

Bioinformatyka — laboratorium 9 i 10

Nazwa pliku: imie_nazwisko_9_10_bio.pdf

Typ ćwiczenia: dwutygodniowe

Cel: Celem ćwiczenia jest nauka programowania w języku wysokopoziomowym Python z wykorzystaniem danych biologicznych.

UWAGA: Tym razem nie oddajesz raportu z wykonania tego ćwiczenia, należy przygotować pliki *.py o nazwach: 1_2.py, 3_5.py, 6_8.py, 9_10.py, 11_12.py, 13_15.py, 16_18.py, 19_21.py, 22_24.py, i zadania dla chętnych: 25_29.py. Umieść je w folderze z Twoim numerem s11111 i spakuj do pliku zip.

Podstawy programowania w Pythonie — zadania

Prowadzący nie narzucają jakiego IDE, IDLE lub innego narzędzia ma używać student celem przygotowania i testowania swojego kodu.

1. **[Task 1-2]** Napisz program w języku Python, który policzy ilość każdego z nukleotydów (adenina, cytozyna, guanina i tymina) w danej sekwencji DNA.

Początek rozwiązania:

```
sekwencja_dna = "ATGCTACGATCGTAGCTAGCTAGCTAGCTAGC"
```

```
# Zainicjowanie liczników nukleotydów
```

```
licznik_adeniny = 0
```

```
licznik_cytozyny = 0
```

```
licznik_guaniny = 0
```

```
licznik_tyminy = 0
```

```
# Przejdźcie przez każdy nukleotyd w sekwencji DNA
```

```
for nukleotyd in sekwencja_dna:
```

```
    if # dalej już Twój kod
```

2. **[Task 3-5]** Napisz program w języku Python, który przekształci sekwencję DNA na RNA, a następnie obliczy długość otrzymanej sekwencji RNA. Program powinien wczytać od użytkownika sekwencję DNA jako łańcuch znaków i przekształcić ją na sekwencję RNA, zastępując wszystkie wystąpienia nukleotydu tyminy (T) przez uracyl (U). Następnie program powinien wyświetlić długość otrzymanej sekwencji RNA. Podpowiedź: użyj w kodzie m.in. `input()` oraz `len()`
3. **[Task 6-8]** Napisz program w języku Python, który oblicza masę cząsteczkową podanego peptydu (peptyd to sekwencja o długości do około 50 aminokwasów) na podstawie sekwencji aminokwasów. Program powinien wczytać od użytkownika sekwencję peptydu jako łańcuch znaków i obliczyć jego masę cząsteczkową, korzystając z podanych mas aminokwasów. Następnie program powinien wyświetlić wynikową masę cząsteczkową peptydu.

Początek rozwiązania:

```
# Słownik zawierający masy aminokwasów
amino_acid_masses = {
    "A": 89.094,
    "C": 121.154,
    "D": 133.104,
    "E": 147.131,
    "F": 165.192,
    "G": 75.067,
    "H": 155.156,
    "I": 131.175,
    "K": 146.189,
    "L": 131.175,
    "M": 149.208,
    "N": 132.119,
    "P": 115.132,
    "Q": 146.146,
    "R": 174.203,
    "S": 105.093,
    "T": 119.120,
    "V": 117.148,
    "W": 204.228,
    "Y": 181.191,
}
```

4. [Task 9-10] Napisz program w języku Python, który obliczy procent zawartości par zasad GC (cytozyna i guanina) w danej sekwencji DNA.

Początek rozwiązania:

```
sekwencja_dna = "ATGCTACGATCGTACGGCGCAAATAGCTAGCTAGCTAGC"
# dalej Twój kod
```

5. [Task 11-12] Napisz program w języku Python, który oblicza częstotliwość alleli w populacji, mając liczbę osobników z homozygotami dominującymi (AA), heterozygotami (Aa) i homozygotami recesywnymi (aa). Oblicz częstotliwość allelu dominującego (A) oraz allelu recesywnego (a). Podpowiedź: częstotliwość wynosiłaby 1 jeśli wszystkie allele byłyby takie same, np. A, a jeśli 50% to A to wtedy jego częstotliwość wynosi 0.5

Początek rozwiązania:

```
# Dane wejściowe
homozygoty_dominujace = 100
heterozygoty = 200
homozygoty_recesywne = 100
# dalej już Twój kod
```

6. [Task 13-15] Napisz program w języku Python, który wczytuje dane dotyczące sekwencji aminokwasowych oraz dwóch właściwości dla każdej sekwencji z pliku Excel, umieszcza je w tabeli (DataFrame) przy użyciu biblioteki Pandas, a następnie oblicza statystyki takie jak

średnia, mediana i odchylenie standardowe dla tych właściwości.

