

## Opis algorytmu

1. Algorytm znak po znaku wczytuje obie sekwencje do wektorów char i int, które znajdują się w struct. Algorytm zapisuje także do struct długość sekwencji oraz numer sekwencji, z której pochodzi ciąg.
2. Z konsoli zostaje sczytany próg, a następnie w pętli usuwane są indeksy pod którymi znajdują się nukleotydy o wiarygodności mniejszej niż zadany próg.
3. Następnie tworzony jest graf. Jest on zapisywany w zmiennej globalnej struct, który zawiera sekwencję w postaci wektora typu char, numer sekwencji, z której pochodzi, miejsce w sekwencji oraz wektor (tupu tego struct) który zapisuje wierzchołki z którymi jest połączenie krawędzią. Każda sekwencja jest dzielona na wszystkie możliwe podciągi (długości wskazanej przez użytkownika w konsoli), a następnie dodawana do grafu z przepisaniem wszystkich potrzebnych parametrów.
4. Algorytm sprawdza każdy wierzchołek z każdym i tworzy między nimi krawędź, jeśli mają taką samą sekwencję, są w różnych sekwencjach oraz miejsce występowanie w grafie nie różni się bardziej niż o dziesięciokrotność podanego podciągu
5. Wyszukiwanie klik w grafie w sposób heurystyczny
  - 1) Sortowanie wierzchołków malejąco według liczby krawędzi z którymi jest połączony
  - 2) Dodanie pierwszego elementu z grafu do klik
  - 3) Algorytm iteruje po wszystkich wierzchołkach i dodaje je do zbioru klik (taki sam typ zmiennej jak graf) jeżeli jest połączony z wszystkimi innymi wierzchołkami w klicie
6. Wyświetlenie wszystkich informacji o wierzchołkach w klicie – numeru sekwencji, z której pochodzi oraz miejsca w sekwencji, a także sekwencji podciągu

Złożoność algorytmu:  $O(n \log(n))$

## Wnioski:

Algorytm wyszukuje klikę w grafie w sposób heurystyczny w czasie wielomianowym. Algorytm działa bardzo szybko, jedyna sytuacja, gdy działa trochę wolniej pojawia się, kiedy graf jest duży i długość podciągów to 4, jednak wynika to z długości tworzenia krawędzi w grafie, a nie wyszukiwania klik. Gdy podciągi są krótsze to jest tworzona klica składająca się z większej ilości wierzchołków. Jeśli motyw ma niską wiarygodność to algorytm ma problem z jego znalezieniem. Przykład takiego zachowania można znaleźć w instancji 1 w teście 1.

## Testowanie

### Instancja 1:

>DOJHLOP01C1SGE length=101 xy=1133\_1724 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

CAAGCT**ACCTTGTGTCAAATGGAAG**TCAGGGCTTGCAACTGTGTAAACACTATGTTCAAACATAAGTAA  
CCTGTCTGCG

ACGCCTATCCCAATGCGCCACCAGAGCCACCTTCCCTATTA

>DOJHLOP01D6ZZX length=104 xy=1602\_3371 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

ATCGCAGTGAA**ACCTTGTGTCAAATGGAAG**TCTACGAGAAATGTTTAGGCTTAAAGTTCCAATAATTGC  
AACTGTAATA

GGGGAAGGTGGCTCTGGTGGCGCATTGGGGATAGGCGTCGCAGA

>DOJHLOP01CUSLY length=109 xy=1053\_3012 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

GGAG**ACCTTGTGTCAAATGGAAG**AGGCAATCGCAGTGAATCTACGAGAAATGTTTAGGCTTAAAGTTTC  
CAATAATTGCA

ACTGTAATAGGGGAAGGTGGCTCTGGTGGCGCATTGGGGATAGGCGTCG

>DOJHLOP01DBA3Z length=105 xy=1241\_3437 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

CCACCT**ACCTTGTGTCAAATGGAAG**TCCCCTATTACAGTTGCAATTATTGGAACTTTAAGCCTAAACATT  
TCTCGTAGA

TTCACCTGCGATTGCCTCTCCTTGCCTTGTTCTTCTGCTATGACG

>DOJHLOP01BGY5C length=99 xy=0486\_1118 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

