

COMPARACIÓN DE SECUENCIAS

SEMANA 03

20 de Abril del 2020



UTEC

Índice

- ¿Por qué estudiar evolución?
- Alineamiento múltiple
- Divergencia intra e inter-específica
- Distancias genéticas

UTEC

1 COMPARACION DE SECUENCIAS




UTEC

¿Por qué estudiar evolución?


Hace millones de años que las flores tienen espinas y hace también millones de años que los corderos, a pesar de las espinas, se comen las flores. ¿Es que sería averiguar por qué las flores pierden el tiempo fabricando unas espinas que no les sirven para nada? ¿Es que no es importante la guerra de los corderos y las flores?...

El Principito (Cap. 7)



UTEC

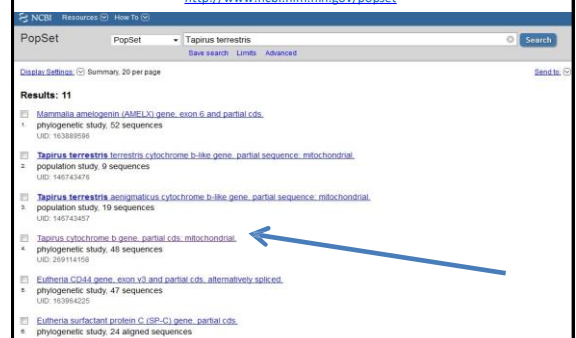
tapir.pdf



UTEC

Obteniendo secuencias: PopSet

<http://www.ncbi.nlm.nih.gov/popset>



UTEC

[illegible]

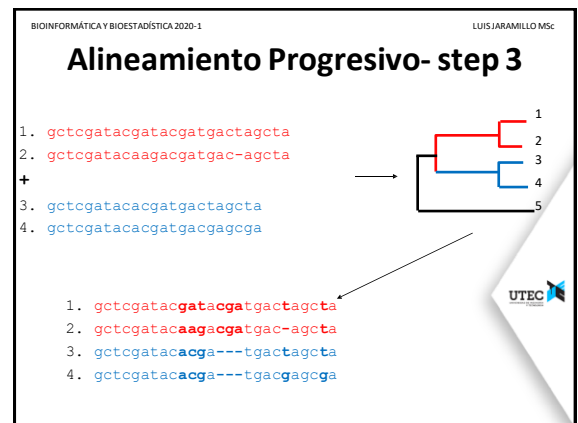
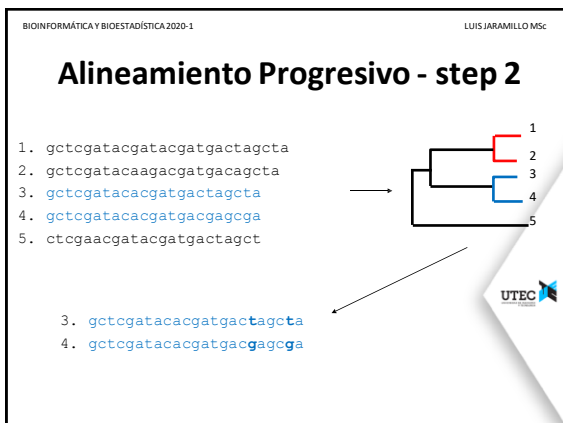
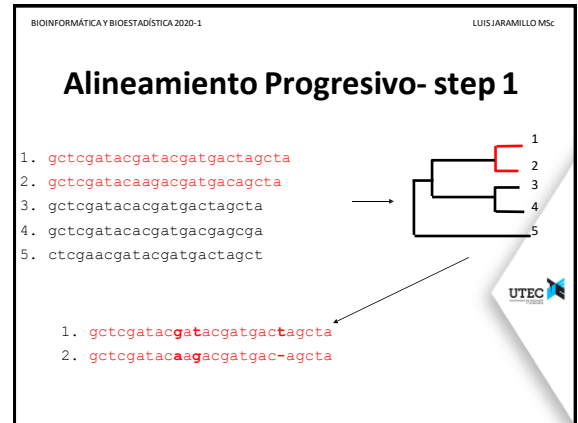
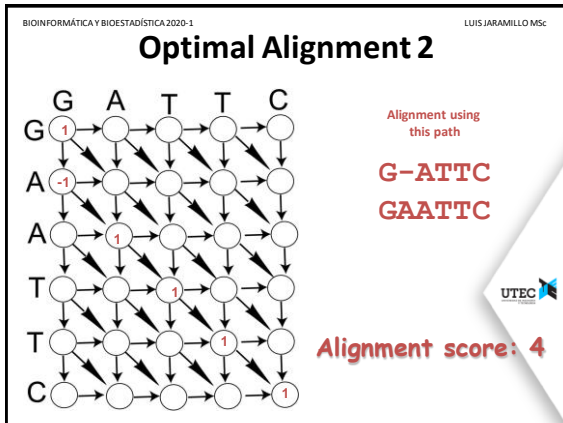
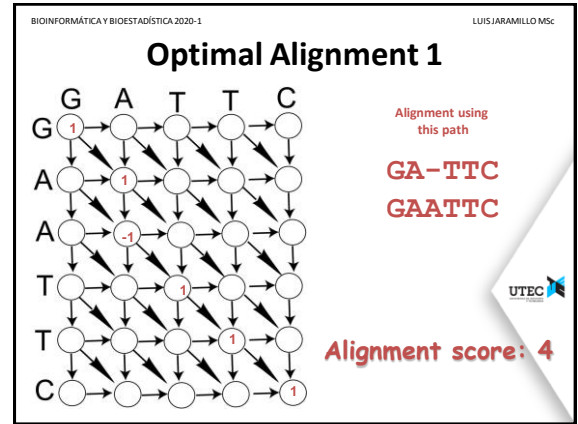
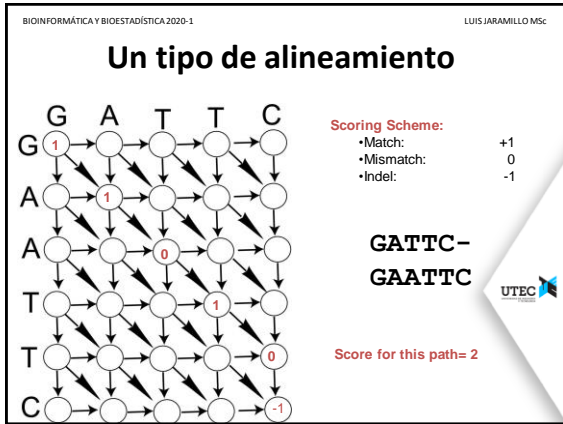
Editando el alineamiento múltiple

