

Dopasowanie sekwencji (sequence alignment)

Co to jest alignment?

Alignment – jest sposobem dopasowania struktur pierwszorzędowych DNA, RNA lub białek do zidentyfikowanych regionów w celu określenia podobieństwa; wynik dopasowania w konsekwencji może pokazać relacje między sekwencjami na poziomie strukturalnym, funkcjonalnym i ewolucyjnym.

```
AAB24882      TYHMCQFHCRVYNNHSGEKLYEONERSKAFSOPSHLQCHRRQIGEKTHEHQGRAFPT 60
AAB24881      -----YECMQCKAKAFAQHSSLKCHYRTHIGKPEVCMQCKRAFSK 40
                ***:  ***:  * ** * :***: * *****.

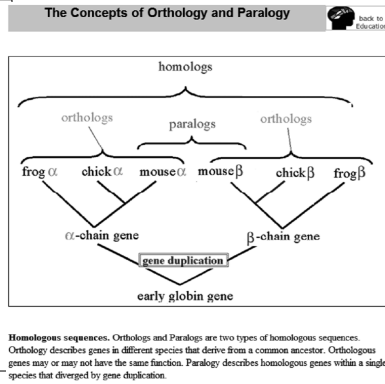
AAB24882      PSHLQVHERHTGKPEVECHQCGQAFKRCSSLQRHRKTHIGKPEVE-CMQCKAKAFAQ- 116
AAB24881      HSHLQCHKRTHIGKPEVCMQCKRAFSQHGLLQRHRKTHIGKPEVMMVNMVKPLHNS 98
                *** * :*****:***: * . :***** : * : :
```

- Podobieństwo – wielkość obserwowalna, wyrażona w procencie identyczności
- Homologia – wniosek wyciągnięty na podstawie podobieństwa dotyczący wspólnego pochodzenia porównywalnych sekwencji; **homologia jest lub jej nie ma**; podobieństwo ze względu na wspólnego przodka (**dywergencja**)
- homoplazja – podobieństwo ze względu na **konwergencję**

- **Dywergencja rozwoju** - w rozwoju ewolucyjnym różnokierunkowe kształtowanie się narządu lub postaci osobników jakiegoś szczepu, wskutek działania odmiennych warunków środowiskowych; np. dywergencja rozwoju kończyn ssaków.
- **Konwergencja** (*łac. convergere, zbierać się, upodabniać się*) - w biologii, proces powstawania morfologicznie i funkcjonalnie podobnych cech (czyli analogicznych) w grupach organizmów odlegle spokrewnionych (niezależnie w różnych liniach ewolucyjnych), odrębnych dla tych grup cech pierwotnych, w odpowiedzi na podobne lub takie same wymagania środowiskowe, np. podobny typ pokarmu, wymagania lokomocyjne

Homologi:

- **ortologi** - różne gatunki, taka sama funkcja (specjacja)
- **paralogi** - podobna funkcja, ale ewoluowały niezależnie (duplikacja, ten sam organizm)
- **analogi** - różne sekwencje, różne motywy, ale identyczna orientacja „ważnych” aminokwasów



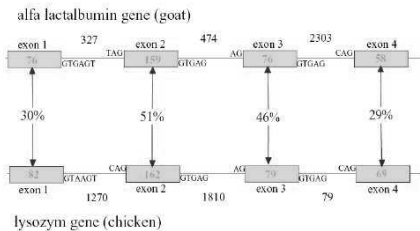
Paralogi a ortologi

- Duplikacje paralogiczne są źródłem nowych genów i nowych funkcji, ale często działających na podobnej zasadzie
Np. poszukiwanie nowych enzymów o funkcji zbliżonej do już znanych
- Geny ortologiczne z reguły (choć nie zawsze) zachowują funkcję
organizmy modelowe – wnioskowanie o funkcji genów na podstawie badań nad innymi organizmami np myszy, a nawet drożdże jako modele do badania chorób człowieka