TTCCC**ACCTTGTGTCAAATGGAAG**CTATTACAGTTGCAATTATTGGAACTTTAAGCCTAAACATTTCTCG  
TAGATTCAC

TGCGATTGCCTCTCCTTGTCTTGTCTTCTGCTATGAG

>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 1:

Próg: 12

Długość podciągów: 5

Wynik:

3, 15

2, 22

1, 17

5, 16

5, 40

4, 17

4, 46

sekwencja: TGGAA

### Test 2:

Próg: 20

Długość podciągów: 7

Wynik:

3, 10

5, 12

1, 13

2, 18

4, 13

sekwencja: CAAGGAG

### Test 3:

Próg: 10

Długość podciągów: 8

Wynik:

2, 18

1, 13

4, 13

3, 11

5, 12

sekwencja: TCAAATGG

Instancia 2:

>DOJHLOP02FU6TN length=110 xy=2288\_1721 region=2 run=R\_2005\_09\_08\_15\_35\_38\_  
AGAAGAACAAGGACAAGGAGAGGCAATCGCAGTGAATCTACGAGAAATTGTTT**ATTGCGTGGAATAG**  
GCTTA

AAGTTCCAATAATTGCAACTGTAAGTAGGGGAAGGTGGCTCTGGGTGGCG

>DOJHLOP02GKA8Z length=102 xy=2574\_2401 region=2 run=R\_2005\_09\_08\_15\_35\_38\_  
CATAGCAGAAGAACAAGGACAAGGAGAGGCAATCGCAGTGAATCTACGAG**ATTGCCTGGAATAAATG**  
TTTAG

GCTTAAAGTTCCAATAATTGCAACTGTAAGTAGGGGAGGTGG

>DOJHLOP01BGY5C length=99 xy=0486\_1118 region=1 run=R\_2005\_09\_08\_15\_35\_38\_  
TTCCCCTATTACAGTTGCAATTATTGGAACCTTAAGCCTAAACATTTCTCGTAGAT**ATTGCATGGTATTCA**  
C

TGCGATTGCCTCTCCTTGTCTTGTCTTCTGCTATGAG

>DOJHLOP01DQU8O length=99 xy=1419\_0278 region=1 run=R\_2005\_09\_08\_15\_35\_38\_  
GAGCTTATGCAGGGCTCATAGCAGAAGAACAAGGACAAGGAGAGGCAATCGCAGTGAAT**ATTGCGTG**  
**GAATC**

TACGAGAAATGTTTAGGCTTAAAGTTCCAATAATTGCAA

>DOJHLOP02IK9M4 length=96 xy=3405\_2430 region=2 run=R\_2005\_09\_08\_15\_35\_38\_  
GCCTAAACATTTCTCGTAGATTCACTGCGATTGCCTCTCCTTGTCTTGTCTT**ATTGCGTGGAATCTGCT**  
A

TGAGCCCTGCATAAGCTCCAGGAGTATCTATGAAAG

>DOJHLOP02FU6TN length=110 xy=2288\_1721 region=2 run=R\_2005\_09\_08\_15\_35\_38\_  
32 32 31 27 32 31 27 32 30 27 31 27 32 31 29 26 30 27 32 24 31 31 27 31 31 27 31 30 31 32  
31 31 31 32 27 21 32 31 32 29 32 20 18 25 23 22 3 18 7 18 29 28 18 **12 34 36 33 32 19 29 25**  
**26 39 35 28** 29 31 26 32 24 22 28

27 18 31 31 27 31 26 25 23 17 26 24 31 27 31 32 25 18 28 31 28 25 31 27 31 32 21 23 23 21  
13 23 14 27 24 23 30 25 27 28 27 24 23 22 3 22 29 26 21 32

>DOJHLOP02GKA8Z length=102 xy=2574\_2401 region=2 run=R\_2005\_09\_08\_15\_35\_38\_  
26 32 31 28 31 31 32 31 31 27 32 30 27 29 31 26 27 21 30 28 30 27 30 24 32 29 28 27 20 29  
26 20 32 28 31 25 32 31 31 22 29 24 29 28 31 17 25 28 17 25 **12 34 36 33 32 19 29 25 26 39**  
**35 28** 26 25 10 19 24 24 23 6 30 21

