

Sprawozdanie z projektu 3 AKwB

1. Wczytanie sekwencji z pliku:

Program otwiera plik o nazwie podawanej przez użytkownika. Pobierana jest nazwa sekwencji i tworzony jest nowy element klasy. W klasie przechowywany jest wektor przechowujący sekwencję, wektor przechowujący wartości poszczególnych nukleotydów, wektor przechowujący miejsce wystąpienia poszczególnych nukleotydów w sekwencji jak również zmienna do zapamiętania nazwy wczytanej sekwencji. Odczytana pierwsza linia pliku to nazwa sekwencji, która jest zapamiętywana dla elementu klasy. Następnie program zaczyna odczyt sekwencji znak po znaku. Odczytane nukleotydy są przekazywane do wektora przechowującego sekwencje. W momencie przekazania nukleotydu do wektora, przekazywana jest również jego pozycja w sekwencji poprzez podanie wartości inkrementatora do odpowiedniego wektora. Wartość inkrementatora jest zwiększana w momencie odczytu nowego nukleotydu. Gdy program natrafi na znak '>' przerywa odczyt. Skonstruowany obiekt klasy przekazywany jest do wektora klasy, w którym przechowywane są obiekty. Następuje wczytanie nowej linii i utworzenie nowego obiektu klasy. Cały proces trwa do momentu natrafienia przez program znaku ';'.

W momencie jego wystąpienia, zamykany jest plik fast. Program przechodzi do odczytu pliku qual. W przypadku przygotowanych sekwencji ich wystąpienia w plikach fasta i qual są takie same. Dzięki temu przy odczycie 1 sekwencji z pliku qual wiadomo, że dana sekwencja w wektorze klasy ma miejsce vector[0]. Następuje pobranie pierwszej linii pliku tekstowego a następnie odczyt znak po znaku wartości nukleotydów. Wartości są odczytywane i przekazywane do wektora przechowującego wartości nukleotydów dla odpowiedniej sekwencji. Odczyt i przekazanie trwają tak długo, dopóki program nie natrafi na 0 które znajduje się na końcu każdej sekwencji jako sygnalizator do zaprzestania odczytu. W momencie wystąpienia 0 przerywana jest pętla wpisująca oraz zwiększana jest wartość inkrementatora, co pozwala na wczytanie kolejnych wartości do następnego obiektu klasy (następnej sekwencji). W momencie natrafienia w trakcie odczytu na symbol ';' program zamyka plik oraz kończy odczyt.

2. Modyfikacja sekwencji:

Modyfikowanie sekwencji rozpoczyna się od podania przez użytkownika wartości wiarygodności. Wartość ta przekazywana jest do funkcji. Modyfikacja złożona jest z dwóch pętli for. Pierwsza pętla odpowiada za pobranie obiektu klasy. Druga natomiast za przeszukanie wektora wartości dla odpowiedniego obiektu. Jeśli wartość wiarygodności jakiegoś nukleotydu jest mniejsza od wartości podanej przez użytkownika program usuwa tą wartość z wektora. Oprócz usunięcia wartości wiarygodności nukleotydu z wektora następuje usunięcie nukleotydu z wektora sekwencji jak również usunięcie jego pozycji z wektora przechowującego te informacje. Po znalezieniu i

usunięciu nukleotydu o wartości wiarygodności mniejszej niż ta podana pętla przeszukująca rozpoczyna się od nowa. Powtarzanie trwa tak długo aż w wektorze wartości zostaną tylko wystąpienia większe od wartości podanej przez użytkownika. Po zakończeniu pętli przeszukującej następuje przejście do następnej sekwencji i jej przeszukiwanie.

