Politechnika Śląska

PROJEKT INŻYNIERSKI

Narzędzie do analiz filogenetycznych.

Filip WSPANIAŁY
Nr albumu: ⟨**306982**⟩

Kierunek: ⟨Inżynieria biomedyczna⟩
Specjalność: ⟨Informatyka i aparatura medyczna⟩

PROWADZĄCY PRACĘ
⟨**dr inż. Anna Tamulewicz**⟩
KATEDRA ⟨**Katedra Informatyki Medycznej i Sztucznej Inteligencji**⟩
Wydział Inżynierii Biomedycznej

Zabrze 2026

Tytuł pracy

Narzędzie do analiz filogenetycznych.

Streszczenie

Celem niniejszej pracy inżynierskiej było zaprojektowanie i zaimplementowanie systemu informatycznego do analiz filogenetycznych sekwencji białkowych. W ramach pracy pogłębiono wiedzę z zakresu filogenetyki oraz zintegrowano uznane algorytmy i narzędzia bioinformatyczne (MAFFT[1], IQ-TREE[2], Biopython[3]), tworząc kompleksowe rozwiązanie do przetwarzania danych. Zaprezentowano architekturę aplikacji umożliwiającą przygotowanie danych wejściowych, wyrównanie sekwencji, analizę filogenetyczną z automatycznym wyborem modelu ewolucyjnego oraz wizualizację drzew filogenetycznych. W pracy opisano szczegółowo strukturę programu, uzasadnienie wyboru narzędzi oraz wyniki testów na rzeczywistych danych z NCBI.

Słowa kluczowe

filogenetyka, bioinformatyka, wyrównanie sekwencji, analiza filogenetyczna, model ewolucyjny, drzewo filogenetyczne

Thesis title

Phylogenetic analysis tool.

Abstract

The objective of this engineering thesis was to design and implement a computational system for phylogenetic analysis of protein sequences. The work deepened knowledge in phylogenetics and integrated established bioinformatics algorithms and tools (MAFFT, IQ-TREE, Biopython), creating a comprehensive solution for data processing. The application architecture is presented, enabling input data preparation, sequence alignment, phylogenetic analysis with automatic evolutionary model selection, and visualization of phylogenetic trees. The thesis provides a detailed description of the program structure, rationale for tool selection, and test results on real datasets from NCBI.

Key words

phylogenetics, bioinformatics, sequence alignment, phylogenetic analysis, evolutionary model, phylogenetic tree

Spis treści

Rozdział 1

Wstęp

1.1 Wprowadzenie do tematu

Analiza filogenetyczna wykorzystuje dane z dziedziny filogenetyki do rekonstrukcji ewolucyjnych zależności i podobieństw międzygatunkowych. Efektem procesu jest drzewo filogenetyczne, generowane w oparciu o wybrany model substytucji, które za pomocą rozgałęzień ilustruje odległości ewolucyjne między analizowanymi taksonami. Kluczowym przełomem w rozwoju tej dyscypliny okazał się postęp informatyki i programowania, umożliwiający automatyzację metod obliczeniowych oraz znaczące przyspieszenie analiz sekwencji genetycznych. Rozwój ten przyczynił się bezpośrednio do powstania nowych algorytmów wyrównywania sekwencji oraz metod inferencji drzew filogenetycznych. Wraz ze wzrostem złożoności algorytmów oraz liczby dostępnych narzędzi, proces analizy filogenetycznej przestał być jednorazowym obliczeniem, a stał się wieloetapowym zadaniem wymagającym doboru metod, parametrów oraz interpretacji wyników. W praktyce badawczej prowadzi to do konieczności łączenia wielu narzędzi programistycznych oraz zarządzania złożonymi procesami obliczeniowymi.

1.2 Osadzenie problemu w dziedzinie

Współczesna analiza filogenetyczna osiągnęła wysoki poziom zaawansowania dzięki ciągłemu doskonaleniu algorytmów i opracowywaniu nowych metod rekonstrukcji drzew ewolucyjnych. Natomiast brak pewności co do optymalności najlepszych algorytmów powoduje, że nawet systemy oceny i porównywania metod opierają się głównie na statystyce. Kluczowym wyzwaniem w analizach filogenetycznych pozostaje pytanie o prawdopodobieństwo, że uzyskane rozwiązanie jest rzeczywiście najlepsze dla danego zbioru danych. Istniejące systemy analityczne mogą wskazać optymalny algorytm, metodę lub model substytucji dla konkretnej sekwencji, jednak ostateczny wybór wymaga integracji wielu narzędzi w spójny workflow. W praktyce badawczej analizy filogenetyczne wymagają

wielokrotnego testowania algorytmów, modeli substytucji oraz parametrów wejściowych, co prowadzi do powstawania złożonych, trudnych do odtworzenia, porównania oraz powtarzalnego uruchamiania workflowów analitycznych. Brakuje systemów, które w sposób zintegrowany umożliwiałyby porównywanie wyników różnych metod, zarządzanie eksperymentami analitycznymi oraz wspomaganie decyzji o wyborze końcowego drzewa filogenetycznego.

1.3 Cel pracy

Celem pracy było zaprojektowanie oraz zaimplementowanie systemu wspomagającego analizy filogenetyczne, umożliwiającego integrację poszczególnych etapów procesu analitycznego w spójny workflow. System ma na celu wsparcie użytkownika w rekonstrukcji drzew filogenetycznych poprzez automatyzację kluczowych kroków analizy oraz uporządkowaną prezentację wyników.

1.4 Zakres pracy

Zakres pracy obejmuje analizę podstaw teoretycznych filogenetyki oraz przegląd wybranych metod i narzędzi wykorzystywanych w analizach filogenetycznych. W ramach pracy dokonano porównania dostępnych podejść do wyrównywania sekwencji, inferencji drzew filogenetycznych oraz doboru modeli ewolucyjnych. Praca obejmuje zaprojektowanie i implementację systemu integrującego wybrane narzędzia analityczne w spójny workflow, umożliwiającego przeprowadzenie analizy filogenetycznej oraz wizualizację uzyskanych wyników. Zakres pracy nie obejmuje opracowywania nowych algorytmów filogenetycznych ani formalnej oceny biologicznej poprawności uzyskanych drzew.

1.5 Zwięzła charakterystyka rozdziałów

1.6 Określenie wkładu autora

Autor odpowiadał za zaprojektowanie architektury oraz implementację systemu wspomagającego analizy filogenetyczne. W ramach pracy autor dokonał integracji wybranych narzędzi do wyrównywania sekwencji, inferencji drzew filogenetycznych oraz doboru modeli ewolucyjnych w spójny workflow analityczny. Autor był również odpowiedzialny za przygotowanie mechanizmów wizualizacji wyników analizy oraz obsługę danych wejściowych, w tym pozyskiwanie sekwencji genetycznych z publicznie dostępnej bazy NCBI.

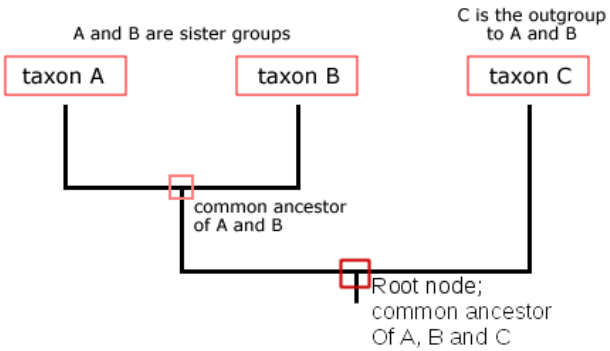
Rozdział 2

Podstawy teoretyczne analizy filogenetycznej

2.1 Filogenetyka i drzewa filogenetyczne

Filogenetyka jest dziedziną biologii zajmującą się badaniem filogenezy, czyli historii rozwoju rodowego organizmów oraz relacji pokrewieństwa pomiędzy taksonami. Obejmuje ona analizę przebiegu procesów ewolucyjnych prowadzących do różnicowania organizmów i powstawania nowych linii rozwojowych. Analiza filogenetyczna umożliwia określanie zależności ewolucyjnych między gatunkami i taksonami na podstawie różnych źródeł danych, takich jak zapisy paleontologiczne, anatomia porównawcza oraz dane molekularne.

W niniejszej pracy wykorzystywane są metody filogenetyki molekularnej, które opierają się na analizie sekwencji DNA lub białek w celu rekonstrukcji relacji ewolucyjnych. Wynikiem takiej analizy jest przedstawione na rysunku 2.1 drzewo filogenetyczne — struktura graficzna przedstawiająca hipotezę pokrewieństwa pomiędzy badanymi taksonami. W zależności od zastosowanej metody rekonstrukcji oraz modelu ewolucyjnego, długości gałęzi drzewa mogą odzwierciedlać miarę zmian genetycznych lub mieć charakter wyłącznie topologiczny. [4][5]



Rysunek 2.1: Drzewo filogenetyczne; źródło: [1]

2.2 Sekwencje genetyczne jako dane wejściowe

Sekwencje genetyczne stanowią podstawowe dane wejściowe wykorzystywane w analizach filogenetycznych realizowanych przez system. Są to uporządkowane ciągi symboli reprezentujących nukleotydy DNA lub aminokwasy budujące białka. W zależności od rodzaju analizy, system może operować na sekwencjach nukleotydowych lub sekwencjach aminokwasowych.

