

Wstęp do bioinformatyki

Laboratorium 4

Dopasowanie wielu sekwencji

Magdalena Trędak

236712

1. Schemat blokowy algorytmu dopasowania wielu sekwencji

Schematy blokowe algorytmów szukania sekwencji centralnej oraz dopasowania wielu sekwencji ze względów dużych rozmiarów i umożliwienia poprawy ich czytelności zamieszczono w repozytorium jako pliki graficzne o nazwach: SchematBlokowySzukaniaSekwencjiCentralnej.png, SchematBlokowyDopasowaniaWieluSekwencji.png.

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

findCentralSequence – 7 przypisań poza pętlami, 11 przypisań w pętlach, każda o rozmiarze ilości możliwych dopasowań sekwencji. Dla liczby sekwencji k ilość tych dopasowań wynosi $x = k(k-1)/2$. Wywoływane są funkcje scoringMatrix oraz traceBackMatrix, każda z nich x razy. Złożoność czasowa $x(n*m) + x(n+m) \rightarrow$ co najwyżej rzędu $O(nm)$

alignAll – 2 przypisania poza pętlami, 4 wewnątrz nich. Dla x będącym ilością porównywanych sekwencji tworzona jest pętla dopasowująca kolejne sekwencje do sekwencji centralnej oraz powtarzająca poprzednie dopasowania. Ilość wywołań funkcji alignTwoSequences wynosi x . Złożoność czasowa $x(n*m)$

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 . Takich macierzy algorytm tworzy $2*x$ – gdzie x to ilość ścieżek dopasowywanych. Pierwsze dopasowanie odbywa się podczas szukania ścieżki centralnej. Następne podczas dopasowywania kolejnych sekwencji do ścieżki centralnej. Stąd wynika, że złożoność pamięciowa programu jest rzędu $2*x*O(mn) \rightarrow$ co najwyżej rzędu $O(mn)$.

3. Przykładowe dopasowanie dla sekwencji powiązanych ewolucyjnie

Porównano sekwencje mitochondrialnego cytochromu b poniższych organizmów:

X75581 - Antarctic minke whale

X75589.1 - sperm whale

X75584.1 - humpback whale

X75586.1 - pygmy right whale

X75583.1 - pygmy Bryde's whale

```
Command Window
>> Tester
#Match: 0
#Mismatch: 1
#Gap: 2
#Length: 1140
#Central sequence: X75584.1
#Alignment cost: 1410
X75583.1 ATGACCAACATCCGAAAAACACACCCACTAATAAAGATTGTCAACGATGCATTTCGTTGAT 60
X75589.1 ATGACCAACATCCGAAAATCACACCCATTAATAAAATCATTACAATGCATTCATCGAC 60
X75586.1 ATGACCAACATCCGAAAAACACACCCACTAATAAAATATCAACAACGCATTCATTGAT 60
X75581.1 ATGACCAACATCCGAAAAACACACCCACTAATAAAATATCAACGACGCATTCGTCGAC 60
X75584.1 ATGACCAACATCCGAAAAACACACCCACTAATAAAATATCAACGACACATTCATTGAT 60
*****

X75583.1 CTCCCCACCCATCAAATATCTCCTCATGATGAAATTCGGCTCCCTACTCGGCCTCTGC 120
X75589.1 CTCCCTACCCATCAAACATTTCTCCTCATGATGAACTTCGGCTCCCTACTTGGACTCTGC 120
X75586.1 CTTCCCACCCATCAAATATCTCCTCATGATGAACTTCGGTTCCCTACTTGGCCTTTGC 120
X75581.1 CTACCCACCCATCAAACATCTCCTCATGATGAACTTCGGCTCCCTACTCGGCCTCTGC 120
X75584.1 CTACCCACCCATCAAATATCTCCTCATGATGAACTTCGGTTCCCTACTCGGCCTTTGC 120
** ** * ***** ** ***** ***** ***** ** ** **

X75583.1 TTAATTACACAAATCCTAACAGGCCTATTCTAGCAATACACTACACACCAGACACAACA 180
X75589.1 CTGATCATACAAATCCTAACAGGCCTATTCTAGCAATACACTACACACCAGACACAACA 180
X75586.1 CTAATCATACAAATCCTAACAGGCCTATTCTAGCAATACATTACACGCCAGACACAACA 180
X75581.1 TTAATTGTACAAATCCTAACAGGTCTATTCTAGCAATACACTACACACCAGACACAACA 180
X75584.1 TTAATTATACAAATCCTAACAGGCCTATTCTAGCAATACACTACACACCAGACACAACA 180
* ** ***** ***** ***** ***** *****

X75583.1 ACCGCCTTCTCATCAGTTGCACACATTTGCCGAGACGTAAACTACGGCTGAGTTATCCGA 240
X75589.1 ACCGCTTTCTCATCAATCACACACATTTGCCGAGACGTCAATTACGGCTGGACCATCCGA 240
X75586.1 ACCGATTCTCATCAGTCACACATATTGCCGAGATGTAAACTACGGCTGAGTTATCCGA 240
X75581.1 ACCGCCTTCTCATCAGTTACACATATCTGCCGAGACGTAAACTACGGCTGAATTATCCGA 240
X75584.1 ACCGCCTTCTCATCAGTCACACACATCTGTGAGACGTAAATTATGGCTGAATTATCCGA 240
***** ***** * ***** ** * ***** ** ** * ***** *****

X75583.1 TACCTACAGCAAACGGAGCCTCCATATTTCTTCATCTGTCTCTACGCTCACATAGGACGA 300
X75589.1 TACCTACATGCAAACGGAGCTTCCATATTTTTCATTTGCCTTTACACTCACATGGGCCGT 300
X75586.1 TATCTACATGCAAATGGAGCCTCCATATTTTTCATCTGCATCTACGCCCACATAGGACGT 300
X75581.1 TATCTACATGCAAACGGAGCCTCCATATTTTTCATTTGCCTTTACGCCCACATAGGACGA 300
X75584.1 TACCTACATGCAAATGGGGCCTCCATATTTCTTCATCTGCCTCTACGCTCACATAGGACGA 300
** ***** ** ** ***** ** ** * ***** ** **
```

X75583.1 GGCCTATACTACGGCTCCTATGCCTTTTCGAGAAACATGAAACATCGGAGTTATCCTACTA 360
X75589.1 GGCTGGTACTACGGCTCCTACATCTTCCAAGAAACATGAAACGTTGGAATGATACTCCTA 360
X75586.1 GGCCTATACTACGGCTCTCATGCTTTCCGAGAGACATGAAATATCGGAGTAATCTTATTA 360
X75581.1 GGCCTATACTATGGAACCCACGCCTTCCGAGAAACATGAAATATCGGAGTTATCCTACTG 360
X75584.1 GGCCTATACTACGGCTCCTACGCCTTTTCGAGAAACATGAAACATCGGAGTTATTCTACTA 360
*** ***** ** * * ** * *** ***** * *** * ** * *

X75583.1 TTCACAGTTATAGCCACCGCATTTCATAGGCTACGTCCTACCCTGAGGACAAATATCATTT 420
X75589.1 ATCACAGTAATAGCCACCGCATTTCGTAGGTTACGTACTGCCCTGAGGACAAATATCATTC 420
X75586.1 TTCACAACCATAGCCACTGCATTTCGTAGGCTATGTCCTGCCCTGAGGACAGATATCATTC 420
X75581.1 TTCACAGTTATAGCCACTGCATTTCGTAGGCTACGTCCTACCCTGAGGACAAATATCATTT 420
X75584.1 TTCACAGTTATAGCCACTGCATTTCGTAGGCTACGTCCTACCCTGAGGACAAATATCATTC 420
***** ***** ***** ***** ** ** ** ***** *****

