

Lab 11

Artur Sak (sak2)

April 13, 2017

1 Sequence Alignment

1. Calculate, by hand, the dynamic programming matrix and an optimal alignment for the DNA sequences CTTAAG and CTAATG, scoring +2 for a match, -1 for a mismatch, and with a gap penalty of 2 (i.e., each gap column contributes -2).

-	-	C	T	T	A	A	G
-	0	-2	-4	-6	-8	-10	-12
C	-2	2	0	-2	-4	-6	-8
T	-4	0	4	2	0	-2	-4
A	-6	-2	2	3	1	2	0
A	-8	-4	0	4	5	3	1
T	-10	-6	-2	2	6	4	2
G	-12	-8	-4	0	4	5	6

Optimal Alignment:

T T A A G - -
- T A A - T G

2. Calculate, by hand, the dynamic programming matrix and an optimal alignment for the DNA sequences CTTAAG and CTAATG, scoring +1 for a match, -1 for a mismatch, and with a gap penalty of 3 (i.e., each gap column contributes -3). Did you get the same result as in a)?

-	-	C	T	T	A	A	G
-	0	-3	-6	-9	-12	-15	-18
C	-3	1	-2	-5	-8	-12	-14
T	-6	-2	2	-1	-4	-7	-10
A	-9	-5	-1	1	-2	-3	-6
A	-12	-8	-4	0	2	1	-4
T	-15	-11	-7	-3	-1	1	-2
G	-18	-14	-10	-6	-2	0	2

Optimal Alignment:

```
T T A A G - -  
- T A A - T G
```

This is the same optimal alignment as in part (1).

3. Run your program on seq1.fna, seq2.fna, sub.txt, gap penalty -8 and include the output in your report.

```
▲ ECE 398/Labs/Lab11-2017-ECE398BD cat align1.txt
The optimal alignment between the given sequences has score 344.
GCAGATTCCDDCTAGACCCGCCGCCACCATGGTCAGGCATGCCCGCTGGGCACAGCCDAG-AGGGGTATAAACAGTGTGGAGGGTGGCCAGGCCAGCTGAGTCCTGAGCAGCAGCCCA
GCCGAGCCACCCGA

--A-A-----A-ACCG-CCTG----ATTTT-AGTCC--CGG-CGGATCTTAGTCCCGTAGCGAAAAATCTTTTCT--ATTCT-T--A-T---TCITAGTCCCG-T-AGC-G--C-
TTTTAG---TCG-
```

4. Run your program on BRCA1 part mutated.fna, BRCA1 part.fna, sub.txt, and gap penalty -8. Include the output in your report.

```
▲ ECE 398/Labs/Lab11-2017-ECE398BD cat align2.txt
The optimal alignment between the given sequences has score 3991.
G-A-A-TT---CT--A---TCT--T---ATTTTCAGGTATACAATAAAGTATTTTAAACTATGTCCACATTAGAGTGCACCCAGATCTGCAGAACATATTCTGCTTTATATAACTGATGTT
GGTCATGAGAATGTGCTAGTTCTCCTTTACCCCAACCCCAACGACTATCCGAGTCAGCGGCTTTTCTCTCTTTGTGGAATCTTCAGAGCATTCGAGATATACAGTCTCAAAAAATCT
-ATAAAATCCCTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGGCAATATGTAATTCAGGAGAAGAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTACCTGCAATGGG
GGCGATAGTAGATACAGGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGCTCTTGGCAATGTAGAGTCCCAATGT-A-----A---T--T--GAG-T---TA-AA-TGCTC
TT-T-C--CAAGAGAAGACCCACCTTAAATCTGAGTGTGCATAAATCTATTTAGCTAAATAACATCACACATTAAAAATTTTCTCAGCT-G--G-A-A--TAATAAAA

GAACATTTAAGATCCACTTTTCTCAGCAAAATTCAGGTATACAATAAAGTATTTTAAACTATGTCCACATTAGAGTGCACCCAGATCTGCAGAACATATTCTGCTTTATATAACTGATGTT
GGTCATGAGAATGTGCTAGTTCTCCTTTACCCCAACCCCAACGACTATCCGAGTCAGCGGCTTTTCTCTCTTTGTGGAATCTTCAGAGCATTCGAGATATACAGTCTCAAAAAATGG
CAT-CTGATTTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGGCAATATGTAATTCAGGAGAAGAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTA-C--C--T---
-GC---A---A-ATADAGAGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGCTCTTGGCAATGTAGAGTCCCAATGTAAATTTTACTTTAATGAGAGTTAAATATAAGTGTCT
TTGGACAGAAAGAGAAGACCCACCTTAAATCTGAGTGTGC-----A--T-TTTTAGCTAAATAACATCACACATTAAAAATTTTCTCAGCTGGAAAGTAGAGGTAATAAAAA
```

