

Sprawozdanie – zadanie 3

Julia Ordecka

Bioinformatyka, II rok, IV semestr, grupa 1

155634

Opis algorytmu

Funkcja `Wczytaj()` sprawdza poprawność plików wejściowych. Następnie iteruje po pięciu sekwencjach i wpisuje dane do zmiennej strukturalnej `sequences`. Do wektora `fasta` wpisuje nukleotydy z sekwencji, sprawdza również, czy plik zawiera litery. Zliczana jest długość sekwencji oraz nadawany jest indeks wejściowy dla nukleotydu (pierwotna pozycja nukleotydu w sekwencji przed operacją usunięcia nukleotydów poniżej progu wiarygodności). Do wektora `qual` wpisywana jest ocena jakości każdego z nukleotydów, sprawdzana jest również poprawność pliku pod względem zawartości. Zapisywana jest także nazwa sekwencji.

Funkcja `Usun()` po wczytaniu progu wiarygodności podanego przez użytkownika oraz sprawdzeniu, czy podana liczba jest zgodna z wymaganym zakresem, usuwa nukleotydy o wiarygodności poniżej tego progu. Z danych usuwana jest także jakość dla tego nukleotydu oraz nadany indeks wejściowy dla nukleotydu. Zmniejszana jest długość sekwencji.

W funkcji `Stworzgraf()` po podaniu przez użytkownika długości podciągu sprawdza czy mieści się ona w wymaganym zakresie. Sprawdza również, czy długość podciągu nie jest większa niż ilość pozostałych nukleotydów po operacji usunięcia - w tym przypadku funkcja nie dopuszcza dalszych obliczeń.

Dla każdego podciągu o zadanej długości tworzony jest wierzchołek grafu, który zawiera informację o sekwencji nukleotydów o długości równej podanemu podciągowi, numerach nukleotydów reprezentujących ich pozycję w sekwencji wejściowej, numerze sekwencji, z której pochodzi oraz zapisywany jest numer wierzchołka. Wierzchołek dodawany jest do grafu.

W funkcji `StworzKrawedz()` krawędziami łączone są wierzchołki które pochodzą z różnych sekwencji, zawierają ten sam podciąg oraz jeżeli różnica w pozycjach podciągów nie jest większa niż dziesięciokrotność długości podciągu.

W funkcji `Znajdzkliki()` wierzchołki grafu są sortowane po ilości krawędzi wychodzących z danego wierzchołka malejąco. Ma to na celu znalezienie najbardziej prawdopodobnego wierzchołka tworzącego wraz z połączonymi wierzchołkami klikę. Przed sortowaniem dane ze zmiennej `graf` kopiujemy do zmiennej pomocniczej `graph_sort`. Oryginalny graf będzie potrzebny w dalszej części funkcji. W dalszej kolejności każdy kolejny wierzchołek z grafu posortowanego przepisujemy do zmiennej pomocniczej `vert_pomocniczy` reprezentującej wierzchołki. Budowa każdego