X75583.1 TGAGGCGCAACCGTCATCACCAACCTCTTATCAGCAATCCCATACATTGGTACTACCCTA 480
X75589.1 TGAGCCGCAACCGTTATCACAAACCTTCTATCAGCAATTCCCTATATCGGCACCAACCTA 480
X75586.1 TGAGGCGCAACCGTCATCACCAACCTCCTATCAGCAATCCCATATATTGGTACCACCTA 480
X75581.1 TGAGGCGCAACCGTCATCACCAACCTCCTATCAGCAATCCCATACATTGGTACCACCTTA 480
X75584.1 TGAGGCGCAACCGTCATCACCAACCTTCTATCAGCAATCCCATACATTGGTACTACCCTA 480
**** ***** ***** ***** ***** ** ** ** **

X75583.1 GTCGAATGAATCTGGGGCGGTTTCTCTGTAGATAAAGCAACACTAACACGCTTTTTTGCC 540
X75589.1 GTAGAGTGAGTTGAGGCGGTTTCTCCGTAGATAAGGCAACACTGACACGCTTCTTCACT 540
X75586.1 GTTGAATGAATCTGGGGTGGCTTCTCCGTAGACAAAGCGACACTAACTCGCTTCTTTGCT 540
X75581.1 GTTGAATGAATCTGAGGTGGCTTCTCTGTAGACAAAGCAACATTAACACGCTTTTTCGCC 540
X75584.1 GTCGAATGAATCTGGGGCGGTTTTCCTAGACAAAGCAACACTAACACGTTTCTTTGCT 540
** ** ** * ** ** ** **

X75583.1 TTCCACTTTATCCTCCCCTTCATTATTCTAGCACTAGCAATGGTCCACCTCATTTTCCTC 600
X75589.1 CTCCACTTCATCCTCCCCTTTATCACCTAACACTAACAAATAGTACATCTCCTATTTCTC 600
X75586.1 TTCCACTTCATCCTCCCCTTTTATTCTAGCGCTAGCAGCTGTTTCATCTCCTTTTCCTC 600
X75581.1 TTCCACTTCATCCTCCCCTTTTATTCTAGCATTAGCAATTGTCCACCTCATTTTCCTC 600
X75584.1 TTCCACTTCATCCTCCCCTTCATCATTACAGCATTAGCAATCGTCCACCTCATTTTCCTC 600
***** ***** ** ** * * ** ** **

X75583.1 CACGAAACAGGATCCAATAACCCACAGGTATTCCATCCAACATAGACAAAATCCCATT 660
X75589.1 CATGAAACAGGATCCAACAACCCACAGGAATTCCTCCAACATAGACAAAATCCCATT 660
X75586.1 CACGAAACAGGATCTAACAACCCACAGGCATCCCATCCAACATAGACAAAATCCCATT 660
X75581.1 CGCGAAACAGGATCCAATAACCCACAGGTATTCCATCTGATATAGACAAAATCCCATT 660
X75584.1 CACGAAACAGGATCCAACAACCCACAGGCATCCCATCCAACATAGACAAAATCCCATT 660
* ***** ** ***** ** ** * ***** *****

X75583.1 CACCCCTATTACACAACCTAAAGACATTCTAGGCGCCCTACTACTAATCCTAACCCCTACTA 720
X75589.1 CACCCCTACCACACAATCAAAGACACCATAGGTGCCCTACTACTAATCCTATCCCTACTT 720
X75586.1 CACCCCTACTACACAATTAAGACATCCTGGGCGTCTACTACTAATCCTGACCCTACTA 720
X75581.1 CACCCCTATTACACAATCAAAGACATTCTAGGCGCCCTACTACTAATCCTAACCCCTACTA 720
X75584.1 CACCCTTACTACACAATCAAAGACACTCTAGGCGCCCTATTACTAATCCTAACCCCTACTA 720
***** ** ***** ***** * ** * ***** ***** ** *****

