Laboratorium ze Wstępu do bioinformatyki

Typ dokumentu:	Zadanie 3	Data oddania zadania: 15.04.2019
Tytuł:	Dopasowanie lokalne par sekwencji	

Autor: Aleksandra Brela, 229636

1. Odnośnik do repozytorium z kodem źródłowym:

   https://github.com/Abrela/LocalAlignment

2. Schemat blokowy algorytmu dopasowania lokalnego
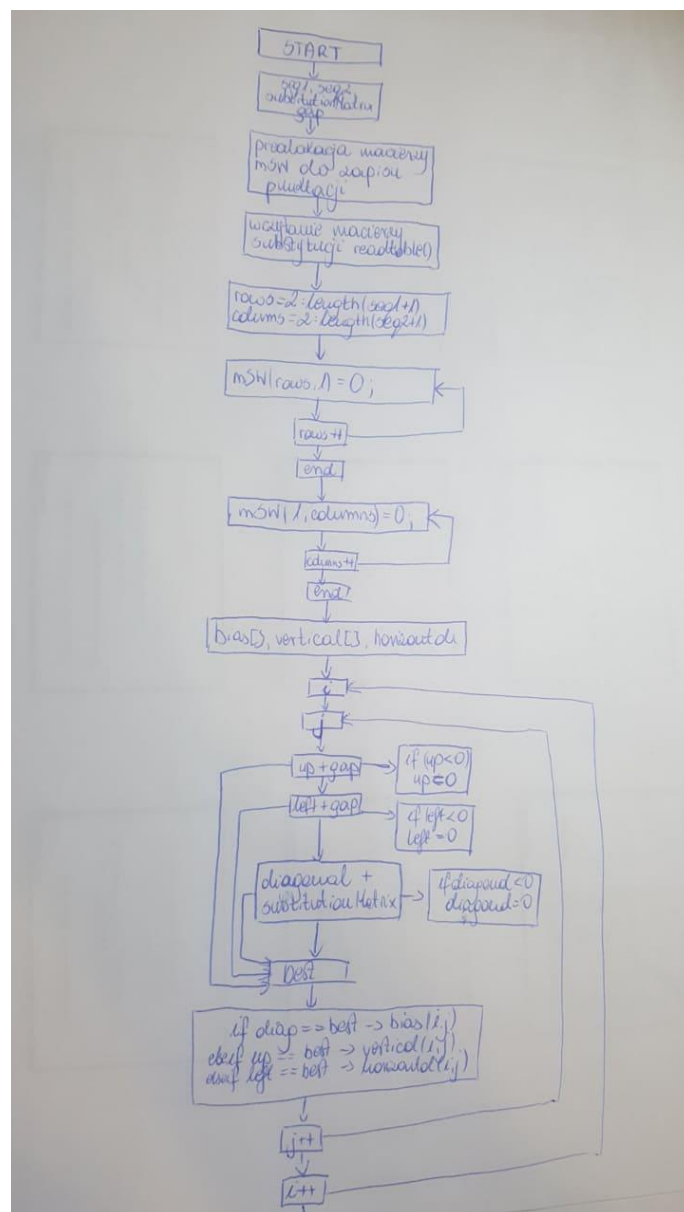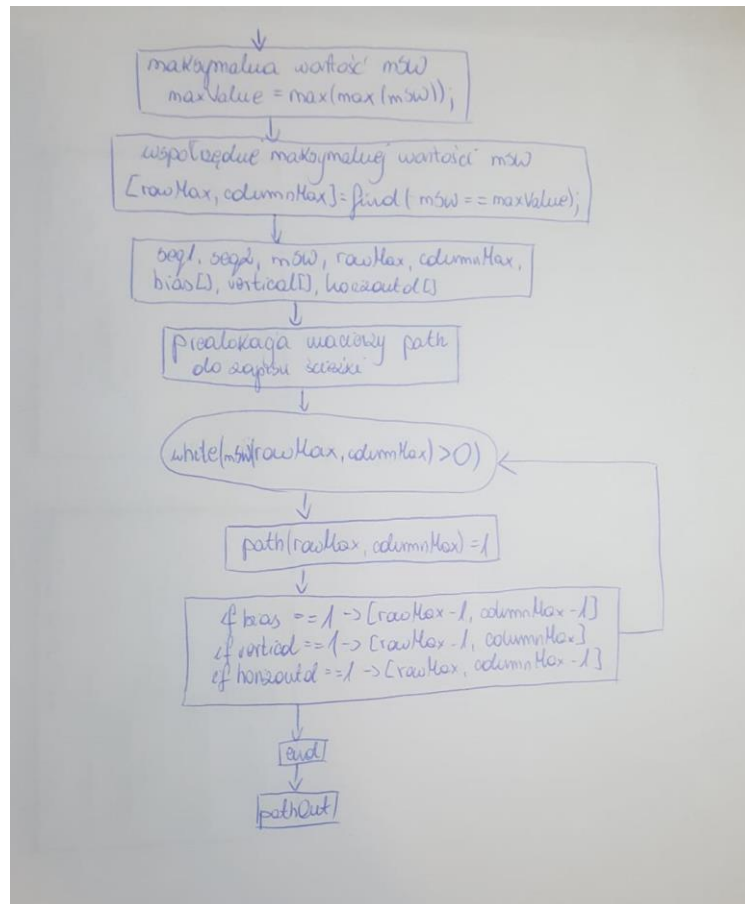
3. Analiza złożoności obliczeniowej algorytmu –
    a. Funkcja **matrixScore():**

```
1:function [ mSWout, coordinateRowMax, coordinateColumnMax, maxValueOut, bias, vertical, horizontal,
2:sOut ] = matrixScore( seq1, seq2, subtitutionMatrix, gap )
3:N1 = length(seq1); przypisanie
4:N2 = length(seq2); przypisanie
5: mSW = zeros( N1+1, N2+1 ); prealokacja, przypisanie
6:[rows, columns]=size(mSW); przypisanie
7:s=readtable(subtitutionMatrix, 'ReadRowNames', true); przypisanie
8:mSW(1,1) = 0; przypisanie
9:for i=2 to (N1+1) #(i=2, i<= (N1+1), i++) inkrementacja, iteracje
10:mSW(i,1) = 0; przypisanie
11:end
12:for  j=2 to (N2+1) #(j=2, j<= (N2+1), j++) inkrementacja, iteracje
13:mSW(1,j) = 0; przypisanie
14:end
15:bias = zeros(size(mSW,1), size(mSW,2)); przypisanie
16:vertical = zeros(size(mSW,1), size(mSW,2)); przypisanie
17:vertical(2:size(vertical,1),1) = ones(size(vertical,1)-1,1);  przypisanie
18:horizontal = zeros(size(mSW,1), size(mSW,2)); przypisanie
19:horizontal(1,2:size(horizontal,2)) = ones(1,size(horizontal,2)-1); przypisanie
20:currentRow = mSW(1,:); przypisanie
21:for i=2 to rows#(i=2, i<=rows, i++) inkrementacja, iteracje
