

Przyrównania wielu sekwencji leżą u podstaw współczesnej biologii obliczeniowej i bioinformatyki. Stanowią one bogate i unikatowe źródło informacji ewolucyjnych na temat sekwencji białek i genomów. Istnieją obecnie różne podejścia do przyrównania sekwencji, które są rutynowo stosowane przez wiele tysięcy badaczy w naukach przyrodniczych. Wydajne przetwarzanie przyrównań wielu sekwencji było jedną z kluczowych cech, które doprowadziły do ogromnego sukcesu metody AlphaFold w przewidywaniu trójwymiarowych struktur białek.

Ostatnie badania wykazały, że funkcjonalne cząsteczki RNA odgrywają kluczowe role w różnych procesach komórkowych. Odkryto szeroką gamę niekodujących RNA, w tym katalityczne rybozomy, ryboprzełączniki wiążące ligandy, długie niekodujące i intronowe RNA, które uczestniczą w różnych aspektach regulacji ekspresji genów. Odkrycia związane z RNA doprowadziły do pojawienia się szczepionek mRNA oraz systemów edycji genomu prowadzonych przez RNA, co znacząco przekształca współczesną biologię i medycynę. Funkcjonalne RNA często tworzą złożone struktury 3D, które definiują ich funkcjonowanie. Badania strukturalne takich RNA umożliwiają zrozumienie mechanizmów leżących u podstaw ich fałdowania.

Dotychczasowe metody przyrównywania sekwencji radzą sobie niestety stosunkowo słabo, gdy chodzi o przyrównywanie sekwencji ustrukturyzowanych cząsteczek RNA. **W przeciwieństwie do 20-literowego alfabetu aminokwasów, 4-literowy alfabet kwasów nukleinowych zazwyczaj nie daje wystarczająco informacji strukturalnych potrzebnych do wygenerowania strukturalnie prawidłowego przyrównania sekwencji. Z tego powodu w tym procesie obliczeniowym powszechnie korzysta się z dodatkowych informacji o strukturze drugorzędowej RNA. Problem polega na tym, że jedynym sposobem na uzyskanie danych o strukturze drugorzędowej RNA jest znajomość prawidłowego przyrównania sekwencji.** Zazwyczaj ten błędny krąg jest przełamany przez wielokrotne rundy naprzemiennych prób przyrównywania sekwencji i przewidywania struktury drugorzędowej, wspomagane różnymi rodzajami danych eksperymentalnych i ręcznym poprawianiem danych przez ekspertów.

Głównym celem proponowanego projektu jest opracowanie pakietu narzędzi obliczeniowych, obejmujących nowe algorytmy do przyrównywania wielu sekwencji RNA, przewidywania struktury drugorzędowej RNA oraz porównywania sekwencji RNA na podstawie struktury, a następnie zastosowanie tego pakietu do zbadania zachowania ewolucyjnego struktur drugorzędowych, pseudowęzłów i alternatywnych zwoi 3D, których podobieństwa pozostają niewykrywalne dla obecnie istniejących metod. Nowa metoda opracowana w ramach niniejszego projektu będzie miała zastosowanie do zestawu nieprzyrównanych homologicznych sekwencji ustrukturyzowanych RNA. Ponadto proponowane podejście umożliwi porównania sekwencji RNA i przyrównań sekwencji z użyciem informacji o ich strukturze. Wyniki zostaną opracowane do generowania ulepszonej klasyfikacji rodzin niekodujących RNA wraz z przeszukiwalną bazą danych.

Nowa metoda będzie wykorzystywać koncepcję przybliżonego przyrównania sekwencji i będzie w stanie efektywnie obsługiwać struktury niekanoniczne i alternatywne.

Oczekiwane wyniki tego projektu będą dwojakie. Postępy metodologiczne proponowanych algorytmów i ich implementacja w oprogramowaniu mogą zainteresować tysiące badaczy zajmujących się nauką o RNA, ponieważ rozwiązują one istotny problem wielu procesów związanych z RNA. Ponadto, kompleksowa analiza porównawcza ustrukturyzowanych RNA przy użyciu sugerowanej metody pozwoli odkryć ewolucyjnie zachowane elementy strukturalne niekodujących RNA, co poprawi nasze zrozumienie roli RNA w ważnych procesach komórkowych.