3. Tworzenie wierzchołków

Tworzenie wierzchołków rozpoczyna się od podania przez użytkownika wartości jakiej długości mają być tworzone podciągi. Wartość musi znajdować się w przedziale od 4 do 9. Wartość przekazywana jest do funkcji. Tworzenie wierzchołków złożone jest z 3 funkcji for. Pierwsza podobnie jak w przypadku modyfikacji odpowiada za wybranie obiektu klasy co pozwala na odczytanie sekwencji jak również jej długości. W drugiej pętli tworzone są nowe obiekty klasy. Długość trwania pętli to wielkość sekwencji nukleotydowej pomniejszona o wartość długości podciagu podaną przez użytkownika. Do każdego obiektu przy wykorzystaniu trzeciej funkcji for wpisywana jest sekwencja o długości podanej przez użytkownika. Jak również do nowo powstałego obiektu przekazywana jest informacja, z której sekwencji pochodzi dany podciąg, numer pierwszego nukleotydu podciagu (numery podawane są po modyfikacji sekwencji oryginalnej), obiektowi nadawany jest również numer id oraz podana jest nazwa sekwencji oryginalnej, z której pochodzi podciąg. Tak utworzony wierzchołek (obiekt klasy) wpisywany jest do wektora wierzchołków. Następuje zwiększenie inkrementatora id i rozpoczynane jest tworzenie wierzchołka dla następnego nukleotydu sekwencji.

4. Tworzenie połączeń w grafie

Tworzenie połączeń pomiędzy wierzchołkami odbywa się poprzez porównanie ze sobą sekwencji podciągów wszystkich wierzchołków. Funkcja porównująca zbudowana jest z 3 funkcji for. Pierwsze dwie odpowiadają za pobranie porównywanych wierzchołków. W drugiej funkcji for dodatkowo znajduje się zmienna do zapamiętania ilości poprawnych wystąpień nukleotydów w sekwencji. Porównywanie wierzchołków odbywa się w momencie gdy nie są to te same wierzchołki ($i \neq j$). Jeśli są to dwa różne wierzchołki to następuje porównanie ich sekwencji przy użyciu trzeciej pętli for. W trzeciej pętli for znajduje się warunek sprawdzający czy wystąpienia nukleotydów porównywanych sekwencji są takie same. Jeśli nukleotydy w sekwencji są takie same zwiększana jest wartość zmiennej mającej pamiętać ilość poprawnych wystąpień nukleotydów. Po sprawdzeniu sekwencji obu wierzchołków sprawdzana jest wartość zmiennej opisanej wcześniej. Jeśli jej wartość jest równa długości sekwencji to oznacza, że porównywane sekwencje były takie same. Do wierzchołka pobranego w pierwszej pętli przekazywany jest numer wierzchołka, z którym był porównywany. Informacja ta przekazywana jest do wektora listy sąsiedztwa, który to wektor jest składową klasy wierzchołków. Następuje przejście do nowego wierzchołka, wyzerowanie zmiennej sprawdzającej poprawność wystąpień nukleotydów i sprawdzenie poprawności sekwencji kolejnego wierzchołka.

5. Poszukiwanie klik

Dla poszukiwania klik tworzony jest obiekt klasy. Do przechowania znalezionych klik użyty został wektor klasy *Cliq*. W klasie przechowywane są takie informacje jak:

- sekwencja nukleotydowa znalezionej klik,
- elementy klik (wierzchołki grafu),

- numery podciągów składających się na kliki,
- nazwy sekwencji, z których pochodzą elementy kliki,
- id kliki,

Funkcja poszukująca klikę bazuje na informacji, że w grafie łączone ze sobą wierzchołki o takiej samej sekwencji nukleotydowej podciągów i aby spełnić warunek znalezienia kliki, należy znaleźć taki zestaw aby w liście sąsiedztwa wystąpiły wierzchołki/podciągi pochodzące ze wszystkich wczytanych sekwencji sekwencji. W funkcji zadeklarowana jest tablica boolowska *sequences[5]*, której każdym miejscu przypisywana jest wartość. Jak również zmienna boolowska *fine* o wartości *true*. W pierwszej pętli wybierany jest wierzchołek dla, którego będzie sprawdzana lista następników. Jak również tworzony jest obiekt klasy. W następnej pętli sprawdzane jest, z której sekwencji pochodzi dany wierzchołek w liście, jak również sprawdzane jest z jakich sekwencji pochodzą wierzchołki obecne w liście następników sprawdzanego wierzchołka. W pętli tej umieszczone jest 5 warunków. Jeśli wystąpił wierzchołek z jednej z badanych sekwencji wtedy na odpowiednim miejscu tablicy *sequences* wartość *false* zamieniana jest na wartość *true*.