22:lastRow = currentRow; przypisanie
23:currentRow = mSW(i,:);  przypisanie
24:best=currentRow(1); przypisanie
```

```
25:for j=2 to columns#(j=2, j<=columns, j++) inkrementacja, iteracje
26:up = lastRow(j) + gap;  przypisanie
27:if (up<0) warunek, porównanie
28:up = 0; przypisanie
29:end
30:left = best + gap;  przypisanie, inkrementacja
31:if (left<0) warunek, porównanie
32:left = 0; przypisanie
33:end
34:if seq1(i-1) == seq2(j-1) warunek, porównanie
35:diagonal = lastRow(j-1) + s{seq1(i-1),seq2(j-1)}; przypisanie, inkrementacja
36:else
37:diagonal = lastRow(j-1) + s{seq1(i-1),seq2(j-1)}; przypisanie, inkrementacja
38:end
39:if (diagonal<0) warunek, porównanie
40:diagonal = 0; przypisanie
41:end
42:MAX = max([up, left, diagonal]); przypisanie
43:if(MAX == diagonal) warunek, porównanie
44:bias(i, j) = 1; przypisanie
45:end
46:if(MAX == up) warunek, porównanie
47:vertical(i, j) = 1; przypisanie
48:end
49:if(MAX == left) warunek, porównanie
50:horizontal(i, j) = 1; przypisanie
51:end
52:if up > left warunek, porównanie
53:best = up; przypisanie
54:else
55:best = left; przypisanie
56:end
57:if diagonal >= best warunek, porównanie
58:best = diagonal; przypisanie
59:end
60:currentRow(j) = best; przypisanie
61:end
62:mSW(i,:) = currentRow; przypisanie
63:end
64:mSWout = mSW; przypisanie
65:sOut = s; przypisanie
66:maxValue = max(max(mSW)); przypisanie
67:[coordinateRowMax,coordinateColumnMax] = find(mSW == maxValue); przypisanie
68:maxValueOut = maxValue; przypisanie
69:end
```

