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) czasowej

findCentralSequence – 7 przypisań poza pętlami, 11 przypisań w pętlach, każda o rozmiarze ilości możliwych dopasowani sekwencji. Dla liczby sekwencji k ilość tych dopasowani 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 n!. Złożoność czasowa x! *(n*m)

3. Przykładowe dopasowanie dla sekwencji powiązanych ewolucyjnie

Porównano sekwencje mitochondrialnego cytochromu b poniższych organizmów:

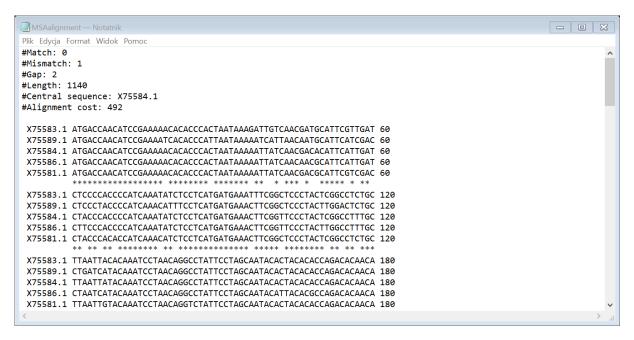
X75581 - Antarctic minke whale

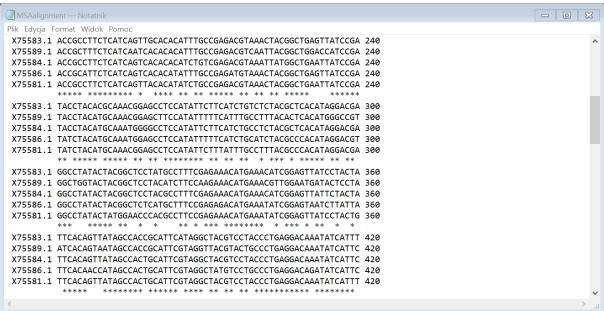
X75589.1 - sperm whale

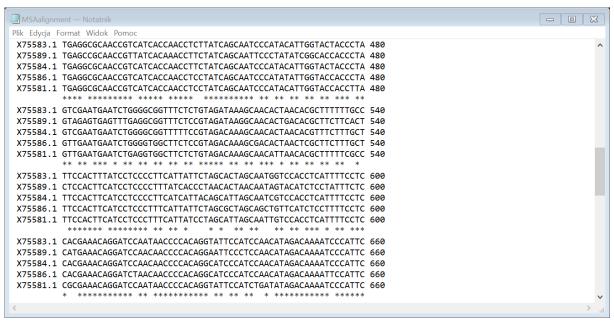
X75584.1 - humpback whale

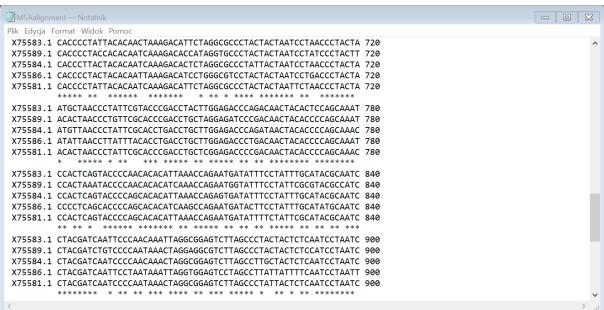
X75586.1 - pygmy right whale

X75583.1 - pygmy Bryde's whale









```
MSAalignment — Notatnik
X75583.1 CTAGCCTTAATCCCAATACTCCACACATCTAAACAACGAAGCATAATGTTCCGACCCTTT 960
X75589.1 CTAGTATTCATCCCAATACTTCATACAGCCAAACAACGAAGCATAATATTCCGACCCTTC 960
X75584.1 CTAGCTTTCATCCCAATACTCCACACATCCAAACAACGAAGCATGATGTTTCGACCCTTT 960
X75586.1 CTAGCCCTCATTCCAATACTCCACACATCCAAACAACGAAGCATAATGTTTCGACCTTTC 960
X75581.1 CTAGCCTTCATCCCAATACTCCACACATCCAAACAACGAAGCATAATATTCCGACCCTTT 960
X75583.1 AGCCAATTCCTATTTTGAGTCCTAATTGCAGACTTACTAACCCTGACATGAATCGGCGGC 1020
X75589.1 AGTCAATTCCTGTTCTGAACACTAATCATAGACCTACTAACCTCTAACATGAATCGGAGGC 1020
X75584.1 AGCCAGTTTCTGTTTTGAATACTAGTAGCAGACCTATTAGCCCTAACATGAATCGGCGGC 1020
X75586.1 AGCCAATTCCTATTCTGAGTGCTAGTCGCAGACCTACTAACCCTAACATGAATTGGTGGC 1020
X75581.1 AGCCAATTTCTATTCTGAGTACTAGTCGCAGACCTATTAACCCTAACATGGATTGGTGGC 1020
X75589.1 CAACCCGTGGAACACCCATATGTAACTGTAGGCCAACTAGCCTCCATCCTATACTTTCTT 1080
****** * ** ****** **
X75583.1 CTAATTCTAGTACTAATACCAGTAACTAGTCTTATCGAGAATAAACTTATAAAATGAAGA 1140
X75589.1 TTAATCCTAATCCTAATACCAACAGCCAGTCTTATCGAGAACAAGCTCTTAAAATGAAGA 1140
