

Wstęp do bioinformatyki

Laboratorium 3

Dopasowanie lokalne par sekwencji

Magdalena Trędak

2367121

1. Schemat blokowy algorytmu dopasowania lokalnego:

Schematy blokowe algorytmów tworzenia macierzy punktowej oraz optymalnej ścieżki dopasowania ze względu na duże rozmiary i umożliwienia poprawy ich czytelności zamieszczono w repozytorium jako pliki graficzne o nazwach: SchematBlokowyGenerowaniaMacierzyPunktów.jpg, SchematBlokowyGenerowaniaŚcieżkiDopasowania.jpg. Do wygenerowania schematów użyto programu online znajdującego się na stronie: <http://www.algorytm.org/narzedzia/edytor-schematow-blokowych.html> (data dostępu 16.05.19)

2. Analiza złożoności programu

a) czasowa

scoringMatrix – 2 pętle $\text{for}(m*n)$, 2 pętle $\text{for}(g*f)$, gdzie g i f to rozmiary macierzy substytucji, znacznie mniejsze od długości sekwencji, 1 warunek if , 5 przypisań wartości w pętlach, 15 poza nimi $O(mn)$ – złożoność czasowa co najwyżej rzędu mn

tracBackMatrix - k -razy pętla for , k to ilość maksimów macierzy punktów > 0 , pętla while – w skrajnym przypadku maksymalny element znajduje się w końcu macierzy punktów a ostatnie 0 na jej początku, wtedy wyszukanie ścieżki odbywa się po całej macierzy $(n+m)$ – rząd co najwyżej $O(m+n)$, 9 porównań, 9 przypisań. Poza pętlą while 11 przypisań. $O(m+n)$ – złożoność czasowa co najwyżej rzędu $m+n$

b) pamięciowa

Macierz punktów jest macierzą o wymiarach $(n * m)$ i tyle też zajmuje miejsca w pamięci – rząd $m*n$. Pozostałe tworzone macierze są takich samych rozmiarów lub mniejsze. Przypisania poszczególnych zmiennych lub wektorów są znacznie mniejsze niż rozmiarów $n*m$. Pozwala to przyjąć założenie, że macierz punktów (scoringMatrix) i macierz ścieżki optymalnego dopasowania (tracBackMatrix) są największymi obiektami, dlatego złożoność pamięciowa programu $O(mn)$ – co najwyżej rzędu mn .

3. Porównanie przykładowych par sekwencji mitochondrialnego cytochromu b dla parametrów : match , mismatch wczytanych z macierzy substytucji, gap = 2

a) niepowiązanych ewolucyjnie

AJ009879.1- Nubian ibex

AY819740.1- African clawed frog

Command Window

```
>> TesterZad3
Warning: DimensionNames property was modified to make the name '#' a
Warning: DimensionNames property was modified to make the name '#' a
Warning: DimensionNames property was modified to make the name '#' a
>seq1 823-918
>seq2 1-96
#Gap -2
#Score: 42
#Length: 121
#Gaps: 50/121 (41.3223 %)
#Identity: 16/121 (13.2231 %)
#seq1 CCGCCCAATCAGCCAATGCATATTCTGAAT-CT--TGG-CAG--CA-GAT-CTACTAACA
#seq2 C-CCC--TCAG--AATG-ATATT-TG--TCCTCATGGT-AGAACAT-ATCCTAC-AA-AT
      *  **          *          *          *  *          *          *

#seq1 --C--TCA-C-ATGAA-T-----T-GGAGGA-CAGC-CAGTCGAA-CATCCTTACAT-TA
#seq2 GCTGT-AGCTAT-AACTAAAAATAGGAGGATCA-CAC---C-AAT-AT--TT-CATGT-T
              *          *  *          *  *          *

#seq1 T
#seq2 T
      *
```