b. Funkcja **matchPath():**

```
1:function [ pathOut, lengthAlignmentOut, identityOut, gapsOut, s1Out, s2Out, last ] = ...
2:matchPath(seq1, seq2, mSW, rowMax, columnMax, bias, vertical, horizontal )
3:path = zeros(size(mSW)); prealokacja, przypisanie
4:last = [rowMax, columnMax]; przypisanie
5:gaps = 0; przypisanie
6:identity = 0; przypisanie
7:s1 = ''; przypisanie
8:s2 = ''; przypisanie
9:while (mSW(last(1),last(2)) > 0) warunek, porównanie
10:path(last(1),last(2)) = 1; przypisanie
11:diagonal = bias(last(1),last(2)); przypisanie
```

```
12:left = vertical(last(1),last(2)); przypisanie
13:up = horizontal(last(1),last(2)); przypisanie
14:if(diagonal == 1)  warunek, porównanie
15:last = [last(1)-1,last(2)-1]; przypisanie, inkrementacja
16:identity = identity+1; przypisanie, inkrementacja
17:s1 = strcat(s1, seq1(last(1))); przypisanie, inkrementacja
18:s2 = strcat(s2, seq2(last(2))); przypisanie, inkrementacja
19:else
20:if(left == 1 || up == 0) warunek, porównanie
21:last = [last(1)-1,last(2)]; przypisanie, inkrementacja
22:gaps = gaps+1; przypisanie, inkrementacja
23:s1 = strcat(s1, seq1(last(1))); przypisanie, inkrementacja
24:s2 = strcat(s2, '-'); przypisanie, inkrementacja
25:end
26:if(left == 0 || up == 1) warunek, porównanie
27:last = [last(1),last(2)-1]; przypisanie, inkrementacja
28:gaps = gaps+1; przypisanie, inkrementacja
29:s1 = strcat(s1, '-'); przypisanie, inkrementacja
30:s2 = strcat(s2, seq2(last(2))); przypisanie, inkrementacja
31:end
32:end
33:end
34:pathOut = path; przypisanie
35:identityOut = identity; przypisanie
36:gapsOut = gaps; przypisanie
37:s1Out = fliplr(s1); przypisanie
38:s2Out = fliplr(s2); przypisanie
39:[r, c] = size(path); przypisanie
40:lengthAlignment = 0; przypisanie
41:for m = 1:r inkrementacja, iteracje
42:for n = 1:c inkrementacja, iteracje
43:if (path(m,n) == 1) warunek, porównanie
44:lengthAlignment = lengthAlignment+1; przypisanie, inkrementacja
45:end
46:end
47:end
48:lengthAlignmentOut = lengthAlignment;  przypisanie
49:end
```