X75583.1 ATGCTAACCCCTATTTCGTACCCGACCTACTTGGAGACCCAGACAACCTACACTCCAGCAAAT 780
X75589.1 ACACTAACCCCTGTTTCGCACCCGACCTGCTAGGAGATCCCGACAACCTACACCCCAGCAAAT 780
X75586.1 ATATTAACCTTATTTACACCTGACCTGCTTGGAGACCCCTGACAACCTACACCCCAGCAAAT 780
X75581.1 ACACTAACCCCTATTTCGCACCCGACCTGCTCGGAGACCCCGACAACCTACACCCCAGCAAAC 780
X75584.1 ATGTTAACCCCTATTTCGCACCTGACCTGCTTGGAGACCCAGATAACTACACCCCAGCAAAC 780
* ***** * ** *** ***** ** ***** ** ** ***** *****

X75583.1 CCACTCAGTACCCCAACACACATTAAACCAGAATGATATTTCTATTTGCATACGCAATC 840
X75589.1 CCACTAAATACCCCAACACACATCAAACCAGAATGGTATTTCTATTCGCGTACGCCATC 840
X75586.1 CCCCTCAGCACCCCAAGCAGACATCAAGCCAGAATGATACTTCTATTTGCATATGCAATC 840
X75581.1 CCACTCAGTACCCCAAGCAGACATTAACCAGAATGATATTTCTATTCGCGTACGCAATC 840
X75584.1 CCACTCAGTACCCCAAGCAGACATTAACCAGAGTGATATTTCTATTTGCATACGCAATC 840
** ** * ***** ***** ** ***** ** ** ** ***** ** ** ** **

X75583.1 CTACGATCAATTCCCAACAAATAGGCGGAGTCTTAGCCCTACTACTCTCAATCCTAATC 900
X75589.1 CTACGATCTGTCCCAATAAACTAGGAGGCGTCTTAGCCCTACTACTCTCCATCCTAATC 900
X75586.1 CTACGATCAATTCCCTAATAAAATAGGTGGAGTCTTAGCCCTATTATTTTCAATCCTAATT 900
X75581.1 CTACGATCAATCCCAATAAACTAGGCGGAGTCTTAGCCCTATTACTCTCAATCCTAATC 900
X75584.1 CTACGATCAATCCCAACAAACTAGGCGGAGTCTTAGCCTTGCTACTCTCAATCCTAATC 900
***** * ** ** *** ***** ** *** ***** * ** * ** *****

X75583.1 CTAGCCTTAATCCCAATACTCCACACATCTAAACAACGAAGCATAATGTTCCGACCCTTT 960
X75589.1 CTAGTATTCATCCCAATACTTCATACAGCCAAACAACGAAGCATAATATTCGACCCTTC 960
X75586.1 CTAGCCCTCATCCCAATACTCCACACATCCAAACAACGAAGCATAATGTTTCGACCCTTC 960
X75581.1 CTAGCCTTCATCCCAATACTCCACACATCCAAACAACGAAGCATAATATTCGACCCTTT 960
X75584.1 CTAGCTTTCATCCCAATACTCCACACATCCAAACAACGAAGCATGATGTTTCGACCCTTT 960
**** * ** ***** ** *** * ***** ***** ** ** *****