5. Run your program on BRCA1 part mutated.fna, BRCA1 part.fna, sub.txt, and gap penalty -1. Include the output in your report.

```
▲ ECE 398/Labs/Lab11-2017-ECE398BD cat align3.txt
The optimal alignment between the given sequences has score 463.
G-A-ATTCT---AT---C--T-T-----ATTTBAGGTATACAATAAAGTATTTTAAACTATGTCCACATTAGAGTGCACCCAGATCTGCAGAACATATTCTGCTTTATATAACTGATGTT
GGTCATGAGAATGTGCTAGTTCTCCTTTACCCCAACCCCAACGACTATCCGAGTCAGCGGCTTTTCTCTCTTTGTGGAATCTTCAGAGCATTCGAGATATACAGTCTCAAAAAATG
-CTAT-AAATCCCTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGGCAATATGTAATTCAGGAGAAGAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTACCTGCAAT
GGGGGGATAGTAGATACAGGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGCTCTTGGCAATGTAGAGTCCCAATGT-A-----A---T--T--GAG-T---TA-AA-TG
CTCTT--TC--CAAGAGAAGACCCACCTTAAATCTGAGTGTGCATAAATCTATTTAGCTAAATAACATCACACATTAAAAATTTTCTCAGCT-G---G-A-A--TAATAAAA

GAACATTTAAGATCCACTTTTCTCAGCAAAATTCAGGTATACAATAAAGTATTTTAAACTATGTCCACATTAGAGTGCACCCAGATCTGCAGAACATATTCTGCTTTATATAACTGATGTT
GGTCATGAGAATGTGCTAGTTCTCCTTTACCCCAACCCCAACGACTATCCGAGTCAGCGGCTTTTCTCTCTTTGTGGAATCTTCAGAGCATTCGAGATATACAGTCTCAAAAAATG
GG-ATGCTAT--TTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGGCAATATGTAATTCAGGAGAAGAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTA-C--C--T
-----GC---A---A-ATADAGAGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGCTCTTGGCAATGTAGAGTCCCAATGTAAATTTTACTTTAATGAGAGTTAAATATAAGTGT
CTCTTGGACAGAAAGAGAAGACCCACCTTAAATCTGAGTGTGC-----A--T-TTTTAGCTAAATAACATCACACATTAAAAATTTTCTCAGCTGGAAAGTAGAGGTAATAAAAA
```

2 BLAST (Basic Local Alignment Search Tool)

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Homo sapiens chromosome 17, GRCh38.p7 Primary Assembly	1035	1699	100%	0.0	100%	NC_000017.11
<input type="checkbox"/>	Homo sapiens chromosome 17 genomic scaffold, GRCh38.p7 alternate locus group ALT_REF_LOCI.2.HSCHR17.6.CT64	1035	1699	100%	0.0	100%	NW_003871093.1
<input type="checkbox"/>	Homo sapiens chromosome 17 genomic scaffold, GRCh38.p7 alternate locus group ALT_REF_LOCI.1.HSCHR17.5.CT64	1014	1014	100%	0.0	99%	NW_003871092.1

Sequences producing significant alignments:

Select: All None Selected: 0

Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	Homo sapiens chromosome 17, GRCh38.p7 Primary Assembly	625	1081	96%	1e-176	89%	NC_000017.11
<input type="checkbox"/>	Homo sapiens chromosome 17 genomic scaffold, GRCh38.p7 alternate locus group ALT_REF_LOCI.2.HSCHR17.6.CT64	625	1081	96%	1e-176	89%	NW_003871093.1
<input type="checkbox"/>	Homo sapiens chromosome 17 genomic scaffold, GRCh38.p7 alternate locus group ALT_REF_LOCI.1.HSCHR17.5.CT64	604	604	96%	1e-170	88%	NW_003871092.1