Homologia – nie zawsze podobne funkcje



... ale nie zawsze

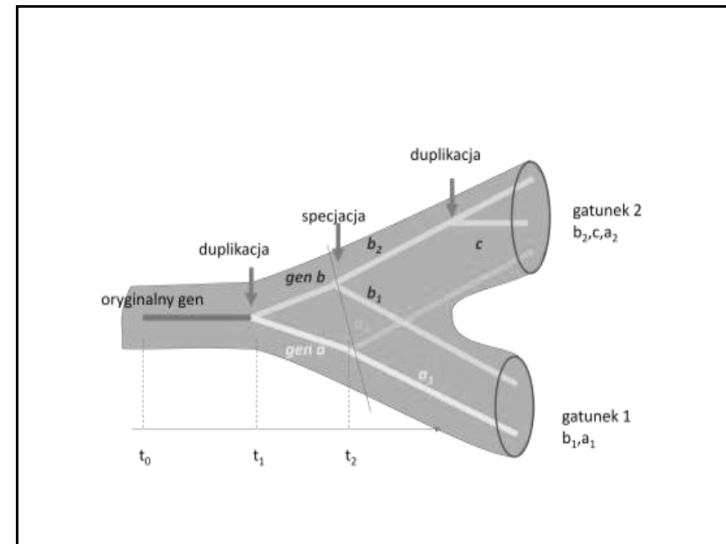
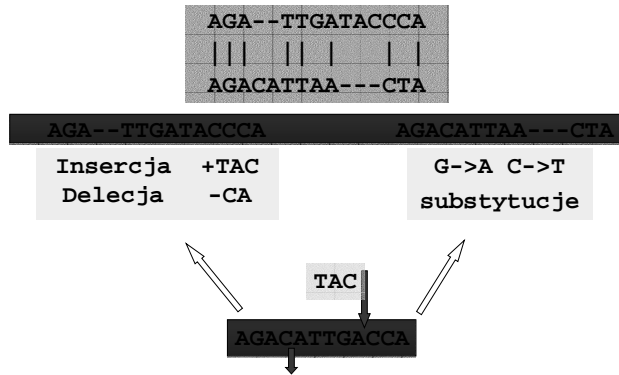


Podobieństwo

- Sekwencje nukleotydowe
 - Zawartość identycznych pozycji między dwoma sekwencjami - % identyczności
 - Długość porównywanych sekwencji
 - Czy identyczne pozycje są zgrupowane, czy też rozproszone w alignmencie
- Sekwencje białkowe
 - Wszystkie powyższe, plus:
 - Podobieństwo pod względem właściwości fizykochemicznych lub kodonów, którymi są kodowane
 - Reszty na konserwatywnych pozycjach – przewidzianych domenach, miejscach katalitycznych.

Mutacje

różnice między sekwencjami świadczą o mutacjach, które zaszły po rozdzieleniu się sekwencji od wspólnego przodka



- Sekwencje są identyczne – nic nowego....
- Sekwencja jest podobna (ma „krewnych”) – nowy członek znanej rodziny
- Sekwencja ma kilka podobnych regionów, motywów lub domen – można zaproponować funkcję
- Nie ma znaczącego podobieństwa – dużo pracy, trzeba eksperymentalnie dowieść funkcji.

Bardzo krótkie lub bardzo podobne sekwencje mogą być dopasowane ręcznie. Bardzo często jednak konieczne jest dopasowanie licznych, bardzo długich i zmiennych sekwencji, które nie mogą być dopasowane tylko i wyłącznie ludzkim wysiłkiem. Zamiast tego, wysiłek wkładany jest w opracowanie algorytmów umożliwiające wysokiej jakości dopasowania, ewentualnie wprowadzanie poprawek do uzyskanych w ten sposób rezultatów, (szczególnie w przypadku sekwencji nukleotydowych).

Dopasowanie globalne a lokalne

- **dopasowanie globalne** – dopasowanie wzdłuż całej sekwencji (zastosowanie: do białek składających się z pojedynczej domeny lub homologicznych słabo zróżnicowanych)



- **dopasowanie lokalne** – uwzględnia domenową naturę białek, szuka subsekwencji (zastosowanie: do białek wielodomenowych, mRNA z sekwencją genomową)



Dopasowanie globalne a lokalne

- Dopasowania globalne, obejmujące pełny zakres wszystkich sekwencji, są najbardziej użyteczne, gdy zestawiane sekwencje są podobne i o zbliżonych rozmiarach.
- Ogólna technika dopasowania globalnego jest znana jako algorytm Needlemana-Wunscha i jest oparta na programowaniu dynamicznym.
- Dopasowania lokalne są bardziej przydatne dla sekwencji nie wykazujących w całości większego podobieństwa, co do których istnieje przypuszczenie, że zawierają podobne subsekwencje czy motywy. Algorytm Smitha-Watermana jest ogólną techniką dopasowania lokalnego, opartą na programowaniu dynamicznym.

Zastosowanie alignmentu

- poszukiwaniu oraz określaniu funkcji i struktury (białek) dla „nowych” sekwencji (nieznanych nam do tej pory)
- określaniu powiązań filogenetycznych między sekwencjami - homologii między sekwencjami oraz w analizach ewolucyjnych

KONIEC