regionów.

Badając zależności filogenetyczne napotykamy wiele trudności spowodowanych przez zjawisko **duplikacji** genów. Zduplikowany gen występuje w organizmie w dwóch (lub więcej) kopiach. Każda kopia ewoluuje niezależnie co prowadzi do powstania całej rodziny spokrewnionych i często podobnych do siebie genów. Ze względu na pochodzenie wyróżniamy następujące rodzaje podobieństwa genów:

- **ortologi** to geny, które są tym samym genem (pełniącym te same funkcje) natomiast pochodzą z różnych organizmów (mających najczęściej wspólnego przodka), ortlologami jest np. hemoglobina czlowieka, konia, królika, etc.

- **paralogi** to geny występujące w tym samym organizmie powstałe w wyniku duplikacji, która dokonała się u pewnego przodka.
- **ksenologi** to geny podobne do siebie z innej przyczyny, np. zostaly przeniesione pomiędzy organizmami poprzez zakażenie wirusem (transfer horyzontalny).

Najczęściej budujemy drzewa w oparciu o ortologi, można jednak znaleźć motywację dla rozważania drzew paralogów.

Każdy węzeł drzewa filogenetycznego reprezentuje gatunek (lub gen) przy czym liscie odpowiadają współcześnie żyjącym gatunkom (zsekwencjonowanym genom) a wierzchołki wewnętrzne hipotetycznym wspólnym przodkom. Rozważa się zarówno drzewa ukorzenione (skierowane) jak i nieukorzenione (pokazują jedynie jak blisko spokrewnione sa dane gatunki/geny). Większości omawianych przez nas metod pozwala skonstruować drzewo nieukorzenione i dopiero później np. poprzez wprowadzenie niespokrewnionego indywiduum (**ang. outgroup**) determinuje się położenie korzenia.

Rozważane przez nas drzewa będą zazwyczaj binarne, ale czasami większy stopień wierzchołka jest nieunikniony (brakuje informacji żeby zrekonstruować dokładną historię ewolucji danej grupy organizmów). Interesującą dla biologa informacją jest zarówno kształt drzewa (jego topologia) jak też wagi na krawędziach, które powinny odzwierciedlać odleglości ewolucyjne lub jeszcze lepiej czas jaki uplynął od wyodrębnienia się rozważanego gatunku.

3.1 Drzewa genów i drzewa gatunków

Drzewo filogenetyczne, które obrazuje historię ewolucyjną pewnej grupy gatunków nazywamy **drzewem gatunków**, natomiast drzewo, które powstało dla grupy genów (po jednym z każdego gatunku) to **drzewo genów**. Drzewa te nie muszą być identyczne. Ponieważ naszym celem jest najczęściej rekonstrukcja drzewa gatunków, a cel ten chcemy osiągnąc używając drzew genów, różnice w tych drzewach powodują wiele problemów natury matematycznej i algorytmicznej, które będziemy omawiać podczas tego wykładu.

3.2 Metody rekonstrukcji drzew

W zależności od typu danych na jakich opieramy metodę rekonstrukcji drzewa filogenetycznego dzielimy rozważane podejścia na dwie grupy:

- oparte na cechach: metoda polega tutaj na badaniu oddzielnie każdej cechy (czyli np każdego nukleotydu w sekwencji DNA), do tej grupy należa metody parsymonii i kompatybilności.

- oparte na odległościach: wejście dla algorytmu stanowi macierz odległości pomiędzy gatunkami (sekwencjami/genami). Odleglości odpowiadają ewolucyjnemu pokrewieństwu genów i są otrzymywane np. na podstawie uliniowien sekwencji lub zostają oszacowane z pomocą matematycznych modeli ewolucji sekwencji.

Literatura

[1] W-H. Li. *Molecular Evolution.* Sinauer Associates, 1997.