3 RNA fold prediction

1. (See end of document for graphics)
2. (See end of document for graphics)
3. The results from parts (1) and (2) do not appear to be the same.
4. The results from part (1) and (2) with the temperature raised to 100 degrees centigrade do not appear to be the same. The increased temperature appears to have a rather substantial impact on the stability of the RNA molecules. It seems that the high temperature prevents the proteins from bonding which keeps the strings from folding (See end of document for graphics).

Range 1: **41276801 to 41277360** [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
1035 bits(560)	0.0	560/560(100%)	0/560(0%)	Plus/Minus

Features: [594 bp at 5' side: keratin-associated protein 9-7](#)
[24466 bp at 3' side: keratin-associated protein 29-1](#)

Query	1	Tgaacattttaagatccactttctcagcaaatcttcaggtatacaataaaagtatttttaaact	60
Sbjct	41277360	TGAACATTTAAGATCCACTTTTCTCAGCAAAATTCAGGTATACAATAAAAGTATTTTAAACT	41277301
Query	61	atgtccacattagagtgcaccagatctgcagaacatattcgtcttatataactgatgttg	120
Sbjct	41277300	ATGTCCACATTAGAGTGCACCAGATCTGCAGAACATATTCGTCTTATATAACTGATGTTG	41277241
Query	121	gtCATGAGAATGTGCTAGTTCCTTTACCCACCACCCAACGACTATCCCAGTCAGCGG	180
Sbjct	41277240	GTCATGAGAATGTGCTAGTTCCTTTACCCACCACCCAACGACTATCCCAGTCAGCGG	41277181
Query	181	CTTTTCTCTCTTTCTGGAATCTTCAGAGCATTCCAGATATACAGTCTCAAAAGATGGCAT	240
Sbjct	41277180	CTTTTCTCTCTTTCTGGAATCTTCAGAGCATTCCAGATATACAGTCTCAAAAGATGGCAT	41277121
Query	241	CTGTATTTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGCCAATATGTATTTCAGGAGAA	300
Sbjct	41277120	CTGTATTTTTTAAGGAATAACTTAAGTTCATTTTGCCTTGCCAATATGTATTTCAGGAGAA	41277061
Query	301	GAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTACCTGCAAATACAGGAGA	360
Sbjct	41277060	GAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTTACCTGCAAATACAGGAGA	41277001
Query	361	TTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGAGTCCCAATGT	420
Sbjct	41277000	TTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGAGTCCCAATGT	41276941
Query	421	AATTTTACTTTAATGAGAGTTAAATATAAGTCTCTTGGACAGAAAGAGAAGACCCAAC	480
Sbjct	41276940	AATTTTACTTTAATGAGAGTTAAATATAAGTCTCTTGGACAGAAAGAGAAGACCCAAC	41276881
Query	481	CTTAAATCTGAGTCTGCATTTTCTAGCTAAATAACATCACACATTAAAATTTTCCTCAGCT	540
Sbjct	41276880	CTTAAATCTGAGTCTGCATTTTCTAGCTAAATAACATCACACATTAAAATTTTCCTCAGCT	41276821
Query	541	GGAAAGTAGAGGTAATAAAA 560	
Sbjct	41276820	GGAAAGTAGAGGTAATAAAA 41276801	

Range 2: **41272408 to 41272963** [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
664 bits(359)	0.0	505/571(88%)	27/571(4%)	Plus/Minus

Features: [6548 bp at 5' side: keratin-associated protein 9-6](#)
[2735 bp at 3' side: keratin-associated protein 9-7](#)

