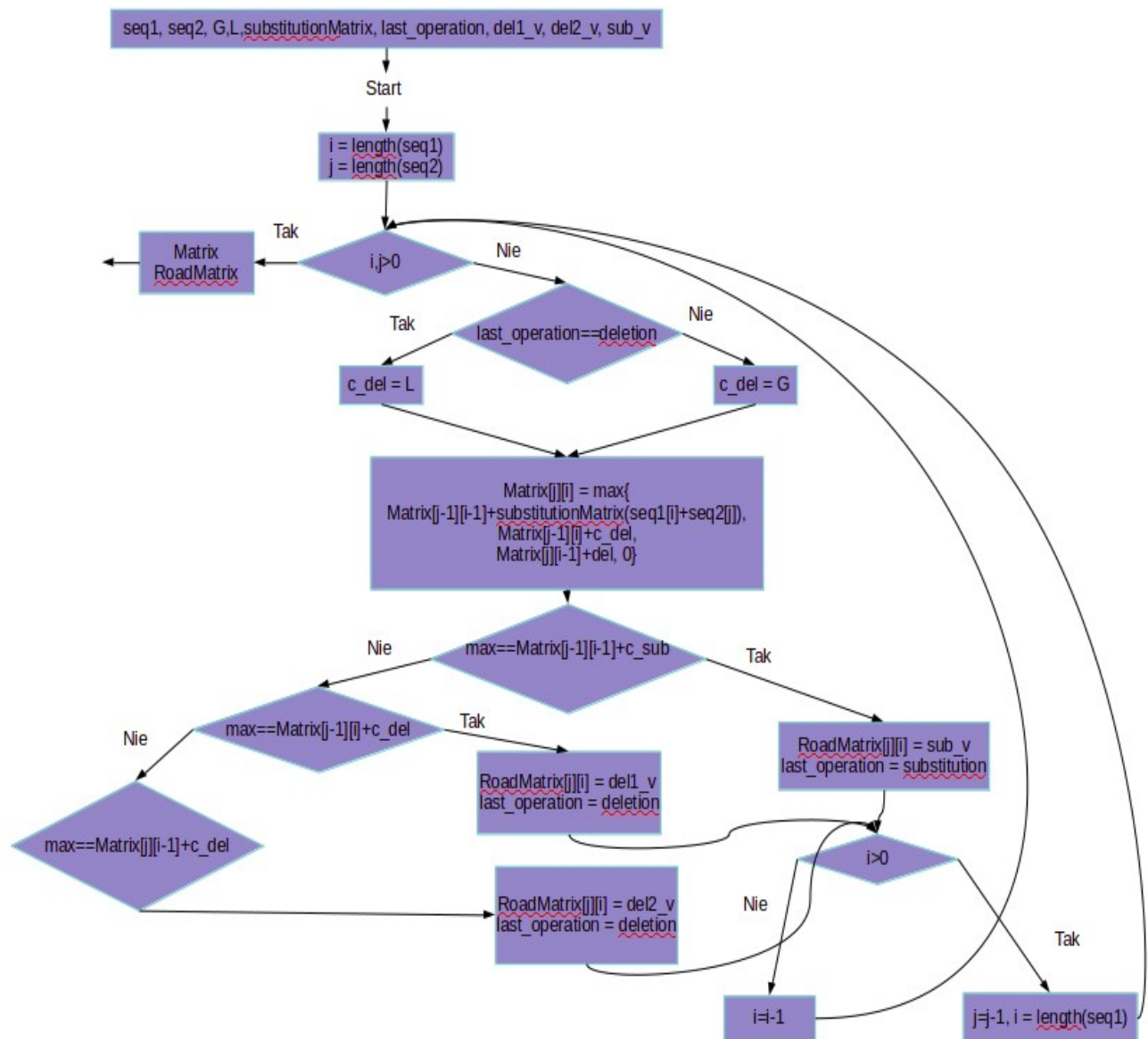


## 1. Schemat blokowy

### 1.1 Tworzenie macierzy drogi oraz macierzy punktacji:

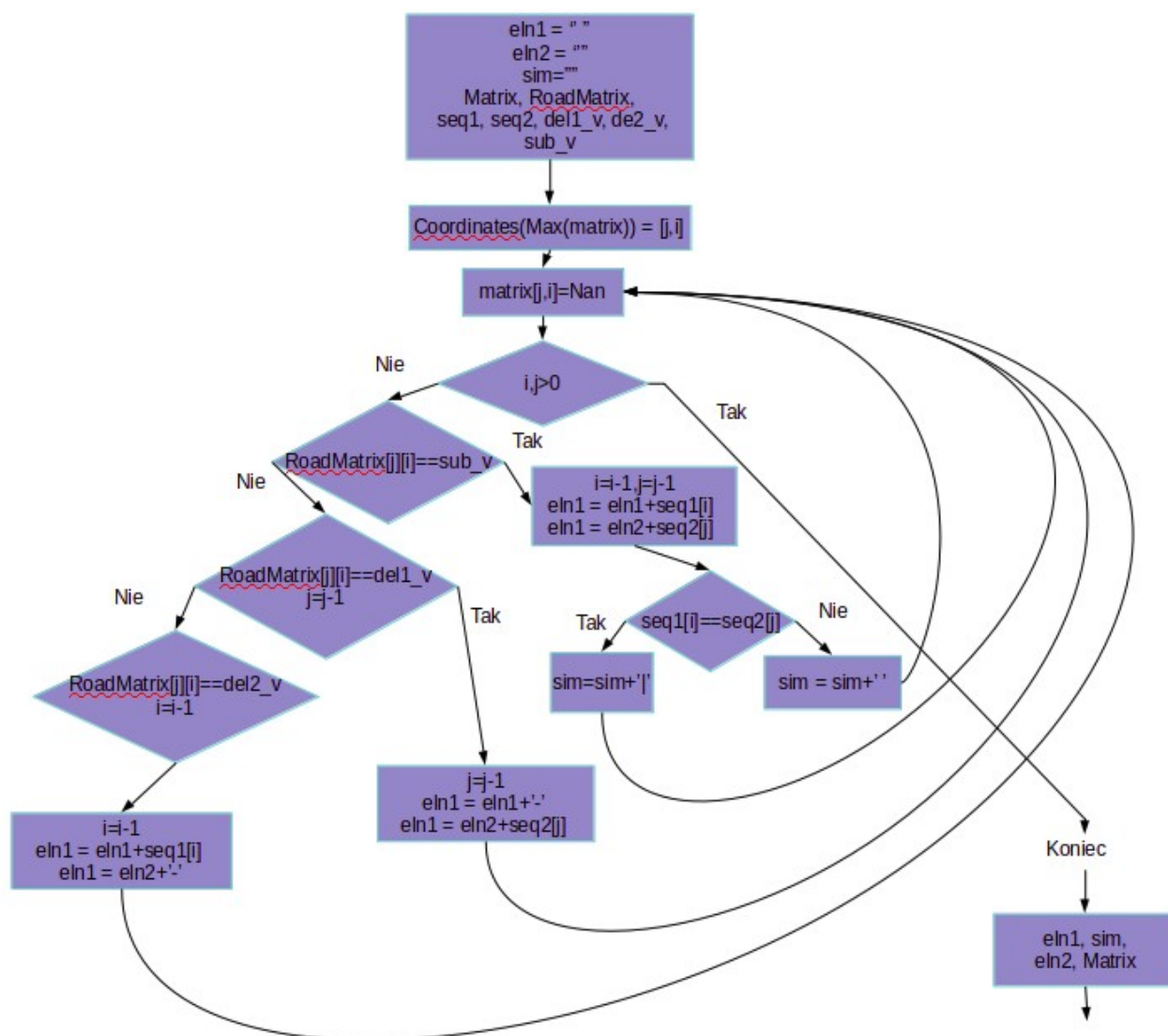


$del1\_v$  – wartość w macierzy drogi świadcząca o tym, iż doszło do delecji (przejście z góry)

$del2\_v$  – wartość w macierzy drogi świadcząca o tym, iż doszło do delecji (przejście z lewej strony)