wierzchołka informuje o tym, że ten wierzchołek jest połączony krawędziami z wierzchołkami, które są zapisane w wektorze edges tego wierzchołka. Jednocześnie w stosunku do tego wierzchołka wierzchołki zapisane w edges spełniają warunki kliku (każdy z wierzchołków zawarty w edges jest połączony z rozpatrywanym wierzchołkiem, pochodzi z innej sekwencji niż rozpatrywany wierzchołek). Wierzchołki zapisane w edges niekoniecznie muszą być połączone między sobą, w szczególności niepołączone są te, które pochodzą z tych samych sekwencji, dlatego w dalszym ciągu algorytm sprawdza, czy wierzchołki umieszczone w wektorze edges wierzchołka `vert_pomocniczy` są między sobą połączone krawędziami. Aby to zrobić, dla każdego wierzchołka zawartego w wektorze edges wierzchołka `vert_pomocniczy` sprawdzamy, czy w jego wektorze edges są zawarte wierzchołki o numerach kolejnych wierzchołków zawartych w wektorze edges wierzchołka `vert_pomocniczy`. W tym celu wykorzystujemy numery wierzchołków, które są jednocześnie indeksami wierzchołków w wektorze `graph` (przed sortowaniem). Jeżeli któryś z wierzchołków zawartych w edges `vert_pomocniczy` nie jest połączony z którymkolwiek z pozostałych wierzchołków zawartych w edges `vert_pomocniczy`, to jest on usuwany z wektora edges `vert_pomocniczy`. W wyniku otrzymujemy wierzchołek `vert_pomocniczy`, który ma krawędzie z wszystkimi wierzchołkami zawartymi w jego edges oraz wszystkie wierzchołki zawarte w jego edges mają połączenia między sobą - czyli znaleziona została klika. Jeżeli rozmiar znalezionej kliku dla rozpatrywanego `vert_pomocniczy` jest maksymalny (to znaczy klika zawiera pięć wierzchołków, rozpatrywany + 4 wierzchołki w wektorze edges, co jest równoznaczne z czterema krawędziami wierzchołka rozpatrywanego), przerywamy obliczenia i wypisujemy klikę. Jeżeli znaleziony rozmiar kliku jest mniejszy niż maksymalny, to porównujemy jego rozmiar z ewentualnie wcześniej znalezionym rozmiarem kliku (na początku funkcji rozmiar ten ustawiamy na 0). W przypadku, gdy jest mniejszy, to przechodzimy do wyszukania kolejnej kliku, a w przeciwnym przypadku wpisujemy nowo znaną klikę do `cliques` i aktualizujemy zmienną informującą o aktualnie znalezionym największym rozmiarze kliku. Na początku funkcji `Znajdzklikę` sprawdzamy, czy poprzednio wyszukana klika ma rozmiar większy lub równy maksymalnemu możliwemu do znalezienia w kolejnych elementach grafu. W takiej sytuacji nie szukamy już innych klik.

Oszacowanie złożoności algorytmu

Funkcja `Wczytaj()`: $O(n)$

Funkcja `Usun()`: $O(n^2)$

Funkcja `Stworzgraf()`: $O(n)$

Funkcja `StworzKrawedz()`: $O(V^2)$

Funkcja `Znajdzklikę()`: $O(V^3)$

Złożoność obliczeniowa algorytmu wynosi $O(V^3)$, gdzie V to liczba wierzchołków w grafie.