4. Porównanie przykładowych par sekwencji:

    a. ewolucyjnie powiązanych

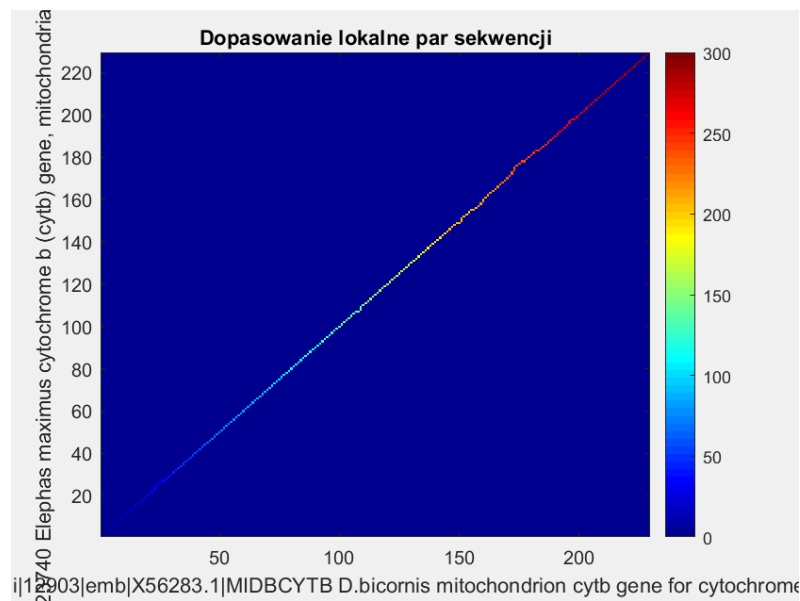### Rhino – Mastodon

```
# 1: GAAATTTTGGCTCTCTACTAGGAATCTGCCTAATCCTACAAATCCTAACCGGACTATTTCTTGCTATACA
# 2: GAAATTTCGGCTCACTACTAGGAGCATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# Mode: distance
# Subtitution matrix:
     2    -7    -5    -7
    -7     2    -7    -5
    -5    -7     2    -7
    -7    -5    -7     2
# Gap: -1
# Score: 296
# Length: 233
# Identity: 184/233 ( 79%)
# Gaps: 88/233 ( 38%)
GAAATTT-TGGCTC-TCTACTAGGA-A-T-CTGCCTAAT-C-C-T-ACAAATCCTAAC-CGGA-CTATT-TCT-T
|||||||__|||||__|||||||||_____|||||||_____||||||||||__|||__||||__||__
GAAATTTC-GGCTCA-CTACTAGGAG-C-A-TGCCTAATT-A-C-C-CAAATCCTAACA-GGAT-TATTC-CTA-
```

Dopasowanie lokalne par sekwencji

Rhino – Asiatic Elephant

```
# 1: GAAATTTTGGCTCTCTACTAGGAATCTGCCTAATCCTACAAATCCTAACCGGACTATTTCTTGCTATACA
# 2: GAAATTTCGGCTCACTACTAGGAGCGTGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# Mode: distance
# Subtitution matrix:
      2    -7    -5    -7
     -7     2    -7    -5
     -5    -7     2    -7
     -7    -5    -7     2
# Gap: -1
# Score: 300
# Length: 235
# Identity: 185/235 ( 79%)
# Gaps: 86/235 ( 37%)
GAAATTT-TGGCTC-TCTACTAGGAA-TC-TGCCTAAT-C-C-T-ACAAATCCTAAC-CGGA-CTATT-TCT-TG
|||||||__|||||__|||||||||_|__|_|||||||||_____||||||||||__|||__||||__||__|
GAAATTTC-GGCTCA-CTACTAGG-AG-CGTGCCTAATT-A-C-C-CAAATCCTAACA-GGAT-TATTC-CTA-G
```

Dopasowanie lokalne par sekwencji

Y-axis: Elephas maximus cytochrome b (cytb) gene, mitochondria (U23740)

