UNIWERSYTET WROCŁAWSKI

WYDZIAŁ BIOTECHNOLOGII

ZAKŁAD MIKROBIOLOGII MOLEKULARNEJ

Przygotowanie bioinformatycznej bazy danych NGS dotyczących bakterii z rodzaju *Streptomyces*

Marek Skrzyński

Praca magisterska

Opiekun pracy : dr Agnieszka Strzałka

Wrocław, 2025 r.

STRONA NA PODZIĘKOWANIA

# Abstract/streszczenie

Wydaje mi się, że tak jak w przypadku pracy licencjackiej jak i sprawozdań itp. to napiszę na końcu.

# Skróty i symbole

Nawet nie wiem czy jakieś konkretne będą, ale raczej okaże się, że tak. Chociażby:  
R – język programowania.

Spis treści

[Abstract/streszczenie 3](#_Toc189146398)

[Skróty i symbole 4](#_Toc189146399)

[Wstęp/Rozwinięcie 6](#_Toc189146400)

[Materiały i metody 7](#_Toc189146401)

[Wyniki 8](#_Toc189146402)

[Dyskusja 9](#_Toc189146403)

[Podsumowanie 10](#_Toc189146404)

[Bibliografia 11](#_Toc189146405)

# Wstęp/Rozwinięcie

* Streptomyces coelicolor i venezuelae:
  + co to za bakterie
  + konkretniej co to za gatunki
  + ich zastosowanie w biotechnologii , z przykładami
  + ich znaczenie jako organizm modelowy, z przykładami
* NGS
  + Czym jest, do czego się wykorzystuje
  + Jakie znaczenie ma wykorzystywanie technik NGS w badaniu bakterii z gatunku streptomyces, z przykładami
  + Z jakich rodzajów NGS pochodzą dane
  + jakie bazy danych już są – Strepdb, ArrayExpress, https://imodulondb.org/index.html
* Dane do aplikacji
  + Kontynuacja z jakich rodzajów badań pochodzą
  + Skąd zaciągnięte są dane/skąd je wzięliśmy
  + Problem z różnorodnością danych
    - Na czym polega
    - Że dane w suplementach
    - Różne rodzaje przeróbki od „raw” do publikowanych danych
  + Co tylko nam przyjdzie na myśl w kwestii „ktoś się może o to przyczepić”
  + Obróbka danych do programu, w kwestii, czemu były obrabiane
  + że są programy do analizy rna-seqów np. iDEP 2.01, https://github.com/federicomarini/GeneTonic
  + spróbuj wziąć surowe, wrzucić na idep a potem do swojego
* Aplikacja
  + Po co
    - Jaki jest jej cel
    - Czemu jest na nią zapotrzebowanie
  + Na co ma pozwalać użytkownikowi
  + Opis jak powstała, raczej po krótce
  + Nie wiem, tu coś powinno być chyba jeszcze ale nie wiem co