Instancje i testy

Instancja 1

```
>DOJHLOP02HA3I7 length=97 xy=2879_2833 region=2 run=R_2005_09_08_15_35_38_
TCAGAAGCCTTGCCGATGAAATCCATCTTTTCAACGTTATATTATTTCTTCCCCAACTGAAGTTGGTATGACA
ATTCTCGAATTCAAGAAGCATTG
>DOJHLOP01D0HHA length=100 xy=1528_2540 region=1
run=R_2005_09_08_15_35_38_
ATATTATTTCTTCCCACTGAAGTTGGTAAATCCATCTTTCATTCAAGAAGCATTGCTTCAGCAAAACAGGCA
AACATTTCTCTTTTAGGATTTTCCAT
>DOJHLOP02GK2UX length=109 xy=2583_1319 region=2
run=R_2005_09_08_15_35_38_
CTTTAGGATTTTCCATCTCAGCCAATTCCATACATGCTCCAACCAATATCTTTAGAAAAATTGAAAAATCCATC
TTTCAATACGTAGTATTCCAGGTCCTGAAACTT
>DOJHLOP02GB8H1 length=131 xy=2482_2423 region=2
run=R_2005_09_08_15_35_38_
TGAACATTCCTCTTTTAATACGTGTATTCCAGGTCCTGAAACTTAATCCATCTTTTCTTAGGGTATCCACC
ATCAATCATTAACAAGGTCTTTTAATTTTAGATGGGTCAATTTCTAAGGTTTTAGG
>DOJHLOP02JHWE0 length=121 xy=3777_1226 region=2
run=R_2005_09_08_15_35_38_
TATTGTTATTTGGGTTGCAAGCATGCCTAAACCTTAGAAATTGACCCATCTAAAATTAAGACCTTGTTTAA
AATCCATCTTTTCATACCCTAAAAACTTGGGTGAGAAGTTTTTCAGGA

>DOJHLOP02HA3I7 length=97 xy=2879_2833 region=2 run=R_2005_09_08_15_35_38_
31 32 32 32 28 22 31 30 25 31 27 32 27 21 32 31 28 32 28 27 14 30 27 21 31
32 31 25 24 20 7 30 24 29 23 25 25 30 25 32 32 32 31 26 31 27 27 13 28 29
23 24 24 18 4 26 20 25 30 32 24 17 31 30 27 30 25 32 31 28 32 32 32 31 27
28 25 28 27 31 28 30 25 30 27 32 29 26 32 25 18 27 31 29 31 27 32
>DOJHLOP01D0HHA length=100 xy=1528_2540 region=1
run=R_2005_09_08_15_35_38_
32 24 21 28 22 31 25 24 9 32 30 25 25 25 18 21 12 29 26 28 24 16 28 20 11
28 23 31 31 27 31 31 28 31 28 31 28 31 26 26 32 29 24 28 25 29 31 28 32 26
20 29 24 29 31 28 28 32 31 28 27 32 24 32 25 25 20 7 31 16 18 6 30 23 22 3
30 28 27 26 13 26 29 18 24 24 19 3 29 20 10 22 25 25 22 11 31 27 32 31
>DOJHLOP02GK2UX length=109 xy=2583_1319 region=2
run=R_2005_09_08_15_35_38_
25 24 23 18 31 25 18 17 24 24 19 5 29 23 31 30 32 30 23 31 32 19 8 24 16
26 19 22 14 31 20 26 31 28 28 29 31 29 30 25 31 27 24 17 31 27 25 28 27 22
26 25 10 28 27 22 22 21 16 6 31 27 26 28 28 18 31 19 29 30 24 29 23 17 24
16 24 24 18 29 26 31 21 24 32 22 29 13 27 25 21 12 24 17 26 25 17 31 30 24
25 27 25 25 18 30 26 20 31
>DOJHLOP02GB8H1 length=131 xy=2482_2423 region=2
run=R_2005_09_08_15_35_38_
31 31 30 25 31 31 25 29 23 29 23 27 28 21 24 23 18 3 30 25 26 28 31 30 32
31 32 28 29 23 31 26 32 31 27 31 31 26 29 32 24 23 18 3 32 31 27 32 30 31
```

```

29 26 25 11 31 26 30 22 22 20 14 2 28 24 23 6 31 29 28 30 26 32 30 24 28
31 31 31 27 32 32 17 28 22 28 27 18 31 30 25 28 21 23 32 25 25 22 13 30 27
25 25 21 8 29 31 31 31 29 28 17 32 32 30 25 28 27 18 30 24 31 27 26 24 25
25 22 12 31 30 25
>DOJHLOP02JHWE0 length=121 xy=3777_1226 region=2
run=R_2005_09_08_15_35_38_
32 31 31 26 32 31 27 32 28 28 16 28 28 16 31 27 31 32 28 22 32 31 32 30 32
31 27 32 26 25 22 11 31 27 31 27 25 31 25 24 9 31 27 31 29 28 27 14 31 32
31 26 24 24 22 13 31 27 24 24 21 13 31 31 31 27 24 23 32 25 25 18 24 22 32
32 31 30 26 32 28 31 29 26 31 30 26 30 32 18 29 28 18 30 22 22 21 16 8 29
31 27 28 28 18 28 32 28 29 28 21 31 24 24 22 13 31 31 30 27 27

```

Test 1.1

Długość podciągu: 9

Próg wiarygodności: 10

największy znaleziony rozmiar kliku: 4 wierzchołki(ow).

sekwencja: 3 (>DOJHLOP02GK2UX), indeks pierwotny pierwszego nukleotydu w podciągu: 67

sekwencja: 1 (>DOJHLOP02HA3I7), indeks pierwotny pierwszego nukleotydu w podciągu: 20

sekwencja: 4 (>DOJHLOP02GB8H1), indeks pierwotny pierwszego nukleotydu w podciągu: 48

sekwencja: 5 (>DOJHLOP02JHWE0), indeks pierwotny pierwszego nukleotydu w podciągu: 75

podciąg: AATCCATCT

Test 1.2

Długość podciągu: 6

Próg wiarygodności: 20

największy znaleziony rozmiar kliku: 5 wierzchołki(ow).