W praktyce analitycznej sekwencje genetyczne pozyskiwane są z badań własnych lub publicznych baz danych, takich jak NCBI, DDBJ, czy ENA i najczęściej zapisywane w formacie FASTA pokaznym na rysunku 2.2. Format ten umożliwia jednoznaczną identyfikację sekwencji oraz jej dalsze przetwarzanie przez narzędzia bioinformatyczne. Sekwencje mogą różnić się długością, stopniem kompletności oraz jakością danych, co wpływa na przebieg kolejnych etapów analizy. [4]

```
>NP_999866.1 cytochrome c oxidase subunit 4 isoform 1, mitochondrial [Danio rerio]
MLATTAFRLVGKRALSTSICLRGAHGVAKVEDYSLPAYFDRRESPLPEIKFVQQLSADQKSLKEKEKGSW
AALSKEEKIALYRISFKESFAEMNQGSGEWKSVVAGIFFFVGLTGLVVLWQRKYVYGDVPNTFDPEYKQK
EIQRMLDMRINPVQGFAAKWDYENNAWKK
```

Rysunek 2.2: Sekwencja białkowa; źródło: [2]

Ze względu na występowanie różnic długości sekwencji oraz obecność insercji i delecji, bezpośrednie porównywanie sekwencji nie jest możliwe. Z tego powodu przed rekonstrukcją drzewa filogenetycznego konieczne jest przeprowadzenie etapu wyrównywania sekwencji, który umożliwia ich porównywanie w ujednoliconej postaci.

2.3 Wyrównywanie sekwencji

Wyrównanie sekwencji jest kluczowym etapem w procesie analizy filogenetycznej, umożliwiającym porównywanie sekwencji genetycznych pochodzących od różnych organizmów. W wyniku procesów ewolucyjnych, takich jak insercje i delecje (indele), sekwencje mogą różnić się długością oraz zawierać przesunięcia pozycji homologicznych, co uniemożliwia ich bezpośrednie porównanie.

Celem wyrównania sekwencji jest identyfikacja pozycji homologicznych pomiędzy sekwencjami poprzez wprowadzenie przerw (ang. gaps), tak aby możliwe było ich dalsze przetwarzanie w kolejnych etapach analizy (rysunek 2.3). Wyrównanie pozwala na ujednolicenie długości sekwencji oraz określenie podobieństw i różnic wynikających z przebiegu ewolucji.

```
>seq1
M---LASRAF-SLIGRR--ALSTSICLR--A--------HGHAGVVKAEDFSLPAYVDRR
DVPLPEAAFVKQLSAQQKALKEKEKASWTALSVDEKVELYRIKFNETYAEMNKGTNEWKT
VLGGVLFFLGLTGVILIWQKIYMYGPIPHTFSDEWVSMQTKRMLDMRINPVEGISSQWDF
EKNEWKK
>seq2
M---LATRVF-NLIGRR--AISTSVCVR--A--------HG--SVVKSEDYALPVYVDRR
DYPLPDVAHVKNLSASQKALKEKEKASWSSLSMDEKVELYRLKFNESFAEMNRSTNEWKT
IVGTALFFIGFTALLLIWEKHYVYGPIPHTFEEEWVAKQTKRMLDMKVAPIQGFSAKWDY
DKNEWKK
```

Rysunek 2.3: Wyrównane sekwencje białkowe; źródło: [2]

W zależności od zastosowanego algorytmu możliwe jest wykonywanie wyrównań globalnych, obejmujących całe sekwencje, lub lokalnych, koncentrujących się na ich fragmentach. Przykładami klasycznych algorytmów wykorzystywanych w tym celu są: algorytm Needlemana–Wunscha dla wyrównań globalnych oraz algorytm Smitha–Watermana dla wyrównań lokalnych. Dobór odpowiedniej metody wyrównania ma istotny wpływ na jakość dalszej analizy filogenetycznej. Wyrównanie może wymagać także poprawek manualnych, zwłaszcza w przypadku sekwencji o niskim stopniu podobieństwa lub zawierających liczne indels. W praktyce badawczej często stosuje się podejście iteracyjne, polegające na wielokrotnym wyrównywaniu sekwencji z różnymi parametrami oraz ręcznej korekcie wyników w celu uzyskania optymalnego wyrównania.[4][6]

2.4 Metody rekonstrukcji drzew filogenetycznych

Rekonstrukcja drzewa filogenetycznego stanowi złożone zagadnienie statystyczne i algorytmiczne, a jej wynik zależy od przyjętych założeń biologicznych oraz zastosowanej metody analizy. W praktyce badawczej dostępnych jest wiele metod rekonstrukcji filogenezy, które mogą prowadzić do odmiennych wyników nawet dla tego samego zestawu danych. Z tego względu często stosuje się podejście polegające na porównywaniu rezultatów uzyskanych z wykorzystaniem różnych metod.

Wśród podstawowych metod rekonstrukcji drzew filogenetycznych wyróżnia się metody:

- metoda największej parsymonii,
- metody odległościowe,
- metody największej wiarygodności,
- metody bayesowskie.

Metody te różnią się sposobem modelowania procesu ewolucyjnego oraz podejściem do oceny najlepszego drzewa filogenetycznego. W zależności od zastosowanej metody, długości gałęzi drzewa mogą reprezentować liczbę zmian ewolucyjnych, estymowaną odległość genetyczną lub mieć charakter wyłącznie topologiczny.

Metody rekonstrukcji drzew filogenetycznych różnią się zakresem przyjmowanych założeń oraz stopniem złożoności obliczeniowej. Metoda największej parsymonii opiera się na minimalizacji liczby zmian ewolucyjnych i nie wykorzystuje jawnych modeli probabilistycznych. Metody odległościowe bazują na macierzy odległości genetycznych pomiędzy sekwencjami, upraszczając analizę kosztem utraty części informacji. Metody największej wiarygodności oraz metody bayesowskie wykorzystują modele probabilistyczne opisujące proces substytucji, co pozwala na bardziej realistyczne modelowanie ewolucji, jednak wiąże się z większym kosztem obliczeniowym. [6]

Rozdział 3

Przegląd istniejących narzędzi i rozwiązań

3.1 Narzędzia do wyrównywania sekwencji

Szeroki wachlarz dostępnych programów do wyrównywania sekwencji umożliwia elastyczne dopasowanie narzędzia do rodzaju oraz rozmiaru analizowanych danych oraz potrzeb badawczych. Programy te wykorzystują różne algorytmy i strategie dopasowania, od klasycznych metod opartych na dynamicznym programowaniu, po bardziej zaawansowane heurystyki optymalizacyjne. Poniżej przedstawiono przegląd wybranych narzędzi do wielosekwencyjnego wyrównywania sekwencji[4]:

- **Clustal W i Clustal Omega**
- **MAFFT**
- **MUSCLE**
- **T-Coffee**
- **PRANK**
- **ProbCons**

Do analizy i wprowadzania poprawek ręcznie mozna wykorzystać edytory wyrównań takie jak[4]:

- **Jalview**
- **AliView**
- **MacVim**

3.2 Narzędzia do inferencji filogenetycznej

Inferencja filogenetyczna polega na rekonstrukcji drzewa filogenetycznego w oparciu o dane sekwencyjne. Dostępne są różne programy i metody, które realizują tę funkcję, wykorzystując odmienne podejścia. Wybór odpowiedniego narzędzia zależy od rodzaju danych, liczby sekwencji oraz wymagań dotyczących dokładności i czasu obliczeń[4].

- **IQ-TREE 2**
- **RAxML**
- **MrBayes**
- **PhyML**
- **PHYLIP**
- **PAUP***
- **MEGA**

Po zakończeniu inferencji filogenetycznej często konieczne jest ocenienie wiarygodności uzyskanego drzewa. Do tego celu stosuje się metody takie jak bootstrap czy analiza bayesowska, które pozwalają na oszacowanie pewności poszczególnych gałęzi drzewa. Wiele z wymienionych narzędzi oferuje wbudowane funkcje do przeprowadzania takich analiz.

Po zakończeniu analizy filogenetycznej otrzymuje się w postaci pliku tekstowego zawierającego opis drzewa w formacie Newick lub Nexus, który może być dalej przetwarzany i wizualizowany za pomocą dedykowanych narzędzi.[4]

3.3 Narzędzia do wizualizacji drzew filogenetycznych

Do wizualizacji drzew filogenetycznych dostępne są różne narzędzia, które umożliwiają graficzne przedstawienie wyników analizy filogenetycznej. Poniżej przedstawiono wybrane programy do wizualizacji drzew:

- **FigTree**
- **Dendroscope**
- **iTOL**
- **ETE Toolkit**
- **Phylo.io**

3.4 Ograniczenia istniejących rozwiązań

Chociaż istnieją systemy i frameworki do automatyzacji analiz bioinformatycznych (np. Galaxy, Snakemake, Nextflow), w praktyce często wymaga się ręcznego ustawiania parametrów poszczególnych narzędzi oraz integracji wyników. Zaprojektowany system automatyzuje pełny łańcuch analizy — od wczytania sekwencji, przez wyrównanie i rekonstrukcję drzewa filogenetycznego, aż po wizualizację wyników — co ułatwia prowadzenie badań, zwiększa powtarzalność wyników i umożliwia prostsze korzystanie z analizy filogenetycznej, także osobom dopiero rozpoczynającym pracę z tymi metodami.

