

Laborator 2 Alignment

Am lucrat in echipa:

Bratu Dorin - ssmaRRR

Șveț Bogdan - banmepls

TASK 1:

Raspuns cod:

=== Matrice de distanțe (triunghi superior) ===

	NM_000546.6	NM_011640.3	NM_131327.2
NM_000546.6	-	0.722066	0.752799
NM_011640.3	-	-	0.725997
NM_131327.2	-	-	-

Cea mai apropiată pereche:

NM_000546.6 vs NM_011640.3 → distanță = 0.72207 (p-distance (truncated))

Cele mai apropiate secvențe sunt NM_000546.6 (om) și NM_011640.3 (șoarece), cu o p-distanță de 0.7221.

Am folosit p-distance trunchiată la lungimea minimă, deoarece lungimile diferă ușor între specii.

Biologic, aceste două gene TP53 sunt omoloage între mamifere și prezintă conservare evolutivă semnificativă comparativ cu gena de pește-zebră (*Danio rerio*).

TASK 2:

Raspuns cod:

=== Aliniere pairwise ===

NM_000546.6 vs NM_011640.3

>>> Global alignment (Needleman-Wunsch)

```
-----CTCA-----AAAGT-CTAG-AGCCACCGTCCA--GGGAGCA----
G---GTAGCT-GCTGGGCTCC-----GGGGACACTTTGCGTTCGGGCTGGGAGCGTGCT--TTC
CAC-G-ACGGTGACACGCTTC-CCTGGATTGGCAGCCAGACTGCCTTCCGGGTCAGTGCCATGG
AGGAGCCGCAGTCAGATCCT-AGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTAT
GGAAACTACTTCCT---GAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTT
GATGCTGTCCCCGGACGATATTGAACAATGGTTCAGTGAAGACCCAGGTCCAGATGAAGCTCC-
CAGAATGCCAGAG-GCTGCTCCCCCGTGGCCCCTG-CACCAGCAG-CTCCTACACCGGCGGCC
CCTGCACCAGCCCCCTCC-TGGCCCCTGTCATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGC
AGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACT-TGCACGTAC
TCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTG
ATT-CCACACCCCCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACA
TGACGGAGGTTGTGAGGCGCTGCCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCCC
TCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACT
TTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCTGAGGTTGGCTCTGACTGTACCACCATCC
ACTACAACACTACATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCAT
CATCACACTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGT
GCCTGTCTGAGAGAGACCGGCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAG-CCTCAC
CAC-GAGCTGCCCCCAGGGAGCACTAAGCGAGCACTGCCCAACAACACCAGCTCCTCTCCCCAG
CCAAAGAAGAAACCACTGGATGGAGAATATTTACCCCTTCA-GATCCGTGGGCGTGAGCGCTTC
GAGATGTTCCGAGAGCTGAATGAGGCCTTGGAACCTCAAGGATGCCAGGCTGGGA-AGGAGCCA
GGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGCCATA
AAAAACTCA-TGTTCAAGACAGAAGGGCCTGACTCAGACTGACAT---TCTCCACTTCTTGTTT
CCCCTGAC-AGCCTCCCACCCCCATCTCTCCCTCCCCTGCCATTTTGGGTTTTGGGTCTT-TG
AACCCTTGCTTGCAATAGGTGTGCGTCAGAAGCACCAGGACTTCCATTTGCTTTGTCCCGGGG
CTCCACTGAACAAGTTGGCCTGCACTGGTGTGTTTGTGTTGTGGGGAGGAGGATGGGGAGTAGGACA
TACCAGCTTAGATTTTAAAGGTTTTTACTGTGAGGGATGTTTGGGAGATGTAAGAAATGTTCTTG
CAGTTAAGGGTTAGTTTACAATCAGCCACATTCTAGGTAGGGGCCCACTTCACCGTACTAACCA
GGGAAGCTGTCCCTCACTGTTGAATTTTCTCTAACTTCAAGGCCCATATCTGTGAAATGCTGGC
ATTTGCACCTACCTCACAGAGTGCATTGTGAGGGTTAATGAAATAATGTACATCTGGCCTTGAA
ACCACCTTTTATTACATGGGGTCTAGAACTTGACCCCCCTTGAGGGTGCTTGTTCCCTCTCCCTG
TTGGTTCGGTGGGTTGGTAGTTTCTACAGTTGGGCAGCTGGTTAGGTAGAGGGAGTTGTCAAGTC
TCTGCTGGCCCAGCCAAACCCTGTCTGACAACCTCTTGGTGAACCTTAGTACCTAAAAGGAAAT
```

