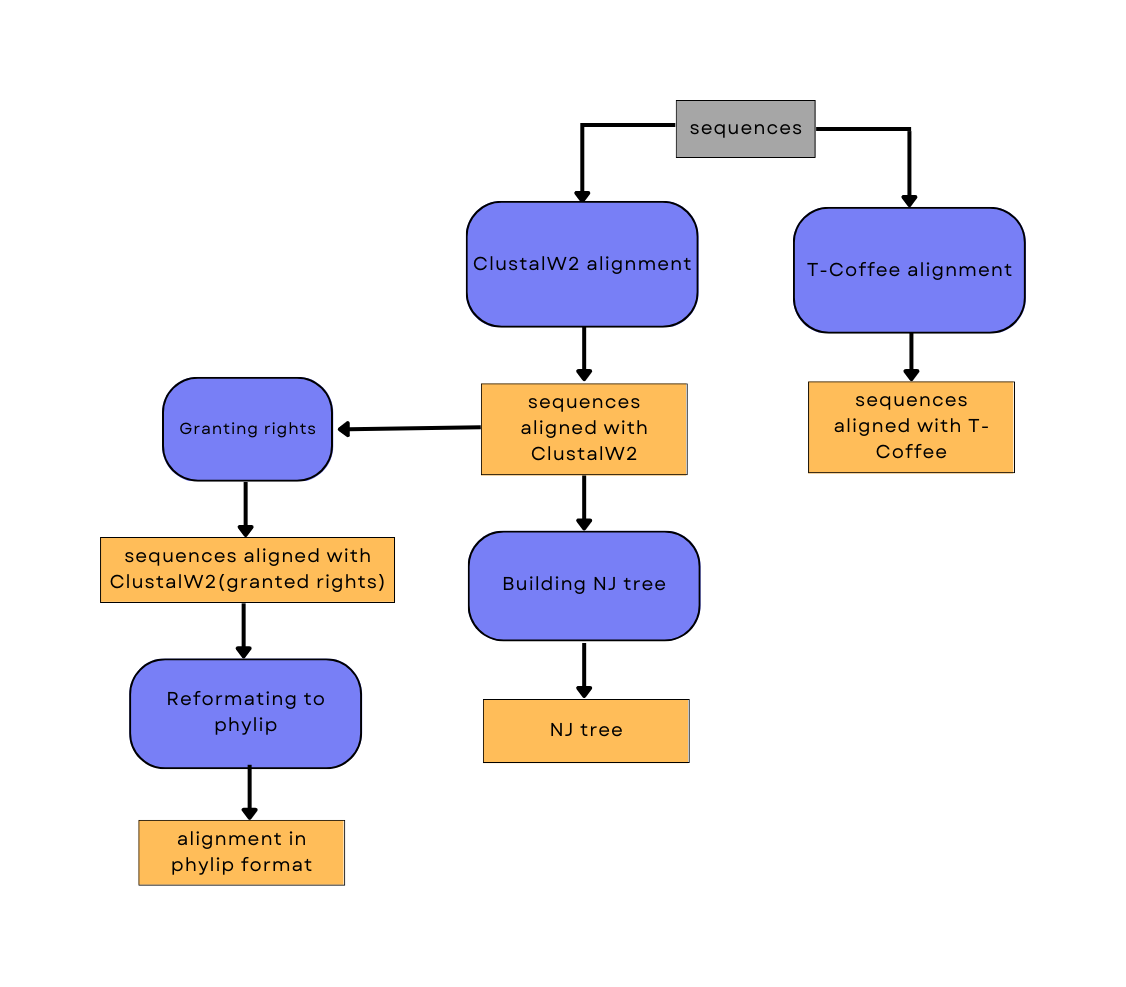
Phylogenetic analysis.nf Autorzy: Paweł Grygielski, Michał Somala