$sub\_v$  – wartość w macierzy drogi świadcząca o tym, iż doszło do substytucji (przejście na skos)

1.2 Wyznaczenie i oznaczenie (za pomocą wartości Nan) optymalnej ścieżki dopasowania:



del1\_v – wartość w macierzy drogi świadcząca o tym, iż doszło do delecji (przejście z góry)

del2\_v – wartość w macierzy drogi świadcząca o tym, iż doszło do delecji (przejście z lewej strony)

sub\_v – wartość w macierzy drogi świadcząca o tym, iż doszło do substytucji (przejście na skos)

## 2. Analiza złożoności obliczeniowej czasowej i pamięciowej

2.1 Tworzenie macierzy drogi oraz macierzy punktacji:

- **czasowa:**  $O(m \cdot n)$
- **pamięciowa:**  $O(m \cdot n)$

Gdzie m i n to odpowiednio długości pierwszej i drugiej sekwencji

2.2 Wyznaczenie i oznaczenie (za pomocą wartości Nan) optymalnej ścieżki dopasowania:

- **czasowa:**  $O(m + n)$
- **pamięciowa:**  $(m + n)$

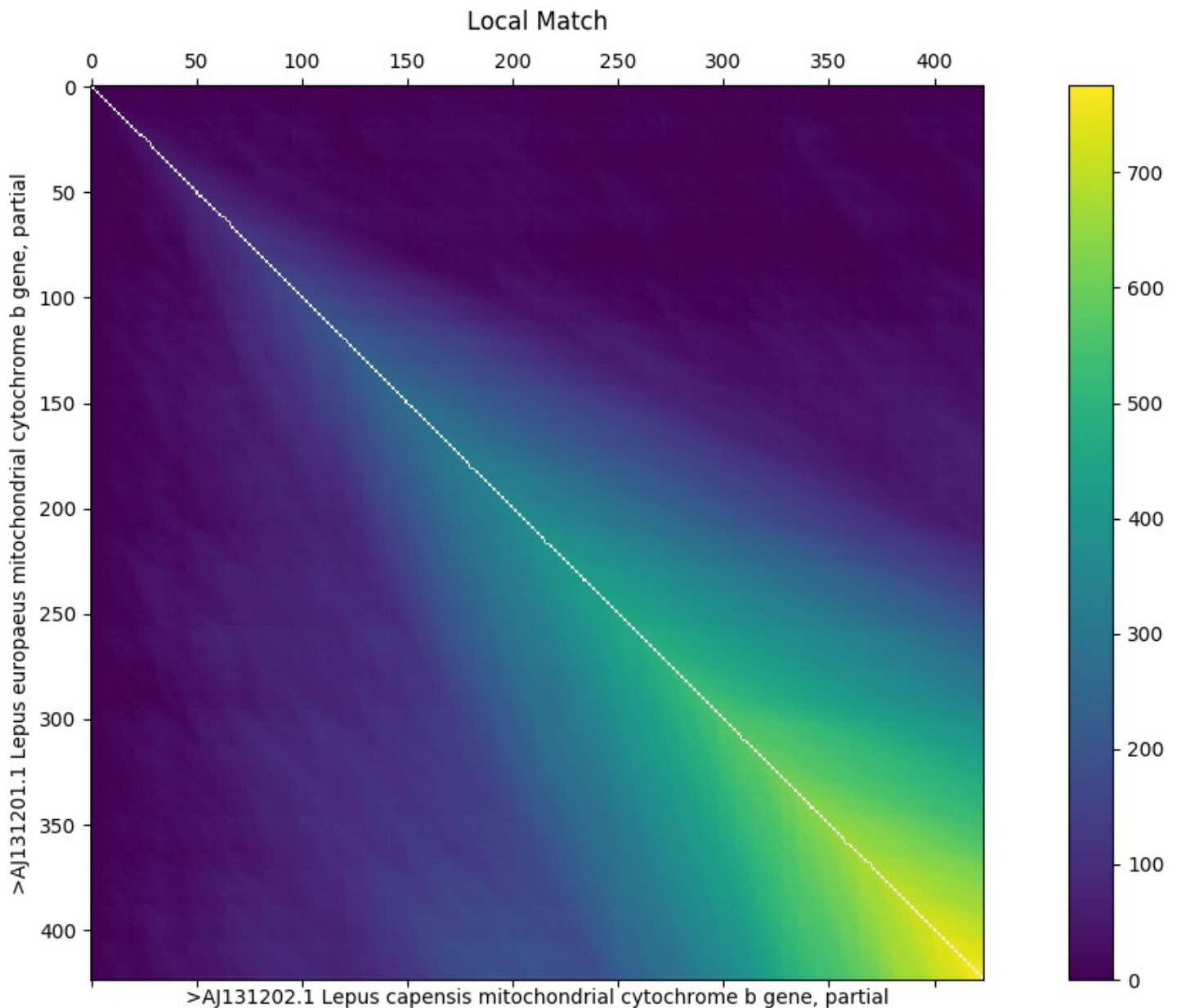
Gdzie m i n to odpowiednio długości pierwszej i drugiej sekwencji