[illegible]

```

||          |||||          ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|||| |  ||  ||||          |  |||||          |||          |.|
||||          |||
|||          ||||  |  |
|||||||.|.|||||||.|.||||.|.||||
||.|.|.|.||||.|.||||.|.||||.|.|||||||.|.||||.|.||||
||| |.|||||          ||  ||  |          |.||          |||||.|.||||
||          |||||
|||.||          ||||  |||          ||  || |.|||| |.|||||          ||  |||
||          ||.||  |||||          ||  |||
||          ||  |||||          |||  ||          |          |.|.|          |          |  |
||          ||          |||  ||          |  |||||.||          |.||
||.|.|  |||||  |  ||          |||  .|.|||||.|  |
|||.||||          |  ||.||||          |

```

```

TTTCCCTCCACGTGCTCACCCTGGCTAAAGTTCT-GTAGCTTCAGTTCATTGGGACCATCCT
GGCTGTAGGTAGC--GACTACAGTTAGGGGGCACCTAGCATTACAGGCCCTCATCCTCCTCCTTC
C-CAGCAGGGTGTACGCTTCTCC--GA-----AGACT-----GGATGACTGCCATGG
AGGAGTCACAGTCGGAT-ATCAGCCTCGAGCTCCCTCTGAGCCAGGAGACATTTTCAGGCTTAT
GGAAACTACTTCCTCCAGAAGATA---TCCTG---CCATCACC-TCAC-TGC-ATGGACGATCT
GTTGCTG-CCCC--AGGATGTTGAGGA---GTTTTTTGAAGGCCCA-----AG-TGAAGC-CCT
CCGAGTGTCAG-GAGCTCCTGCAGCACAGGACCCTGTCACC-G-AGACCCCTGGGCCAGTGGCC
CCTGCCCCAG-CCACTCCATGGCCCCTGTCATCTTTTGTCCCTTCTCAAAAAACTTACCAGGGC
AACTATGGCTTCCACCTGGGCTTCCTGCAGTCTGGGACAGCCAAGTCTGTTA-TGTGCACGTAC
TCTCCTCCCCTCAATAAGCTATTCTGCCAGCTGGCGAAGACGTGCCCTGTGCAGTTGTGGG-TC
AGCGCCACACCTCCAGCTGGGAGCCGTGTCCGCGCCATGGCCATCTACAAGAAGTCACAGCACA
TGACGGAGGTCTGTGAGACGCTGCCCCCACCATGAGCGCTGCTCCGATGGTGATGGCCTGGCTCC
TCCCCAGCATCTTATCCGGGTGGAAGGAAATTTGTATCCCGAGTATCTGGAAGACAGGCAGACT
TTTCGCCACAGCGTGGTGGTACCTTATGAGCCACCCGAGGCCGGCTCTGAGTATACCACCATCC
ACTACAAGTACATGTGTAATAGCTCCTGCATGGGGGGCATGAACCGCCGACCTATCCTTACCAT
CATCACACTGGAAGACTCCAGTGGGAACCTTCTGGGACGGGACAGCTTTGAGGTTTCGTGTTTGT
GCCTGCCCTGGGAGAGACCGCCGTACAGAAGAAGAAAATTTCCGCAA-AAA-GGAAGTCCTTTG
CCCTGAACTGCCCCCAGGGAGCGCAAAGAGAGCGCTGCCCACCTGCACAAGCGCCTCTCCCCCG
CAAAAGAAAAAACCCTTGATGGAGAGTATTTACCC-TCAAGATCCGCGGGCGTAAACGCTTC
GAGATGTTCCGGGAGCTGAATGAGGCCTTAGAGTTAAAGGATGCCCATGCT-ACAGAGGAGTCT
GGAGACAGCAGGGCTCACTCCAGCTACCTGAAGACCAAGAAGGGCCAGTCTACTTCCCGCCATA
AAAAA-ACAATGGTCAAGAAAGTGGGGCCTGACTCAGACTGAC-TGCCTCTGCA--TCCCG-TC
CCCA-TCACCAGCCTCCC-C-----CTCT-CCT----TGC--T-----GTCTTATG
A---CTT-C-----AGG---GC-TGAGA--CA--CA--A--TCC-----TCCCGG--
-TCC-CT-----T----CTG--CTG-----C-
-----CT-----TTTTTAC-----CTTG