3.5 Opis zastosowanych technologii i narzędzi

3.5.1 Python

Python jest wysokopoziomowym językiem programowania, powszechnie wykorzystywanym w bioinformatyce oraz do automatyzacji analiz danych. W projekcie Python został użyty jako główny język implementacji systemu, odpowiadający za sterowanie przebiegiem analizy filogenetycznej, obsługę interfejsu użytkownika oraz integrację z zewnętrznymi narzędziami bioinformatycznymi. Implementacja została wykonana w środowisku programistycznym Visual Studio Code.

Biblioteki

W projekcie wykorzystano następujące biblioteki Pythona:

- **Biopython** – analiza danych biologicznych
- **subprocess** – integracja z zewnętrznymi narzędziami (MAFFT, IQ-TREE)
- **Matplotlib** – wizualizacja drzew.
- **tkinter** – interfejs użytkownika do tworzenia aplikacji okienkowych.

3.5.2 Ubuntu - Linux

System wykorzystuje środowisko systemu operacyjnego Ubuntu (Linux). Wybór systemu Linux podyktowany był wysoką kompatybilnością z narzędziami bioinformatycznymi, takimi jak MAFFT oraz IQ-TREE 2, które są natywnie rozwijane i testowane w tym środowisku.

3.5.3 MAFFT

Narzędzie do wyrównywania sekwencji DNA i białek. MAFFT oferuje różne algorytmy, które można dostosować do rozmiaru i charakterystyki danych wejściowych. W projekcie MAFFT został wykorzystany do przeprowadzenia etapu wyrównywania sekwencji przed rekonstrukcją drzewa filogenetycznego.

3.5.4 IQ-TREE 2

Narzędzie do rekonstrukcji drzew filogenetycznych metodą największej wiarygodności (Maximum Likelihood). Program umożliwia automatyczny dobór modelu substytucji przy użyciu `ModelFinder`[7], oferuje wydajne algorytmy optymalizacji drzewa oraz pozwala na ocenę stabilności węzłów drzewa przy użyciu ultrafast bootstrap `UFBoot2`[8], co czyni go jednym z najczęściej wykorzystywanych narzędzi do inferencji filogenetycznej

3.5.5 System kontroli wersji: Git

Git pozwala na monitorowanie zmian oraz zarządzanie historią w kodzie źródłowym.

3.5.6 Repozytorium zdalne: Github

Github to zdalne repozytorium w pełni zintegrowane z Git. Zostało użyte do bezpiecznego przechowywania projektu oraz umożliwienia nadzoru nad postępem prac.

Rozdział 4

Specyfikacja zewnętrzna systemu

4.1 Wymagania sprzętowe i systemowe

Minimalne wymagania sprzętowe:

- Procesor: dwurdzeniowy, 2 GHz lub szybszy
- Pamięć RAM: minimum 4 GB
- Przestrzeń dyskowa: co najmniej 5 GB wolnego miejsca

Minimalne wymagania systemowe:

- System operacyjny: Windows 10 (64-bit) lub nowszy
- Windows Subsystem for Linux (WSL) z zainstalowaną dystrybucją Ubuntu 20.04 LTS lub nowszą
- MAFFT w wersji 7.475 lub nowszej, zainstalowany w środowisku WSL
- IQ-TREE 2 w wersji 2.1.3 lub nowszej, zainstalowany w środowisku WSL

4.2 Instalacja

Aby uruchomic aplikację, należy wykonać następujące kroki:

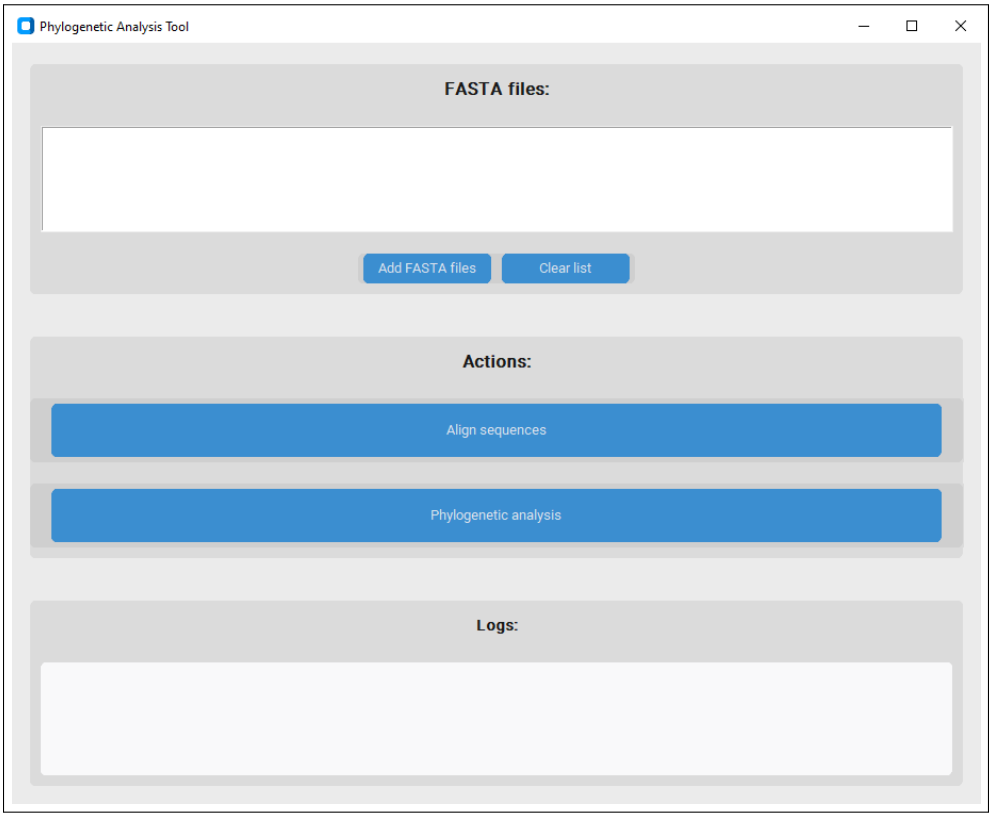
1. Zainstalować Windows Subsystem for Linux (WSL) oraz dystrybucję Ubuntu 20.04 LTS lub nowszą.
2. W środowisku WSL zainstalować narzędzia MAFFT oraz IQ-TREE 2.
3. Upewnić się, że narzędzia są dostępne w ścieżce systemowej (PATH) w WSL.
4. Pobrać kod źródłowy aplikacji z repozytorium GitHub github.com/FilipWspanialy/PhylogeneticAnalyses i uruchomić plik `main.exe` w systemie Windows.

4.3 Instrukcja obsługi

4.3.1 Interfejs użytkownika

Graficzny interfejs użytkownika umożliwia użytkownikowi:

- wczytanie sekwencji w formacie FASTA,
- Wyrównanie sekwencji przy użyciu MAFFT,
- Przeprowadzenie analizy filogenetycznej z IQ-TREE 2,
- Wizualizację uzyskanego drzewa filogenetycznego.
- Podgląd logów z poszczególnych etapów analizy.

