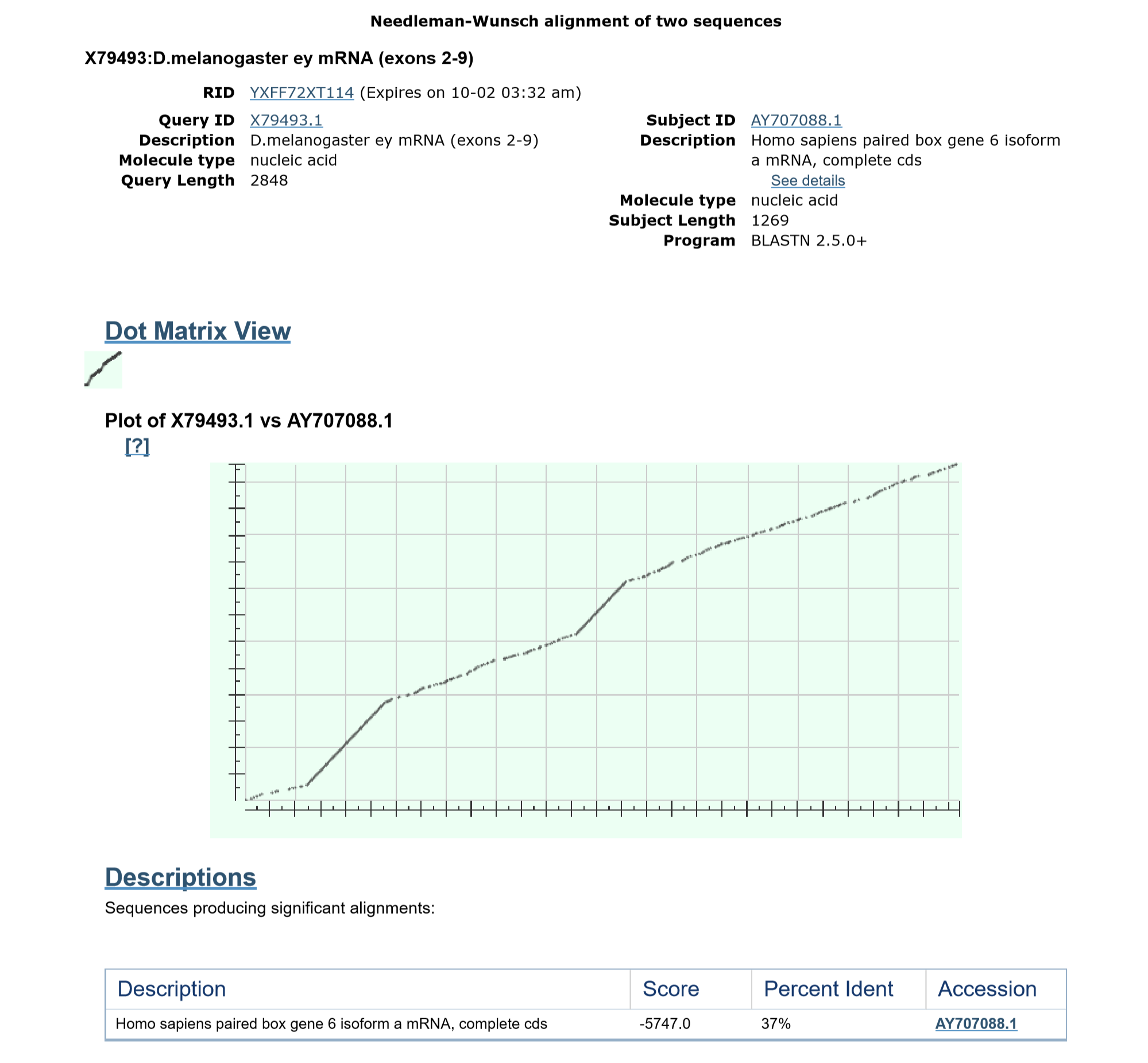
Modulo 3

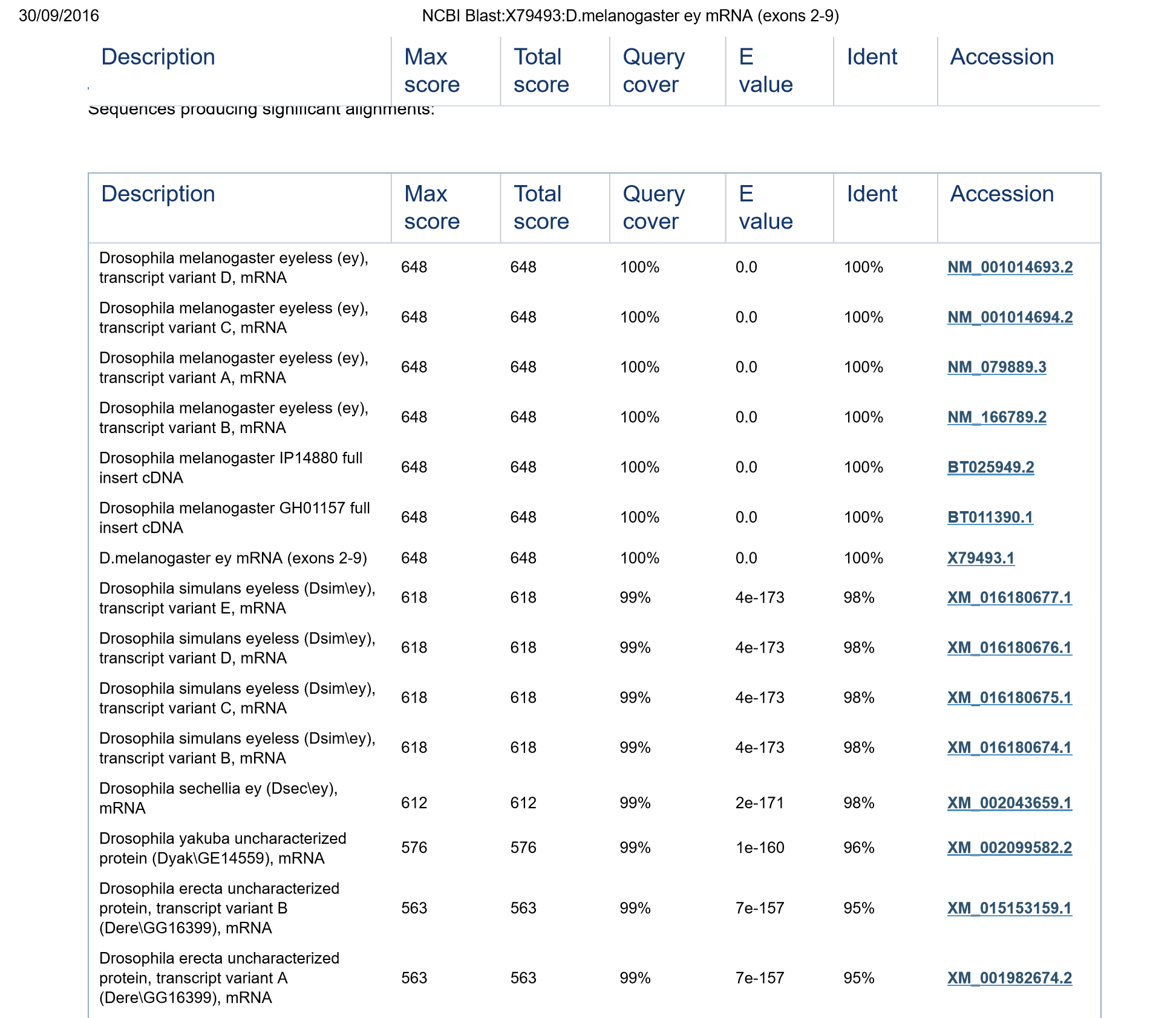
Exercise 1:

By doing the alignment between the D. melanogaster and Homo sapiens paired box genes, the percentage indent was 37% and the score was -5747.0. That means that there is a evolutionary gap between the organisms that this gene sequence is being evaluated. The whole report is hosted in the following link on GitHub: [Report 