### 2.3 Całokształt:

- **czasowa:**  $O(m*n)$
- **pamięciowa:**  $O(m*n)$

## 3. Porównanie przykładowych par sekwencji ewolucyjnie powiązanych i niepowiązanych

### 3.1 Powiązanych:



Rys 1, Dopasowanie lokalne dla genów kodujących cytochrom b dla zająca szara i zająca pustynnego

Sekwencje zostały wybrane nie ze względu na duże pokrewieństwo danych gatunków, lecz ze względu na to, że są tych samej długości.

```
ATGACCAACATTTCGTAACGCGATCCCC-TACTAAAAATTGTTAACCCTC-CCTAATCGACCTTCCC-GCCCCCTCAAACATCTCTGCCTGATGAACTTCGGCTCTCTATTGGGATTATGCCTAATAATCCAGATCCTAACTGGCTTATT-  
|||||  
ATGACCAACATTTCGTAACGCGA-CCCCCTACTAAAAATTGTTAACCCTCTC-TAATCGACCT-CCCCGCCCCCTCAAACATCTCTGCCTGATGAACTTCGGCTCTCTATTGGGATTATGCCTAATAATCCAGATCCTAACTGGCTTATT
```

C-TAGGCCATACACTACACATCAGACACAGC-AACAGCATTCTCCTCAGTTACACATATTTGCCGAGA-CGTAAACTACGGCTGACT-TATTGTTACTTACACGCTAATGGAGCATCAATATTTCTTTATTTGCTTATATATACATGTAGGCCGTGGAATCTA  
 |||  
 C-TAGGCCATACACTACACATCAGACACAGCTA-CAGCATTCTCCTCAGTTACACATATTTGCCGAGATG-TAAACTACGGCTGACTC-ATTGTTACTTACACGCTAATGGAGCATCAATATTTCTTTATTTGCTTATATATACATGTAGGCCGTGGAATCTA  
 |||  
 -TTACGGCTCATATACTTACCTAGAAACCTGAAA-TATTGGCATTATTCTA-TTATTGCGAGTTA-TGGCTACAGCATTATAGGCTA-TGTCCTCCCATGAGGACAAATATCATTCTGA  
 |||  
 CT-ACGGCTCATATACTTACCTAGAAACCTGAAAC-ATTGGCATTATTCTACT-ATTGCGAGT-AATGGCTACAGCATTATAGGCTAC-GTCCTCCCATGAGGACAAATATCATTCTGA

### Fragmenty statystyk:

```
# Mode: similarity
```

```
# Punctuation: {'AA': 2, 'CA': -7, 'GA': -5, 'TA': -7, 'AC': -7, 'CC': 2, 'GC': 7, 'TC': -5, 'AG': -5, 'CG': -7, 'GG': 2, 'TG': -7, 'AT': -7, 'CT': -5, 'GT': -7, 'TT': 2}
```

```
# Gap: -2
```