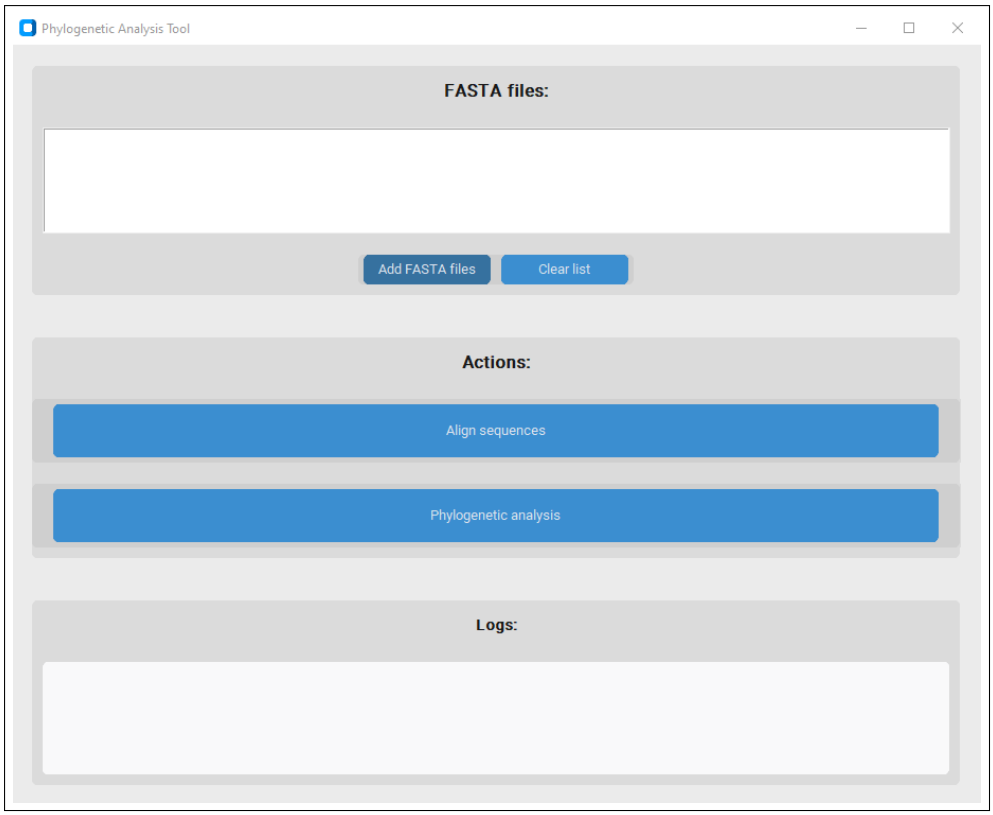


Rysunek 4.1: Interfejs użytkownika aplikacji

4.3.2 Przykład działania

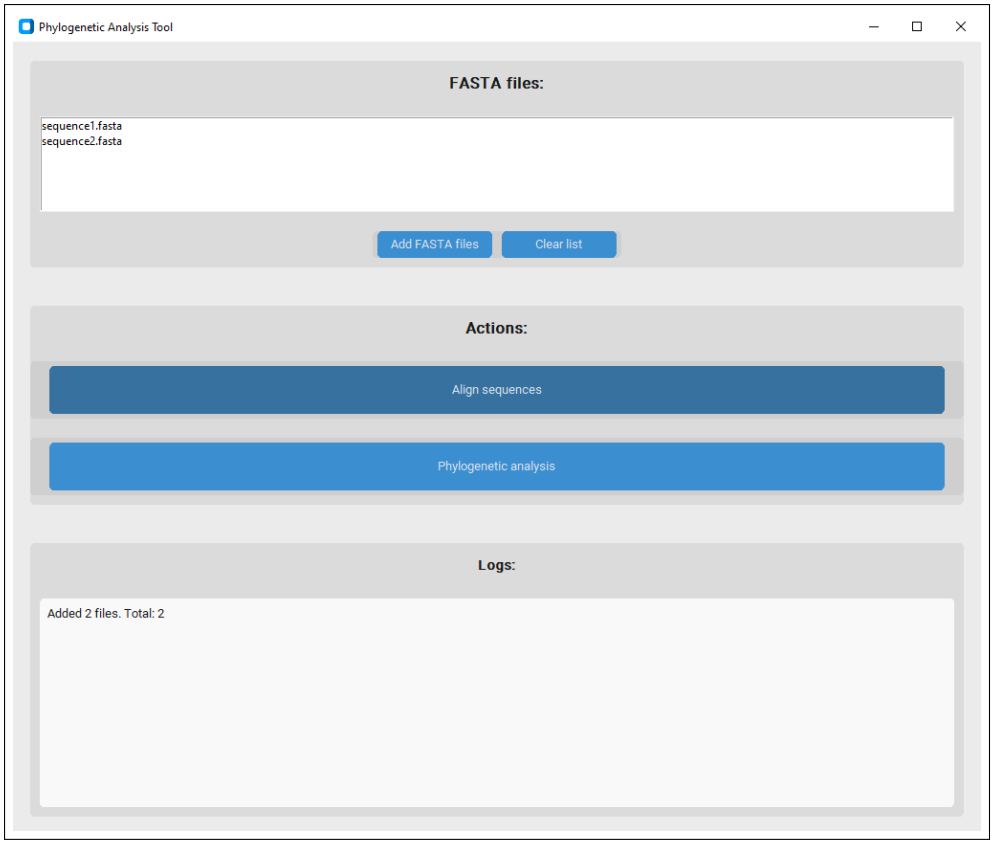
Poniżej przedstawiono przykładowy przebieg analizy filogenetycznej z wykorzystaniem aplikacji:

1. Użytkownik uruchamia aplikację
2. Wczytuje plik FASTA z sekwencjami białkowymi.



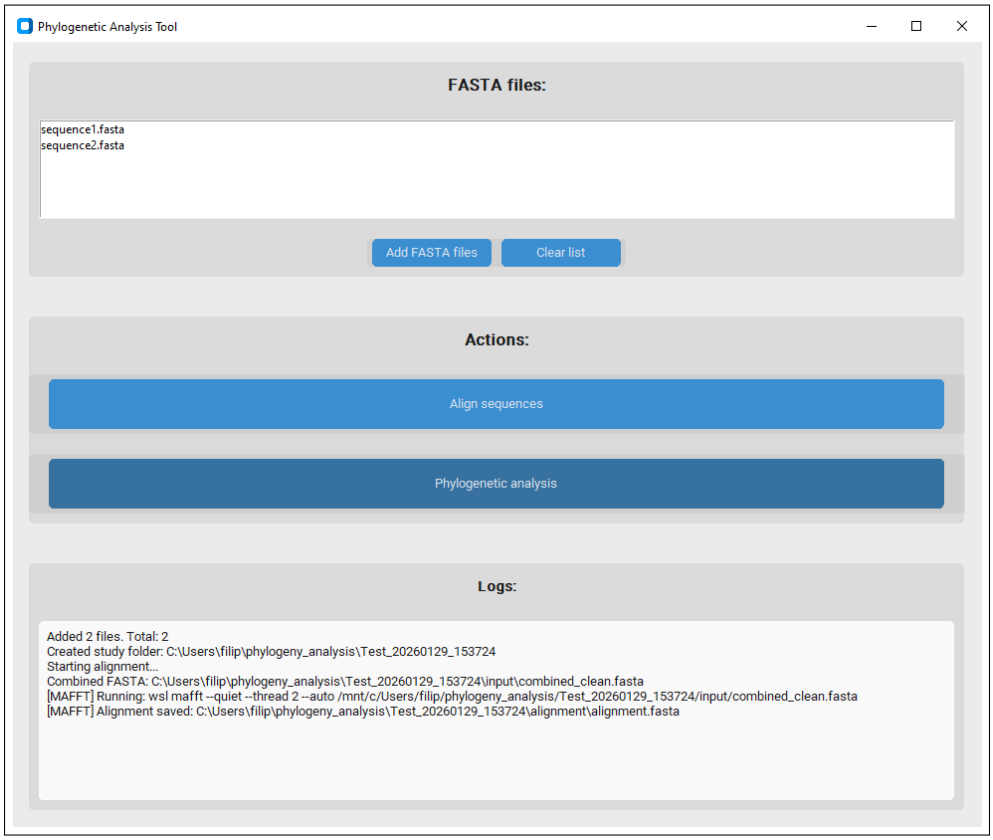
Rysunek 4.2: Wybór plików wejściowych

3. Inicjuje wyrównanie sekwencji za pomocą MAFFT.



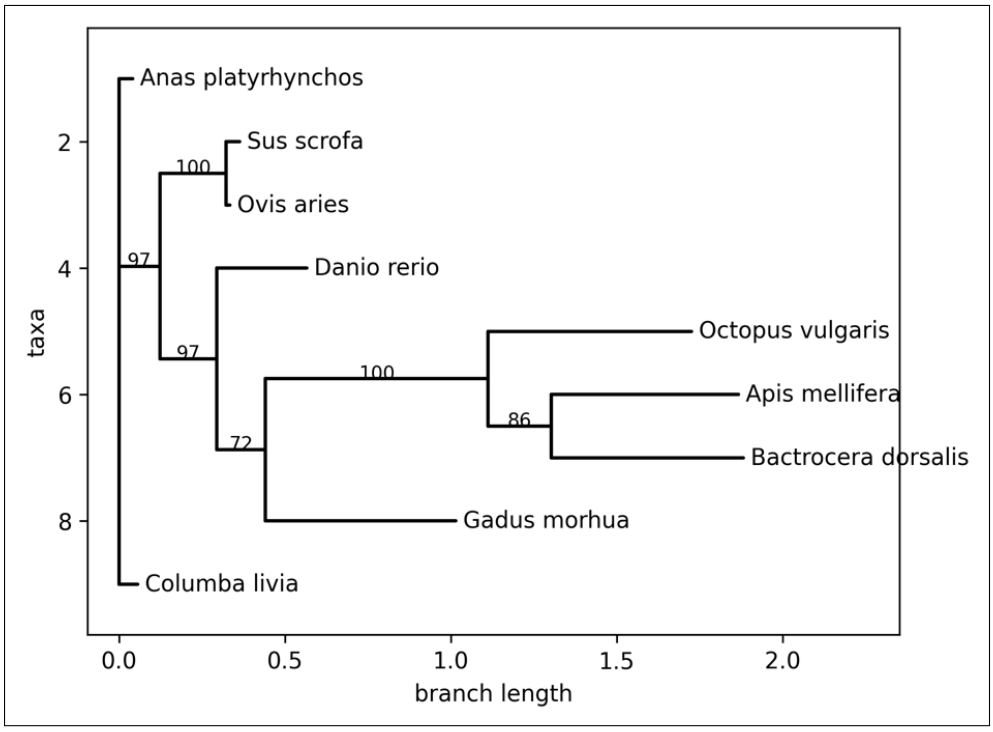
Rysunek 4.3: Wyrównanie sekwencji

4. Po zakończeniu wyrównania, uruchamia analizę filogenetyczną z IQ-TREE 2.



Rysunek 4.4: Analiza filogenetyczna

5. Po zakończeniu analizy następuje wizualizacja uzyskanego drzewa filogenetycznego.



Rysunek 4.5: Wizualizacja drzewa filogenetycznego

Wszystkie czynności są rejestrowane w logach dostępnych w interfejsie użytkownika, a pliki wynikowe z każdego etapu zapisywane są w katalogu, którego ścieżka zostaje podana w trakcie działania aplikacji.

4.3.3 Bezpieczeństwo

Aplikacja nie przechowuje żadnych danych osobowych użytkownika.

Rozdział 5

Specyfikacja wewnętrzna systemu

5.1 Idea systemu

Celem aplikacji jest integracja istniejących, sprawdzonych narzędzi do analiz filogenetycznych w jeden spójny workflow umożliwiający przeprowadzenie analizy bez konieczności posiadania wiedzy eksperckiej z zakresu filogenetyki. Użytkownik ma możliwość intuicyjnej pracy z plikami wejściowymi, natomiast dobór parametrów analizy oraz modeli ewolucyjnych odbywa się w sposób automatyczny.