- Remover extremos (colas en 5' y 3')
- Realinear

The screenshot displays a multiple sequence alignment editor interface. On the left, a list of sequence identifiers is shown: TF11, TF5, TQ24, TQ55, TE20, TC184, TE17, T97, TV44, TA13, TF10, TB058, TQ28, TB02, TA12, TV48, TB01, TA10, and TE19. The main area shows the alignment of these sequences, with gaps (represented by dashes) and indels (represented by colored blocks) visible. A UTEC logo is present in the bottom right corner.

- Observar el polimorfismo: mutaciones, en tamaño (gaps / indels).

Path Graph for aligning two sequences



BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSC

Alineamiento Progresivo - final step

1. gctcgatacgaatacgtatgactagcta
 2. gctcgatacagaagacgtatgac-agcta
 3. gctcgatacacga---tgactagcta
 4. gctcgatacacga---tgacagcgga
 +
 5. ctccaagcatcgtatgactagct


1
2
3
4
5

UTEC
UNIVERSIDAD TECNOLÓGICA DE CALABAZA DE LA VIEJA


BIOMÉTRICA Y BIOESTADÍSTICA 2020-1
LUIS JARAMILLO MSc

Divergencia de las secuencias en especies y géneros diferentes

Gen mitocondrial 16S rRNA en especies y géneros diferentes



Bostryx agulari



[illegible]

BIOMINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSc

Divergencia intraespecífica utilizando el gen mitocondrial 16S rRNA dentro de una especie

lachay_71d
lachay_72d
_de_Aguila_71
_de_Aguila_84
lachay_74d
_de_Aguila_72
_de_Aguila_7a
lachay_73d
_de_Aguila_83
_de_Aguila_8a
_de_Aguila_3a
_de_Aguila_4a
lachay_67d
lachay_66d
BeordAnt312
BeordAnt311
Parasaway_29e
Parasaway_80
BeordAnt21c

UTEC

The image displays two screenshots of the MEGA 5.0.4 software interface. The top screenshot shows the 'File' menu with 'Convert File Format to MEGA...' selected. The bottom screenshot shows the 'MS: Select File and Format' dialog box with the file path 'C:\Users\Vipoo\Desktop\conferencial\practical\ape3.ahn' and 'Data Format: ahn (CLUSTAL)' selected.