Query	2	gaacattttaagatccactttctcagcaaatcttcaggtatacaataaaagtattttaaacta	61
Sbjct	41272963	GAATATTTAAGATCCACTTTTCTCAGCAAAATTCAGGTGTACAAAAAGTATTTTAAACTA	41272904
Query	62	tgtccacattagagtgcaccagatctgcagaacatattcgtcttatataactgatgttgg	121
Sbjct	41272903	TATCCAAATTAGTGTAACACAGATCTGCAAACTTATTTGTCTTATATAACTGATGTAGG	41272844
Query	122	tCA-TGAGAATGTGCTAGTTC-----CCTTTACCCACCACCCAACGACTATCCCAGT	174
Sbjct	41272843	TCAAAGAGAATGTGCTAGTTCCTCGTCTCCTTTACCCCATCACCAACGACTATCCCAGT	41272784
Query	175	CAGCGGCTTTTCTCTCTTTCTGGAATCTTCAGAGCATTCCAGATATACAGTCTCAAAAGA	234
Sbjct	41272783	CAGCGGCTTTTCTCTCTCTCTGGAATCTTCACAGCATCCAGATATACAGTCTCAAAAGA	41272724
Query	235	TGGCATCTGTATTTTAAAGGAATAACTTAAGTTCATTTTGCCTTGCCAATATGTATTCA	294
Sbjct	41272723	TGGCATCTGTATTTTATAAGGAATAATTAAGTTCATTTTGCCTTGCAATATGTATTCA	41272664
Query	295	GGAGAAGAAG---GAAGCCATATAAGAATGGAAATGTTATCTGCAAAATGTACCTGCAAA	351
Sbjct	41272663	GGAGAAGAAGAAGGAAGCCATATAAGAATGGAAATGTTATCTGCAAAAT-----	41272616
Query	352	TACAGGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGAG	411
Sbjct	41272615	-ACAGGAGATTATTCCTCCATTGTTAAGTAAACTTAAGTGCATATCTTGCTAATGCTGAG	41272557
Query	412	TCCCAATGTAATTTTACTTTAATGAGAGTTAAATATAAGTGTCTTTGGACAGAAAGAGA	471
Sbjct	41272556	TCCCAATATAGTTTTTCTTTATGAGAGTTAAATGTAAGTGCTTTGGACAGAAAAA-A	41272498
Query	472	AGACCCAACCTTAAATCTGAGTCTG--CATTTTCTAGCTAAATAACATCACACATTAAAAT	529
Sbjct	41272497	AGACCCAACCTTGCATCTGAGTCTGTCCATTTTCTAGCTAAATAACATCACACATTAAAAT	41272438
Query	530	TTTCCTCAGCTGGAAAGTAGAGGTAATAAAA 560	
Sbjct	41272437	TTTT-TCAGCTGGAAAGTAGAGGTAATAAAA 41272408	

Range 1: **41276817 to 41277331** [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
625 bits(338)	1e-176	482/541(89%)	52/541(9%)	Plus/Minus

Features: [610 bp at 5' side: keratin-associated protein 9-7](#)
[24495 bp at 3' side: keratin-associated protein 29-1](#)

Query	14	atttcaggtatacaataaagtatttttaaactatgtccacattagagtgaccagatctgc	73
Sbjct	41277331	ATTTCAGGTATACAATAAAGTATTTTAAACTATGTCCACATTAGAGTGCACCAGATCTGC	41277272
Query	74	agaacatattcgctcttatataactgatgttggtCATGAGAATGTGCTAGTTCTCCTTTAC	133
Sbjct	41277271	AGAACATATTCTGCTCTTATATAACTGATGTGTGTCATGAGAATGTGCTAGTTCTCCTTTAC	41277212
Query	134	CCCACCACCAACGACTATCCAGTCAGCGGCTTTTCTCTCTTCTGGAATCTTCAGAGC	193
Sbjct	41277211	CCCACCACCAACGACTATCCAGTCAGCGGCTTTTCTCTCTTCTGGAATCTTCAGAGC	41277152
Query	194	ATTCCAGATATACAGTCTCAAAAAAT--C-TATAAAATCCCTTTTAAGGAATAACTTAAG	250
Sbjct	41277151	ATTCCAGATATACAGTCTCAAAAGATGGCATCTGTATT---TTTAAGGAATAACTTAAG	41277095
Query	251	TTCAATTTGCCTTGCCAAATATGTATTTCAGGAGAAGAAGGAAGCCATATAAAGATGGAAT	310
Sbjct	41277094	TTCAATTTGCCTTGCCAAATATGTATTTCAGGAGAAGAAGGAAGCCATATAAAGATGGAAT	41277035
Query	311	GTTATCTGCAAAATGTTACCTGCAATGGGGCGATAGCTAGATACAGGAGATTACTCCTCC	370
Sbjct	41277034	GTTATCTGCAAAATGTTACCTGCAA-----ATACAGGAGATTACTCCTCC	41276991
Query	371	ATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGAGTCCCAATGTAATT-----	424
Sbjct	41276990	ATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGAGTCCCAATGTAATTTTACT	41276931
Query	425	-----GAG--TTAAAT-----G--CTCTTTC-C---AAGAGAAGACCAACCTTAAATCTG	467
Sbjct	41276930	TTAATGAGAGTTAAATATAAGTGCTCTTGGACAGAAAGAGAAGACCAACCTTAAATCTG	41276871
Query	468	AGTCTGCATAAATCTATTTTAGCTAAATAACATCACACATTAAATTTTCTCAGCTGGA	527
Sbjct	41276870	AGTCTGCAT-----TTTAGCTAAATAACATCACACATTAAATTTTCTCAGCTGGA	41276818
Query	528	A 528	
Sbjct	41276817	A 41276817	