# G: -15

# L: -2

# Score: 779.0

```
# Length: 435
```

```
# Identity: 410/435 (94%)
```

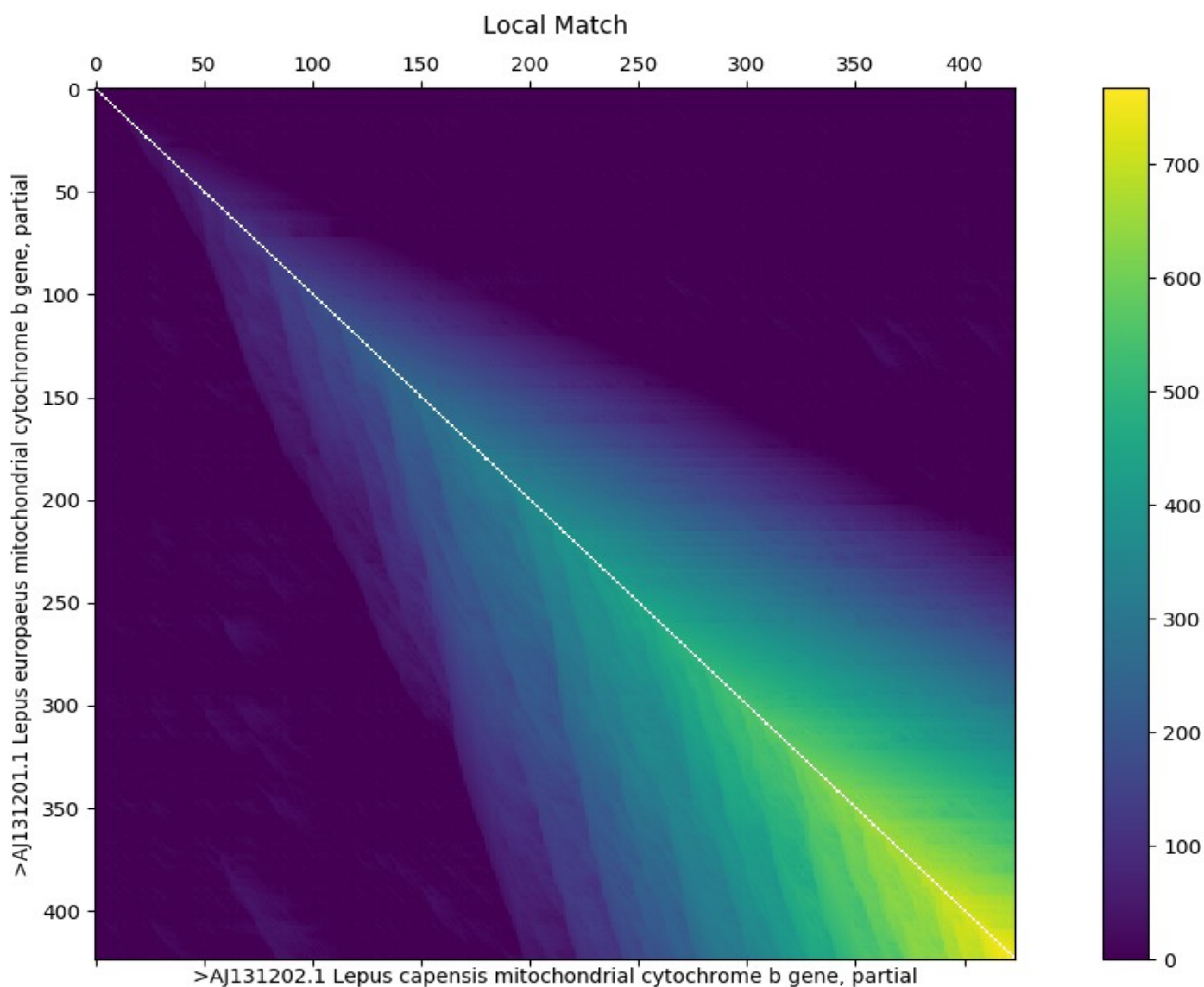
# Gaps: 24/435 (5%)

### Fasta Format:

```
>seq1 1-423
```

```
>seq2 1-423
```

Dopasowanie za pomocą algorytmu rozróżniającego karę rozpoczęcia oraz kontynuacji przerwy:



Rys 2, Dopasowanie llokalne dla genów kodujących cytochrom b dla zająca szara i zająca pustynnego za pomocą algorytmu rozróżniającego karę rozpoczęcia oraz kontynuacji przerwy

### Fragmenty statystyk:

#### # Mode: similarity

```
# Punctuation: {'AA': 2, 'CA': -7, 'GA': -5, 'TA': -7, 'AC': -7, 'CC': 2, 'GC': 7, 'TC': -5, 'AG': -5, 'CG': -7, 'GG': 2, 'TG': -7, 'AT': -7, 'CT': -5, 'GT': -7, 'TT': 2}
# Gap: -2
# G: -15
# L: -2
# Score: 771.0
# Length: 434
# Identity: 411/434 (94%)
# Gaps: 22/434 (5%)
```

#### Fasta Format:

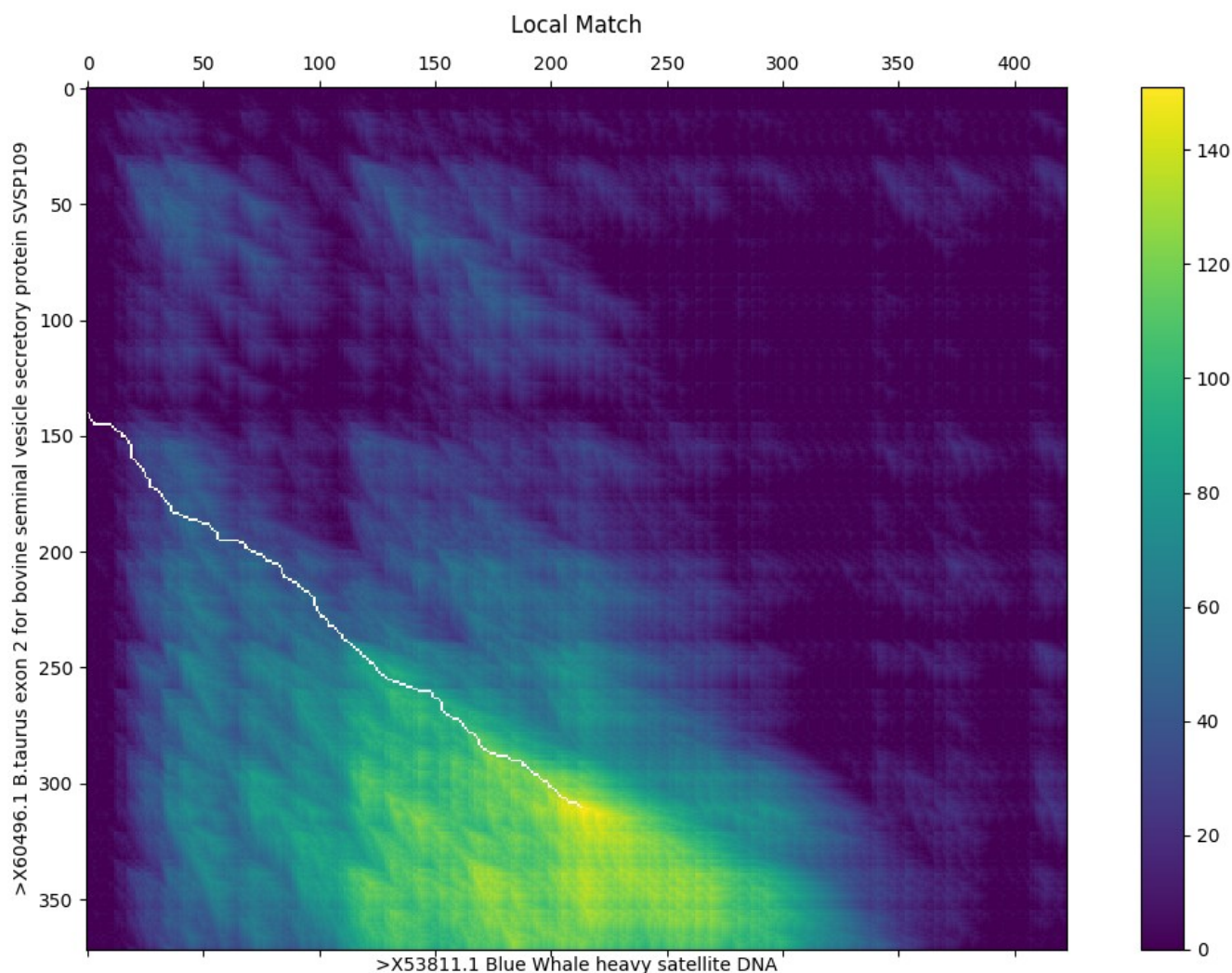
```
>seq1 1-423
```

```
>seq2 1-423
```