[illegible]

[illegible]

Data Explorer

- Desde esta ventana podemos observar los sitios conservados, variables, e informativos.
- Otros datos: grado de degeneración, transformar la secuencia de nucleótidos a secuencia de aminoácidos.
- Exportar el archivo en formatos de tabla.

BIOPROFORMA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MS

Jvarkit JMS Sequence Data Explorer

Display Settings Highlight Statistics Help

T A C C A A T C A C C C A T T A T C G A C C T A C C A A C C C A T C A A A

✓ 1. TC18
✓ 2. TC18
✓ 3. TC18
✓ 4. TC18
✓ 5. TC18
✓ 6. TC18
✓ 7. TC18
✓ 8. TC18
✓ 9. TC18
✓ 10. TC18
✓ 11. TC18
✓ 12. TC18
✓ 13. TC18
✓ 14. TC18
✓ 15. TC18
✓ 16. TC18
✓ 17. TC18
✓ 18. TC18
✓ 19. TC18
✓ 20. TC18
✓ 21. TC18
✓ 22. TC18
✓ 23. TC18
✓ 24. TC18
✓ 25. TC18
✓ 26. TC18
✓ 27. TC18
✓ 28. TC18
✓ 29. TC18
✓ 30. TC18
✓ 31. TC18
✓ 32. TC18
✓ 33. TC18
✓ 34. TC18
✓ 35. TC18
✓ 36. TC18
✓ 37. TC18
✓ 38. TC18
✓ 39. TC18
✓ 40. TC18
✓ 41. TC18
✓ 42. TC18
✓ 43. TC18
✓ 44. TC18
✓ 45. TC18
✓ 46. TC18
✓ 47. TC18
✓ 48. TC18
✓ 49. TC18
✓ 50. TC18
✓ 51. TC18
✓ 52. TC18
✓ 53. TC18
✓ 54. TC18
✓ 55. TC18
✓ 56. TC18
✓ 57. TC18
✓ 58. TC18
✓ 59. TC18
✓ 60. TC18
✓ 61. TC18
✓ 62. TC18
✓ 63. TC18
✓ 64. TC18
✓ 65. TC18
✓ 66. TC18
✓ 67. TC18
✓ 68. TC18
✓ 69. TC18
✓ 70. TC18
✓ 71. TC18
✓ 72. TC18
✓ 73. TC18
✓ 74. TC18
✓ 75. TC18
✓ 76. TC18
✓ 77. TC18
✓ 78. TC18
✓ 79. TC18
✓ 80. TC18
✓ 81. TC18
✓ 82. TC18
✓ 83. TC18
✓ 84. TC18
✓ 85. TC18
✓ 86. TC18
✓ 87. TC18
✓ 88. TC18
✓ 89. TC18
✓ 90. TC18
✓ 91. TC18
✓ 92. TC18
✓ 93. TC18
✓ 94. TC18
✓ 95. TC18
✓ 96. TC18
✓ 97. TC18
✓ 98. TC18
✓ 99. TC18
✓ 100. TC18

Sitios conservados

T A C C A A T C A C C C A T T A T C G A C C T A C C A A C C C A T C A A A

✓ 1. TC18
✓ 2. TC18
✓ 3. TC18
✓ 4. TC18
✓ 5. TC18
✓ 6. TC18
✓ 7. TC18
✓ 8. TC18
✓ 9. TC18
✓ 10. TC18
✓ 11. TC18
✓ 12. TC18
✓ 13. TC18
✓ 14. TC18
✓ 15. TC18
✓ 16. TC18
✓ 17. TC18
✓ 18. TC18
✓ 19. TC18
✓ 20. TC18
✓ 21. TC18
✓ 22. TC18
✓ 23. TC18
✓ 24. TC18
✓ 25. TC18
✓ 26. TC18
✓ 27. TC18
✓ 28. TC18
✓ 29. TC18
✓ 30. TC18
✓ 31. TC18
✓ 32. TC18
✓ 33. TC18
✓ 34. TC18
✓ 35. TC18
✓ 36. TC18
✓ 37. TC18
✓ 38. TC18
✓ 39. TC18
✓ 40. TC18
✓ 41. TC18
✓ 42. TC18
✓ 43. TC18
✓ 44. TC18
✓ 45. TC18
✓ 46. TC18
✓ 47. TC18
✓ 48. TC18
✓ 49. TC18
✓ 50. TC18
✓ 51. TC18
✓ 52. TC18
✓ 53. TC18
✓ 54. TC18
✓ 55. TC18
✓ 56. TC18
✓ 57. TC18
✓ 58. TC18
✓ 59. TC18
✓ 60. TC18
✓ 61. TC18
✓ 62. TC18
✓ 63. TC18
✓ 64. TC18
✓ 65. TC18
✓ 66. TC18
✓ 67. TC18
✓ 68. TC18
✓ 69. TC18
✓ 70. TC18
✓ 71. TC18
✓ 72. TC18
✓ 73. TC18
✓ 74. TC18
✓ 75. TC18
✓ 76. TC18
✓ 77. TC18
✓ 78. TC18
✓ 79. TC18
✓ 80. TC18
✓ 81. TC18
✓ 82. TC18
✓ 83. TC18
✓ 84. TC18
✓ 85. TC18
✓ 86. TC18
✓ 87. TC18
✓ 88. TC18
✓ 89. TC18
✓ 90. TC18
✓ 91. TC18
✓ 92. TC18
✓ 93. TC18
✓ 94. TC18
✓ 95. TC18
✓ 96. TC18
✓ 97. TC18
✓ 98. TC18
✓ 99. TC18
✓ 100. TC18