sekwencja: 3 (>DOJHLOP02GK2UX), indeks pierwotny pierwszego nukleotydu w podciągu: 65

sekwencja: 1 (>DOJHLOP02HA3I7), indeks pierwotny pierwszego nukleotydu w podciągu: 19

sekwencja: 2 (>DOJHLOP01D0HHA), indeks pierwotny pierwszego nukleotydu w podciągu: 30

sekwencja: 4 (>DOJHLOP02GB8H1), indeks pierwotny pierwszego nukleotydu w podciągu: 48

sekwencja: 5 (>DOJHLOP02JHWE0), indeks pierwotny pierwszego nukleotydu w podciągu: 75

podciąg: AATCCA

Test 1.3

Długość podciągu: 4

Próg wiarygodności: 30

największy znaleziony rozmiar kliki: 4 wierzchołki(ow).

sekwencja: 3 (>DOJHLOP02GK2UX), indeks pierwotny pierwszego nukleotydu w podciągu: 15

sekwencja: 2 (>DOJHLOP01D0HHA), indeks pierwotny pierwszego nukleotydu w podciągu: 31

sekwencja: 4 (>DOJHLOP02GB8H1), indeks pierwotny pierwszego nukleotydu w podciągu: 49

sekwencja: 5 (>DOJHLOP02JHWE0), indeks pierwotny pierwszego nukleotydu w podciągu: 49

podciąg: ATCT

Instancja 2

>DOJHLOP01C1SGE length=101 xy=1133_1724 region=1
run=R_2005_09_08_15_35_38_

CAAGCTACCTTGTATCAAATGGAAGTCAGGGCTTGCAACTGTGTAAACACTATGTTCAAACA
TAAGTAACCTGTCTGCGACGCCTATCCCAATGCGCCACCAGAGCCACCTTCCCTATTA

>DOJHLOP01D6ZZX length=104 xy=1602_3371 region=1 run=R_2005_09_08_15_35_38_

ATCGCAGTGAAACCTTGTGTCAAATGGAAGTCTACGAGAAATGTTTAGGCTTAAAGTTCCAA
TAATTGCAACTGTAATAGGGGAAGGTGGCTCTGGTGGCGCATTGGGGATAGGCGTCGCAG
A

>DOJHLOP01CUSLY length=109 xy=1053_3012 region=1
run=R_2005_09_08_15_35_38_

GGAGACCTTGTGTCAAATGGAAGAGGCAATCGCAGTGAATCTACGAGAAATGTTTAGGCTT
AAAGTTCCAATAATTGCAACTGTAATAGGGGAAGGTGGCTCTGGTGGCGCATTGGGGATA
GGCGTCG

>DOJHLOP01DBA3Z length=105 xy=1241_3437 region=1
run=R_2005_09_08_15_35_38_

CCACCTACCTTGTGTCAAATGGAAGTCCCCTATTACAGTTGCAATTATTGGAACCTTAAGCC
TAAACATTTCTCGTAGATTCACTGCGATTGCCTCTCCTTGTCTTCTTCTGCTATGACG

>DOJHLOP01BGY5C length=99 xy=0486_1118 region=1 run=R_2005_09_08_15_35_38_

TTCCCACCTTGTGTCAAATGGAAGCTATTACAGTTGCAATTATTGGAACCTTAAGCCTAAACA
TTTCTCGTAGATTCACTGCGATTGCCTCTCCTTGTCTTCTTCTGCTATGAG

>DOJHLOP01C1SGE length=101 xy=1133_1724 region=1
run=R_2005_09_08_15_35_38_

29 28 23 30 28 29 15 16 7 34 36 23 22 20 36 23 29 11 15 18 34 31 17 35 27 23 30 31 24 23
6 32 25 18 31 30 27 21 28 31 31 31 28 26 23 22 3 31 32 31 25 31 31 23 20 10 29 24 23 5
30 28 31 31 27 27 28 30 26 24 17 31 30 31 21 26 32 30 31 31 32 13 29 24 31 32 14 28 28
19 23 15 23 32 30 19 30 25 29 19 9 24 31 12 12 31 26 31 31 27 28 22 29 28 17 26 29 21 13
31