Po sprawdzeniu listy następników pętla *for* sprawdza, czy na którymś z miejsc tablicy *sequence[5]* nie występuje wartość *false*. Jeśli na, którymkolwiek miejscu w tablicy występuje wartość *false* wtedy zmieniana jest wartość zmiennej *fine* na *false*. Jeśli natomiast na wszystkich miejscach tablicy była wartość *true* wtedy spełniony jest warunek, mówiący że w tablicy następników występują elementy ze wszystkich sekwencji. Po spełnieniu tego warunku do obiektu klasy *cliquee* przekazywane jest:

- sekwencja podciągu,
 - wpisywane są elementy kliki (wierzchołki grafu),
 - nadawane jest id kliki,
- Utworzony obiekt przekazywany jest do wektora klasy *cliquee*.

Po wykonaniu powyższych operacji do kliki przekazywane są informacje o numerze podciągu jak również nazwie sekwencji, z której pochodzi każdy podciąg należący do kliki. Całość wykonywana jest w 3 pętlach *for* gdzie dzięki znajomości id. wierzchołka należącego do kliki odnajdywany jest odpowiedni wierzchołek. Po jego znalezieniu do wektorów klasy *cliquee* przekazywana jest nazwa sekwencji jak również numer podciągu, które są zapisane w danym wierzchołku.

6. Złożoność obliczeniowa

Najbardziej złożonymi funkcjami w programie są:

- funkcja tworząca wierzchołki – 3 pętle *for*,
- funkcja tworząca połączenia w grafie – 3 pętle *for*,
- funkcja wyszukująca kliki – 3 pętle *for*,

W samej funkcji wyszukującej kliki umieszczone jest więcej pętli *for*, lecz nie są to pętle stricte zagnieżdżone w sobie lecz występujące po sobie. Dlatego też określiłbym je jako potrójne wystąpienie 2 pętli *for*.

Na tej podstawie szacuję złożoność wymienionych funkcji jako $O(n^3)$ każda.

Pozostałe funkcje:

- funkcja modyfikująca sekwencje – 2 pętle *for*,

Złożoność obliczeniową całego programu szacuję jako: $O(n^3)$

7. Wyniki

```
Fasta.1
Motyw: TCAGAAAGGCCTGTTT
Motyw lenght: 16nt

>DOJHLOP01E2OD5
TGTATTCCAGGTCTGAAAACTCTCACCAAGTTTTAGGGTATCCACCATCAATCATT
AAACAAGGTCTTTTAATTTAGATGGGTCAATTTCTAAGTTTTAGGCATGCTTGCATCAGAAAGGCCTGTTTACCCAAA
>DOJHLOP02HHDJI
GTGGCTTCAGAAAGGCCTGTTTTTTACATCAATAATTTTTGAGAATTTTAATTTGCTCCAAAATCAACAAGTTAGAA
ATACGACTCTTGATTGGCAAAGGTTTACCCTGGTAATACTCATACAGCTTGGGTG
>DOJHLOP02G02S7
GTCTAATTTCAGAAAGGCCTGTTTGGACATTCAACAAGTTTTGAAGATGCCAAGCGAAAGGCAATTAGGCTTGGGG
TATGACCATATCGCTCAAGGGGATTGGACGTATGGTGTACTGCA
>DOJHLOP02JGLS0
GATGGCATCAGAAAGGCCTGTTTTCATCCATATCGGCATATTTCTGGCAACCTATTAAGAGTCTTGATTAATAA
TTTTCAGCAAAAAAGTATCAGTAATCAAAGGTAATGTGGTCCA
>DOJHLOP01DNKEA
ATAACGTTGGAAAATCAGAAAGGCCTGTTTGATGGATTCATCGGCAAGGCTTCTGAAAGGCACGGGTTTTCGGCT
GTTGGTTTGAAATCAATATTCAGACATTAAACCGTCTGAACATTTTGGTTTACTTTTTT
TTTTATGG
```

```
Qual.1

>DOJHLOP01E2OD5
32 31 32 31 31 28 31 28 28 31 28 31 31 28 28 30 24 24 19 3 31 31 28 32 31 29 30 28 27 14 31 28 29 22 22 21 16 7 32 27
27 13 32 28 31 31 28 31 31 27 31 29 32 31 26 32 32 27 29 24
27 26 13 30 25 19 25 19 24 30 25 25 22 11 30 25 25 25 21 9 32 31 27 29 29 28 18 32 30 30 25 29 28 18 31 17 31 28 26 24
24 23 21 12 31 31 27 29 31 31 31 29 24 32 17 28 32 31 31 31 26 25 11 31 27 31 26 28 32 29 | 28 16 26 28 27
15 29 28 18 0
>DOJHLOP02HHDJI
27 31 31 27 31 32 31 31 31 26 25 11 31 27 31 26 28 32 29 28 16 29 28 17 30 32 31 31 31 27 30 29 26 22 22 21 16 6 31
31 30 31 27 25 25 22 12 31 27 29 28 17 29 31 32 31 26 25 25 22 13 31 32 27 24 31 30 27 32 31 27 21 29 29 28
16 31 29 32 31 31 32 30 32 31 27 31 24 29 23 31 27 30 27 27 18 30 25 27 26 12 29 31 27 32 28 29 31 26 32 31 27 32 32
31 27 29 28 26 25 32 24 32 32 28 25 26 26 18 25 26 0
>DOJHLOP02G02S7
25 31 28 30 31 26 28 21 32 31 31 31 26 25 11 31 27 31 26 28 32 29 28 16 30 24 30 31 27 31 26 32 29 23 32 30 27 31 24
24 19 4 29 28 22 28 31 29 32 28 21 29 23 24 30 25 28 28 16 30 25 28 23 14 31 27 29 31 26 32 31 27 25 25 21 9
31 31 29 28 22 27 21 31 26 26 31 32 27 30 29 28 31 27 23 23 21 13 20 27 26 18 30 27 22 27 32 32 32 31 30 27 29 31 32
31 25 27 29 18 31 0
>DOJHLOP02JGLS0
27 30 29 31 27 31 30 32 31 31 31 26 25 11 31 27 31 26 28 32 29 28 16 32 31 31 28 31 27 32 30 32 28 30 31 26 30 30 31
32 25 25 20 7 27 31 26 19 30 27 20 31 26 31 30 30 24 31 26 31 25 28 32 29 31 27 28 32 31 26 25 25 22 10 28 32
25 25 21 9 28 28 30 27 15 15 15 14 12 8 3 31 31 32 28 31 31 28 32 30 25 26 29 29 28 18 31 26 31 24 23 7 32 30 32 23 15
15 22 13 22 0
>DOJHLOP01DNKEA
32 31 30 27 31 32 31 28 31 27 25 24 20 5 32 31 31 31 26 25 11 31 27 31 26 28 32 29 28 16 29 24 31 29 24 24 29 28 16 28
31 29 28 30 26 32 23 15 29 24 31 30 27 32 28 31 23 22 3 30 25 31 24 29 24 23 6 25 25 20 6 31 27 20 31 25
29 28 26 28 23 24 23 7 27 26 26 11 32 31 28 27 14 18 31 30 25 31 28 30 31 31 30 29 26 30 28 31 27 29 29 29 31 32 30 26
31 11 31 24 24 19 5 30 26 29 28 18 18 29 7 7 6 6 6 5
4 3 2 1 31 29 30 27 0
```

```
Podaj nazwe pliku fastq do odczytu: fastal
Podaj nazwe pliku qual do odczytu: qual1
Podaj wartosc wiarygodnosci:
30
Podaj dlugosc podciagow w zakresie od 4 do 9:
6
Ilosc znalezionych klik: 1
Klika nr: 1
-----
Sekwencja kliku
-----
T C A G G C
-----
Elementy kliku:
-----
Nazwa sekwencji: >DOJHLOP01E2OD5 |Numer podciagu z sekwencji oryginalnej: 118
-----
Nazwa sekwencji: >DOJHLOP02HHDJI |Numer podciagu z sekwencji oryginalnej: 6
-----
Nazwa sekwencji: >DOJHLOP02G02S7 |Numer podciagu z sekwencji oryginalnej: 9
-----
Nazwa sekwencji: >DOJHLOP02JGLS0 |Numer podciagu z sekwencji oryginalnej: 8
-----
Nazwa sekwencji: >DOJHLOP01DNKEA |Numer podciagu z sekwencji oryginalnej: 15
-----
```

```
Fasta.2
Motyw: ATTGTGCAATCTGA
Motyw length: 14nt

>DOJHLOP01CPWLL
CAGGTCTCTGAATTGTCAATCTGAAACCTCTCACCCCAAGTTTTAGGGTATCCACCATCAATCATTAACAAAGG
TCITTTAATTITAGATGGGTCAATTTCTAAGGTTTTAGGCATGCTTGCAACCCAAATAACAA
>DOJHLOP01ATP7M
CCCAAGCTGATAGATTGTGCAATCTGATATTACACGTGTAACCTTTGGCCAATCAAGAGTCGATATTTCTAA
CTTGTTGATTTTGGAGCAAAAATAAAATTTCTACAAAAAATTATTGATGTAAGCCA
>DOJHLOP01C1OKC
CCAAATGTTTGGTCTAATTGTGCAATCTGAATTGGACATTCACAAAGTTTGAAGATGCCAAGCGAAAGGCATT
AGGCTTGGGGTATGACCATATCGCTCAAGGGGATTTGGACGTATGGTG
>DOJHLOP01ATT30
CATGGACCACATTGTCAATCTGAATTTACCTTTGATTACTGATACTTTTTTCTGAAAAATATTTTATCAAA
GACTCTTAATAGGTTGCCAGAAAATATGCCGATATGGATGATG
>DOJHLOP01B3GME
TAAAAAAA ATTGTGCAATCTGAAGTAACCAAAATGTTCAGACGGTTAATGTCTGAATATTTGATTCAAACCAAC
AGCCGAAAAACCCGTGCTTTTCAGAAAGCTTGCCGATGAAATCCATCTTTTCAACGTTAT
ATTA
;
```

```
Qual.2
>DOJHLOP01CPWLL
29 29 22 14 31 25 18 28 31 27 31 25 24 9 31 23 31 29 24 31 29 31 31 31 27 19 32 27 20 28 24 23 27 27 26 12 29 27 31 22
22 21 15 5 31 23 22 5 31 21 24 31 26 31 28 22 28 26 31 30 26 22 30 21 31 27 25 25 18 29 31 28 27 21
25 29 26 25 22 10 31 28 26 25 22 10 32 31 26 27 24 23 7 24 29 29 23 24 23 7 29 32 31 28 27 21 24 24 18 2 31 29 23 31 31
29 31 32 25 19 32 29 30 28 27 26 12 28 27 19 26 29 26 31
28 26 0
>DOJHLOP01ATP7M