Range 2: **41272616 to 41272935** [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
366 bits(198)	6e-99	282/321(88%)	12/321(3%)	Plus/Minus

Features: [6756 bp at 5' side: keratin-associated protein 9-6](#)
[2763 bp at 3' side: keratin-associated protein 9-7](#)

Query	14	atttcaggtatacaataaagtatttttaaactatgtccacattagagtgaccagatctgc	73
Sbjct	41272935	ATTTCAGGTGTACAAAAAGTATTTTAAACTATATCCAAATTAGTGACACCAGATCTGC	41272876
Query	74	agaacatattcgctcttatataactgatgttggtCA-TGAGAATGTGCTAGTTCT-----	126
Sbjct	41272875	AAAACCTATTGTCTTATATAACTGATGTAGGTCAAAGAGAATGTGCTAGTTCTTCGTCT	41272816
Query	127	CCTTTACCCCAACCAACGACTATCCAGTCAGCGGCTTTTCTCTCTTCTGGAATCT	186
Sbjct	41272815	CCTTTACCCCATCACCAACGACTATCCAGTCAGCGGCTTTTCTCTCTCTGGAATCT	41272756
Query	187	TCAGAGCATTCAGATATACAGTCTCAAAAAAT-CTATAAAATCCCTTTTAAGGAATAAC	245
Sbjct	41272755	TCACAGCATCCAGATATACAGTCTCAAAAGATGGCAT-CTGTATTTATAAGGAATAAT	41272697
Query	246	TTAAGTTCATTTTGCCCTTGCCAATATGTATTTCAGGAGAAGAAG---GAAGCCATATAAGA	302
Sbjct	41272696	TTAAGTTCATTTTGCCCTTGCTCAATATGTATTTCAGGAGAAGAAGAAGGAAGCCATATAAGA	41272637
Query	303	ATGGAAATGTTATCTGCAAAT 323	
Sbjct	41272636	ATGGAAATGTTATCTGCAAAT 41272616	

Range 3: **41272424 to 41272617** [GenBank](#) [Graphics](#)