>DOJHLOP01D6ZZX length=104 xy=1602_3371 region=1 run=R_2005_09_08_15_35_38_

32 32 32 31 32 30 31 30 25 31 27 15 16 7 34 36 23 22 20 36 23 29 11 15 18 34 31 17 35 27
29 30 32 28 29 31 32 31 27 27 13 19 24 27 27 13 29 31 26 30 31 27 29 28 17 30 29 24 30
25 28 25 28 31 28 28 23 32 30 27 21 32 25 29 32 31 28 29 30 25 25 21 9 23 15 31 28 29 28
23 31 28 30 29 27 25 31 31 28 23 26 30 29 30 27 26 25 22 10 29 30 29 27 25 31 25 31 32
31 31 31 26 30

>DOJHLOP01CUSLY length=109 xy=1053_3012 region=1
run=R_2005_09_08_15_35_38_

28 18 27 30 15 16 7 34 36 23 22 20 36 23 29 11 15 18 34 31 17 35 27 26 31 27 31 29 24 29
31 31 31 21 32 31 32 24 17 30 32 26 31 26 28 32 29 24 23 6 31 31 25 24 9 30 25 19 28 31
27 26 25 10 30 27 20 31 26 31 27 32 30 27 28 23 31 31 30 26 32 31 31 31 31 27 28 31 25
25 22 12 23 15 27 25 30 31 28 31 31 26 31 29 27 29 31 28 32 31 29 31 30 27 25 25 22 11
31 30 26 30 27 31 32 32 32 32

>DOJHLOP01DBA3Z length=105 xy=1241_3437 region=1
run=R_2005_09_08_15_35_38_

31 27 21 30 27 25 15 16 7 34 36 23 22 20 36 23 29 11 15 18 34 31 17 35 27 23 25 25 22 12
27 32 30 25 29 32 21 28 31 27 31 27 23 16 26 24 32 31 27 30 25 30 25 31 29 28 18 31 28
26 31 28 28 28 27 19 31 25 29 28 18 28 29 31 32 32 31 32 25 27 20 29 32 31 29 29 31 31

31 30 25 28 31 28 31 31 31 31 27 31 27 29 31 21 13 29 26 27 29 26 24 30 25 31 27 14 31
26 29 30 23 20 9 31

>DOJHLOP01BGY5C length=99 xy=0486_1118 region=1 run=R_2005_09_08_15_35_38_

29 24 24 24 18 15 16 7 34 36 23 22 20 36 23 29 11 15 18 34 31 17 35 27 2 30 31 29 24 23
31 24 29 29 24 31 30 30 25 31 27 27 31 26 29 24 25 18 31 27 26 13 31 27 30 29 24 32 29
28 17 32 31 27 26 13 32 29 30 31 28 30 28 31 29 24 32 31 31 31 30 29 28 31 28 22 32 30
25 32 32 31 27 20 29 24 31 32 26 20 31 28 30 31 27 30 31 28 32 28 31 32 30 29 26 25 21
27

Test 2.1

Próg wiarygodności: 30

Długość podciągu: 7

największy znaleziony rozmiar kliki: 2 wierzchołki(ow).

sekwencja: 2 (>DOJHLOP01D6ZZX), indeks pierwotny pierwszego nukleotydu w podciągu: 15

sekwencja: 1 (>DOJHLOP01C1SGE), indeks pierwotny pierwszego nukleotydu w podciągu: 10

podciąg: TTTGGAC

Test 2.2

Próg wiarygodności: 7

Długość podciągu: 7

największy znaleziony rozmiar kliki: 5 wierzchołki(ow).

sekwencja: 3 (>DOJHLOP01CUSLY), indeks pierwotny pierwszego nukleotydu w podciągu: 5

sekwencja: 1 (>DOJHLOP01C1SGE), indeks pierwotny pierwszego nukleotydu w podciągu: 7

sekwencja: 2 (>DOJHLOP01D6ZZX), indeks pierwotny pierwszego nukleotydu w podciągu: 12

sekwencja: 4 (>DOJHLOP01DBA3Z), indeks pierwotny pierwszego nukleotydu w podciągu: 7

sekwencja: 5 (>DOJHLOP01BGY5C), indeks pierwotny pierwszego nukleotydu w podciagu: 6

podciag: ACCTTGT

Test 2.3

Próg wiarygodności: 20

Długość podciagu: 6

największy znaleziony rozmiar kliki: 5 wierzchołki(ow).