X-axis: i|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene for cytochrome

Mastodon – Asiatic Elephant

```
# 1: GAAATTTCGGCTCACTACTAGGAGCATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACAT
# 2: GAAATTTCGGCTCACTACTAGGAGCGTGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACAT
# Mode: distance
# Subtitution matrix:
     2    -7    -5    -7
    -7     2    -7    -5
    -5    -7     2    -7
    -7    -5    -7     2
# Gap: -1
# Score: 400
# Length: 228
# Identity: 214/228 ( 94%)
# Gaps: 28/228 ( 12%)
GAAATTTCGGCTCACTACTAGGAGC-ATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACATTA-
||||||||||||||||||||||||||__||||||||||||||||||||||||||||||||||||||||||||||||__
GAAATTTCGGCTCACTACTAGGAGCG-TGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACATTAC-

>>NorthAmericanMastodon gi|924716|gb|U23737.1|MAU23737 Mammut americanum cy
GAAATTTCGGCTCACTACTAGGAGC-ATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACATTA-

>>AsiaticElephant gi|924712|gb|U23740.1|EMU23740 Elephas maximus cytochrome
GAAATTTCGGCTCACTACTAGGAGCG-TGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACATTAC
```

Dopasowanie lokalne par sekwencji
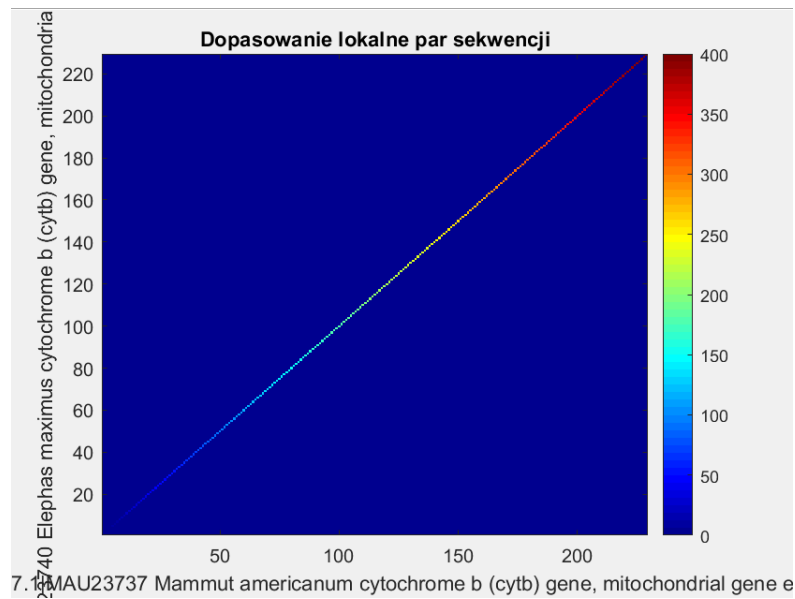
b. ewolucyjnie niepowiązanych