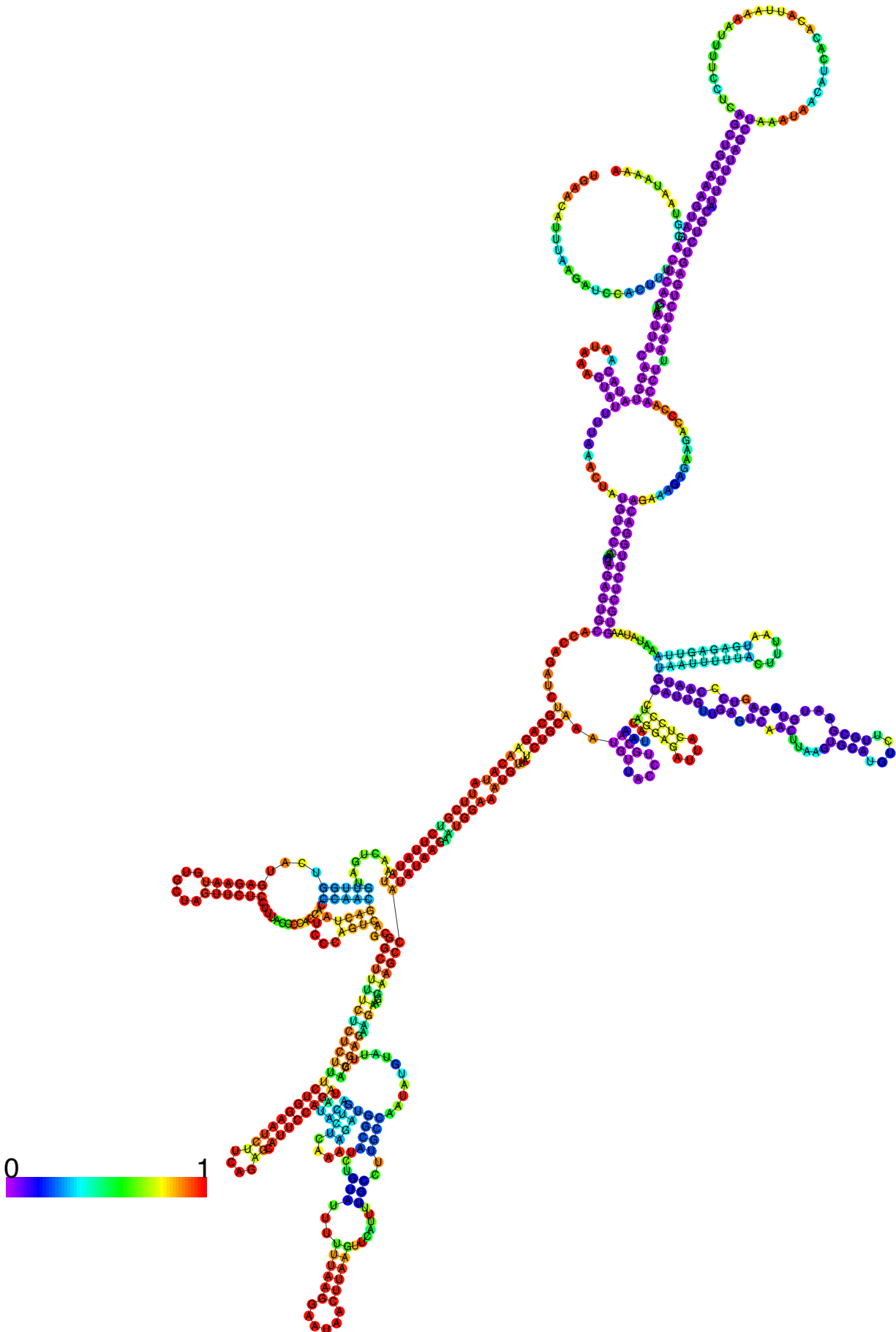
▼ Next Match ▲ Previous Match ▲ First Match

Score	Expect	Identities	Gaps	Strand
89.8 bits(48)	2e-15	154/200(77%)	28/200(14%)	Plus/Minus

Features: [6564 bp at 5' side: keratin-associated protein 9-6](#)
[3081 bp at 3' side: keratin-associated protein 9-7](#)

Query	351	ATACAGGAGATTACTCCTCCATTGTTGAGTCAACTTAAGTGCATGTCTTGCGAATGTAGA	410
Sbjct	41272617	ATACAGGAGATTATCTCCTCCATTGTTAAGTAAACTTAAGTGCATATCTTGCTAATGCTGA	41272558
Query	411	GTCCC-A-AT-G-----TAATT--GAGTTAAATG---CT--CTTT---C-CAAGAGA	448
Sbjct	41272557	GTCCCAATATAGTTTTTCTTTATTGAGAGTTAAATGTAAGTGCCTTTGGACAGAAAAAA	41272498
Query	449	AGACCCAACCTTAAATCTGAGTCTGCATAAATCTATTTTAGCTAAATAACATCACACATT	508
Sbjct	41272497	AGACCCAACCTTGCATCTGAGTCTG--TCCA--T-TTTTAGCTAAATAACATCACACATT	41272443
Query	509	AAAATTTTCTCAGCTGGAA 528	
Sbjct	41272442	AAAATTTT-TCAGCTGGAA 41272424	