X75584.1 TTAATCCTAGTATTAATACCAATAACTAGTCTTATCGAGAACAACTTATAAAATGAAGA 1140
X75586.1 CTGATCTTAATCCTAATGCCAGTAACCAGTCTTATCGAAAATAAACTTATAAAATGAAGA 1140
X75581.1 TTAATCCTAGTACTAATACCAGTAGCTAGCCTTATCGAGAACAAGCTTATAAAATGAAGA 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: 428
  Z96066.1 ATGAAAAACATACGAAAATCACCCCACTACTAAAAATTATTAATCACGCATTTATTGAC 60
  Z96057.1 ATGAAAATTATACGAAAAACACCCCACTCATAAAAATTATCAACCATGCATTCATCGAC 60
  Z96062.1 ATGAAAATCATACGAAAAACACACCCACTCCTAAAAATCATCAACCACGCATTCGTCGAC 60
          Z96064.1 CTCCCAGCACCCTCTAANATCTCATCATGATGAAACTTTGGCTCCCTACTGGGAGTCTGC 120
  Z96066.1 CTACCCGCACCATCCAATATCTCATCATGATGAAACTTTGGGTCCCTTCTAGGAGTCTGT 120
  Z96057.1 CTCCCTGCACCCTCCAACATCTCATCATGATGAAACTTCGGCTCCCTATTGGGAGTATGC 120
  Z96062.1 CTCCCTGCACCCTCCAACATCTCATCATGATGAAACTTCGGCTCTCTATTAGGGGTATGC 120
          ** ** **** ** ** ** ** **********
```

Command Window

25000111	CTAGTAATACAAATCCTAACAGGCTTATTCCTAGCATTACATTACACCTCTGACACCACC CTAGTAGCACAAATTATCACAGGACTATTTCTAGCAATACACTACACCTCTGATACCACA	
Z96057.1	CTAATAATCCAAATCCTTACAGGCCTATTCTTAGCAATACACTACACCTCCGATACTACA	180
Z96062.1	$\tt CTAGTAATACAAATCCTCACAGGCCTATTCTTAGCAATACACTATACCTCAGATACTATA$	180
	*** **	
Z96064.1	${\tt ACAGCGTTCTCATCCGTAACCCACATCTGCCGAGACGTAAACTACGGATGACTAATTCGT}$	240
Z96066.1	${\tt ACAGCATTTCATCTGTAACTCATATCTGCCGAGATGTAAACTACGGTTGATTGA$	240
Z96057.1	ACAGCATTCACATCCGTAACCCACATCTGCCGAGACGTAAACTACGGCTGATTAATTCGA	240
Z96062.1	ACAGCATTCTCATCCGTAACCCACATCTGCCGAGACGTAAACTACGGATGACTAATTCGA	240
	***** ** **** **** ***** ** ******* ****	
Z96064.1	TACCTCCATGCAAACGGGGCCTCCATATTCTTCATATGCCTATTCATACACGTAGGACGA	300
Z96066.1	TATCTTCATGCAAACGGAGCCTCAATATTCTTCATGTGTCTATTTATACACGTAGGACGA	300
Z96057.1	TATCTCCATGCAAACGGAGCCTCCATATTCTTCGTATGCTTGTTTATACACGTAGGACGA	300
Z96062.1	TATCTCCATGCAAACGGAGCCTCCATATTCTTCGTATGTTTGTT	300
	** ** ******** **** ***** * ** * * * * *	