System realizuje kompletny proces analizy filogenetycznej, obejmujący przygotowanie danych wejściowych, wyrównywanie sekwencji, budowę drzewa filogenetycznego oraz wizualizację wyników. Całość została zaprojektowana z myślą o środowisku Windows, przy jednoczesnym wykorzystaniu narzędzi Linuksowych dostępnych poprzez Windows Subsystem for Linux.

5.2 Wymagania funkcjonalne

System spełnia następujące wymagania funkcjonalne:

- import plików sekwencji w formacie FASTA,
- łączenie oraz wstępne czyszczenie danych wejściowych,
- uruchamianie narzędzi do wyrównywania sekwencji,
- budowę drzewa filogenetycznego z automatycznym doborem modelu,
- automatyczne tworzenie struktury katalogów wynikowych,
- wizualizację drzewa filogenetycznego,
- logowanie przebiegu analizy w interfejsie użytkownika.

5.3 Architektura

Architektura systemu została zaprojektowana w sposób modułowy, co zapewnia czytelny podział odpowiedzialności pomiędzy poszczególne komponenty aplikacji oraz umożliwia jej dalszy rozwój. System składa się z logicznych warstw odpowiadających za interakcję z użytkownikiem, realizację analizy, zarządzanie danymi oraz prezentację wyników.

5.3.1 Warstwy systemu

1. **Warstwa interfejsu użytkownika** – odpowiada za interakcję z użytkownikiem, obsługę plików wejściowych oraz prezentację wyników i logów.
2. **Warstwa logiki aplikacji** – realizuje proces analizy filogenetycznej oraz steruje uruchamianiem narzędzi analitycznych.
3. **Warstwa zarządzania danymi** – odpowiada za przygotowanie, organizację i spójność danych wykorzystywanych w analizie.
4. **Warstwa wizualizacji** – generuje oraz udostępnia użytkownikowi wizualną reprezentację drzewa filogenetycznego.

5.3.2 Opis modułów systemu

Wszystkie funkcjonalności systemu zostały zaimplementowane w klasie `PhylogenyApp`. Ze względu na skalę projektu architektura systemu została zrealizowana w postaci jednej klasy, w obrębie której wydzielono logiczne moduły funkcjonalne grupujące powiązane metody. Takie podejście zapewnia czytelny podział odpowiedzialności przy jednoczesnym zachowaniu prostoty implementacyjnej.

Moduł interfejsu użytkownika

Moduł interfejsu użytkownika został zaimplementowany z wykorzystaniem biblioteki `tkinter`. Struktura GUI tworzona jest w metodzie `setup_ui` i obejmuje elementy umożliwiające wybór plików FASTA, uruchamianie poszczególnych etapów analizy oraz podgląd logów i wyników. Moduł obsługuje zdarzenia generowane przez użytkownika i pełni rolę warstwy pośredniczącej pomiędzy użytkownikiem a logiką aplikacji.

Moduł logiki aplikacji

Moduł logiki aplikacji realizuje właściwy proces analizy filogenetycznej, obejmujący wyrównywanie sekwencji oraz budowę drzewa filogenetycznego. Proces wyrównywania sekwencji realizowany jest w metodzie `align_sequence`, natomiast analiza filogenetyczna i konstrukcja drzewa wykonywane są w metodzie `phylogenetic_analysis`. W ramach tych

metod wykorzystywane są funkcje `run_mafft` oraz `run_iqtree`, które za pomocą biblioteki `subprocess` uruchamiają zewnętrzne narzędzia analityczne w środowisku Windows Subsystem for Linux.

Parametry wykorzystywane w MAFFT[9]

```
cmd = [
    "wsl", "mafft",
    "--quiet", "--thread",
    "2", "--auto",
    wsl_input
]
```

- `--quiet` – wyciszenie komunikatów informacyjnych,
- `--thread 2` – wykorzystanie dwóch wątków procesora,
- `--auto` – automatyczny dobór strategii wyrównania,
- `wsl_input` – ścieżka do pliku FASTA w formacie zgodnym z WSL.

Parametry wykorzystywane w IQ-TREE 2 [10]

```
cmd = [
    "wsl", "iqtree2",
    "-s", wsl_alignment,
    "-m", "MFP",
    "-bb", "1000",
    "-alrt", "1000",
    "-nt", "AUTO",
    "-pre", wsl_prefix,
    "-redo"
]
```

- `-s wsl_alignment` – plik wyrównanych sekwencji,
- `-m MFP` – automatyczny wybór modelu ewolucyjnego,
- `-bb 1000` – ultrafast bootstrap,
- `-alrt 1000` – test ALRT,
- `-nt AUTO` – automatyczne wykorzystanie wątków CPU,
- `-pre wsl_prefix` – prefiks plików wynikowych,
- `-redo` – nadpisanie istniejących wyników.

Moduł zarządzania danymi

Moduł zarządzania danymi odpowiada za przygotowanie i organizację danych wejściowych oraz wynikowych. W jego ramach wykorzystywane są metody `combine_fasta`, `sanitize_fasta_headers` oraz `build_label_map`, które umożliwiają łączenie plików FASTA, ujednolicanie nagłówków sekwencji oraz zachowanie mapowania etykiet wykorzystywanych na etapie wizualizacji drzewa. Dodatkowo metoda `create_study_dirs` odpowiada za automatyczne tworzenie struktury katalogów wynikowych. Moduł ten zapewnia spójność danych pomiędzy kolejnymi etapami analizy oraz niezależność pozostałych komponentów od szczegółów operacji plikowych.

Moduł wizualizacji danych

Moduł wizualizacji danych odpowiada za graficzną prezentację wyników analizy filogenetycznej. Do wizualizacji drzewa filogenetycznego wykorzystano bibliotekę `Matplotlib`. Na podstawie pliku wynikowego w formacie Newick generowana jest statyczna reprezentacja drzewa, uwzględniająca przemapowane etykiety sekwencji zgodnie z wcześniej utworzoną mapą nagłówków. Proces ten realizowany jest w metodzie `plot_tree`, która odpowiada za odczyt struktury drzewa, jego renderowanie oraz zapis wizualizacji do pliku graficznego.

Rozdział 6

Testy i analiza działania systemu

6.1 Cel testów

Celem przeprowadzonych testów było zweryfikowanie poprawności działania systemu na różnych zestawach danych wejściowych oraz ocena jakości uzyskiwanych wyników filogenetycznych. Testy miały na celu sprawdzenie, czy aplikacja prawidłowo integruje narzędzia MAFFT oraz IQ-TREE 2, a także czy uzyskane drzewa filogenetyczne są zgodne z oczekiwaniami biologicznymi.

6.2 Testy dla sekwencji białek z bazy NCBI

6.2.1 Test 1: Cytochrome c

Charakterytyka danych:

- Białko: cytochrome c
- Liczba sekwencji: 10
- Długość sekwencji: około 100 aminokwasów
- Gatunki: ssaki
- Źródło: NCBI Protein Database[2]

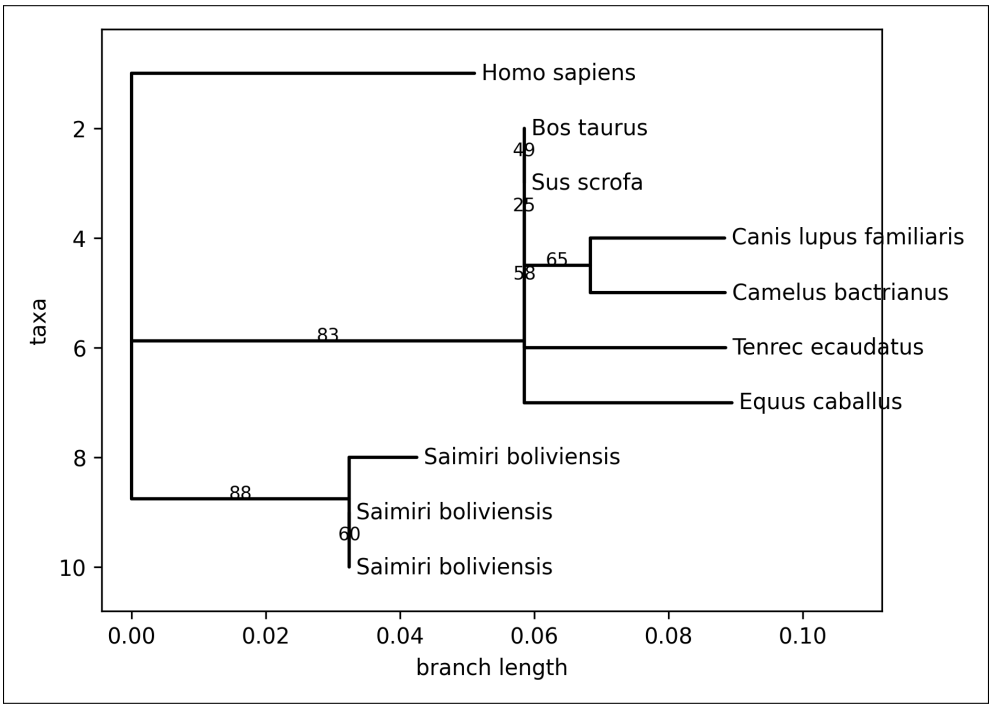
Cytochrom c jest białkiem silnie konserwatywnym, powszechnie wykorzystywanym w analizach filogenetycznych jako marker referencyjny.

```
>gi|11128019|ref|NP_061820.1| cytochrome c [Homo sapiens]
-------------------------MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLH
GLFGRKTGQAPGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERAD
LIAYLKKATNE
>gi|114051487|ref|NP_001039526.1| cytochrome c [Bos taurus]
-------------------------MGDVEKGKKIFVQKCAQCHTVEKGGKHKTGPNLH
GLFGRKTGQAPGFSYTDANKNKGITWGEETLMEYLENPKKYIPGTKMIFAGIKKKGERED
LIAYLKKATNE
```

Rysunek 6.1: Przykładowe wyrównane sekwencje białkowe - test 1; źródło: [2]

Cel testu:

- Weryfikacja poprawności działania stworzonej aplikacji pod względem integracji narzędzi MAFFT oraz IQ-TREE 2.
- Sprawdzenie działania systemu na niewielkim zbiorze sekwencji białkowych o dobrze poznanej filogenezie w celu oceny poprawności uzyskiwanych wyników.