12 24 31 26 23 22 3 25 21 11 26 19 27 25 31 31 27 24 16 29 30 22 14 27 30 29 32 31 27 12 30  
27 25 24 20 7 19 26 19 32 23 14

>DOJHLOP01BGY5C length=99 xy=0486\_1118 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

29 24 24 24 18 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 **12 34 36 33 32**  
**19 29 25 26 39 35 28** 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

>DOJHLOP01DQU8O length=99 xy=1419\_0278 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

27 26 31 30 30 26 26 30 30 30 29 26 25 11 31 29 32 31 26 30 30 29 28 32 30 27 29 31 27 31  
31 28 26 20 30 26 31 28 30 26 31 31 31 27 21 31 30 25 27 31 31 29 30 29 27 28 28 22 31 **12**  
**34 36 33 32 19 29 25 26 39 35 28** 27

26 22 26 30 20 24 26 25 9 20 26 23 22 3 31 22 13 29 31 28 27 27 13 22 22 14 27 21 29 26 29  
27 25 21 11 30 31 28 22

>DOJHLOP02IK9M4 length=96 xy=3405\_2430 region=2 run=R\_2005\_09\_08\_15\_35\_38\_

18 26 19 24 24 23 6 31 32 29 28 17 31 25 32 25 30 32 31 23 27 21 31 31 28 32 28 31 27 30 24  
16 28 27 21 30 32 29 25 17 26 19 28 31 26 20 31 27 31 30 24 31 30 24 **12 34 36 33 32 19 29**  
**25 26 39 35 28** 28 30 28 32 31 26

27 32 27 21 27 27 13 32 31 27 23 21 31 27 25 19 32 29 24 31 29 26 30 18 31 30 31 29 32 21  
27 21 27 27 18 27

#### Test 1:

Próg: 25

Długość podciągów: 4

Wynik:

4, 56

2, 43

1, 48

3, 45

3, 83

5, 62

5, 52

5, 42

5, 23

sekwencja: TGCT

Test 2:

Próg: 11

Długość podciągów: 4

Wynik:

4, 59

5, 53

1, 51

2, 50

1, 81

3, 75

2, 85

3, 55

5, 28

3, 21

sekwencja: ATTG

Test 3:

Próg: 22

Długość podciągów: 9

Wynik:

2, 45

1, 49

5, 44

4, 56

sekwencja: TTGCTGGAA

Instancia 3:

>DOJHLOP01C6EES length=99 xy=1185\_3810 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

AAGCTCCAGGAGTATCTATGAAAGAAATTATTGGAAGATTAAAGCGATCGGCATGGTCCA

TCAGCCTCAAAGCCTTTCT**GGTCTCAATGCTAG**GTAGCCTCCAGGTTTGGCCA

>DOJHLOP01DZTX4 length=95 xy=1521\_0714 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

TTGGCATGGCAAAACCTGGAGGCTACAGAAAGGCTTTGAGGCTGATGGACCATGCCGATC

GCTTTAATCTTCCAATAATTTCTTTCAT**GGTCTCATTGCTAG**AGATACT

>DOJHLOP01BP4NK length=113 xy=0590\_2174 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

CATGGTCCATCAGCCTCAAAGCCTTTCTGTAGCCTCCAGGTTTGGCATGCCAAAGTTTC

TTGCAACATTCTCTTTTGTATCTCTTACCTTTTTTGTGTCCAAATTAAA**GGTCTCAATGCTAGGG**

>DOJHLOP01ANIOR length=122 xy=0150\_3017 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

GGAGGGTTAGCTCGAATAGGTGAAAAATCCGTCCTTTTAATTGGACAACAAAAAGGTAGA

GATAACAAAAGAGAATGTTGCAAGAACTTTGGCATGGACAAAACCTGGAGGCTACAG**GGCCTCAAT**  
**GCTAGAA**

GG

>DOJHLOP02IMA9L length=114 xy=3417\_2039 region=2 run=R\_2005\_09\_08\_15\_35\_38\_

GGAGACAGGAACGGAAGTATGATCAAGCTCTAATAGGAGGGTTAGCTCGAATAGGTGAA

AATCCGTCCTTTTAATTGGACAACAAAAAGGTAGAGATACAAAAG**GGTCTCAAAGCTAG**AGAATGTTG

>DOJHLOP01C6EES length=99 xy=1185\_3810 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

