[Supplementary Essay 2]

BME32401

Prof. Jong BHAK

20131107 YeongJae Kim

Scientific Essay about "What is alignment?"

Abstract

This paper is a scientific essay for the topic; "What is alignment?". It includes introducing several DNA sequence alignment strategies, my own way and apply the 'Phylo' game.

Summary

Sequence alignment is an arranging the sequences of DNA or protein in bioinformatics. The important thing is how to arrange correct and fast. So, a strategy is most important of alignment because computer program replaces handwork in present, the amount of data does not affect to time of alignment. Therefore, many biologists developed alignment program and published.

One of these is 'Dot Plot Sequence alignment'. This way compares two sequences and can be easy to identify insertion, deletions and so on. First step of this method is making 2D-plot. And then, marks dot on matching sequence (Figure 1) and finds the deletion or insertion (Figure 2).

Figure 1. Dot marking on the 2D-plot

Comparing two sequences (x axis and y axis). When Nth sequence on X-axis same with Mth sequence on Y-axis, mark dot on the (N, M) site in 2D-plot.

Figure 2. Analyzing sequence

Following dot to right-down direction and find deletion or insertion sequence. In this figure, Y-axis sequence has 7 deletion (DOROTHY_ _ _ _ _ _ _ _HODGKIN).

Second, to align several sequences, 'Clustal Omerga' program is useful. It also pairwise sequence alignment tools. I already used it so show an example and some tips. For instance, I analyses PCNA protein sequences of human, zebrafish and xenopus. First step is to find sequence on PubMed site (Figure 3).
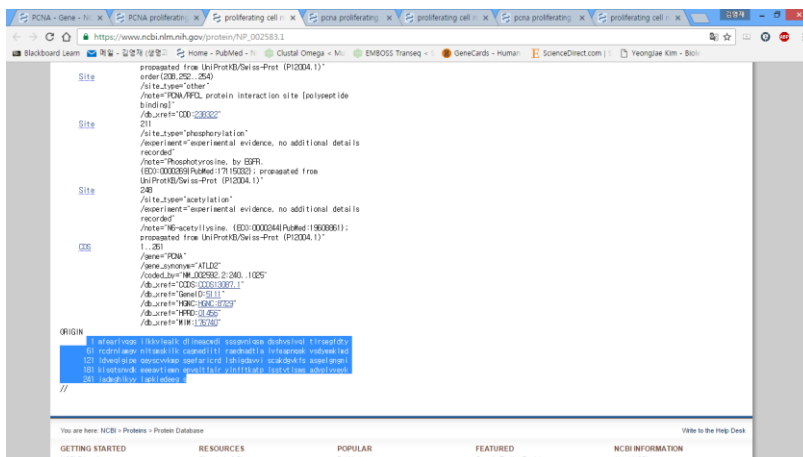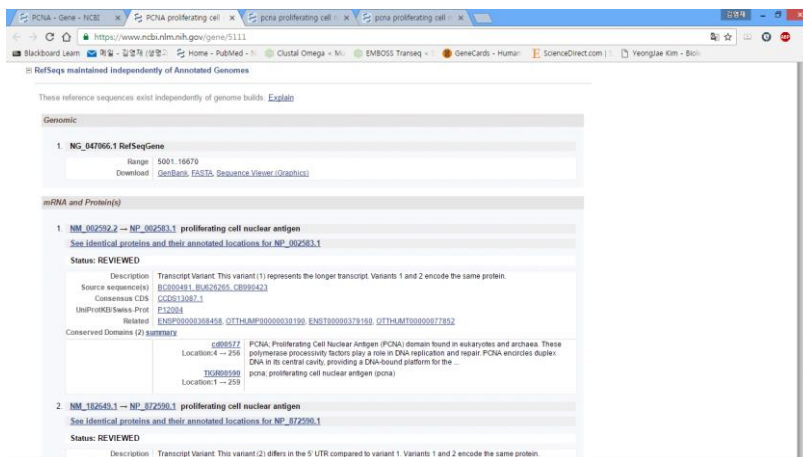
Scientific Essay about "What is alignment?"