ATGACCAACATTCTGTAACGCA-TCCCCTACTAAAAATTGTTAACCACTC-CCTAATCGACCT-TCCCGCCCCCTCAACATCTCTGCCTGATGAACTTCGGCTCTCTATTGGGATTATGCCTAATAATCCAGATCCTAACTGGCTTATT-  
 |||||  
 ATGACCAACATTCTGTAACGCA-CCCCTACTAAAAATTGTTAACCACTCT-CTAATCGACCTC-CCCCCCCCCTCAACATCTCTGCCTGATGAACTTCGGCTCTCTATTGGGATTATGCCTAATAATCCAGATCCTAACTGGCTTATT-  
 |||||  
 CCTAGCCATACACTACACATCAGACACAGC-AACAGCATTCTCCTCAGTTACACATATTGCCGAGA-CGTAAACTACGGCTGACT-TATTCGTTACTTACACGCTAATGGAGCATCAATATTCTTTATTGCTTATATACATGTAGGCCGTGGAATCTA-  
 |||||  
 -CTAGCCATACACTACACATCAGACACAGCT-ACAGCATTCTCCTCAGTTACACATATTGCCGAGAT-GTAAACTACGGCTGACTC-ATTCGTTACTTACACGCTAATGGAGCATCAATATTCTTTATTGCTTATATACATGTAGGCCGTGGAATCTAC-  
 |||||  
 TTACGGCTCATATACTTACCTAGAACCTGAAA-TATTGGCATTATTCTA-TTATTCGAGTTATGGCTACAGCATTTCATAGGCTA-TGTCCTCCCATGAGGACAAATATCATTCTGA-  
 |||||  
 -TACGGCTCATATACTTACCTAGAACCTGAAA-ATTGGCATTATTCTAC-TATTCGAGTAATGGCTACAGCATTTCATAGGCTAC-GTCCTCCCATGAGGACAAATATCATTCTGA-  
 |||||

### 3.2 Niepowiązanych



Rys 4, Dopasowanie lokalne dla przypadkowo wybranych genów niepowiązanych ewolucyjnie ze sobą

#### Fragment statystyk:

# Mode: similarity

# Punctuation: {'AA': 2, 'CA': -7, 'GA': -5, 'TA': -7, 'AC': -7, 'CC': 2, 'GC': 7, 'TC': -5, 'AG': -5, 'CG': -7, 'GG': 2, 'TG': -7, 'AT': -7, 'CT': -5, 'GT': -7, 'TT': 2}

# Gap: -2

# G: -15

# L: -2

# Score: 151.0

# Length: 258

# Identity: 94/258 (36%)

```
# Punctuation: {'AA': 2, 'CA': -7, 'GA': -5, 'TA': -7, 'AC': -7, 'CC': 2, 'GC': 7, 'TC': -5, 'AG': -5, 'CG': -7, 'GG': 2, 'TG': -7, 'AT': -7, 'CT': -5, 'GT': -7, 'TT': 2}
```

### Fasta Format:

```
>seq2 210-333
```

### 3.3 Wnioski

Zastosowanie kary rozpoczęcia i kontynuacji przerwy znacznie bardziej wpływa na dopasowanie lokalne sekwencji niepowiązanych ze sobą ewolucyjnie. Może zmniejszyć długość dopasowanych sekwencji prawie dwukrotnie (dla sekwencji niepowiązanych). Sekwencje niepowiązane genetycznie cechują się mniejszą liniowością dopasowania i większą zmianą pod wpływem zmiany sposobu punktacji i algorytmu dopasowania lokalnego.