27 21 31 31 31 29 23 28 29 24 26 30 32 32 31 31 31 31 29 32 23 22 5 30 23 22 2 30 26 30 24

17 31 26 21 13 31 20 30 25 29 28 16 32 29 31 24 31 28 31 26 30 29 31 28 22 31 29 23 31

29 31 28 32 31 27 32 32 23 22 3 30 30 26 29 28 18 31 31 **34 29 34 36 23 27 34 12 14 26 38 11**

**19 29** 28 29 28 24 31 28 27 27 21 30 31 28 25 25 20 6 28 28 22 31

>DOJHLOP01DZTX4 length=95 xy=1521\_0714 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

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

15 31 26 29 29 28 16 25 24 31 27 31 32 32 30 29 31 28 31 31 28 31 28 29 31 27 31 31 28 31

30 31 29 28 17 31 28 31 28 28 23 19 9 30 25 27 29 24 28 28 19 31 28 27 14 26 31 29 **34 29 34**

**36 23 27 34 12 14 26 38 11 19 29** 32 31 32 31 32 21 31

>DOJHLOP01BP4NK length=113 xy=0590\_2174 region=1 run=R\_2005\_09\_08\_15\_35\_38\_



31 29 29 31 28 32 31 28 31 29 29 31 32 30 27 31 30 28 27 15 32 31 28 29 28 18 32 32 31 31  
31 32 29 26 31 31 28 25 27 25 24 24 22 12 31 28 25 29 31 31 31 27 28 27 19 31 28 28 19 31  
  
28 22 32 29 30 26 31 29 27 25 31 18 31 24 24 21 12 30 31 31 32 28 31 28 23 16 31 30 25 20  
20 19 16 9 1 31 28 23 31 29 30 26 23 22 2 31 28 24 23 21 12 **34 29 34 36 23 27 34 12 14 26**  
**38 11 19 29** 25 23

>DOJHLOP01ANIOR length=122 xy=0150\_3017 region=1 run=R\_2005\_09\_08\_15\_35\_38\_

24 9 15 24 23 6 21 12 31 31 29 28 28 31 27 21 31 24 31 28 32 31 22 22 20 13 2 31 31 28 31  
21 29 23 26 25 22 10 27 21 31 27 27 21 31 32 25 23 30 22 22 21 16 7 29 24 29 31 31 27

30 31 28 17 6 31 26 25 22 9 25 29 23 19 9 27 31 18 8 29 26 31 28 31 23 22 3 32 29 28 17 31  
26 32 32 31 30 27 21 18 26 25 22 11 29 24 28 30 28 31 25 18 32 24 31 24 28 27 **34 29 34 36**  
**23 27 34 12 14 26 38 11 19 29** 24 22 28 23

>DOJHLOP02IMA9L length=114 xy=3417\_2039 region=2 run=R\_2005\_09\_08\_15\_35\_38\_

28 18 26 29 28 31 28 31 27 30 25 32 30 27 31 26 32 30 32 30 27 32 31 31 31 31 27 32 32 30  
31 31 31 27 27 32 28 26 25 29 28 17 31 27 29 31 31 27 31 31 30 26 32 32 31 27 24 30 23 23  
  
21 13 32 31 27 31 27 28 21 25 25 22 10 30 25 30 24 31 27 29 32 25 23 28 21 21 20 16 9 30 27  
28 31 32 32 30 32 29 18 28 25 25 22 12 32 **34 29 34 36 23 27 34 12 14 26 38 11 19 29** 31 25  
29 23 32 27 30 24 30

#### Test 1:

Próg: 29

Długość podciągów: 4

Wynik:

5, 16

1, 0

1, 47

3, 9

2, 20

2, 12

3, 14

3, 23

2, 25

sekwencja: AGCT

Test 2:

Próg: 22

Długość podciągów: 6

Wynik:

2, 81

1, 71

3, 99

5, 96

sekwencja: GTCTCA

Test 3:

Próg: 19

Długość podciągów: 7

Wynik:

2, 85

3, 104

1, 75

4, 106

sekwencja: CTCATGC