Human – Mastodon

```
# 1: GAAATTTCGGCTCACTACTAGGAGCATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2     -7     -5     -7
     -7      2     -7     -5
     -5     -7      2     -7
     -7     -5     -7      2
# Gap: -1
# Score: 248
# Length: 236
# Identity: 174/236 ( 74%)
# Gaps: 104/236 ( 44%)
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG
||||__||||||||||||__||__||__||__|||||__||_____|||||||_|___||||||__||||||||
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG

# 1: GAAATTTCGGCTCACTACTAGGAGCATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2     -7     -5     -7
     -7      2     -7     -5
```

```
        -5    -7     2    -7
        -7    -5    -7     2
# Gap: -1
# Score: 248
# Length: 238
# Identity: 175/238 ( 74%)
# Gaps: 106/238 ( 45%)
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG
||||__||||||||||||__||__||__||__|||||__||_____|||||||_|___||||||__||||||||
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG

# 1: GAAATTTCGGCTCACTACTAGGAGCATGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance |
# Subtitution matrix:
        2    -7    -5    -7
       -7     2    -7    -5
       -5    -7     2    -7
       -7    -5    -7     2
# Gap: -1
# Score: 248
# Length: 242
# Identity: 176/242 ( 73%)
# Gaps: 108/242 ( 45%)
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG
||||__||||||||||||__||__||__||__|||||__||_____|||||||_|___||||||__||||||||
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG


>>NorthAmericanMastodon gi|924716|gb|U23737.1|MAU23737 Mammut americanum cy
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG

>>human cytochrome b
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG

>>NorthAmericanMastodon gi|924716|gb|U23737.1|MAU23737 Mammut americanum cy
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG

>>human cytochrome b
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG

>>NorthAmericanMastodon gi|924716|gb|U23737.1|MAU23737 Mammut americanum cy
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-ATGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG

>>human cytochrome b
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG
```
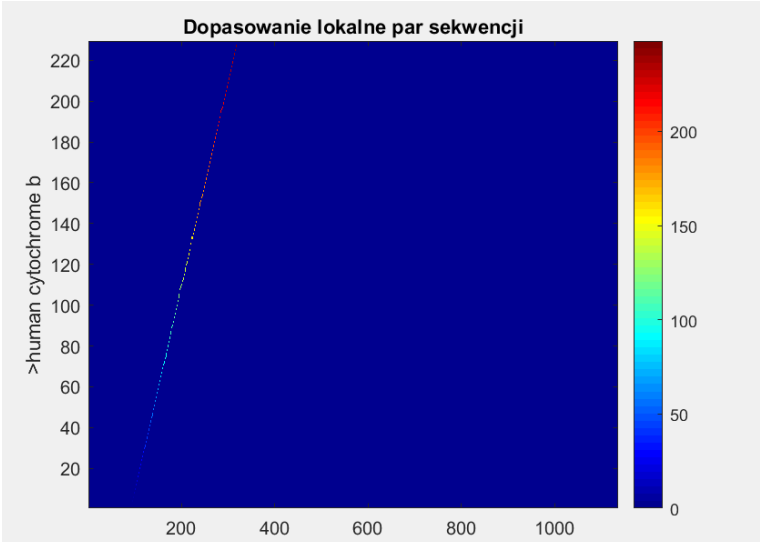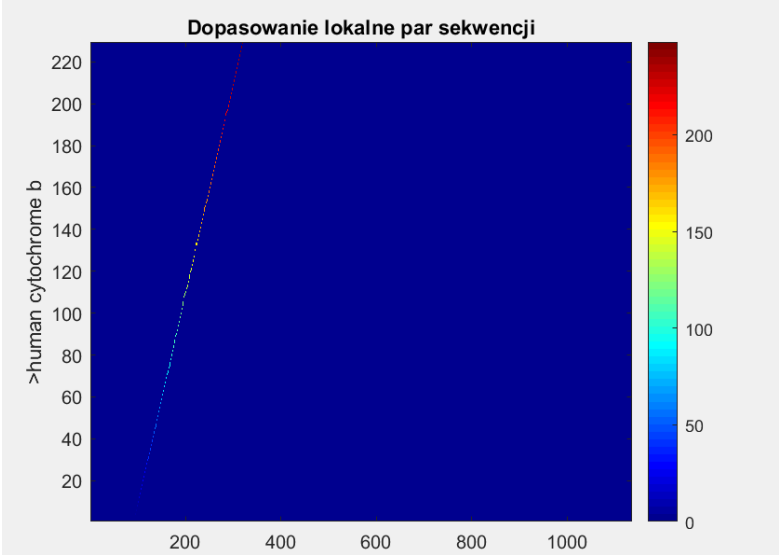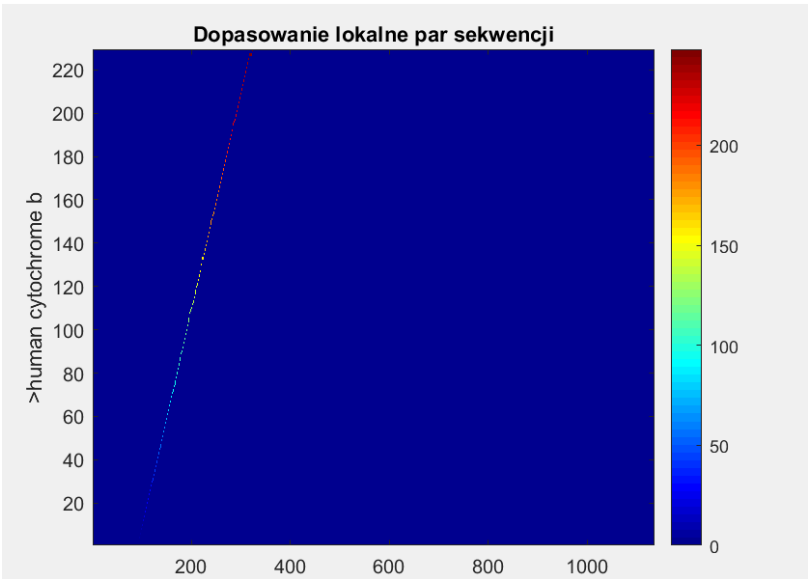
Dopasowanie lokalne par sekwencji