26 25 11 29 24 31 30 31 31 31 28 29 31 29 31 25 24 9 31 23 31 29 24 31 29 31 31 31 31 27 21 23 15 31 31 31 25
18 30 24 23 6 28 23 26 25 11 29 28 22 28 23 32 32 31 27 32 27 30 32 28 31 25 32 27 26 13 31 32 29 24
29 31 28 29 26 20 31 29 26 25 22 10 30 27 31 28 31 24 24 19 4 24 22 24 24 21 12 30 27 25 31 26 31 20 20 19 16 9 1 25 24
29 28 22 31 32 27 27 21 27 26 13 23 28 26 20 0
>DOJHLOP01C1OKC
31 27 29 28 18 31 31 28 28 15 31 27 31 31 32 31 31 25 24 9 31 23 31 29 24 31 29 31 31 31 28 31 27 31 26 26 31 31 30 27
32 30 28 32 31 28 31 25 25 21 8 31 29 24 31 30 28 29 29 24 31 27 32 32 31 28 28 16 28 23 29 31 29 23
30 28 22 29 31 26 24 24 19 4 25 25 29 31 26 27 20 22 32 29 29 29 31 32 32 30 31 28 25 24 20 5 24 29 28 17 30 27 31 31
29 23 31 30 31 28 31 28 0
>DOJHLOP01ATT30
32 32 32 31 28 28 31 27 29 31 31 25 24 9 31 23 31 29 24 31 29 31 31 31 31 29 28 18 29 30 26 28 28 18 32 32 31 28 26 31
28 31 29 22 30 32 11 11 11 10 9 8 5 2 32 31 30 32 24 24 19 4 27 19 25 25 22 12 31 28 32 31 31 28
32 30 32 26 32 31 28 30 27 31 31 31 28 30 25 32 30 25 20 32 25 25 21 9 31 28 26 30 29 24 30 26 30 30 30 27 27 31 28
30 31 30 0
>DOJHLOP01B3GME
31 11 11 11 11 9 8 5 31 25 24 9 31 23 31 29 24 31 29 31 31 31 2 30 31 27 25 31 27 29 28 18 32 32 31 27 29 31 29 31 30
31 27 30 25 31 28 32 32 32 31 31 31 24 17 31 29 29 28 17 30 24 27 26 13 30 28 28 16 30 28 29 26 30
31 30 29 26 30 26 25 22 10 29 28 18 31 31 28 31 27 28 27 14 32 32 31 24 17 26 28 23 31 28 29 27 21 30 29 30 28 26 25 9
31 29 24 31 16 17 25 25 22 11 31 27 28 23 19 22 31 27 30 32
30 31 28 32 0
;
```

```
Podaj nazwe pliku qual do odczytu: qual2
Podaj wartosc wiarygodnosc:
20
Podaj dlugosc podciagow w zakresie od 4 do 9:
7
Ilosc znalezionych klik: 7
Klika nr: 1
-----
Sekwencja kliku
A T T G T C A
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 11
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 16
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 17
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 11
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 9
-----
Klika nr: 2
-----
Sekwencja kliku
T T G T C A A
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 12
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 17
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 18
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 12
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 10
-----
```

```
Klika nr: 3
-----
Sekwencja kliku
T G T C A A T
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 13
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 18
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 19
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 13
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 11
-----
Klika nr: 4
-----
Sekwencja kliku
G T C A A T C
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 15
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 20
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 21
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 15
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 13
-----
```

```
Klika nr: 5
-----
Sekwencja kliku
T C A A T C T
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 16
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 21
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 22
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 16
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 14
-----
Klika nr: 6
-----
Sekwencja kliku
C A A T C T G
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 17
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 22
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 23
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 17
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 15
-----
```

```
Klika nr: 7
-----
Sekwencja kliku
A A T C T G A
-----
Elementy kliku:
Nazwa sekwencji: >DOJHLOP01CPWLL |Numer podciagu z sekwencji oryginalnej: 18
-----
Nazwa sekwencji: >DOJHLOP01ATP7M |Numer podciagu z sekwencji oryginalnej: 23
-----
Nazwa sekwencji: >DOJHLOP01C1OKC |Numer podciagu z sekwencji oryginalnej: 24
-----
Nazwa sekwencji: >DOJHLOP01ATT30 |Numer podciagu z sekwencji oryginalnej: 18
-----
Nazwa sekwencji: >DOJHLOP01B3GME |Numer podciagu z sekwencji oryginalnej: 16
-----
```

```

Fasta.5
Motyw: TCAAAATTG
Motyw lenght: 9nt

>DOJHLOP02HSG8F
GAAAACTCTCACCCAAGTTTTAGTCAAAATTGGGTATCCACCATCAATCATTAAACAAGGCTTTTA
ATTTTAGATGGGCAATTTCTAAGGTTTTAGGCATGCTTGAACCCAAATAACAATATCTG
>DOJHLOP01AD4GJ
AAGCTGTATGAGTATTACCAAGTGGTAAACCTTTGCCAATCAAGAGTCGTATTTCTAACTT
GTTGATTTTGGAGCAAAATTTAAATTTCTACAAAATTTATTGATGTAA TCAAAATTGAGCCA
>DOJHLOP01CNMJ2
AGTTCAATTTGATCTTGCTTGGTGCCAAATGTTGGTCTAATTGGACATTCAACAAGTT
TTGAAGACTGCCAAGACGAAGGCATTAGGCTTGGGGTATGAACCATCAAAATTGTA
>DOJHLOP01CY96H
ATCGGCATATTTCTGGCAACCTATTAAGAGTCTTGATTAATAATTTTCAAGCAAAAAA
AGTATCAGTAATCAAAGGTAATGTGGTCCATGTTGTTACATGTCAAAATTGCA
>DOJHLOP01AWB72
TATTTGATTTCAAACCAACA TCAAAATTGCGCCGAAAACCCGTGCCTTCAGAGCCTTGCCGATGAAAT
CCATCTTTTCCAACGTTATATTATTTCTCCCAACTGAAGTTGGTATGACAATT
;