Rysunek 6.2: Drzewo filogenetyczne dla cytochromu c - test 1

Wyniki testu:

Test przeprowadzony na krótkich sekwencjach białkowych cytochromu c zakończył się powodzeniem. Aplikacja poprawnie zrealizowała pełny pipeline analizy filogenetycznej, integrując narzędzia MAFFT oraz IQ-TREE 2 – od etapu wyrównania sekwencji po rekonstrukcję drzewa filogenetycznego.

Uzyskane drzewo filogenetyczne jest zgodne z ogólnymi oczekiwaniami biologicznymi i odzwierciedla znane relacje filogenetyczne pomiędzy analizowanymi gatunkami ssaków. Celowe uwzględnienie trzech sekwencji pochodzących od tego samego gatunku skutkowało ich bliskim sąsiedztwem na drzewie wynikowym, co potwierdza poprawność działania algorytmów wyrównywania oraz rekonstrukcji drzewa.

Wysokie wartości wsparcia bootstrap (>70) obserwowane są dla węzłów odpowiadających głównym podziałom taksonomicznym, natomiast dla rozgałęzień pomiędzy blisko spokrewnionymi gatunkami wartości bootstrap są niższe (25–65). Wynika to z wysokiej konserwatywności białka cytochromu c, które dostarcza ograniczonej ilości sygnału filogenetycznego.

Test potwierdza poprawność działania zaprojektowanego systemu na prostych, kontrolowanych danych wejściowych oraz poprawną integrację wykorzystanych narzędzi bioinformatycznych.

6.2.2 Test 2: Cytochrome c oxidase subunit 4

Charakterytyka danych:

- Białko: cytochrome c oxidase subunit 4 (COX4)
- Liczba sekwencji: 9
- Długość sekwencji: około 170 aminokwasów
- Gatunki: głównie ssaki i ryby, ale także owady i mięczaki
- Źródło: NCBI Protein Database[2]

COX4 jest integralną podjednostką enzymu cytochrome c oxidase (kompleks IV), uczestniczącego w końcowym etapie łańcucha oddechowego. W przeciwieństwie do cytochromu c, sekwencja COX4 jest mniej konserwatywna ewolucyjnie i wykazuje większą zmienność pomiędzy odlegle spokrewnionymi gatunkami.

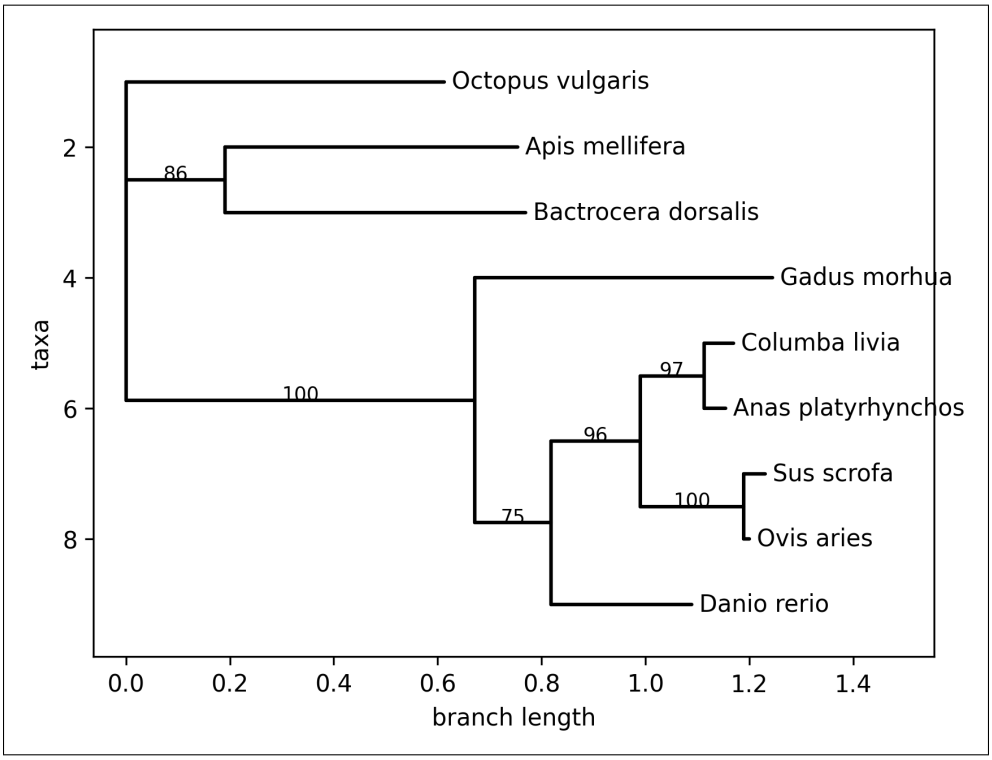
```
>CAI9715882.1 cytochrome c oxidase subunit 4 isoform 1, mitochondrial [Octopus vulgaris]
MRSCQCVNNEPTACQSAVLGQEARDKIHPRIGNREIVGFGINGSASYFDIPEIPFPAIRFKESTGENLAL
LDKQAADWKTLTLQEKKDLYRISFCQTYAEMHAPTGDWKRIISFFLMSCSLTGWIIIWMKLYVYPPVPHS
LTQEWQDAMLEKMIAQRVNPIEGVSSQYDYENERWKN

>XP_006572317.1 cytochrome c oxidase subunit 4 isoform 1, mitochondrial [Apis mellifera]
MANKLLLSYLRQNTSMCVRGLSAMPEFPNKIGNRDVVGNGWNGEEAYLDRSDFPLPAIRFKANTPDIMAL
REKEKGDWKKLSIEEKKILYRASFRQTFSEFLAPTGEWRGHIGIALIGVSFSLLTYILLKIYAFPPLPES
FNEENRLAQLERMKLLQVNPIAGISSKN
```

Rysunek 6.3: Przykładowe sekwencje białkowe - test 2; źródło: [2]

Cel testu:

- Weryfikacja działania aplikacji na białkach integralnych błony mitochondrialnej o większej zmienności sekwencji niż cytochrom c.
- Sprawdzenie jakości wyrównania i stabilności drzewa filogenetycznego dla białek o różnych długościach i izoformach.



Rysunek 6.4: Drzewo filogenetyczne dla COX4 - test 2

Wyniki testu:

Test przeprowadzono na sekwencjach białka COX4 pochodzących z różnych grup taksonomicznych, w tym ssaków, ryb, owadów i mięczaków. Aplikacja poprawnie zintegrowała narzędzia MAFFT oraz IQ-TREE 2, generując wyrównanie sekwencji oraz drzewo filogenetyczne.

Uzyskane drzewo wykazuje logiczną strukturę filogenetyczną. Octopus vulgaris został wyraźnie odseparowany jako najbardziej odległy takson w zbiorze. Apis mellifera i Bactrocera dorsalis tworzą klad owadów z umiarkowanym wsparciem bootstrap (86). Gatunki ryb i ptaków grupują się w stabilne klady, np. Columba livia i Anas platyrhynchos (bootstrap 97) oraz Sus scrofa i Ovis aries (bootstrap 100).

Wysokie wartości bootstrap dla głównych kladów (75–100) wskazują na stabilność rekonstrukcji, natomiast niższe wartości dla owadów odzwierciedlają większą zmienność sekwencji w tej grupie. Test potwierdza poprawne działanie pipeline'u na białkach mitochondrialnych o zróżnicowanej długości i zmienności sekwencji.