X75583.1 AGCCAATTCCTATTTTGAGTCCTAATTGCAGACTTACTAACCCCTGACATGAATCGGCGGC 1020
X75589.1 AGTCAATTCCTGTTCTGAACACTAATCATAGACCTACTAACTCTAACATGAATCGGAGGC 1020
X75586.1 AGCCAATTCCTATTCTGAGTGCTAGTCGCAGACCTACTAACCCCTAACATGAATTGGTGGC 1020
X75581.1 AGCCAATTTCTATTCTGAGTACTAGTCGCAGACCTATTAACCCTAACATGGATTGGTGGC 1020
X75584.1 AGCCAGTTTCTGTTTTGAATACTAGTAGCAGACCTATTAGCCCTAACATGAATCGGCGGC 1020
** ** ** ** ** ** ***** **** * ***** ** ** * ** ***** ** ** **

```

X75583.1 CAACCCGTAGAACACCCCTACGTAATCGTAGGCCAATTTCGCATCCATCCTCTATTTTCCTC 1080
X75589.1 CAACCCGTGGAACACCCATATGTAACGTAGGCCAACTAGCCTCCATCCTATACTTTCTT 1080
X75586.1 CAACCCGTAGAACACCCCTTATGTTATAGTAGGCCAACTCGCATCCATCCTTTATTTCTTC 1080
X75581.1 CAACCCGTAGAACACCCCTACATAATCGTAGGCCAACTCGCATCCATCCTTTATTTCTC 1080
X75584.1 CAACCCGTAGAACACCCATACATAATCGTAGGCCAACTCGCATCCATCCTCTACTTCCTC 1080
*****

```

```

X75583.1 CTAATTCTAGTACTAATACCAGTAACTAGTCTTATCGAGAATAAACTTATAAAATGAAGA 1140
X75589.1 TTAATCCTAATCCTAATACCAACAGCCAGTCTTATCGAGACAAGCTCTTAAATGAAGA 1140
X75586.1 CTGATCTAATCCTAATGCCAGTAACCAGTCTTATCGAAAATAAACTTATAAAATGAAGA 1140
X75581.1 TTAATCCTAGTACTAATACCAGTAGCTAGCCTTATCGAGACAAGCTTATAAAATGAAGA 1140
X75584.1 TTAATCCTAGTATTAATACCAATAACTAGTCTTATCGAGACAACTTATAAAATGAAGA 1140
* * * * *

```

- Z96064.1 – *Acomys ignitus*
- Z96066.1 – *Acomys russatus*
- Z96057.1 – *Acomys spinosissimus*
- Z96062.1 – *Acomys dimidiatus*

```

Command Window

>> Tester
#Match: 0
#Mismatch: 1
#Gap: 2
#Length: 1141
#Central sequence: Z96057.1
#Alignment cost: 956
Z96064.1 ATGAAAAACATACGAAAAACACACCCGCTCCTAAAAATTATCAACCACGCATTTCATTGAT 60
Z96066.1 ATGAAAAACATACGAAAAATCACACCCACTACTAAAAATTATTAATCACGCATTTATTGAC 60
Z96057.1 ATGAAAATTATACGAAAAACACACCCACTCATAAAAAATTATCAACCATGCATTTCATCGAC 60
Z96062.1 ATGAAAATCATACGAAAAACACACCCACTCCTAAAAATCATCAACCACGCATTTCGTCGAC 60
*****
Z96064.1 CTCCCAGCACCCCTCTAANATCTCATCATGATGAAACTTTGGCTCCCTACTGGGAGTCTGC 120
Z96066.1 CTACCCGCACCATCCAATATCTCATCATGATGAAACTTTGGGTCCCTTCTAGGAGTCTGT 120
Z96057.1 CTCCCTGCACCCCTCCAACATCTCATCATGATGAAACTTCGGCTCCCTATTGGGAGTATGC 120
Z96062.1 CTCCCTGCACCCCTCCAACATCTCATCATGATGAAACTTCGGCTCTCTATTAGGGGTATGC 120
* * * * *