```

```

-----TAG-----CTA-----GGGC-----TCA--G-----CC-
-----CCCTC-----TCT-----
-----GAGT--A--GT--GGTT-----C--CTGGC-----
-CCA-----AGT--TGGG-----GAA-----
-----TAG-----GTTG--A--T-----AGTTGTCAGGTC
TCTGCTGGCCCAGCGAAATTCTATC---CAGCCAGTTGTTGGACCCTGGCACCTACAATGAAAT
CTCACCTACCCACACCCTGTAAGA-TTC-TATCTTG-----GG--CC-C-----
-----TCAT----AGGGTCCATATC-----CT-----
-----CCAG-----GGCCTA-----CTT
T-CCT-----TC---CA-TTCTG-CAAAGCCT-----GT--CTG---CA---TTTAT-CC
ACC----CC--CCA-----CC----CT-G
TCTC-----CCT---CT-----T-----TTTTT-----T-----T--T---TT---TT
-----ACC----CC-----T-TTTTATAT-----ATCA-ATTTCTATT
TT-A---CA-----ATA---AAATTTTGT-T--ATCACTTA----
A-AAAAAA-----A

```

Score=2010

Global score: 2010.0

>>> Local alignment (Smith-Waterman)

1

```

CT--CAAAAGT-CTAG-AGCCACCGTCCA--GGGAGCA----G---GTAGCT-GCTGGGCTCC-
----GGGGACACTTTGCGTTCGGGCTGGGAGCGTGCT--TTCCAC-G-ACGGTGACACGCTTC-
CCTGGATTGGCAGCCAGACTGCCTTCCGGGTCAGTGCCATGGAGGAGCCGCAGTCAGATCCT-A
GCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCT---GAAAA
CAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATT
GAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCC-CAGAATGCCAGAG-GCTGCTCC
CCCCGTGGCCCCCTG-CACCAGCAG-CTCCTACACCGGCGGCCCTGCACCAGCCCCCTCC-TGG
CCCCTGTCATCTTCTGTCCCTTCCCAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCT
TCTTGCAATTCTGGGACAGCCAAGTCTGTGACT-TGCACGTACTCCCCTGCCCTCAACAAGATGT
TTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATT-CCACACCCCCGCCCCGCA
CCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTG
CCCCCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCCCCTCCTCAGCATCTTATCCGAGTG
GAAGGAAATTTGCGTGTGGAGTATTTGGATGACAGAAACACTTTTCGACATAGTGTGGTGGTGC
CCTATGAGCCGCCTGAGGTTGGCTCTGACTGTACCACCATCCACTACAACCTACATGTGTAACAG
TTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCATCATCACTGGAAGACTCCAGT
GGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGTGCCTGTCCTGGGAGAGACCGGC
GCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAG-CCTCACCAC-GAGCTGCCCCCAGGGAGC
ACTAAGCGAGCACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTGGATG

```



```

| | . | | | | | | . | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
. | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | . | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
. | . | | | | | | | | | | | | | | | | | | | | | | | | | | |
. | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
. | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
. | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| . | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

23