sekwencja: 3 (>DOJHLOP01CUSLY), indeks pierwotny pierwszego nukleotydu w podciagu: 12

sekwencja: 1 (>DOJHLOP01C1SGE), indeks pierwotny pierwszego nukleotydu w podciagu: 25

sekwencja: 2 (>DOJHLOP01D6ZZX), indeks pierwotny pierwszego nukleotydu w podciagu: 19

sekwencja: 4 (>DOJHLOP01DBA3Z), indeks pierwotny pierwszego nukleotydu w podciagu: 14

sekwencja: 5 (>DOJHLOP01BGY5C), indeks pierwotny pierwszego nukleotydu w podciagu: 13

podciag: GTCAGG

Instancja 3

```
>DOJHLOP01BQX95 length=100 xy=0599_3707 region=1 run=R_2005_09_08_15_35_38_
GGGACAAAGTTCTCGGAAAGCTTGATGTTTCAGGAGAATAACCAATCGATTGAGATGGCATCATCCATATCG
GCATATTGTCTGGCAACCTATTAAGAGT
>DOJHLOP02GIMV0 length=94 xy=2555_1998 region=2 run=R_2005_09_08_15_35_38_
AGTTCCTGAAAGCTTGATGTTTCATGAAGAAAGACCAATCGATTGAGATGGCATCATCCATATCGGCATTATT
ATCTGGCAACCTATTAAGAGT
>DOJHLOP01BN95F length=101 xy=0569_2001 region=1 run=R_2005_09_08_15_35_38_
TCCTGGGAAAGCTTGATGTTTCAGGAGAGAAAGCTTGATGTTTCAGGAGCATCATCCATATCGGCATATTGTCT
GGCAACCTATTAAGAGTCTTGATTAACA
>DOJHLOP01DE8MX length=107 xy=1286_2535 region=1 run=R_2005_09_08_15_35_38_
GGGAAAGCTTGATGTTTCAGGAGAAAAGTAAGCTTGATGTTTCAGGAATCATCCATATCGGCATATTATCTGGCA
ACCTAATTAAGAGTCTTGAATTAACATTATATT
>DOJHLOP01BS77S length=99 xy=0625_3398 region=1 run=R_2005_09_08_15_35_38_
TTGCTGAACATATTGAAAGCTTGATGTTTCAGGAGTTGCCAGAACATATGCCGATATGGATGATGCCATCTCA
ATCGATTGTTCTTCTCCTGAACAT
```

```
>DOJHLOP01BQX95 length=100 xy=0599_3707 region=1 run=R_2005_09_08_15_35_38_
25 22 11 28 32 28 27 14 31 31 28 31 27 32 28 28 15 23 22 4 29 32 31 28 31 26 31 30 31 28 31 30 31
28 32 31 18 22 21 15 6 31 28 31 28 32 32 32 31 31 28 29 26 32 30 32 30 27 32 27 32 32 32 31 31 28
```


29 32 31 28 32 31 28 30 32 20 31 25 25 22 12 32 24 31 27 30 30 25 31 28 30 30 30 26 30 27 31 19 27 31

>DOJHLOP02GIMV0 length=94 xy=2555_1998 region=2 run=R_2005_09_08_15_35_38_

30 31 31 27 31 27 31 28 27 18 28 27 14 32 32 31 27 31 28 32 32 31 27 31 32 31 27 29 28 22 22 21 16 7 31 27 31 26 31 32 32 32 31 27 29 30 32 27 31 31 27 32 31 32 32 28 31 31 27 27 31 32 27 32 31 27 28 31 17 5 32 24 24 22 13 30 20 27 25 23 23 15 25 17 32 30 20 9 29 23 31 9 17 31

>DOJHLOP01BN95F length=101 xy=0569_2001 region=1 run=R_2005_09_08_15_35_38_

31 31 27 31 29 28 17 26 26 11 31 32 30 27 31 28 32 31 31 28 31 31 30 27 31 32 22 22 21 16 6 31 28 31 28 31 31 32 31 28 26 26 32 29 31 31 28 32 32 31 32 32 31 31 28 24 32 31 30 32 29 24 24 29 24 32 25 25 20 6 31 14 28 22 31 26 20 27 20 32 32 24 17 28 22 32 9 15 29 31 29 24 28 18 31 28 26 25 22 10