6.3 Testy dla sekwencji nukleotydowych z bazy NCBI

6.3.1 Test 3: COX2–COX1

Charakterytyka danych:

- Sekwencje: COX2–COX1 intergenic spacer oraz częściowa sekwencja genu COX1, mitochondrialna
- Liczba sekwencji: 10
- Długość sekwencji: około 700 nukleotydów
- Gatunki: koralowce głębinowe
- Źródło: NCBI Protein Database[2]

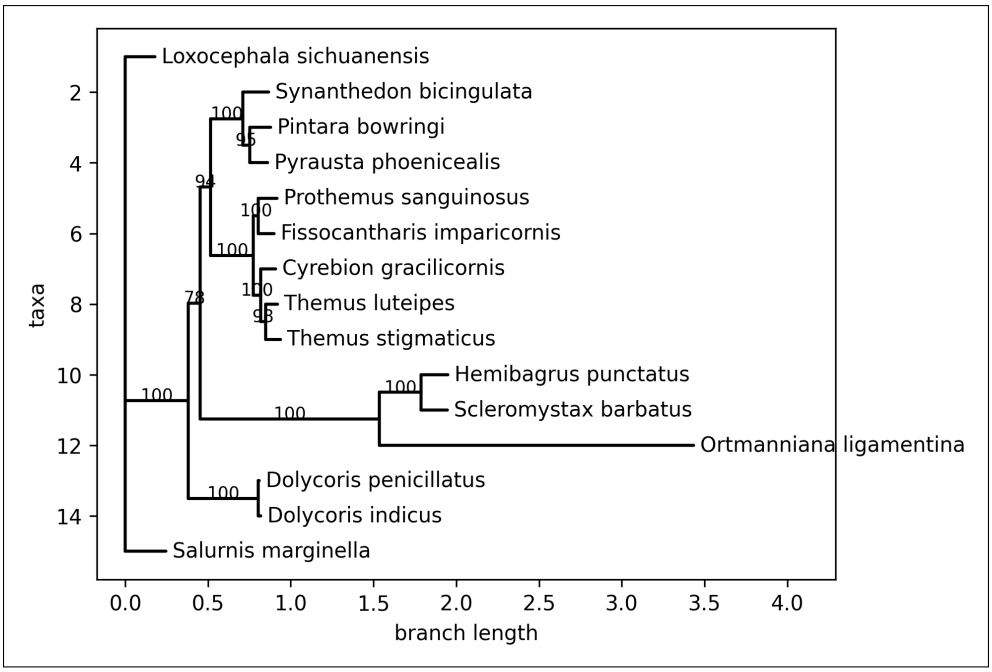
Sekwencje nukleotydowe wykorzystano do rekonstrukcji filogenezy na poziomie gatunków i bliskich rodzin koralowców. Ze względu na większą zmienność w porównaniu z białkami mitochondrialnymi, sekwencje te dostarczają bardziej szczegółowego sygnału filogenetycznego między taksonami blisko spokrewnionymi

```
>NC_088742.1 Loxocephala sichuanensis mitochondrion, complete genome
AGTAAGATGCCTGATTAAAGGATTATTTTGATATAATAAATAATGTAAAATTTACTCTTACTATTATAAA
TCTTGGGTTTAAACCAAGCCTAAAGAAATCAAAATTCTTAATGCATCATACATTTATTTATAAAAAGATA
AGCTAATTTAAGCTTATGGGTTCATACCCCATCCATGAATAAAATTCTCTTTTTAATTAAAATTAATTCA
ACTAAAACAATATTTTTAATAATTTTATTTTTATCAACAATTATAGTTTTATCATCAAATAATATTTTAA
TATCATGAATAGCAATAGAAATTAATTTATTTATATTTATACCAATAATAACAAAAAAACAAAAAATAAA
AGATCAGCCAATAAAATATTTTATTATTCAAAGATTATCCTCATCAACAATATTAATATCAATTTTAATA
AATTTAAATTCTGAGATCCCCCTCAGAATAAGAATTTTATTAATAACAAGAATATTAATAAAAATAGGTA
TAATCCCATTCCATATATGACTTCCAACTATTATAAATAAAATATCATGAAATAACTGTTTTATTTTATC
AACATGACAAAAAATTGCACCTATTAATATTTTATCACAATTAATTGAATTAAAAATATTAATTTTACCT
TTATGTTTATCTTTACTAGTAGCACCAGTTACAGCAATTAAACAACTTTCAAGAAAAAAAATTATAGCTT
ATTCATCAATTTCAAATTCACCTTGAATATTAATTTCAATAATAATATCAAAATTTTTTTTTTATATATT
TATAATAATTTATTCAACACTAACATTTATAATAATAAAAACAATAAAAAATATAAATTTAAACTTTATT
AACCAAATTTTAACTAGAACAAAAATACAAAAATTAAATTTAATTATTTTAACTTTATCAATAAGAGGTT
TACCTCCTCTAACAGGATTTTTACCAAAATGAATAATTTTACAACAAATAATTAATTTTTCTGAAATAGT
```

Rysunek 6.5: Przykładowa sekwencja nukleotydowa - test 3; źródło: [2]

Cel testu:

- Weryfikacja działania pipeline'u analizy filogenetycznej na sekwencjach nukleotydowych.
- Sprawdzenie, czy narzędzia MAFFT i IQ-TREE 2 radzą sobie z fragmentami genów mitochondrialnych o różnej długości i zmienności.



Rysunek 6.6: Drzewo filogenetyczne dla COX2–COX1 - test 3

Wyniki testu:

Test przeprowadzono na sekwencjach nukleotydowych COX2–COX1 pochodzących od gatunków głębinowych koralowców. Aplikacja poprawnie zrealizowała pełen pipeline analizy filogenetycznej, obejmujący wyrównanie sekwencji oraz rekonstrukcję drzewa filogenetycznego.

Uzyskane drzewo wykazuje logiczną organizację taksonomiczną. Calyptrophora clinata została odseparowana jako samodzielny takson, natomiast blisko spokrewnione gatunki, takie jak Thouarella grasshoffi, Narella versluysi, Candidella imbricata oraz Narella bellissima, grupują się w zagnieżdżone klady. Wartości bootstrap dla tych rozgałęzień są zróżnicowane (29–99), co odzwierciedla zmienność sekwencji nukleotydowych.

Najwyższe wsparcie bootstrap obserwuje się dla kladów dobrze określonych genetycznie, natomiast niskie wartości dla części rozgałęzień wskazują na ograniczoną pewność relacji w obszarach o niewielkiej różnorodności lub fragmentarycznych danych. Test potwierdza poprawność działania pipeline'u na sekwencjach nukleotydowych oraz zdolność systemu do rozróżniania taksonów o różnym stopniu pokrewieństwa.

6.3.2 Test 4: Mitochondrialne sekwencje nukleotydowe

Charakterytyka danych:

- Sekwencje: pełne genomu mitochondrialnego
- Liczba sekwencji: 15

- Długość sekwencji: około 17000 nukleotydów
- Gatunki: różne, głównie owady i ryby
- Źródło: NCBI Protein Database[2]

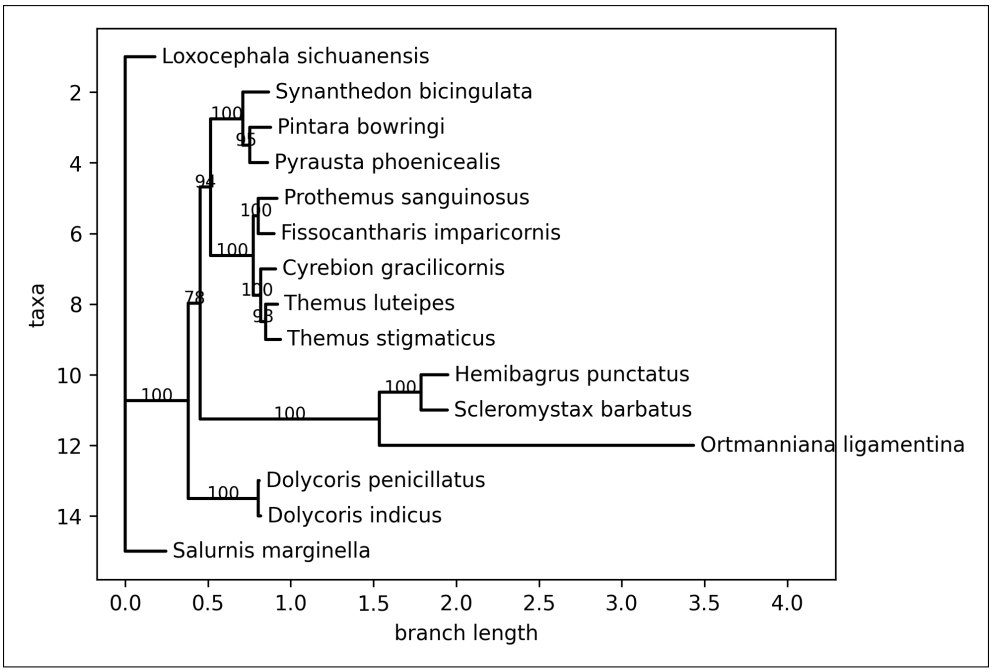
Sekwencje obejmują kompletne genomu mitochondrialne poszczególnych gatunków, co zapewnia bogaty sygnał filogenetyczny i umożliwia analizę relacji ewolucyjnych na poziomie całych genomów.

```
>NC_088742.1 Loxocephala sichuanensis mitochondrion, complete genome
AGTAAGATGCCTGATTAAAGGATTATTTTGATATAATAAATAATGTAAAATTTACTCTTACTATTATAAA
TCTTGGGTTTAAACCAAGCCTAAAGAAATCAAAATTCTTAATGCATCATACATTTATTTATAAAAAGATA
AGCTAATTTAAGCTTATGGGTTCATACCCCATCCATGAATAAAATTCTCTTTTTAATTAAAATTAATTCA
ACTAAAACAATATTTTTAATAATTTTATTTTTATCAACAATTATAGTTTTATCATCAAATAATATTTTAA
TATCATGAATAGCAATAGAAATTAATTTATTTATATTTATACCAATAATAACAAAAAAACAAAAAATAAA
AGATCAGCCAATAAAATATTTTATTATTCAAAGATTATCCTCATCAACAATATTAATATCAATTTTAATA
AATTTAAATTCTGAGATCCCCCTCAGAATAAGAATTTTATTAATAACAAGAATATTAATAAAAATAGGTA
TAATCCCATTCCATATATGACTTCCAACTATTATAAATAAAATATCATGAAATAACTGTTTTATTTTATC
AACATGACAAAAAATTGCACCTATTAATATTTTATCACAATTAATTGAATTAAAAATATTAATTTTACCT
TTATGTTTATCTTTACTAGTAGCACCAGTTACAGCAATTAAACAACTTTCAAGAAAAAAAATTATAGCTT
ATTCATCAATTTCAAATTCACCTTGAATATTAATTTCAATAATAATATCAAAATTTTTTTTTTATATATT
TATAATAATTTATTCAACACTAACATTTATAATAATAAAAACAATAAAAAATATAAATTTAAACTTTATT
AACCAAATTTTAACTAGAACAAAAATACAAAAATTAAATTTAATTATTTTAACTTTATCAATAAGAGGTT
TACCTCCTCTAACAGGATTTTTACCAAAATGAATAATTTTACAACAAATAATTAATTTTTCTGAAATAGT
```

Rysunek 6.7: Przykładowy fragment sekwencji nukleotydowej - test 4; źródło: [2]

Cel testu:

- Weryfikacja poprawności działania aplikacji na dużych sekwencjach nukleotydowych.
- Ocena wydajności i stabilności pipeline'u przy pracy z pełnymi genomami mitochondrialnymi.
- Sprawdzenie jakości wyrównania oraz stabilności rekonstrukcji drzewa filogenetycznego przy sekwencjach o dużej długości.