```

CTGGCTAAAGTTCT-GTAGCTTCAGTTCATTGGGACCATCCTGGCTGTAGGTAGC--GACTACA
GTTAGGGGGGCACCTAGCATTTCAGGCCCTCATCCTCCTCCTTCC-CAGCAGGGTGTACGCTTCT
CC--GA-----AGACT-----GGATGACTGCCATGGAGGAGTCACAGTCGGAT-ATCA
GCCTCGAGCTCCCTCTGAGCCAGGAGACATTTTCAGGCTTATGGAACTACTTCCTCCAGAAGA
TA---TCCTG---CCATCACC-TCAC-TGC-ATGGACGATCTGTTGCTG-CCCC--AGGATGTT
GAGGA---GTTTTTTGAAGGCCCA-----AG-TGAAGC-CCTCCGAGTGTGTCAG-GAGCTCCTGC
AGCACAGGACCCTGTCACC-G-AGACCCCTGGGCCAGTGGCCCCTGCCCCAG-CCACTCCATGG
CCCCTGTGTCATCTTTTGTCCCTTCTCAAAAACTTACCAGGGCAACTATGGCTTCCACCTGGGCT
TCCTGCAGTCTGGGACAGCCAAGTCTGTTA-TGTGCACGTACTCTCCTCCCCCTCAATAAGCTAT
TCTGCCAGCTGGCGAAGACGTGCCCTGTGCAGTTGTGGG-TCAGCGCCACACCTCCAGCTGGGA
GCCGTGTCCGCGCCATGGCCATCTACAAGAAGTCACAGCACATGACGGAGGTGCTGAGACGCTG
CCCCCACCATGAGCGCTGCTCCGATGGTGATGGCCTGGCTCCTCCCCAGCATCTTATCCGGGTG
GAAGGAAATTTGTATCCCGAGTATCTGGAAGACAGGCAGACTTTTCGCCACAGCGTGGTGGTAC
CTTATGAGCCACCCGAGGCCGGCTCTGAGTATAACCACCATCCACTACAAGTACATGTGTAATAG
CTCCTGCATGGGGGGCATGAACCGCCGACCTATCCTTACCATCATCACACTGGAAGACTCCAGT
GGGAACCTTCTGGGACGGGACAGCTTTGAGGTTGCTGTTTGTGCCTGCCCTGGGAGAGACCGCC
GTACAGAAGAAGAAAATTTCCGCAA-AAA-GGAAGTCCTTTGCCCTGAACTGCCCCCAGGGAGC
GCAAAGAGAGCGCTGCCCACCTGCACAAGCGCCTCTCCCCCGCAAAGAAAAAACCCTTGATG
GAGAGTATTTACCC-TCAAGATCCGCGGGCGTAAACGCTTCGAGATGTTCCGGGAGCTGAATG
AGGCCTTAGAGTTAAAGGATGCCCATGCT-ACAGAGGAGTCTGGAGACAGCAGGGCTCACTCCA
GCTACCTGAAGACCAAGAAGGGCCAGTCTACTTCCCGCCATAAAAAA-ACAATGGTCAAGAAAG
TGGGGCCTGACTCAGACTGAC-TGCCTCTGCA--TCCCG-TCCCA-TCACCAGCCTCCC-C--
----CTCT-CCT----TGC--T-----GTCTTATGA---CTT-C-----AGG---
GC-TGAGA--CA--CA--A--TCC-----TCCCG--TCC-CT-----TCTG--CT
GC-CT--T-TTTTACCTTGTAGCTAG--GGCT---CAG-----C--CCCCTCTCTGAGT---A-
GTGTTTCCTG---GCCCAAGTTGGGGA-AT---AG----GTTGAT--AGTT----GTCAGGTCT
CTGC--TGGCCCAGCGAAATTCTATCCAG----CCAGTT----GT--T-----GGACCCTGGC
ACCT-AC-AATGAA--ATCTC--AC-CCTA-CCCCACACCCTGT-AAGATTCT--ATCTTGGG

```

```
C---CCTCATAGGGTCCATATCCTCCAGGGCCTACT---TTCCT-TCCATTCT-GC----AAA-
-GCCTGT--CTGCAT-----T-TA-----TCCACCCCC--CA----CCCTGTCTCCCTCT---TTT
T--T---T---TT-----TTTTTAC-----C-CCTTTTTA--TA-----TATCAA-TTT
C--CTATTTTA--CAA-----T-AAAA--TTTTG-----TTA-TCACTTAAAAAAAAA
```

Score=2107.5

Local score: 2107.5

>>> Fragmente selectate pentru comparație

Global fragment (cu gap-uri):

NM_000546.6:

```
-----CTCA-----AAAGT-CTAG-AGCCACCGTCCA--GGGAGCA----
G---GTAGCT-GCTGG
```

NM_011640.3:

```
TTTCCCCCTCCACGTGCTCACCCTGGCTAAAGTTCT-GTAGCTTCAGTTCATTGGGACCATCCT
GGCTGTAGGTAGC--G
```

Local fragment (fără gap-uri):

NM_000546.6:

```
CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGT
TCGGGCTGGGAGCGTG
```

NM_011640.3:

```
TTTCCCCCTCCACGTGCTCACCCTGGCTAAAGTTCTGTAGCTTCAGTTCATTGGGACCATCCTG
GCTGTAGGTAGCGACT
```

=== Comparație ===

Local alignment are un scor mai mare pe regiunea conservată.

În alinierea globală (Needleman–Wunsch), algoritmul a forțat potrivirea completă a celor ~1780 nt, introducând multiple gap-uri la începutul și la sfârșitul secvențelor pentru a alinia întreaga lungime. În schimb, alinierea locală (Smith–Waterman) a identificat o regiune centrală conservată de ~90 nt fără gap-uri, cu un scor mai mare datorită potrivirii perfecte pe segmentul comun.

Fragmente selectate pentru comparație

Global fragment (cu gap-uri):

NM_000546.6:

-----CTCA-----AAAGT-CTAG-AGCCACCGTCCA--GGGAGCA----
G---GTAGCT-GCTGG

NM_011640.3:

TTTCCCTCCACGTGCTCACCCTGGCTAAAGTTCT-GTAGCTTCAGTTCATTGGGACCATCCT
GGCTGTAGGTAGC--G

Local fragment (fără gap-uri):

NM_000546.6:

CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTTGCGT
TCGGGCTGGGAGCGTG

NM_011640.3:

TTTCCCTCCACGTGCTCACCCTGGCTAAAGTTCTGTAGCTTCAGTTCATTGGGACCATCCTG
GCTGTAGGTAGCGACT

Global are multe întreruperi: -----, -, --, ----

Local le aliniază curat, fără întreruperi, pentru același fragment de bază

În alinierea globală, potrivirea este “forțată” cu multiple gap-uri pentru a acoperi întreaga secvență, în timp ce alinierea locală găsește regiunea reală omoloagă fără întreruperi.

TASK 3:

Extras din MSA (regiune conservată)

NM_131327.2: TTTTGTAAGCAGCATTT

NM_000546.6: -----TTT

NM_011640.3: -----TTT

În alinierea multiplă (MSA) realizată cu Clustal Omega (EBI) pentru secvențele NM_000546.6 (uman), NM_011640.3 (șoarece) și NM_131327.2 (pește-zebră) din gena TP53, s-a observat o regiune conservată la pozițiile ~1816–1832. Segmentul conține motivul TTT, prezent identic la toate cele trei specii, sugerând o regiune funcțională conservată posibil implicată în terminarea transcrierii sau stabilitatea ARNm-ului. Regiunile conservate indică constrângeri evolutive — mutațiile aici ar putea afecta funcția genei. Spre deosebire de aliniamentele pairwise, MSA permite observarea simultană a coloanelor conservate și a variațiilor specifice fiecărei linii evolutive, oferind o imagine globală a conservării și divergenței.

TASK Bonus:

Semiglobal este preferat când capetele diferă sistematic (de ex. UTR-uri variabile, secvențe parțiale, read-uri de NGS/primeri) și vrem să comparăm suprapunerea internă fără să penalizăm porțiunile nealiniat de la extremități.

Față de global, evită scoruri artificial scăzute din cauza capetelor neomoloage.

Față de local, păstrează întreaga suprapunere relevantă (nu doar sub-fragmentul optim), ceea ce ajută la estimarea mai realistă a diferențelor pe zona comună.