Wstęp

Chcemy uzyskać przyrównania sekwencji mitochondrialnych oraz stworzyć z nich drzewo filogenetyczne W tym celu na pliku zawierającym sekwencje w formacie FASTA, wykonujemy przyrównanie za pomocą ClustalW2 oraz drugie przyrównanie, za pomocą T-Coffee. Aby utworzyć drzewo musimy skorzystać z wcześniej otrzymanego przyrównania. Drzewo zostanie zbudowane metodą NJ (neighbourhood joining). Nadamy prawa do przyrównania ClustalW2 wszystkim użytkownikom(jest to związane z problemem dotyczącym odczytywania pliku w dalszej części pipeline’u, oraz z zamiarem utworzenia możliwości pracy na otrzymanym przyrównaniu dla pozostałych użytkowników). Chcielibyśmy również uzyskać alignment w formacie phylyp, w tym celu przeformatujemy alignment z ClustalW2

Schemat blokowy pipeline’u



Materiały i Metodyka

W naszym przypadku, chcieliśmy poddać analizie 94 sekwencje mitochondrialne fragmentów genu COI u Theraphosidae.

Aby pipeline zadziałał, sekwencje muszą znajdować się w jednym pliku, dodatkowo by połączyć sekwencje w odpowiedni plik, użyjemy załączonego skryptu merge.sh.

Sekwencje umieszczamy w folderze “dane” w bibliotece domowej i uruchamiamy skrypt, który utworzy folder merged, gdzie znajdować się będzie nasz plik z sekwencjami. Z tego folderu i zawartego w nim pliku będzie korzystał nasz pipeline

Pipeline wykorzystuje Nextflow, w związku z czym to środowisko jest niezbędne do jego uruchomienia. Skrypty wewnątrz pipeline’u korzystają z języka programowania “bash”.

Aby pipeline działał poprawnie, poza Nextflowem musimy zainstalować konkretne paczki potrzebne do wykonania przyrównań, zbudowania drzewa i przeformatowania plików:

-ClustalW2, -T-Coffee, -HMMER, -easel

Rezultaty

W trakcie swojego działania, pipeline stworzy w bibliotece domowej folder *“wyniki”*, a w nim kolejne foldery:

-“*clustal\_allignment*” - gdzie znajdować się będzie przyrównanie stworzone przez ClustalW2, w formacie clustal file

-“*tcoffee*” - gdzie znajdować się będzie przyrównanie stworzone przez T-Coffee, formacie clustal file

-*“phylips*” - zawierać będzie przyrównanie z ClustalW2 przeformatowane na phylips, w formacie txt

-“*clustal\_tree*” - w którym znajdziemy drzewo NJ. W formacie PH.

Dyskusje

Obserwacje dotyczące stworzonego oprogramowania:

* Przyrównywanie za pomocą T-Coffee trwa znacznie dłużej niż przyrównywanie za pomocą ClustalW2.
* Największymi problemami okazało się: przeformatowanie pliku na format phylip. Wymagało dwóch paczek, o czym nie mogliśmy znaleźć informacji w dokumentacji esl, a także znalezienie sposobu, aby proces czekał, aż inny proces zakończy pracę.

Porównanie z konwencjonalnym trybem tworzenia pipeline:

Zalety:

* Nextflow umożliwia wykonywanie procesów jednocześnie.
* Wskazywanie, które procesu już zakończyły pracę.
* Znacznie lepsza czytelność.
* Łatwe tworzenie wielu outputów.

Wady:

* Trudność w znalezieniu informacji o rozwiązaniach konkretnych problemów, oraz informacji o błędach i ich naprawie, szczególnie dla DSL2 z którego korzystaliśmy.
* Trudności z implementacją niektórych języków, np. AWK i sed, gdy próbowaliśmy wykorzystać backslash(\).

Zakres obowiązków

kod został napisany przez Michała Somalą i Pawła Grygielskiego

prezentacja Pitch została wykonana przez Michała Somalę

dokumentacja została wykonana przez Pawła Grygielskiego