Figure 3

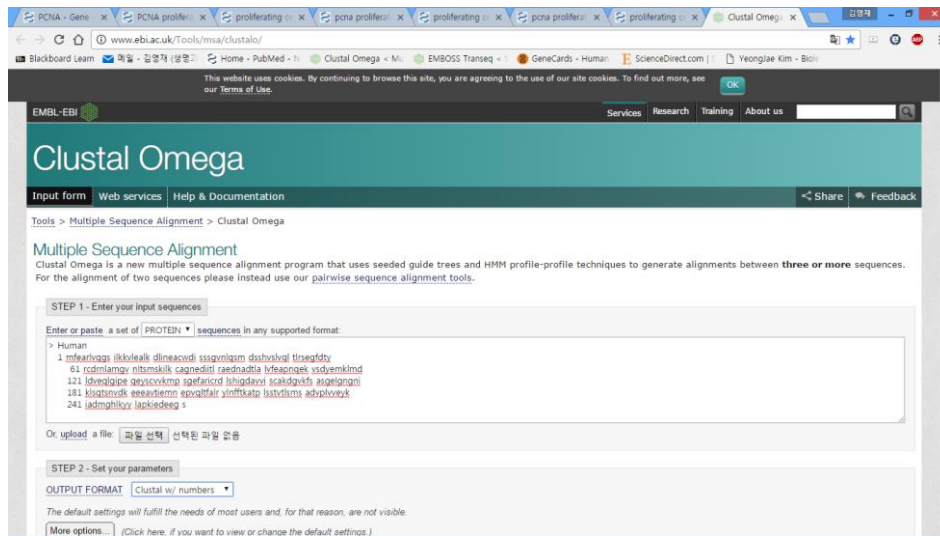a. Select 'gene' and type in wanted protein name
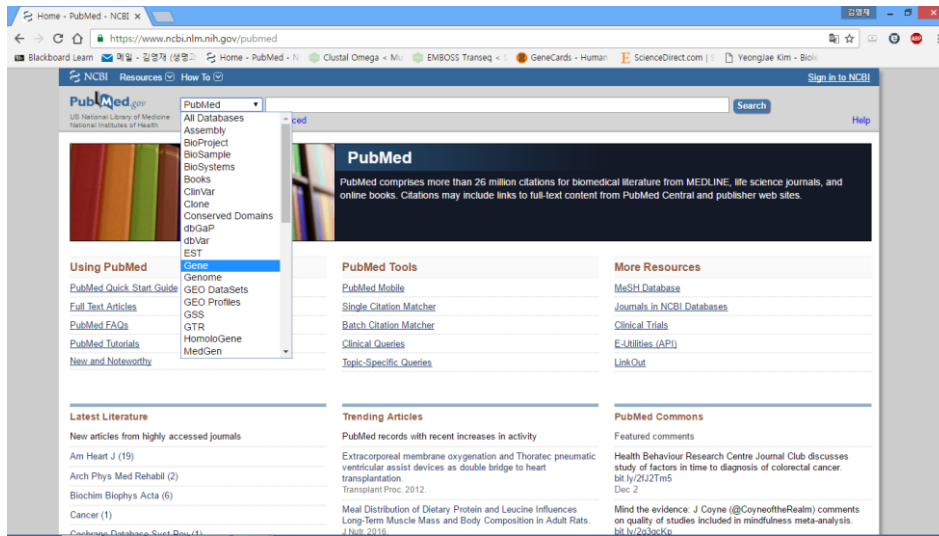
b, c.   Find protein sequence

Scientific Essay about "What is alignment?"

d. copy and paste protein sequences after type in '> (species)' and change the OUTPUT FORMAT from 'Clustal w/o numbers' to 'Clustal w/ numbers'. It can be easy to comparing the sequences.