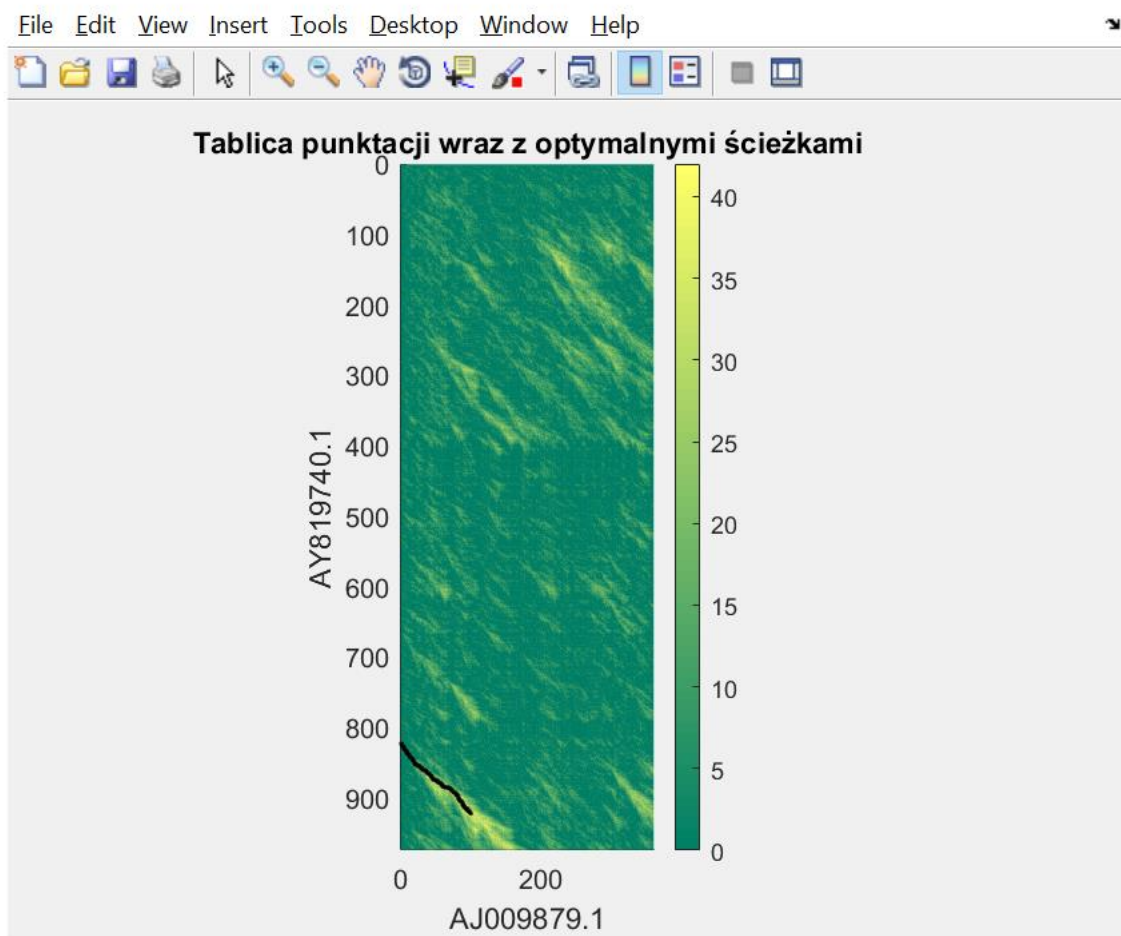
fx

Command Window

```
>seq1 823-921
>seq2 1-99
#Gap -2
#Score: 42
#Length: 125
#Gaps: 52/125 (41.6 %)
#Identity: 18/125 (14.4 %)
#seq1 CCGCCCAATCAGCCAATGCATATTCTGAAT-CT--TGG-CAG--CA-GAT-CTACTAACA
#seq2 C-CCC--TCAG--AATG-ATATT-TG--TCCTCATGGT-AGAACAT-ATCCTAC-AA-AT
      *  **          *          *          *  *          *          *

#seq1 --C--TCA-C-ATGAA-T-----T-GGAGGA-CAGC-CAGTCGAA-CATCCTTACAT-TA
#seq2 GCTGT-AGCTAT-AACTAAAAATAGGAGGATCA-CAC---C-AAT-AT--TT-CATGT-T
              *          *  *          *  *          *

#seq1 TT-AT
#seq2 TC-TT
      *  *  *
```



b) powiązanych ewolucyjnie

X75584.1 - humpback whale

X75583.1 - pygmy Bryde's whale

Command Window

>> TesterZad3

Warning: DimensionNames property was modified to make the name '#' a valid

Warning: DimensionNames property was modified to make the name '#' a valid

Warning: DimensionNames property was modified to make the name '#' a valid

>seq1 2-1140

>seq2 2-1140

#Gap -2

#Score: 1782

#Length: 1223

#Gaps: 166/1223 (13.5732 %)

#Identity: 1057/1223 (86.4268 %)