Dane wejściowe (plik Excel o nazwie dane_sekwencji.xlsx):

Sekwencja	Własność1	Własność2
AAKLPLAR	1.2	3.5
KKLPARAA	0.8	3.1
LLPKARAA	1.1	2.9
LARPKKAA	1.4	4.2
KKALPRRA	0.9	3.8
RLPKALAA	1.6	4.0
LARAKKPA	1.3	3.3
KAAARPLR	1.0	2.7
ALRARKPA	1.5	3.9
RALKLPKA	1.7	4.1

7. [Task 16-18] Napisz program w języku Python, który wczyta przykładowe dane o liczbie osób zaszczepionych przeciwko COVID-19 w różnych grupach wiekowych i stworzy wykres słupkowy prezentujący te dane. Skorzystaj z bibliotek Pandas i Matplotlib.

Dane wejściowe (plik CSV o nazwie dane_szczepienia.csv):

```
Grupa wiekowa,Liczba zaszczepionych
0-17,150000
18-29,500000
30-49,750000
50-64,850000
65+,900000
```

Początek rozwiązania:

```
import pandas as pd
import matplotlib.pyplot as plt

# Wczytanie danych z pliku CSV
dane = pd.read_csv('dane_szczepienia.csv')
# dalej już Twój kod
```

8. [Task 19-21] Napisz program w języku Python, który wczyta dwa zbiory danych biologicznych z plików CSV, obliczy średnią i odchylenie standardowe dla każdego zbioru danych, a następnie przeprowadzi test t (t-test), aby porównać średnie wartości w obu zbiorach danych. W zadaniu użyj bibliotek Scipy i Pandas. Podpowiedź: Wczytaj dwa pliki CSV z danymi biologicznymi (np. 'dane_biologiczne1.csv' i 'dane_biologiczne2.csv') do dwóch DataFrame'ów Pandas.

```
dane_biologiczne1.csv:
wartosc
12.5
```

14.6
11.8
15.2
13.7
16.1
14.8
12.9

dane_biologiczne2.csv:
wartosc
17.3
18.6
19.2
20.1
17.9
15.8
18.4
16.7

9. [Task 22-24] Napisz program w języku Python, który wczyta sekwencję DNA od użytkownika, przeprowadzi transkrypcję sekwencji DNA na sekwencję RNA, a następnie przeprowadzi translację sekwencji RNA na łańcuch białkowy. W zadaniu użyj biblioteki Biopython.

Początek rozwiązania:

```
from Bio.Seq import Seq

# Wczytaj sekwencję DNA za pomocą funkcji input()
# Dalej Twój kod:
```

Przykładowe dane wejściowe:

Wprowadź sekwencję DNA: AATGCGCTTGCGTATTACGAT

Przykładowe dane wyjściowe dla tego zadania:

Sekwencja DNA: AATGCGCTTGCGTATTACGAT

Sekwencja RNA: AAUGCGCUUGCGUAUUACGAU

Łańcuch białkowy: NALAYYD

Dla chętnych:

10. [Task 25-29] Napisz program w języku Python, który przeszuka bazę danych PubMed (Entrez) w poszukiwaniu publikacji naukowych związanych z rakiem jelita. W zadaniu użyj biblioteki Biopython oraz modułu Entrez. Wypisz tytuły, autorów i daty publikacji dla każdej znalezionej publikacji. Wypisz 10 pierwszych znalezionych publikacji. Podpowiedź: rak jelita to “colon cancer”

Początek rozwiązania:

```
from Bio import Entrez
```

```
# Ustaw swój adres e-mail (wymagane dla korzystania z API Entrez)
Entrez.email = "your_email@example.com"
# Dalej Twój kod:
```