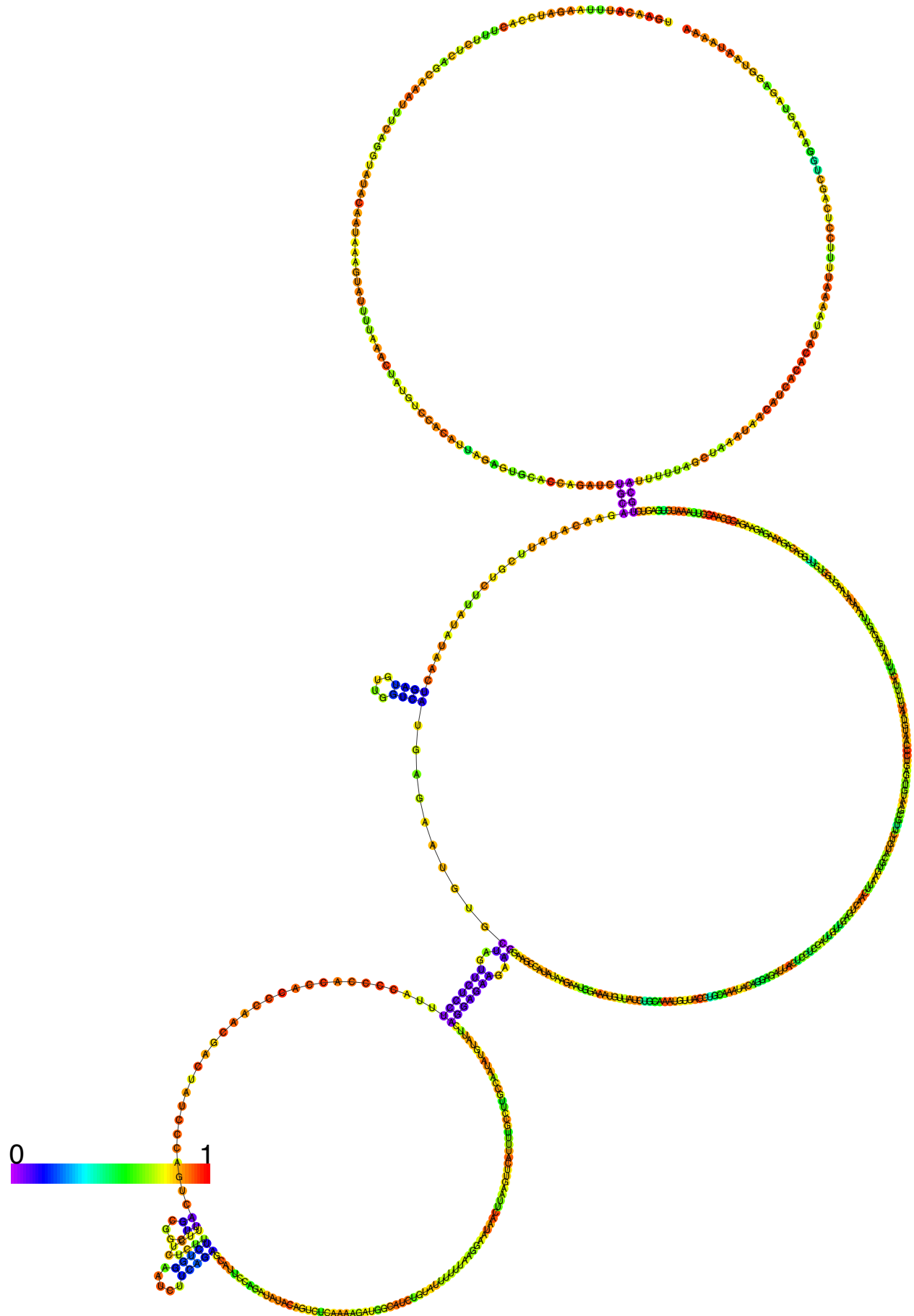
BRCA_part.fna



BRCA_part_mutated.fna



BRCA_part.fna (100 degrees centigrade)



BRCA_part_mutated.fna (100 degrees centigrade)

