

Ustanowienie uniwersalnego modelu pangenu

Termin *pangenom* został pierwotnie zaproponowany jako struktura danych do łącznej analizy grupy genów bakteryjnych. Wobec coraz szerszej dostępności różnorodnych sekwencji genomowych znaczenie tego terminu ewoluowało i obecnie odnosi się do modelu umożliwiającego wspólną analizę genomów blisko spokrewnionych organizmów. Taki model powinien umożliwiać wykonywanie różnorodnych operacji, w tym konstrukcję, wizualizację, adnotację, analizę danych eksperymentalnych itd.

Dotychczas zaproponowano wiele modeli pangenu, począwszy od nieustrukturyzowanych kolekcji genomów po skomplikowane obiekty, wymagające złożonego wstępnego przetwarzania danych sekwencyjnych. Do najpopularniejszych rozwiązań należą różnego rodzaju *grafy sekwencji*, w których poszczególne genomy są reprezentowane przez ścieżki posiadające wspólne fragmenty w identycznych lub podobnych obszarach i rozwidlające się w rejonach o odmiennych sekwencjach. Modele te oferują różne funkcjonalności i podlegają określonym ograniczeniom. Wybór modelu determinuje zakres dostępnych narzędzi, co z kolei wpływa na jakość wyników analizy, jej skuteczność, a nawet możliwość jej przeprowadzenia.

Celem tego projektu jest rozwiązanie tego problemu. Opracujemy potoki analizy danych, które przełamią bariery pomiędzy różnymi modelami i połączą ich zalety. Cel ten zostanie osiągnięty w następujących krokach:

1. **Zaprojektowanie i implementacja metod transformacji modeli pangenu.** Skupimy się na transformacjach prowadzących od modeli, które są stosunkowo łatwe do zbudowania, do innych modeli.
2. **Implementacja algorytmów i struktur danych do przenoszenia adnotacji między modelami pangenu.** Adnotacje pangenu obejmują genomy indywidualne, odziedziczone adnotacje tych genomów oraz dane eksperymentalne. Będą one przenoszone zgodnie z transformacjami modelu, co będzie wymagało efektywnej reprezentacji tych ostatnich.
3. **Opracowanie efektywnych potoków analizy danych opartych na transformacjach modeli pangenu.** Przeprowadzimy testy porównawcze narzędzi realizujących podobne zadania na różnych modelach pangenu. Następnie zaproponujemy potoki łączące różne narzędzia i modele.

Podsumowując, nasz projekt doprowadzi do powstania uniwersalnego modelu pangenu.