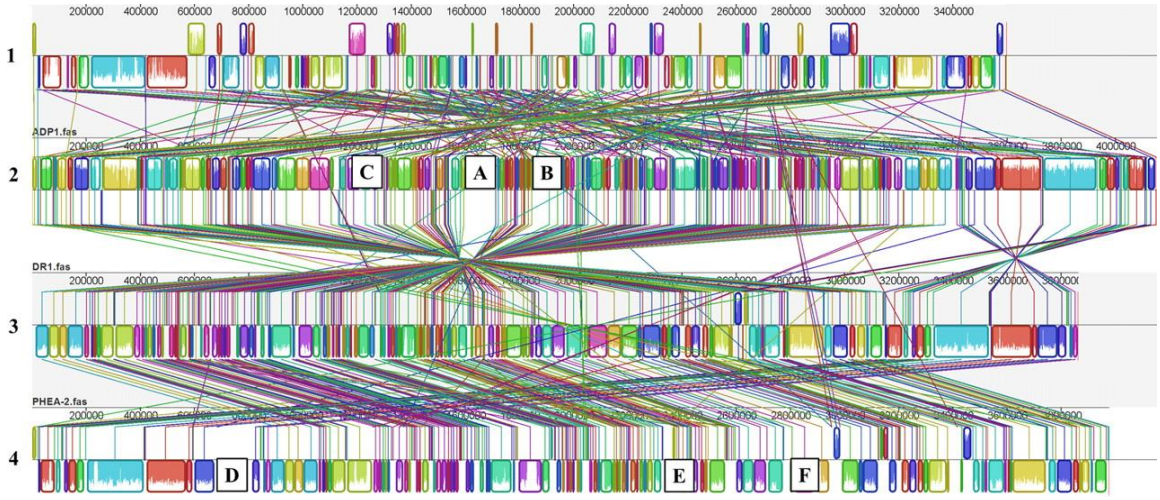
Sekwencję genomu człowieka opublikowano po raz pierwszy w 2003 roku. Program *Human Genome Project* realizowany był przez 13 lat przy udziale ośrodków naukowych z całego świata i pochłonął niemal 3 miliardy dolarów. Dzisiejsze technologie umożliwiają zsekwencjonowanie ludzkiego genomu w ciągu kilku dni, przy koszcie nieprzekraczającym tysiąca dolarów. Wynikająca z tego gwałtownie wzrastająca dostępność danych sekwencyjnych, zarówno genomowych, jak i białkowych, otworzyła ogrom nowych możliwość w wielu dziedzinach współczesnej nauki. Dopasowywanie wielu sekwencji (ang. *multiple sequence alignment*) jest bez wątpienia jedną z najpopularniejszych analiz wykonywanych na danych sekwencyjnych. Polega ona na zidentyfikowaniu występujących w procesie ewolucji sekwencji zdarzeń. Obejmują one proste modyfikacje (wstawienie bądź usunięcie fragmentu, czy zamiana symbolu na inny), jak i bardziej skomplikowane zjawiska (rearanżacje, duplikacje, horyzontalne transfery genów). Wynik dopasowania często reprezentowany jest graficznie (Rys.1).



Rys.1. Reprezentacja graficzna dopasowania genomów czterech gatunków (reprezentowanych przez wiersze). Widoczna jest obecność wielu skomplikowanych zjawisk ewolucyjnych takich jak rearanżacje czy duplikacje. Odpowiadające sobie bloki w różnych organizmach (regiony homologiczne) oznaczono kolorami i połączono liniami (Jung et al., 2011).

Dopasowanie sekwencji ma kluczowe znaczenie dla zrozumienia budowy i procesów zachodzących w żywych organizmach. Przykładowo, wzbogaca wiedzę o strukturze i funkcjach białek w ludzkim ciele, czy przybliża mechanizmy ekspresji genów poprzez identyfikację miejsc wiązań cząsteczek regulatorowych. Tym samym jest niezwykle istotne dla odkrycia przyczyn wielu chorób i opracowania metod ich leczenia. Inny obszar zastosowania to badanie zależności ewolucyjnych pomiędzy organizmami, co stanowi ważny wkład w poznanie historii wyodrębniania się gatunków.

W ramach projektu planowana jest praca nad doskonalszymi algorytmami dopasowania wielu sekwencji genomowych i białkowych. W badania poruszany będzie zarówno aspekt jakości wyników, ocenianej w odniesieniu do dostępnych publicznie dopasowań wzorcowych, jak i szybkości działania algorytmów. Drugi z wymienionych elementów jest szczególnie istotny w obliczu nieustannie rosnącej dostępności danych. Problemem jest liczba sekwencji (jak ma to miejsce w przypadku dopasowań rodzin białkowych posiadających dziesiątki tysięcy przedstawicieli), a także ich długość i obecność skomplikowanych zjawisk ewolucyjnych (co występuje w przypadku analiz genomów, których chromosomy składają się z nawet z setek milionów nukleotydów). Powszechność procedury dopasowania sprawia, że wyniki badań będą mogły znaleźć zastosowanie nie tylko w biologii i biotechnologii, ale także w medycynie, farmakologii, rolnictwie, ochronie środowiska i wielu innych dziedzinach.