Sitios informativos

[illegible][illegible]

BIOMÉTRICA Y BIOESTADÍSTICA 2020-1

LUIS IARAMILLO MSC

Distancias genéticas

- Número de mutaciones.

	1	2	3	4	5	6	7	8	9
1. TE19									
2. TE16	1.00								
3. TE24	4.00	5.00							
4. TE28	10.00	11.00	10.00						
5. TE11	16.00	17.00	16.00	14.00					
6. TE1	17.00	18.00	17.00	15.00	1.00				
7. TE6	15.00	16.00	15.00	13.00	1.00	2.00			
8. TE5	16.00	17.00	16.00	14.00	2.00	3.00	1.00		
9. TE2	16.00	17.00	16.00	14.00	2.00	2.00	1.00	0.00	
10. TE3	15.00	16.00	15.00	13.00	1.00	2.00	1.00	1.00	

- Distancia – p

	1	2	3	4	5	6	7	8	9
1. TE19	0.0000								
2. TE16	0.0009								
3. TE24	0.0008	0.0047							
4. TE28	0.0004	0.0103	0.0094						
5. TE11	0.0151	0.0160	0.0151	0.0132					
6. TE1	0.0160	0.0158	0.0160	0.0141	0.0009				
7. TE6	0.0141	0.0151	0.0141	0.0122	0.0009	0.0019			
8. TE5	0.0151	0.0160	0.0151	0.0132	0.0101	0.0020	0.0009		
9. TE2	0.0151	0.0160	0.0151	0.0132	0.0101	0.0020	0.0009	0.0000	
10. TE3	0.0141	0.0151	0.0141	0.0122	0.0009	0.0019	0.0000	0.0000	0.0009

- Distancia utilizando un modelo de sustitución

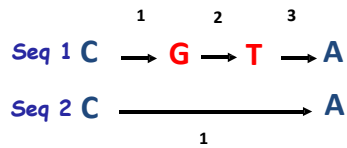
	1	2	3	4	5	6	7	8	9
1. TE19									
2. TE16	0.0000								
3. TE24	0.0008	0.0047							
4. TE28	0.0009	0.0105	0.0095						
5. TE11	0.0151	0.0163	0.0153	0.0134					
6. TE1	0.0163	0.0173	0.0163	0.0143	0.0006				
7. TE6	0.0141	0.0153	0.0143	0.0124	0.0009	0.0019			
8. TE5	0.0151	0.0163	0.0153	0.0134	0.0009	0.0020	0.0009		
9. TE2	0.0153	0.0163	0.0153	0.0134	0.0019	0.0020	0.0009	0.0000	
10. TE3	0.0141	0.0153	0.0143	0.0124	0.0009	0.0019	0.0009	0.0000	0.0009

Los cambios escondidos

Seq 1 AGCGAG

Seq 2 GCGGAC

Number of changes



Conclusiones

- ✓ El alineamiento es, finalmente, una hipótesis de homología entre las secuencias de nuestro set de datos.
- ✓ Homología: Los caracteres son homólogos cuando provienen de un mismo ancestro común
- ✓ Divergencia intraespecífica utilizando el gen mitocondrial 16S rRNA dentro de una especie
- ✓ De ClustalX a MEGA



Gracias

NUESTRA PACIENCIA
CONSEGUIRÁ MÁS
COSAS QUE NUESTRA
FUERZA!!