e. result


```
CLUSTAL O(1.2.3) multiple sequence alignment


xenopus      mfearlvqgsilkkvlealkdlideacwditssgislqsmdsshvslvqltlrsdgfdty      60
Human        mfearlvqgsilkkvlealkdlineacwdisssgvnlqsmdsshvslvqltlrsegfdty      60
Zebrafish    mfearlvqgsilkkvlealkdliteacwdvsssgislqsmdsshvslvqltlrsdgfdsy      60
             ********************* *****::***:.*****************:***:*

xenopus      rcdrnqsigvkmssmskilkcaasediitlraednadtvtmvfespnqekvsdyemklmd     120
Human        rcdrnlamgvnltsmskilkcagnediitlraednadtlalvfeapnqekvsdyemklmd     120
Zebrafish    rcdrnlamgvnlssmskilkcagnediitlraednadalalvfetlnqekvsdyemklmd     120
             ***** ::**:::********* .*************::::***: *************

xenopus      ldveqlgipeqeyscvikmpsgefaricrdlsqigdavviscakdgvkfsasgelgtgnv     180
Human        ldveqlgipeqeyscvvkmpsgefaricrdlshigdavviscakdgvkfsasgelgngni     180
Zebrafish    ldveqlgipeqeyscvvkmpsgefaricrdlsqigdavmiscakdgvkfsasgelgtgni     180
             ***************:**************:*****:************** .**:

xenopus      klsqtsnvdkeeeavtiemnepvqltfalrylnfftkatplsptvtlsmsadiplvveyk     240
Human        klsqtsnvdkeeeavtiemnepvqltfalrylnfftkatplsstvtlsmsadvplvveyk     240
Zebrafish    klsqtsnvdkedeavtiemnepvqlifalnylnfftkatplsktvtlsmsadiplvveyk     240
             ***********:************ ***.*********** ********:*******

xenopus      iadmghvkyylapkiedeeas       261
Human        iadmghlkyylapkiedeegs       261
Zebrafish    iadmghvkyylapkideess-       260
             ******:********::*..
```

Third, I developed some ideas of how to perfect alignment. I think most people find instinctively matched big fragment of sequences to align sequences. Therefore, I select sequential short fragment (10~15bp) from 1st sequence to last letter. And give a score which show how many sequence match to others. When I find best score, do pairwise sequence alignment based on the best score fragment. Advantage of this method, if the sequences' length is very different, it cannot easy to find start point.
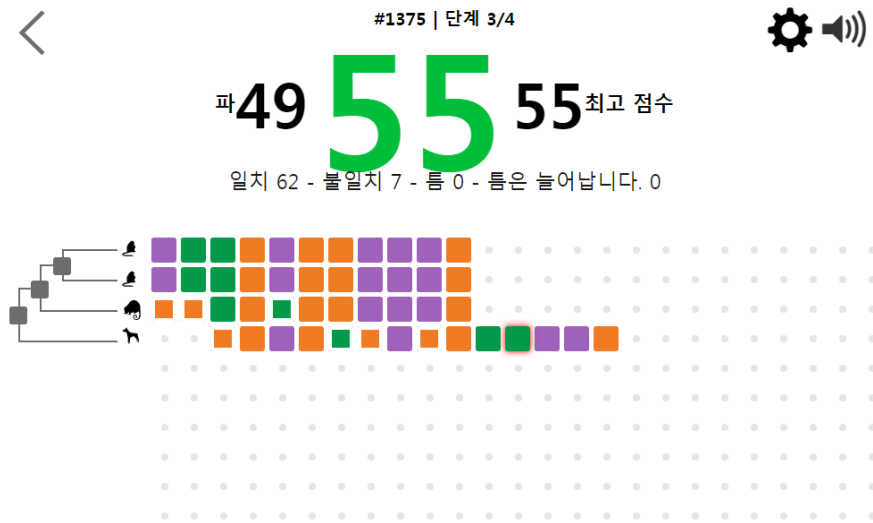
Scientific Essay about "What is alignment?"

So I notice initial point first and pairwise alignment.

  Finally, how to develop this method and apply? If many people play the game and work on alignment in the same time, it is nice sound. So some people make games named 'Phylo'. It is alignment game and data to analyze human alignment mechanism(Figure 4).

  Figure 4.

  a.   Playing the Phylo game

  b.   Using the solving mechanism, makers can easily solve the alignment.

Scientific Essay about "What is alignment?"

Reference

1. https://www.youtube.com/watch?v=6Udqou3vmng

2. http://www.globalspec.com/reference/65984/203279/dotplots-and-sequence-alignments

3. http://bioinfoalzheimer.weebly.com/chapter-3.html

4. http://www.ebi.ac.uk/Tools/msa/clustalo/

5. https://www.youtube.com/watch?v=tq-bWYVnQCM

6. https://www.youtube.com/watch?v=gsL6TFiqAx0

7. Horse genome analyses: https://www.youtube.com/watch?v=doODsjIJPoM

8. Phylo game: http://phylo.cs.mcgill.ca/