Rysunek 6.8: Drzewo filogenetyczne dla pełnych genomów mitochondrialnych - test 4

Wyniki testu:

Aplikacja poprawnie wykonała pełen pipeline analizy filogenetycznej dla pełnych genomów mitochondrialnych, obejmujący wyrównanie sekwencji oraz rekonstrukcję drzewa filogenetycznego.

W porównaniu z testami na krótszych fragmentach genów, czas obliczeń był znacząco dłuższy, co bezpośrednio wynika z dużej długości analizowanych sekwencji. Jednocześnie pipeline zachował stabilność działania, potwierdzając możliwość analizy dużych zestawów danych.

Uzyskane drzewo filogenetyczne wykazuje wysoką spójność z oczekiwanymi relacjami taksonomicznymi. Węzły odpowiadające głównym grupom charakteryzują się wysokimi wartościami bootstrap (>95), natomiast niższe wsparcie (70–85) obserwowane dla części rozgałęzień pomiędzy blisko spokrewnionymi gatunkami stanowi naturalny efekt ograniczonej różnorodności mitochondrialnej.

Test potwierdza, że system radzi sobie poprawnie z dużymi sekwencjami nukleotydowymi, zapewniając stabilne i wiarygodne wyniki analizy filogenetycznej.

6.4 Podsumowanie wyników

Wyniki analiz filogenetycznych uzyskane w trakcie testów wskazują, że aplikacja poprawnie integruje narzędzia MAFFT oraz IQ-TREE 2, umożliwiając pełen proces analizy filogenetycznej zarówno dla danych białkowych, jak i nukleotydowych.

Celem testów była weryfikacja poprawności działania systemu, a nie uzyskanie wyników naukowych. Uzyskane drzewa filogenetyczne wykazały spójność z oczekiwanymi relacjami taksonomicznymi, co potwierdza skuteczność zastosowanych algorytmów wyrównywania sekwencji i rekonstrukcji drzewa.

Zaobserwowano również przewidywaną zależność pomiędzy długością sekwencji a czasem analizy – dłuższe sekwencje wymagają więcej czasu obliczeniowego, a wraz ze wzrostem liczby danych konieczne jest zwiększenie mocy obliczeniowej.

6.5 Analiza działania systemu

W trakcie testów aplikacja wykazała stabilność działania na różnych zestawach danych wejściowych, pod warunkiem, że kod był odpowiednio dostosowany do specyfiki sekwencji. Konieczne były drobne modyfikacje, ponieważ różnorodność sposobów zapisu danych wejściowych czasami utrudniała automatyczne wyodrębnienie nazw sekwencji, co ograniczało uniwersalność systemu.

Przy próbach przetworzenia bardzo dużych zestawów pełnych genomów mitochondrialnych system napotkał ograniczenia sprzętowe związane z niewystarczającą ilością pamięci RAM. Pokazuje to, że pipeline jest stabilny i funkcjonalny dla większości typowych danych, jednak jego wydajność i skalowalność zależą od dostępnych zasobów sprzętowych oraz przygotowania danych wejściowych.

6.6 Potencjalny rozwój systemu

1. Rozszerzenie obsługi różnych formatów plików wejściowych (np. Clustal, Phylip), aby zwiększyć uniwersalność aplikacji.

2. Implementacja dodatkowych narzędzi do analizy filogenetycznej, takich jak RAxML czy MrBayes, umożliwiająca użytkownikowi wybór preferowanej metody rekonstrukcji drzewa.

3. Optymalizacja wydajności poprzez równoległe przetwarzanie sekwencji lub wykorzystanie chmury obliczeniowej do obsługi większych zestawów danych.

4. Dodanie funkcji automatycznego generowania raportów analizy w formacie PDF lub HTML, zawierających podsumowanie statystyk oraz wizualizacje.

5. Udoskonalenie interfejsu użytkownika poprzez możliwość konfiguracji parametrów analizy i lepszą wizualizację postępu pracy.

Rozdział 7

Podsumowanie i wnioski

Analizy filogenetyczne opierają się na modelach ewolucyjnych oraz założeniach statystycznych, które mają bezpośredni wpływ na uzyskane wyniki. Metody te bazują na prawdopodobieństwie, które ocenia najbardziej prawdopodobny scenariusz ewolucyjny w oparciu o dostępne dane i matematyczne modele zmian sekwencji.

W pracy przeanalizowano relatywnie niewielkie zbiory danych, jednak w przypadku większych zestawów sekwencji zapotrzebowanie na moc obliczeniową rośnie znacząco. Wyniki mogą się różnić w zależności od zastosowanej metody i parametrów analizy, co wymaga ostrożnej interpretacji.

Pomimo wiedzy na temat podobieństw międzygatunkowych i możliwości porównywania ich z cechami morfologicznymi, należy pamiętać, że proces ewolucji zachodzi przez bardzo długi czas i może obejmować zmiany trudne do uchwycenia przy użyciu dostępnych modeli. W związku z tym, nawet jeśli dwa organizmy wydają się podobne, analiza filogenetyczna opiera się na przyjętym wzorcu zmian sekwencji, który najlepiej odzwierciedla ewolucyjny sygnał, ale nie zawsze oddaje pełną złożoność historii ewolucyjnej.

Zaleca się, aby wyniki analiz traktować w kontekście biologicznym, uzupełniając je wiedzą morfologiczną i ekologiczną oraz stosując różne metody rekonstrukcji drzewa filogenetycznego w celu oceny stabilności wniosków.

Bibliografia

[1] MAFFT Manual [online] https://mafft.cbrc.jp/alignment/software/manual/manual.html, data dostępu: 29.01.2026.

[2] Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, and Robert Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, in press. https://doi.org/10.1093/molbev/msaa015, data dostępu: 29.01.2026.

[3] Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., de Hoon, M.J.L. (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423.

[4] Grzegorz Góralski *Wstęp do filogenetyki molekularnej i tworzenia drzew filogenetycznych* [online] https://ggoralski.github.io/proba_mdbook/print.html#wst%C4%99p-do-filogenetyki-molekularnej-i-tworzenia-drzew-filogenetycznych, data dostępu: 09.01.2026.

[5] PWN Encyklopedia [online] https://encyklopedia.pwn.pl/haslo/filogeneza;3900970.html, data dostępu: 09.01.2026.

[6] Krzysztof Spalik, Marcin Piwczyński *Rekonstrukcja filogenezy i wnioskowanie filogenetyczne w badaniach ewolucyjnych* [artykuł] KOSMOS, Problemy Nauk Biologicznych, t. 58, nr 3–4, 2009, s. 485–498.

[7] Subha Kalyaanamoorthy, Bui Quang Minh, Thomas KF Wong, Arndt von Haeseler, Lars S. Jermiin *ModelFinder: Fast model selection for accurate phylogenetic estimates* [artykuł] Nature Methods, 14:587–589, 2017. https://doi.org/10.1038/nmeth.4285, data dostępu: 31.01.2026.

[8] Diep Thi Hoang, Olga Chernomor, Arndt von Haeseler, Bui Quang Minh, Le Sy Vinh *UFBoot2: Improving the ultrafast bootstrap approximation* [artykuł] Mol. Biol. Evol., 35:518–522, 2018. https://doi.org/10.1093/molbev/msx281, data dostępu: 31.01.2026.

[9] MAFFT Manual [online] https://mafft.cbrc.jp/alignment/software/manual/manual.html, data dostępu: 29.01.2026.

[10] IQ-TREE 2 Command Reference [online] https://iqtree.github.io/doc/Command-Reference, data dostępu: 29.01.2026.

Dodatki

Źródła

[1] Georgia Tech Biological Sciences *Phylogenetic Trees* [online] https://organismalbio.biosci.gatech.edu/biodiversity/phylogenetic-trees/, data dostępu: 09.01.2026.

[2] National Center for Biotechnology Information [online] https://www.ncbi.nlm.nih.gov/guide/data-software/, data dostępu: 09.01.2026.

Załączniki

Lista dodatkowych plików uzupełniających tekst pracy

W systemie do pracy dołączono dodatkowe pliki zawierające:

Spis rysunków