```


Command Window

```
Z96064.1 CTAGTAATACAAATCCTAACAGGCTTATTCTAGCATTACATTACACCTCTGACACCACC 180
Z96066.1 CTAGTAGCACAAATTATCACAGGACTATTTCTAGCAATACACTACACCTCTGATACCACA 180
Z96057.1 CTAATAATCCAAATCCTTACAGGCCTATTCTTAGCAATACACTACACCTCCGATACTACA 180
Z96062.1 CTAGTAATACAAATCCTCACAGGCCTATTCTTAGCAATACACTATACCTCAGATACTATA 180
*** **      ***** * ***** ***** ***** ** ***** ** ** *
```

```
Z96064.1 ACAGCGTTCTCATCCGTAACCCACATCTGCCGAGACGTAAACTACGGATGACTAATTCGT 240
Z96066.1 ACAGCATTTCATCTGTAACCTCATATCTGCCGAGATGTAAACTACGGTTGATTGATTCTGA 240
Z96057.1 ACAGCATTACATCCGTAACCCACATCTGCCGAGACGTAAACTACGGCTGATTAATTCGA 240
Z96062.1 ACAGCATTCTCATCCGTAACCCACATCTGCCGAGACGTAAACTACGGATGACTAATTCGA 240
***** **      **** ***** ** ***** ***** ***** ***** ** * *****
```

```
Z96064.1 TACCTCCATGCAAACGGGGCCTCCATATTCTTCATATGCCTATTTCATACACGTAGGACGA 300
Z96066.1 TATCTTCATGCAAACGGAGCCTCAATATTCTTCATGTGTCTATTTATACACGTAGGACGA 300
Z96057.1 TATCTCCATGCAAACGGAGCCTCCATATTCTTCGTATGCTTGTTTATACACGTAGGACGA 300
Z96062.1 TATCTCCATGCAAACGGAGCCTCCATATTCTTCGTATGTTTGTTTCATACACGTAGGACGA 300
** ** ***** ***** ***** ***** * ** * ** ***** *****
```

Command Window

```
Z96064.1 GGAATCTACTACGGGTCCTATACCTTCATAGAAACCTGAAACATTGGCATCATTCTACTG 360
Z96066.1 GGAATCTACTACGGATCCTATACCTTTACAGAAACCTGAAATATTGGCATCATTCTACTG 360
Z96057.1 GGCATCTACTATGGATCTTACACCTTTACAGAAACATGAAACCTTGGTGTTATTCTACTA 360
Z96062.1 GGCATTTATTATGGATCCTACACCTTTACAGAAACATGAAACCTCGGCGTTATTCTATTG 360
** ** ** ** ** ** ** ** ** ** ***** * ***** ***** * ** * ***** *
```

```
Z96064.1 TTCGCCGTAATAGCAACTGCATTTATAGGATACGTTCTCCCATGGGGACAAATATCCTTC 420
Z96066.1 TTCGCCGTAATAGCAACTGCATTTATAGGATATGTGCTTCCCTGAGGACAAATATCCTTC 420
Z96057.1 TTTGCCGTAATAGCAACTGCATTTATAGGATATGTCCTTCCATGAGGACAAATATCCTTC 420
Z96062.1 TTCGCCGTAATAGCAACTGCATTTATAGGATATGTCCTCCCATGGGGACAAATATCCTTC 420
** ***** ***** ***** ** ** ** ** ***** *****
```

```
Z96064.1 TGAGGGGCCACAGTCATTACAAATCTACTCTCAGCTATCCCTTACATAGGAACTAATCTA 480
Z96066.1 TGAGGAGCCACAGTTATTACTAACCTTCTCTCAGCAGTTCCCTACATCGGTACTAATTTA 480
Z96057.1 TGAGGAGCCACAGTCATTACAAACCTCCTTTTACGCTATTCCCTACATCGGTACTAACCTA 480
Z96062.1 TGAGGAGCTACAGTCATTACAAACCTACTATCAGCTATTCCCTATATCGGCACCTAACCTA 480
***** ** ***** ***** ** ** ***** * ** ** ** ***** **
```