# Materiały i metody

* Wszystko tu będzie obrzucone niezliczoną ilością screenów
* Dane
  + Format danych
    - Ujednolicony format, dlaczego i po co
    - Jaki? Wybór nowego „zaprojektowanego przez nas”
    - Które wartości w danych są ważne i dlaczego, stąd wychodzi jaki format stworzyć
* Wczytywanie danych do programu
  + W jakiej postaci, wektor, tabela, lista, czemu, który
* Wyświetlanie danych
  + Wykresy
  + W każdym opisane co to za wykres, screen jak wygląda, czemu ma służyć, jakimi wartościami z danych się posługuje
    - Strona wyświetlania danych
    - Strzałkowy genom
    - Strzałkowy RNA-seq/microarray
    - Chip-seq
    - Tabele z danymi pod wykresami
    - Opisanie wszystkich przycisków, wyborów, zmiennych które można zastosować do modyfikacji wykresu
      * Zaprezentowanie screenami jak to działa, jakie są i jakie modyfikacje można wprowadzać na wykresie
        + Zmiana genomu
        + Wybór wyświetlanych wykresów
        + Wybór danych do wykresów
        + Wybór genu
        + Oddzielny wybór danych do chip-seq
        + Przyciski „przesuwania się” po wykresie
        + Granice wykresu
        + Granice logFC z tym jak się zmieniają
        + Przycisk filtra FDR
      * Dodanie, że jest opcja pobrania wykresu
    - Dodawanie własnych danych przez użytkownika
      * Opis jak dane należy dodawać, z dodaniem, że trzeba je sobie nazwać
      * Jakie manipulacje są wprowadzane w danych (przyrównanie do genomu z dodaniem start, end itp.) (trzeba gdzieś też będzie zamieścić wszystkie genomy których używaliśmy)
    - Strona porównania danych
    - Diagram venna, co pokazuje, jakie modyfikacje wprowadzane są do danych
    - Heat mapa
    - Wykres intime, jakie modyfikacje wprowadzane są do danych
    - Wszelkie przyciski modyfikacji wprowadzaych na wykresach
    - Opcja pobrania wykresów w formie obrazu
    - HELP$Info
      * Co jest w nim zawarte
        + Pobranie szablonu formatu danych
        + Może dodać do aplikacji żeby umiała pivot zrobić z danymi, to by było spoko, że jak ktoś wrzuci dane, to jeśli ktoś dobrze ponazywa kolumny, to żeby mogło wybrać kolumny, zrobić pivot, bo przestawianie danych sobie samemu w excelu jest problematyczne. Wiadomo że ultra uniwersalnie się tego nie zrobi ale może by działało
        + Trzeba tu też wstawić jakie dane są w aplikacji, taką bardziej rozbudowaną bibliografię
        + Podkreślić że jest tu opis tego co się dzieje w aplikacji, myślę, że zamiast opisywać wszystko w tym małym help&info, można będzie po prostu skopiować dużą część z sekcji materiały i metody która powstanie tutaj i wstawić jako plik do pobrania, taki pomysł mi do głowy przyszedł

# Wyniki

* Prezentacja kilku przykładów jak można zastosować aplikację pewnie też w wersji bardzo obrazkowej
  + Co mi pierwsze do głowy przychodzi, to:
    - Opcja, ktoś bazuje tylko na tym co jest w aplikacji,
      * Ogląda któreś zbiory które go zainteresowały na wykresach strzałkowych, przechodzi do zakładki comparison i tam ogląda dokładniej poszczególne geny/zbiory genów na wykresie i wyciąga jakieś wnioski, jakie? To się okaże, trzeba spróbować samemu coś tak pomyśleć, to znaczy ja

możliwe przykłady zastosowania:

analiza jednego białka np. chcemy się dowiedzieć w czy i w jakich warunkach zachodzi ekspresja genu, czy należy do konkretnego regulonu, operonu

analiza całego klastra genów np. chcemy sprawdzić w jakich warunkach aktywny jest klaster genów odpowiedzialnych za produkcję metabolitów wtórnych np. aktynorodyna lub red u S. coelicolor lub chloramfenikol u S. venezuelae

* + - Opcja, ktoś dodaje własne,
      * Podobnie jak z przykładem poniżej, tyle że wychodzi ktoś od swoich danych, znajduje, którąś publikacje, która ma podobne założenia co jego i ogląda co się dzieje w jego w porównaniu z tą drugą, może też sobie pobrać wykresy do zamieszczenia w publikacji

# Dyskusja

* Problematyka ujednolicania danych i jak to jest że porównujemy dane które zostały bazowo inaczej przygotowane
* Szczerze to nie wiem co tu jeszcze może być, pewnie więcej pomysłów przyjdzie do głowy na koniec

# Podsumowanie

* Jak sobie poradziliśmy z problemami dotyczącymi danych
* Jakie korzyści może przynieść używanie bazy
  + Jakie wyniki i wnioski można wyciągać
  + Jak to pomoże ludziom
* Podkreślenie, że aplikacja jest w pełni uaktualnialna, to znaczy, będzie można dodawać nowe zbiory danych
* I pewnie też coś jeszcze co przyjdzie do głowy po napisaniu reszty

# Bibliografia