#seq1 ATGACCAACATCCGAAAAACACCCCACTAAT-AAAGATT-GTCAACG--ATGCATTC-G

#seq2 ATGACCAACATCCGAAAAACACCCCACTAATAAAA-ATTA-TCAACGACA--CATTCA-

***** *** ** ***** *

#seq1 TTGATCT-CCCCACCCCATCAAATATCTCCTCATGATGAAA-TTTCGG-CTCCCTACTCG

#seq2 TTGATCTA-CCCACCCCATCAAATATCTCCTCATGATGAAAC-TTCGGT-TCCCTACTCG

***** ***** ***** *****

#seq1 GCC-TCTGCTTAATTA-CACAAATCCTAACAGGCCTATTCCTAGCAATACACTACACACC

#seq2 GCCTT-TGCTTAATTAT-ACAAATCCTAACAGGCCTATTCCTAGCAATACACTACACACC

*** * ***** *****

fx

Command Window

#seq1 AGACACAACAACCGCCTTCTCATCAGT--TGACACAT-T-TGCCGAGACGTAAA-CTA-

#seq2 AGACACAACAACCGCCTTCTCATCAGTCA--CACACATCTGT--CGAGACGTAAAT-TAT

***** ***** *

#seq1 CGGCTG-AGTTATCCGATACCTACA-CGCAAA--CGGAGCCTCCATATTCTTCATCTG-T

#seq2 -GGCTGAA-TTATCCGATACCTACAT-GCAAAATG-GG-GCCTCCATATTCTTCATCTGC-

***** * ***** ***** ** *****

#seq1 CTCTACGCTCACATAGGACGAGGCCTATACTACGGCTCCTA-TGCCTTTCGAGAAACATG

#seq2 CTCTACGCTCACATAGGACGAGGCCTATACTACGGCTCCTAC-GCCTTTCGAGAAACATG

***** *****

#seq1 AAACATCGGAGTTA-TCCTACTATTACAGTTATAGCCAC-CGCATTG-ATAGGCTACGT

#seq2 AAACATCGGAGTTATT-CTACTATTACAGTTATAGCCACT-GCATTG-TAGGCTACGT

***** * ***** ***** *****

#seq1 CCTACCCTGAGGACAAATATCATT-TTGAGGCGCAACCGTCATCACCACCC-TCTTATCA

#seq2 CCTACCCTGAGGACAAATATCATTG-TGAGGCGCAACCGTCATCACCACCC-TATCA

***** ***** ** *****

Command Window

```
#seq1 GCAATCCCATACATTGGTACTACCCTAGTCGAATGAATCTGGGGCGG--TTTCTCTGTAG
#seq2 GCAATCCCATACATTGGTACTACCCTAGTCGAATGAATCTGGGGCGGTTTTTC-C-GTAG
*****

#seq1 A-TAAAGCAACACTAACACGCTTT-TTTGC-CTTCCACTT-TATCCTCCCCCTTCAT-TAT
#seq2 AC-AAAGCAACACTAACACG-TTTCTTTGCT-TTCCACTTC-ATCCTCCCCCTTCATC-AT
* *****

#seq1 T-CTAGCA-CTAGCAAT-GGTCCACCTCATTTTCCTCCACGAAACAGGATCCAA-TAACC
#seq2 TAC-AGCAT-TAGCAATC-GTCCACCTCATTTTCCTCCACGAAACAGGATCCAAC-AACC
* * *****

#seq1 CCACAGG-TAT-TCCATCCAACATAGACAAAATCCCATTCCACCC-CTA-TTACACAA-C
#seq2 CCACAGGC-ATC-CCATCCAACATAGACAAAATCCCATTCCACCCT-TAC-TACACAATC
***** **

#seq1 TAAAGACA-TTCTAGGCGCCCTA-CTACTAATCCTAACCCTACTAATG-CTAACCCTATT
#seq2 -AAAGACAC-TCTAGGCGCCCTAT-TACTAATCCTAACCCTACTAATGT-TAACCCTATT
*****
```

Command Window

```
#seq1 AGC-CTT-AATCCCAATACTCCACACAT-CTAAACAACGAAGCAT-AATG-TTCCGACCC
#seq2 AGCT-TTC-ATCCCAATACTCCACACATCC-AAACAACGAAGCATG-ATGTTT-CGACCC
      *** ** ***** * *****

#seq1 TTTAGCCA--ATTCTT-ATTTTG-AGT-CCTA--ATTGCAGA-CT-TACTA-ACCCT-GA
#seq2 TTTAGCCAGT-TT-CTG-TTTTGAA-TA-CTAGTA--GCAGACCTAT--TAG-CCCTA-A
      ***** ** ** ***** * * *** * ***** ** * ** ***** *

#seq1 CATGAATCGGCGGCCAACCCGTAGAACACCC-CTAC-GTAATCGTAGGCCAA-TTCGCAT
#seq2 CATGAATCGGCGGCCAACCCGTAGAACACCCA-TACA-TAATCGTAGGCCAAC-TCGCAT
      ***** *** ***** *****

#seq1 CCATCCTCTA-TTTCCTC-CTAAT-TCTAGTA-CTAATACC-AGTAACTAGTCTTATCGA
#seq2 CCATCCTCTAC-TTCCTCT-TAATC-CTAGTAT-TAATACCAA-TAACTAGTCTTATCGA
      ***** ***** **** ***** ***** * *****

#seq1 GAA-TAAACTTATAAAATGAAGA
#seq2 GAAC-AAACTTATAAAATGAAGA
      *** *****
```

File Edit View Insert Tools Desktop Window Help