```

```

Qual.5

>DOJHLOP02HSG8F
27 25 25 22 12 27 31 27 26 30 30 27 29 28 18 31 27 31 22 21 20 16 8 29 28 32 31 24 23 21 12 30 27 30 28 18 26 32 31 30
27 32 31 27 29 32 32 30 27 31 32 28 30 25 26 26 18 32 30 27 31 27 27 32 24 23 21 13 31
27 23 23 21 13 32 31 32 29 28 27 18 31 31 27 26 26 18 32 26 30 27 28 25 24 24 21 13 31 31 27 32 31 32 31 31 27
30 24 29 26 28 27 18 28 28 18 30 28 26 31 28 25 30 25 31 32 31
32 0
>DOJHLOP01AD4GJ
30 26 31 30 29 30 31 31 31 30 25 31 32 29 30 25 32 27 21 31 26 32 28 23 28 24 23 5 28 23 27 27 13 29 26 19 30 26 30 32
31 27 28 31 31 30 26 30 17 29 29 28 17 27 25 32 31 24 23 21 12 30 27 30 29 24 31 31 28 0
>DOJHLOP01CNMJ2
25 17 27 25 25 24 17 24 24 19 3 32 27 28 30 18 7 19 24 21 13 26 20 18 25 24 16 28 27 15 29 24 26 25 11 23 15 16 30 21
31 27 29 24 29 24 31 32 23 31 28 31 31 28 25 30 27 32 25 25
21 9 28 31 27 31 20 10 16 28 24 17 26 24 25 30 29 31 28 27 15 28 23 24 18 24 23 32 30 25 32 31 28 24 24 18 2 27 24 29
30 17 6 26 20 29 32 31 24 23 21 12 30 27 30 26 22 0
>DOJHLOP01CY96H
29 27 29 24 28 31 18 29 25 25 22 11 29 28 23 15 31 24 17 31 28 26 30 25 18 31 28 31 29 29 29 26 24 29 28 31 27
24 24 22 12 27 29 25 25 21 9 32 17 6 31 32 14 14 13 11 8
4 24 29 32 32 30 24 31 31 27 20 26 29 29 28 17 31 28 29 26 25 11 18 23 32 18 8 22 22 13 31 29 29 30 26 28 29 24 31 28
17 21 29 30 32 31 24 23 21 12 30 27 30 31 24 0
>DOJHLOP01AWB72
28 31 26 25 11 30 27 24 23 7 32 27 26 12 30 26 31 28 32 32 32 31 24 23 21 12 30 27 30 31 31 27 31 24 24 18 3 27 27 13
32 31 29 29 23 26 25 10 32 32 32 24 16 30 25 18 28 23 29 27 21 30 31 29 30 24 23 6 31
27 21 32 17 20 24 24 19 3 26 19 28 22 22 28 29 24 27 31 28 31 28 31 25 24 9 18 25 18 25 24 19 5 29 24 24 31 27 25 18 31
31 28 28 22 26 31 26 31 24 32 21 12 31 26 0
;

