CS426: Introduction to Computational Biology

Section 4

TA.

- My name: Liviu Popescu
- email: liviup@cs.cornell.edu
- Office: 424 Rhodes Hall
- Office hours: MW 3:00 – 4:00

Multi Sequence Alignment

- Consider the following 4 strings: MASH, MESH, SQUAMISH, and SQUASH.
- The goal is to perform a multiple alignment for these strings. We'll use the following simplified scoring matrix (based on distance rather than similarity).
  - d(x,y) = 0 if x and y are identical
  - d(x,y) = 1 if x and y are nonidentical vowels
  - d(x,y) = 2 if x and y are nonidentical consonant
  - d(x,y) = 2 if one of x and y is a space
  - d(x,y) = 3 if one of x and y is a consonant and the other is a vowel

Sequence Pairwise Distances

- Find all the pairwise distances for the 4 strings.
  - D(MASH,MESH) = 1
  - D(MASH,SQUAMISH) = 9
  - D(MASH,SQUASH) = 6 (more than one path)
  - D(MESH,SQUAMISH) = 9
  - D(MESH,SQUASH) = 7 (more than one path)
  - D(SQUAMISH,SQUASH) = 4

Center Star Method

- Which string makes the best center for the Center Star Method?
  - With MASH at the center, the sum of distances to MASH is 16.
  - With MESH at the center, the sum of distances to MESH is 17.
  - With SQUAMISH at the center, the sum of distances to SQUAMISH is 22.
  - With SQUASH at the center, the sum of distances to SQUASH is 17.
  - Thus the best center is MASH.

Center Star Method

- What is the resulting multiple alignment?
  - Here is the resulting multiple alignment:

```
----M--ASH
----M--ESH
SQUAM--ISH
----SQUASH
```

Costs of the alignment

- 3. What is the tree-based cost for this multiple alignment? What is its sum-of-pairs cost?
  - The cost given above (16) is the tree-based cost.
  - The sum-of-pairs cost is 1 + 9 + 6 + 9 + 7 + 15 = 47.

Center Star Method

- Does the scoring matrix satisfy the Triangle Inequality? Note that the letter Y can be counted as both a consonant and a vowel. What does this tell us about the sum-of-pairs cost for the optimal multiple alignment?
  - Yes, although a small change (distance for nonidentical constants = 1) would make if fail.
  - The true optimal aligment has sum-of-pairs cost > 47/2.

Consensus String

- What is the consensus string for the above alignment?
  - `----M--ASH`

Iterative Pairwise Alignment

- Build a multiple alignment using Iterative Pairwise Alignment in its simplest form (a single multiple alignment is maintained and the distance to a group is the minimum of all distances to members of the group).

Iterative Pairwise Alignment

- The two closest strings are MASH and MESH, so these are aligned first.

```
MASH
MESH
```

Iterative Pairwise Alignment

- The next closest is SQUASH (distance 6 from MASH), so it is added next.
  - `M--ASH`
  - `M--ESH`
  - `SQUASH`

Iterative Pairwise Alignment

- The profile of the alignment

	1	2	3	4	5	6
A	0	0	0	2/3	0	0
E	0	0	0	1/3	0	0
H	0	0	0	0	0	0
M	2/3	0	0	0	0	1
Q	0	2/3	0	0	0	0
U	0	0	1/3	0	0	0
S	1/3	0	0	0	1	0

Iterative Pairwise Alignment

- Finally, SQUAMISH is included.

```
M--A--SH
M--E--SH
SQUA--SH
SQUAMISH
```

Iterative Pairwise Alignment

- This last alignment is computed by comparing a string (SQUAMISH) against a profile. Here is the Dynamic Programming table that is used.

	M M S	- - Q	- - U	A E A	S S S	H H H
S	4/3	2	7/3	3	0	2
Q	2	4/3	7/3	3	2	2
U	3	7/3	4/3	1	3	3
A	3	7/3	5/3	1/3	3	3
M	2/3	2	7/3	3	2	2
I	3	7/3	5/3	1	3	3
S	4/3	2	7/3	3	0	2
H	2	2	7/3	3	2	0

Iterative Pairwise Alignment

- It is also possible to build a multiple alignment using a different form of Iterative Pairwise Alignment. This time, we always combine the two closest items where an item is either a string or a profile. Is the resulting alignment any different than the previous one?

