

Dane treningowe do projektu zaliczeniowego nr 2

wersja 2.1

Zestaw składa się z:

- katalogu `reference/` zawierającego plik `reference.fasta` z sekwencją referencyjną oraz pliki indeksu programu `bowtie2` dla tej sekwencji,
- katalogu `reads/` z plikami zawierającymi symulowane odczyty z sekwencji referencyjnej:
 - `reads1.fasta` – 1000 odczytów zawierających $\sim 1\%$ błędnie odczytanych nukleotydów,
 - `reads2.fasta` – 1000 odczytów zawierających 1-3% błędnie odczytanych nukleotydów,
 - `reads3.fasta` – 1000 odczytów zawierających 3-5% błędnie odczytanych nukleotydów,
- skryptów do ewaluacji wyników asemblacji.

Skrypty uruchamia się poleceniem:

```
./evaluate.sh contigs.fasta
```

gdzie `contigs.fasta` jest plikiem z predykcjami contigów. Skrypty używają:

- programu `bowtie2`,
- pythonowego modułu `pysam`.

Na zbiorach odczytów zostały przetestowane dwa algorytmy referencyjne, poniższa tabelka przedstawia otrzymane oceny (m oznacza maksymalną liczbę punktów za projekt przewidzianą w zasadach punktacji):

	algorytm1	algorytm2
liczba punktów za rozwiązanie	$m - 4$	$m - 2$
ocena dla zbioru		
- <code>reads1.fasta</code>	0.06	0.59
- <code>reads2.fasta</code>	0.03	0.21
- <code>reads3.fasta</code>	0.03	0.08