a) the sequence provided here is nucleotide sequence, so we should use blastn for alignment. It codes cytochrome c oxidase subunit II [Paracoccus denitrificans]. Tblastx is the most computational intensive, because

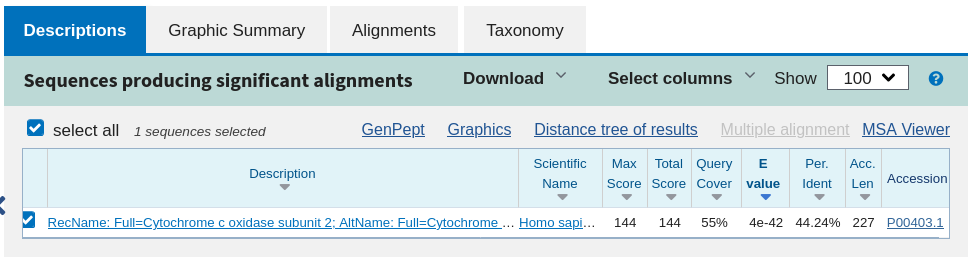
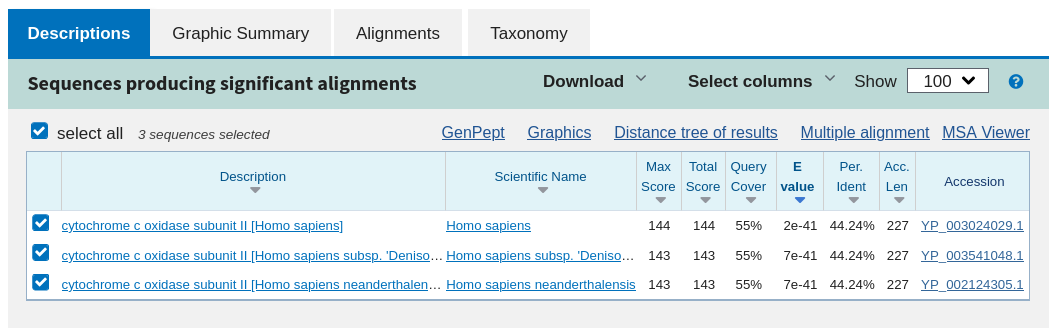
b) done

c) I should consider using the protein sequence since it is more sensitive than DNA sequence, one amino acid could have several condons for it, it would be time consuming.

d) if I don’t choose the specific organism for human there would be a lots of similarity proteins.

e) the e value is 4e-42 when selecting the organism of homo sapiens, but when I used the filter by adding homo sapiens after the blast is done, the e value is 6e-41, the database size would affect the e valule.

g) the first one used the database of refseq protein.



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| position | 1 | 2 | 3 | 4 | 5 | 6 | Overall |
| A | 0.33 | 0.22 | 0.11 | 0.22 | 0.56 | 0.11 | 0.1 |
| T | 0.44 | 0.11 | 0.56 | 0.44 | 0.22 | 0.22 | 0.1 |
| C | 0.11 | 0.11 | 0.11 | 0.22 | 0.11 | 0.44 | 0.4 |
| G | 0.11 | 0.56 | 0.11 | 0.11 | 0.11 | 0.22 | 0.4 |

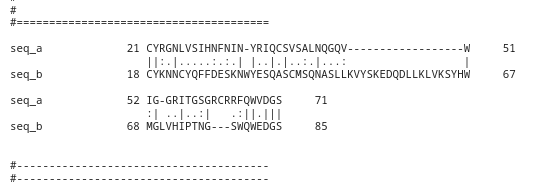
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | 3.3 | 2.2 | 1.1 | 2.2 | 5.6 | 1.1 |
| T | 4.4 | 1.1 | 5.6 | 4.4 | 2.2 | 2.2 |
| C | 0.275 | 0.275 | 0.275 | 0.55 | 0.275 | 1.1 |
| G | 0.275 | 1.4 | 0.275 | 0.275 | 0.275 | 0.55 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| A | 1.72 | 1.14 | 0.14 | 1.14 | 2.49 | 0.14 |
| T | 2.14 | 0.14 | 2.49 | 2.14 | 1.14 | 1.14 |
| C | -1.86 | -1.86 | -1.86 | -0.86 | -1.86 | 0.14 |
| G | -1.86 | 0.49 | -1.86 | -1.86 | -1.86 | -0.86 |

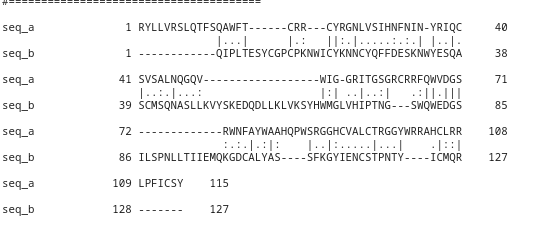
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | 1.58 | 1 | 0 | 3.17 | 0 | 0 | 0 | 0 | 0 | 1.58 |
| C | 2.32 | 1 | 3.17 | 0 | 3.17 | 0 | 0 | 0 | 1 | 0 |
| G | 1.58 | 2.8 | 0 | 0 | 0 | 3.17 | 0 | 2.8 | 2.32 | 2 |
| T | 0 | 0 | 0 | 0 | 0 | 0 | 3.17 | 1.58 | 2 | 2 |

AACACGCGAT = 1.58+1+3.17+3.17+3.17+3.17+0+2.8+0+2

3. Water alignment indentity; 23.9%



Needle alignment: 17.8%



these two sequences have low similarity, there are some residues are matched which means they are conserved. I don’t see any evolutionary relationship between these two sequences.

e) Overall RMSD: 1.778