```

```

Podaj nazwe pliku fastq do odczytu: fasta5
Podaj nazwe pliku qual do odczytu: qual5
Podaj wartosc wiarygodnosci:
25
Podaj dlugosc podciagow w zakresie od 4 do 9:
5
Ilosc znalezionych klik: 1
Klika nr: 1
-----
Sekwencja kliku
-----
T C T T G
-----
Elementy kliku:
-----
Nazwa sekwencji: >DOJHLOP02HSG8F |Numer podciagu z sekwencji oryginalnej: 26
-----
Nazwa sekwencji: >DOJHLOP01AD4GJ |Numer podciagu z sekwencji oryginalnej: 25
-----
Nazwa sekwencji: >DOJHLOP01CNMJ2 |Numer podciagu z sekwencji oryginalnej: 107
-----
Nazwa sekwencji: >DOJHLOP01CY96H |Numer podciagu z sekwencji oryginalnej: 105
-----
Nazwa sekwencji: >DOJHLOP01AWB72 |Numer podciagu z sekwencji oryginalnej: 21
-----

```

8. Wnioski

Można zauważyć, że znalezione kliki pokrywają się z motywami obecnymi w sekwencjach. Co było dla mnie ciekawe nie zawsze podanie mniejszej wiarygodności i krótszej sekwencji podciągów skutkowało znalezieniem większej ilości kliku.

Program był bardzo ciekawy do napisania. Było to swego rodzaju namacalne wykorzystanie umiejętności pisania kodu do tworzenia programu do obróbki wyniku z sekwencjonowania (nawet jeśli miał to być tylko przykład i swego rodzaju wyzwanie-zabawa). Tak jak poprzednio największy problem sprawiło mi początkowe zrozumienie treści zadania. Lecz w momencie gdy tą treść na spokojnie rozumiałem i przeanalizowałem to napisanie funkcji nie stanowiło już większego problemu. Ponieważ w większości opierało się to na odpowiednim znalezieniu obiektu/wektora/elementu w wektorze a następnie edycji czy przekazaniu wartości.