Command Window

U	minana winac	JVV	
	Z96064.1	$\tt GGAATCTACTACGGGTCCTATACCTTCATAGAAACCTGAAACATTGGCATCATTCTACTG$	360
	Z96066.1	$\tt GGAATCTACTACGGATCCTATACCTTTACAGAAACCTGAAATATTGGCATCATTCTACTG$	360
	Z96057.1	$\tt GGCATCTACTATGGATCTTACACCTTTACAGAAACATGAAACCTTGGTGTTATTCTACTA$	360
	Z96062.1	$\tt GGCATTTATTATGGATCCTACACCTTTACAGAAACATGAAACCTCGGCGTTATTCTATTG$	360
		** ** ** ** ** ** ** ** ** ** ** ** **	
	Z96064.1	$\tt TTCGCCGTAATAGCAACTGCATTTATAGGATACGTTCTCCCATGGGGACAAATATCCTTC$	420
	Z96066.1	$\tt TTCGCCGTAATAGCAACTGCATTTATAGGATATGTGCTTCCCTGAGGACAAATATCCTTC$	420
	Z96057.1	$\tt TTTGCCGTAATAGCAACTGCATTTATAGGATATGTCCTTCCATGAGGACAAATATCCTTC$	420
	Z96062.1	$\tt TTCGCCGTAATAGCAACTGCATTTATAGGATATGTCCTCCCATGGGGACAAATATCCTTC$	420
		** ********************************	
	Z96064.1	${\tt TGAGGGGCCACAGTCATTACAAATCTACTCTCAGCTATCCCTTACATAGGAACTAATCTA}$	480
	Z96066.1	TGAGGAGCCACAGTTATTACTAACCTTCTCTCAGCAGTTCCCTACATCGGTACTAATTTA	480
	Z96057.1	TGAGGAGCCACAGTCATTACAAACCTCCTTTCAGCTATTCCCTACATCGGTACTAACCTA	480
	Z96062.1	TGAGGAGCTACAGTCATTACAAACCTACTATCAGCTATTCCCCTATATCGGCACTAACCTA	480
		····· ·· ·· ··· · · · · · · · · · · ·	

Command Window				
	Z96064	.1 GTAGAATGGATCTGAGGTGGCTTTTCAGTAGACAAAGCCACTTTAACACGTTTCTTCGCC	540	
	Z96066	.1 GTAGAATGAATCTGAGGGGGGTTCTCAGTAGACAAAGCCACTCTTACACGCTTCTTCGCA	540	
	Z96057	.1 GTAGAATGAATCTGAGGTGGGTTTTCAGTAGACAAAGCCACCCTCACACGCTTCTTCGCA	540	
	Z96062	.1 CTACAATGAATTTGAGGTGGGTTCTCAGTAGACAAAGCCACCCTTACACGCTTCTTCGCA	540	
		** **** ** ***** ** ** ********** * ****		
	Z96064	.1 TTTCACTTCATNCTACCNTTTATCATTGCAGCTCTAGCTATGGTCCACCTACTATTTCTT	600	
	Z96066	.1 TTCCATTTTATCCTCCCATTCATCATTGCAGCCCTAGCTATAGTCCACCTACTTTTTCTC	600	
	Z96057	.1 TTCCACTTCATCCTTCCATTTATCATTGCAGCCCTAGCCATAGTACACCTACTATTCCTC	600	
	Z96062	.1 TTCCACTTTATCCTCCCCTTTATCATTGCAGCCCTAGCCATAGTACACCTTCTATTCCTT	600	
		** ** ** ** ** ** ** ** ******* **** ** ** ** ** **		
	Z96064	.1 CACGAAACAGGATCTAATAANCCCTCCGGAATAAACTCCGATGCAGACAAAATCCCCTTT	660	
	Z96066	.1 CACGAAACTGGATCCAATAACCCCACCGGAATCAACTCTGATGCAGACAAAATTCCATTC	660	
	Z96057	.1 CACGAAACAGGCTCTAATAACCCCCACAGGAATGAACTCTGACGCAGACAAAATCCCCTTC	660	
	Z96062	.1 CACGAAACTGGCTCTAATAACCCCCACAGGAGTGAACTCCGACGCAGACAAAATCCCATTC	660	
		****** ** ** **** *** * *** * **** * * *		
7.0	96064 1	CACCCCTATTACACAATTAAAGATCTCCTAGGCATATTCCTAATATTAGCAGCCCTACT	rc 720	
		CACCCGTACTACACTATAAAAGACTTCCTAGGGGCCCTACTCATATTAGTAGTTCTACT		
		CACCCCTACTATACAATCAAAGACCTCTTAGGCGCACTCCTCATACTACTAGTTCTCCT		
43	96062.1	CACCCCTACTACACAATTAAAGACCTTTTAGGTGTATTCCTAATACTATTAATTCTCCT		
		***** * * * * * * * * * * * * * * * * *		
_,			~ 500	
		CTATTAGTCCTCTTCTCCCCAGACCTGTTAGGAGACCCCGACAACTACACCAGCCAA		
		CTACTTGTTTTATTTTCCCCAGATATTTTAGGCGACCCAGATAATTATACACCAGCTAA		
ZS	96057.1	CTATTAGTACTATTTTCCCAGACTTATTAGGAGACCCCGACAACTACACCAGCCAA	C 780	
ZS	96062.1	CTATTAGTATTATTCTCCCCAGACCTATTAGGAGACCCAGACAACTATACACCGGCTAA	C 780	
		*** * ** * * * * * **** * **** * * * * *	*	
ZS	96064.1	$\tt CCCCTAAACACTCCCCCCACATCAAACCAGAATGATATTTCCTATTTGCATACGCCATGATATTTCCTATTTGCATACGCCATGATATTTCCTATTTTGCATACGCCATGATATTTCCTATTTTGCATACGCCATGATATTTTTTTT$	C 840	
ZS	96066.1	CCCCTAAACACACCTCCTCACATCAAACCAGAGTGATACTTCCTATTCGCCTACGCTAT	C 840	
ZS	96057.1	CCTCTAAATACCCCTCGCCACATTAAACCAGAATGATACTTTCTATTCGCATACGCTAT	C 840	
ZS	96062.1	CCCCTAAACACCCCTCCCCACATCAAACCAGAGTGATACTTCCTATTCGCATACGCTAT	C 840	
		** **** ** ** * **** ***** **** ** **** ** ****	*	