7.1|MAU23737 Mammut americanum cytochrome b (cytb) gene, mitochondrial gene e

Dopasowanie lokalne par sekwencji

7.1|MAU23737 Mammut americanum cytochrome b (cytb) gene, mitochondrial gene e


Dopasowanie lokalne par sekwencji

7.1|MAU23737 Mammut americanum cytochrome b (cytb) gene, mitochondrial gene e

Human – Asiatic Elephant

```
# 1: GAAATTTCGGCTCACTACTAGGAGCGTGCCTAATTACCCAAATCCTAACAGGATTATTCCTAGCCATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2    -7    -5    -7
     -7     2    -7    -5
     -5    -7     2    -7
     -7    -5    -7     2
# Gap: -1
# Score: 250
# Length: 253
# Identity: 178/253 ( 70%)
# Gaps: 118/253 ( 47%)
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-GTGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG
||||__||||||||||||__||__||__||__|||||__||_____||||||||_|___||||||__||||||||||
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG


>>AsiaticElephant gi|924712|gb|U23740.1|EMU23740 Elephas maximus cytochrome
GAAA-TTTCGGCTCACT-ACT-AGG-AGC-GTGCCT-AAT-T-A-CCCAAATC-CT-AACAGGA-TTATTCCTAG

>>human cytochrome b
GAAAC-TTCGGCTCACTC-CTT-GGC-GCC-TGCCTG-ATC-C-T-CCAAATCAC-C-ACAGGAC-TATTCCTAG
```