Command Window

```
Z96064.1 GTAGAATGGATCTGAGGTGGCTTTTCAGTAGACAAAGCCACTTTAACACGTTTCTTCGCC 540
Z96066.1 GTAGAATGAATCTGAGGGGGTTCTCAGTAGACAAAGCCACTCTTACACGTTTCTTCGCA 540
Z96057.1 GTAGAATGAATCTGAGGTGGGTTTTCAGTAGACAAAGCCACCCTCACACGTTTCTTCGCA 540
Z96062.1 CTACAATGAATTTGAGGTGGGTTCTCAGTAGACAAAGCCACCCTTACACGTTTCTTCGCA 540
      ** ***** ** ***** ** ** ***** ***** * ***** *****
```

```
Z96064.1 TTTCACTTCATNCTACCNTTTATCATTGCAGCTCTAGCTATGGTCCACCTACTATTTCTT 600
Z96066.1 TTCCATTTTATCCTCCCATTTCATCATTGCAGCCCTAGCTATAGTCCACCTACTTTTTCTC 600
Z96057.1 TTCCACTTCATCCTTCCATTTCATCATTGCAGCCCTAGCCATAGTACACCTACTATTCTC 600
Z96062.1 TTCCACTTTATCCTCCCCTTTATCATTGCAGCCCTAGCCATAGTACACCTTCTATTCTT 600
      ** ** ** ** ** ** ** ** ** ** ** ** ***** ***** ** ** ***** ** ** **
```

```
Z96064.1 CACGAAACAGGATCTAATAANCCCTCCGGAATAAACTCCGATGCAGACAAAATCCCCTTT 660
Z96066.1 CACGAAACTGGATCCAATAACCCACCGGAATCAACTCTGATGCAGACAAAATTCATTTC 660
Z96057.1 CACGAAACAGGCTCTAATAACCCACAGGAATGAACTCTGACGCAGACAAAATCCCCTTC 660
Z96062.1 CACGAAACTGGCTCTAATAACCCACAGGAGTGAAGTCCGACGCAGACAAAATCCCATTTC 660
      ***** ** ** ***** *** * *** * ***** ** ***** ***** ** **
```

```
Z96064.1 CACCCCTATTACACAATTAAGATCTCCTAGGCATATTCCTAATATTAGCAGCCCTACTC 720
Z96066.1 CACCCGTACTACACTATAAAAGACTTCCTAGGGGCCCTACTCATATTAGTAGTTCTACTC 720
Z96057.1 CACCCCTACTATACAATCAAAGACCTCTTAGGCGCACTCCTCATACTACTAGTTCTCCTA 720
Z96062.1 CACCCCTACTACACAATTAAGACCTTTTAGGTGTATTCCTAATACTATTAATTCTCCTC 720
      ***** ** ** ** ***** * ***** * ** ** ** * ** **
```

```
Z96064.1 CTATTAGTCCTCTTCTCCCCAGACCTGTTAGGAGACCCCGACAACCTACACACCAGCCAAC 780
Z96066.1 CTACTTGTTTTATTTTCCCCAGATATTTTAGGCGACCCAGATAATTATACACCAGCTAAC 780
Z96057.1 CTATTAGTACTATTTTCCCCAGACTTATTAGGAGACCCCGACAACCTACACACCAGCCAAC 780
Z96062.1 CTATTAGTATTATTCTCCCCAGACCTATTAGGAGACCCAGACAACCTATACACCCGGCTAAC 780
      *** * ** * ** * ***** * ***** ***** ** ** ** ***** ** **
```

```
Z96064.1 CCCCTAAACACTCCCCCCCACATCAAACCAGAATGATATTTCTATTTGCATACGCCATC 840
Z96066.1 CCCCTAAACACACCTCCTCACATCAAACCAGAGTGATACTTCCTATTTCGCCTACGCTATC 840
Z96057.1 CCTCTAAATACCCCTCGCCACATTAAACCAGAATGATACTTTCTATTTCGCATACGCTATC 840
Z96062.1 CCCCTAAACACCCCTCCCCACATCAAACCAGAGTGATACTTCCTATTTCGCATACGCTATC 840