Z96064.1	$\tt CTACGCTCTATTCCTAACAAACTCGGAGGAGTACTAGCCCTAGTATTATCCATCTTAATC$	900
Z96066.1	$\tt CTACGTTCTATCCCCAATAAACTCGGAGGAGTGTTAGCTTTAGTACTATCCATCC$	900
Z96057.1	$\tt CTACGCTCTATCCCCAATAAACTTGGAGGAGTACTAGCCCTAGTACTATCTAT$	900
Z96062.1	$\tt CTACGCTCTATCCCCAATAAACTTGGAGGCGTACTAGCCCTAGTATTATCCATCC$	900
	***** **** ** ** ** **** **** ** *** *** *** *** ***	
Z96064.1	CTGATCTTCTTACCCTTCATCCACACCTCCAAACAACGAAGCCTAATATTCCGCCCAATC	960
Z96066.1	CTAGCCTTACTACCTCTCATGCACACCTCAAAACAACGAAGCCTTATGTTTCGCCCAATT	960
Z96057.1	CTAATTCTCATACCCTTAATCCACACCTCAAAACAACGAAGCTTAATATTCCGCCCAATT	960
Z96062.1	CTGATTCTCATACCCCTCATCCACACATCAAAACAACGAAGCTTAATGTTCCGCCCAATT	960
	**	
796064 1	AGCCAAACCATATTTTGAATCTTAGCAGCAAACCTTCTTATCCTAACCTGAATCGGAGGA	1020
	AGCCAAATCCTATTTTGAATACTAACAGCTAATCTTCTAATTCTAACTTGAATTGGAGGA	
	AGCCAAACCATATTCTGAATCCTAACAGCTAACCTACTCATCCTAACCTGAATCGGAGGA	
	AGCCAGACCATATTCTGAATCCTAACAGCTAACCTACTTATCCTAACCTGAATCGGAGGG	
296062.1	** **	1020
796064 1 (CAACCTGTAGAACACCCCTTCATTATTATCGGCCAACTAGCCTCAATCAGCTATTTCTCC 1080	
	CAACCAGTAGAACACCCCTTCATTATCATTGGACAAGTAGCCTCAATCAGCTACTTCTCT 1080	
	CAACCCGTAGAACATCCCTTTATCATTATTGGCCAACTAGCTTCAATCAGCTATTTCTCC 1080	
Z96062.1	CAACCCGTAGAACACCCCTTTATCATCATCGGCCAACTAGCTTCAATCAGCTATTTCTCC 1080	
	**** ****** **** ** ** ** ** ** *** ****	
	ATTATTNTAATTCTAATACCAATCTCAGGCCTAATCTAAAACAAAATANTAAAAATAAAA 1140	
	ATTATTTTAATCCTAATACCAATCTCAGGACTAATCGAAAACAAAATACTAAAACTT 1137	
	ATCATTTTAATCCTAATACCAATCTCAGGACTAATTGAAAATAAAATACTAAAATTAAAC 1140 ATTATTTTAATTTTAATACCAATCTCAGGACTAATTGAAAACAAAATACTAAAATTAAAC 1140	
	** *** **** ************* **** **** ****	
Z96064.1	Т 1141	
Z96066.1	T 1138	
Z96057.1	T 1141	

Z96062.1 T 1141