>DOJHLOP01DE8MX length=107 xy=1286_2535 region=1 run=R_2005_09_08_15_35_38_

24 19 4 23 22 4 32 31 25 18 30 31 32 30 31 28 26 29 30 26 17 27 22 22 21 16 7 30 25 31 28 26 28 28 32 29 24 31 29 21 27 29 28 22 15 32 20 28 29 31 26 20 32 30 26 21 31 20 11 25 29 17 29 25 24 20 5 32 30 27 20 17 30 26 31 28 17 20 11 31 28 25 24 8 27 24 31 17 28 25 24 27 19 9 29 27 25 24 22 12 21 12 32 25 25 21 8

>DOJHLOP01BS77S length=99 xy=0625_3398 region=1 run=R_2005_09_08_15_35_38_

29 23 31 29 29 31 25 24 20 5 27 24 25 25 20 7 31 27 31 31 31 28 32 26 32 28 31 31 28 31 28 26 32 31 28 29 24 30 29 24 25 32 25 24 19 5 22 32 31 31 30 26 32 30 28 29 31 29 26 31 31 23 31 29 31 31 27 26 27 29 31 31 28 23 31 32 29 31 23 22 3 31 27 22 22 21 16 7 22 28 31 26 32 32 25 18 31 28 29

Test 3.1

Próg wiarygodności: 15

Długość podciagu: 7

największy znaleziony rozmiar kliki: 5 wierzchołki(ow).

sekwencja: 2 (>DOJHLOP02GIMV0), indeks pierwotny pierwszego nukleotydu w podciagu: 17

sekwencja: 1 (>DOJHLOP01BQX95), indeks pierwotny pierwszego nukleotydu w podciagu: 25

sekwencja: 3 (>DOJHLOP01BN95F), indeks pierwotny pierwszego nukleotydu w podciagu: 37

sekwencja: 4 (>DOJHLOP01DE8MX), indeks pierwotny pierwszego nukleotydu w podciagu: 35

sekwencja: 5 (>DOJHLOP01BS77S), indeks pierwotny pierwszego nukleotydu w podciagu: 24

podciąg: GATGTTC

Test 3.2

Próg wiarygodności: 27

Długość podciagu: 6

największy znaleziony rozmiar kliki: 4 wierzchołki(ow).

sekwencja: 1 (>DOJHLOP01BQX95), indeks pierwotny pierwszego nukleotydu w podciągu: 27

sekwencja: 2 (>DOJHLOP02GIMV0), indeks pierwotny pierwszego nukleotydu w podciągu: 19

sekwencja: 3 (>DOJHLOP01BN95F), indeks pierwotny pierwszego nukleotydu w podciągu: 17

sekwencja: 5 (>DOJHLOP01BS77S), indeks pierwotny pierwszego nukleotydu w podciągu: 26

podciąg: TGTTC A

Test 3.3

Próg wiarygodności: 17

Długość podciągu: 5

największy znaleziony rozmiar kliki: 5 wierzchołki(ow).

sekwencja: 1 (>DOJHLOP01BQX95), indeks pierwotny pierwszego nukleotydu w podciągu: 24

sekwencja: 2 (>DOJHLOP02GIMV0), indeks pierwotny pierwszego nukleotydu w podciągu: 16

sekwencja: 3 (>DOJHLOP01BN95F), indeks pierwotny pierwszego nukleotydu w podciągu: 36

sekwencja: 4 (>DOJHLOP01DE8MX), indeks pierwotny pierwszego nukleotydu w podciągu: 34

sekwencja: 5 (>DOJHLOP01BS77S), indeks pierwotny pierwszego nukleotydu w podciągu: 61

podciąg: TGATG

Wnioski

Wyszukiwane podciągi są różne i zależą od wartości podanych przez użytkownika. Wysoki próg wiarygodności powoduje usuwanie większej ilości nukleotydów z sekwencji i wyszukany podciąg może różnić się od wprowadzonego motywu. Im wyższe są obie wartości, tym trudniej znaleźć wspólny motyw w sekwencjach. Dla krótkich długości podciągów wyszukane mogą zostać motywy inne niż wprowadzone.