      ** ***** ** ** * ***** ***** ***** ** ***** ** ***** **
```

```

Z96064.1 CTACGCTCTATTCCCTAACAACTCGGAGGAGTACTAGCCCTAGTATTATCCATCTTAATC 900
Z96066.1 CTACGTTCTATCCCCAATAAACTCGGAGGAGTGTTAGCTTTAGTACTATCCATCCTTGTC 900
Z96057.1 CTACGCTCTATCCCCAATAAACTTGGAGGAGTACTAGCCCTAGTACTATCTATCCTAATC 900
Z96062.1 CTACGCTCTATCCCCAATAAACTTGGAGGCGTACTAGCCCTAGTATTATCCATCCTAATC 900
***** ***** ** ** ***** ***** ** ***** ***** ***** ** **

Z96064.1 CTGATCTTCTTACCCTTCATCCACACCTCCAAACAACGAAGCCTAATATTCCGCCCAATC 960
Z96066.1 CTAGCCTTACTACCTCTCATGCACACCTCAAAACAACGAAGCCTTATGTTTCGCCCAATT 960
Z96057.1 CTAATTCTCATACCCTTAATCCACACCTCAAAACAACGAAGCTTAATATTCCGCCCAATT 960
Z96062.1 CTGATTCTCATACCCTTCATCCACACATCAAAACAACGAAGCTTAATGTTCCGCCCAATT 960
**          * ***** ** ***** ***** * ** ** *****

Z96064.1 AGCCAAACCATATTTTGAATCTTAGCAGCAAACCTTCTTATCCTAACCTGAATCGGAGGA 1020
Z96066.1 AGCCAAATCCTATTTTGAATACTAACAGCTAATCTTCTAATTCTAACTTGAATTGGAGGA 1020
Z96057.1 AGCCAAACCATATTTCTGAATCCTAACAGCTAACCTACTCATCCTAACCTGAATCGGAGGA 1020
Z96062.1 AGTCAGGCCATATTTCTGAATCCTAACAGCTAACCTACTTATCCTAACCTGAATCGGAGGG 1020
** **      * ***** ***** ** ***** ** ** ** ***** ***** *****

Z96064.1 CAACCTGTAGAACACCCCTTCATTATTATCGGCCAACTAGCCTCAATCAGCTATTTCTCC 1080
Z96066.1 CAACCCAGTAGAACACCCCTTCATTATCATTGGACAAGTAGCCTCAATCAGCTACTTCTCT 1080
Z96057.1 CAACCCGTAGAACATCCCTTTATCATTATTGGCCAACCTAGCTTCAATCAGCTATTTCTCC 1080
Z96062.1 CAACCCGTAGAACACCCCTTTATCATCATCGGCCAACTAGCTTCAATCAGCTATTTCTCC 1080
***** ***** ***** ** ** ** ** ***** ***** ***** *****

Z96064.1 ATTATTNTAATTCTAATACCAATCTCAGGCCTAATCTAAAACAAAATANTAAAAATAAAA 1140
Z96066.1 ATTATTTTAATCCTAATACCAATCTCAGGACTAATCGAAAACAAAATACTAAACT---T 1137
Z96057.1 ATCATTTTAATCCTAATACCAATCTCAGGACTAATTGAAAATAAAATACTAAAATTAAAC 1140
Z96062.1 ATTATTTTAATTTTAATACCAATCTCAGGACTAATTGAAAACAAAATACTAAAATTAAAC 1140
** *** ***** ***** ***** ***** ***** ***** ***** *

Z96064.1 T 1141
Z96066.1 T 1138
Z96057.1 T 1141
Z96062.1 T 1141
*
```

W pierwszym porównaniu nie występują przerwy, w drugim jest ich 3. Sekwencje porównywane są tej samej lub podobnej długości. W znacznej większości niezakonserwowanych kolumn występuje tylko jeden mismatch, co świadczy o dużym podobieństwie porównanych sekwencji.