Dopasowanie lokalne par sekwencji

|EMU23740 Elephas maximus cytochrome b (cytb) gene, mitochondrial gene encoding
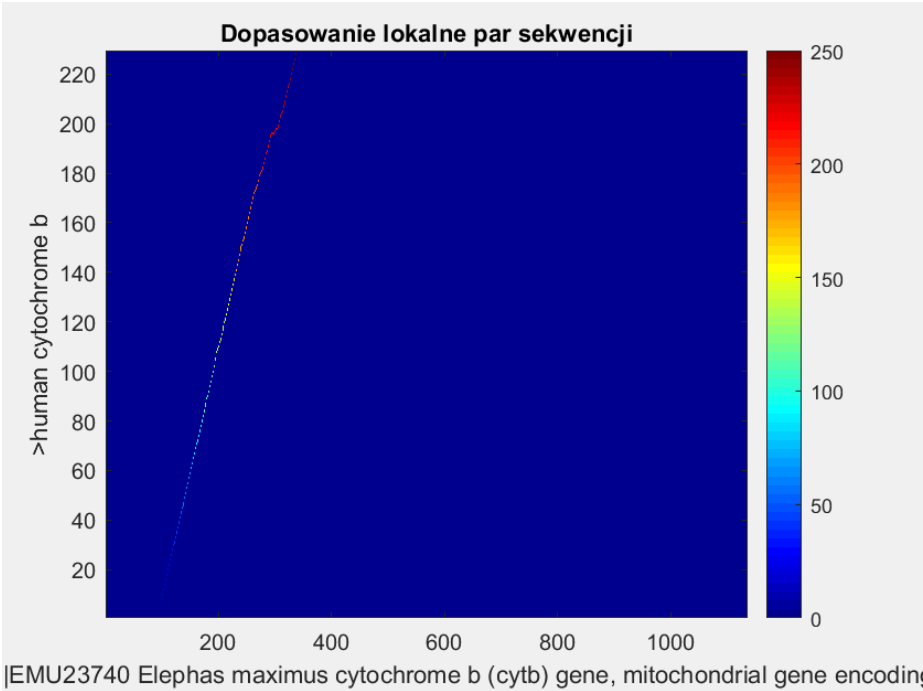
Human – Rhino

```
# 1: GAAATTTTGGCTCTCTACTAGGAATCTGCCTAATCCTACAAATCCTAACCGGACTATTTCTTGCTATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2     -7    -5    -7
     -7      2    -7    -5
     -5     -7     2    -7
     -7     -5    -7     2
# Gap: -1
# Score: 264
# Length: 229
# Identity: 179/229 ( 78%)
# Gaps: 94/229 ( 41%)
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT
||||__||__|||||__||__||__||_____||||||__|||||__||||||_|___||__||||||||__||
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CT
```

```
# 1: GAAATTTTGGCTCTCTACTAGGAATCTGCCTAATCCTACAAATCCTAACCGGACTATTTCTTGCTATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2     -7    -5    -7
     -7      2    -7    -5

     -5     -7     2    -7
     -7     -5    -7     2
# Gap: -1
# Score: 264
# Length: 231
# Identity: 180/231 ( 78%)
# Gaps: 96/231 ( 42%)
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT
||||__||__|||||__||__||__||_____||||||__|||||__||||||_|___||__||||||||__||
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CT
```

```
# 1: GAAATTTTGGCTCTCTACTAGGAATCTGCCTAATCCTACAAATCCTAACCGGACTATTTCTTGCTATACA
# 2: ATGACCCCAATACGCAAAATTAACCCCCTAATAAAATTAATTAACCACTCATTCATCGACCTCCCCACCC
# Mode: distance
# Subtitution matrix:
      2     -7    -5    -7
     -7      2    -7    -5
     -5     -7     2    -7
     -7     -5    -7     2
# Gap: -1
# Score: 264
# Length: 235
# Identity: 181/235 ( 77%)

# Gaps: 98/235 ( 42%)
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT
||||__||__|||||__||__||__||_____||||||__|||||__||||||_|___||__||||||||__||
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CT
```

```
>>Rhino gi|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene f
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT-

>>human cytochrome b
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CTA

>>Rhino gi|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene f
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT-

>>human cytochrome b
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CTA

>>Rhino gi|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene f
GAAA-TTT-TGGCTC-TCT-ACT-AGG-A-A-TCTGCCT-AATCCT-ACAAATC-CT-AAC-CGGACTATT-TCT-

>>human cytochrome b
GAAAC-TTC-GGCTCA-CTC-CTT-GGC-G-C-CTGCCTG-ATCCTC-CAAATCAC-C-ACA-GGACTATTC-CTA
```


ıi|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene for cytochrome

ıi|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene for cytochrome

Dopasowanie lokalne par sekwencji

>human cytochrome b

i|12903|emb|X56283.1|MIDBCYTB D.bicornis mitochondrion cytb gene for cytochrom