Iterative Pairwise Alignment

- Again, the two closest strings are MASH and MESH.
  - MASH
  - MESH
- The next closest pair is SQUAMISH and SQUASH so these are combined next.
  - SQUAMISH
  - SQUA--SH

Iterative Pairwise Alignment

- Now these two groups are combined using profile/profile alignment to produce a final multiple alignment.
- Profile for the first group

	1	2	3	4
A	0	1/3	0	0
E	0	1/3	0	0
H	0	0	0	1
M	1	0	0	0
S	0	0	1	0

Iterative Pairwise Alignment

- Second alignment profile:

	1	2	3	4	5	6	7	8
A	0	0	0	1	0	0	0	0
H	0	0	0	0	0	0	0	1
I	0	0	0	0	0	1/2	0	0
M	0	0	0	0	1/2	0	0	0
Q	0	1	0	0	0	0	0	0
U	0	0	1	0	0	0	0	0
S	1	0	0	0	1	0	1	0

Iterative Pairwise Alignment

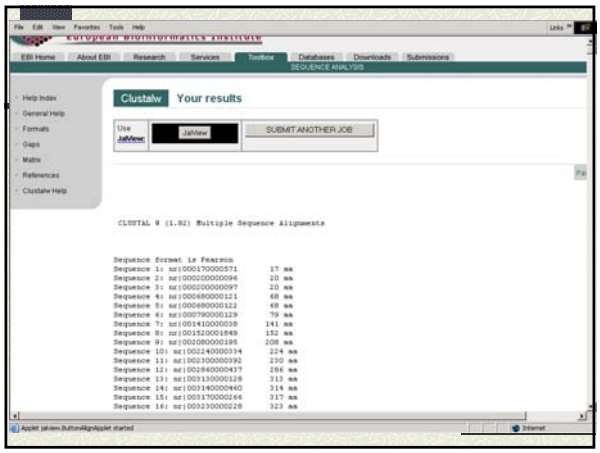
- The dynamic programing matrix is:

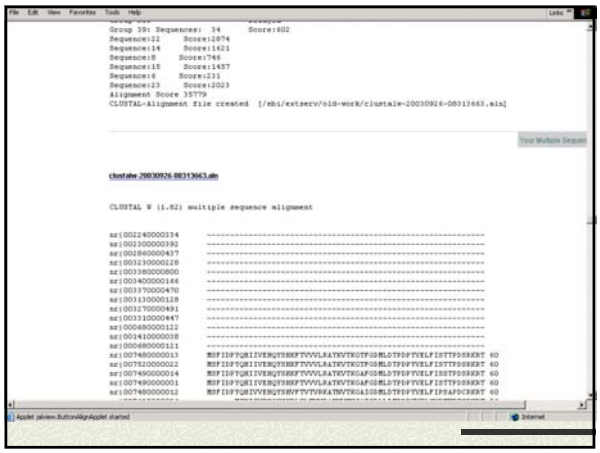
		S	Q	U	A	M	I	S	H
	-	S	Q	U	A	-	-	S	H
-	0	2	4	6	8	9	10	12	14
MM	2	2	4	6	8	9	11	12	14
AE	4	4	5	5	13/2	17/2	19/2	23/2	27/2
SS	6	4	6	7	8	17/2	21/2	19/2	23/2
HH	8	6	6	8	10	10	11	25/2	19/2

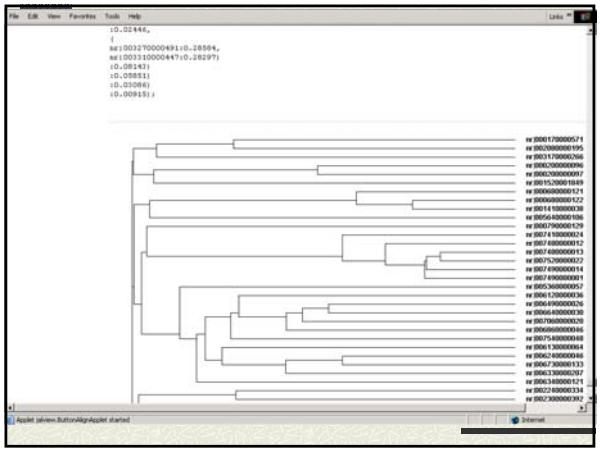
Iterative Pairwise Alignment

- Alignment is:
  - SQUAMISH
  - SQUA--SH
  - -M-A--SH
  - -M-E--SH

ClustalW @ EBI






File Edit View Favorites Tools Help
:0.02446,
(
nz:003270000491:0.28504,
nz:003310000447:0.28297)
:0.08143)
:0.05851)
:0.03086)
:0.00915);
nz:000170000571
nz:002090000195
nz:003170000266
nz:000200000096
nz:000320000097
nz:001520001849
nz:000680000121
nz:000690000122
nz:001410000038
nz:000540000106
nz:000790000129
nz:000410000024
nz:000480000012
nz:000480000013
nz:000529000022
nz:000490000014
nz:000490000001
nz:000360000057
nz:000529000036
nz:000490000026
nz:000640000030
nz:000960000020
nz:000960000046
nz:000540000048
nz:000130000064
nz:000240000066
nz:000730000133
nz:000330000207
nz:000340000121
nz:000220000334
nz:000300000392
Applet jalview.ButtonAlignApplet started
Internet