1](https://github.com/ICMC/Computation-Biology/blob/master/modulo3/exercise1.pdf)



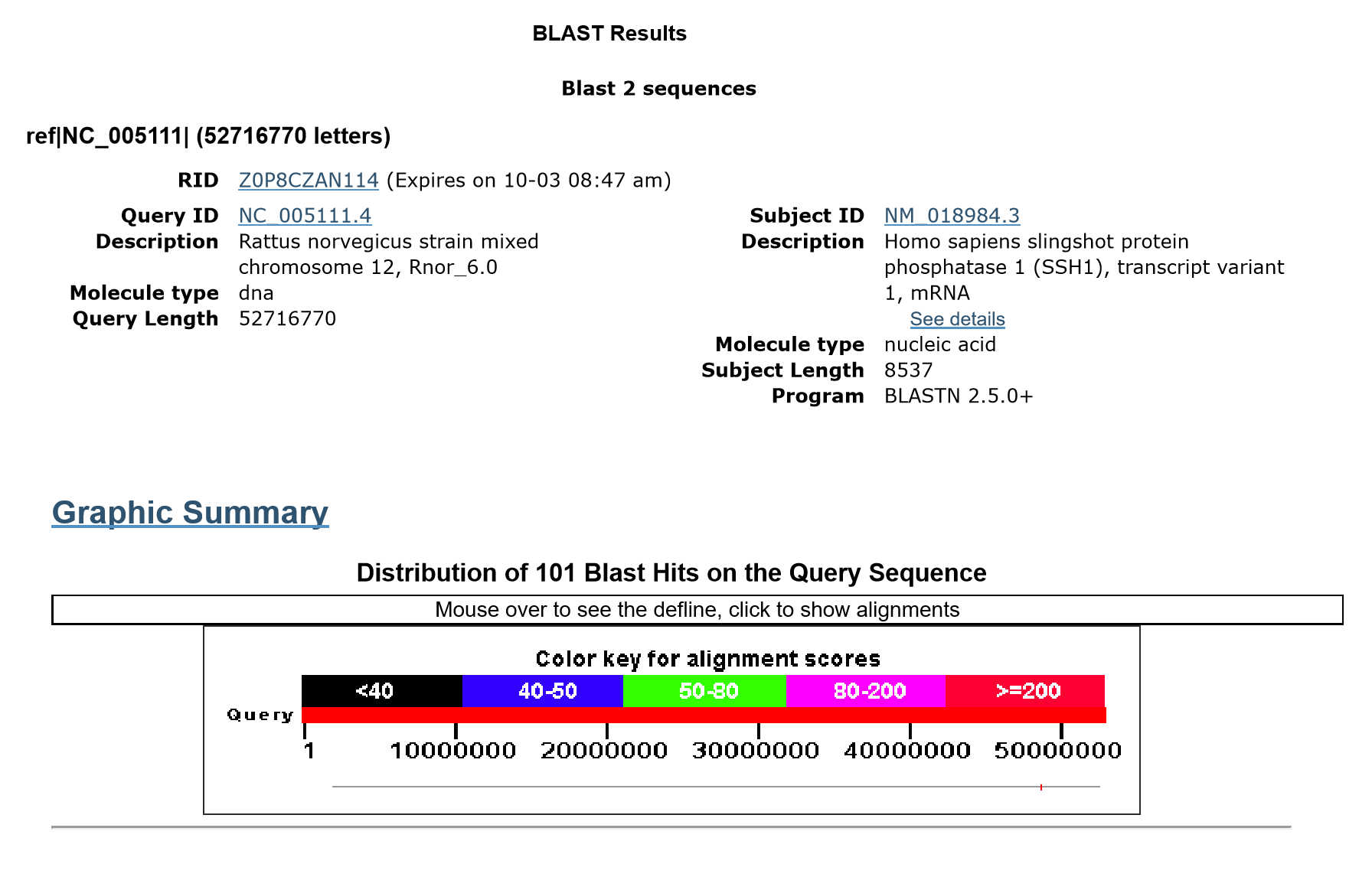
Exercicio 2:

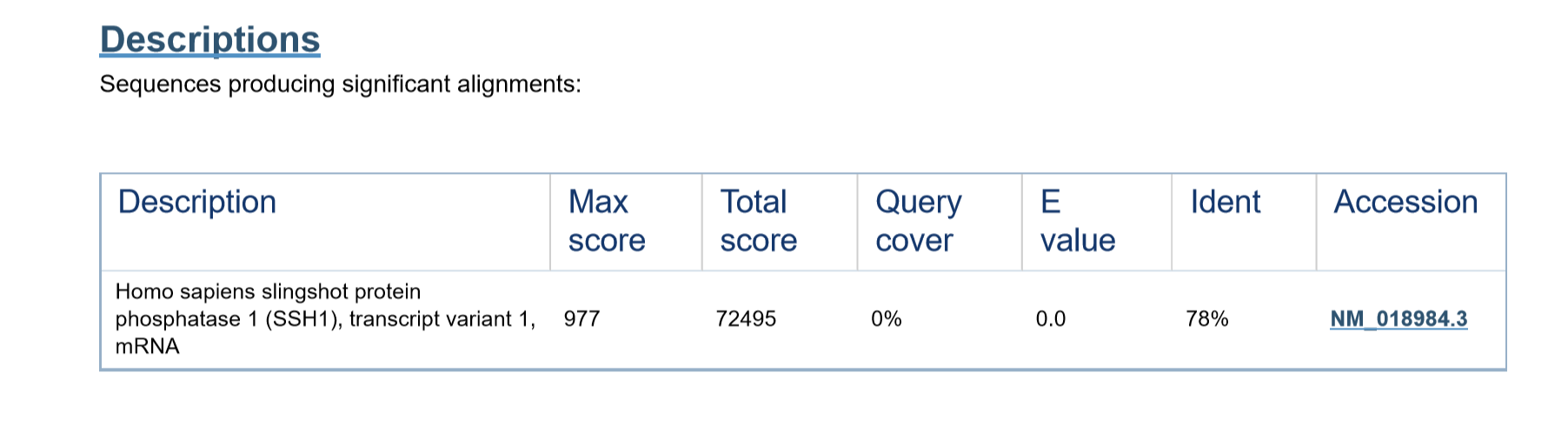
The task for this assignment was to compare the segment that had more similarity between X79493 (eyless) and AY707088 (aniridia). The segment with bigger similarity was between 241 and 599. So X79493 was compared across other species in that specific segment[241:599] porving that it is a gene that is similar and present in a significant number of other species and that has an ident of 100% and a score of about 648. That means that it is potentially a very important gene that was kept suring evolution, not only for one species, but for a diverse number. The full report can be found on the following link: [Report 2](https://github.com/ICMC/Computation-Biology/blob/master/modulo3/exercise2.pdf)



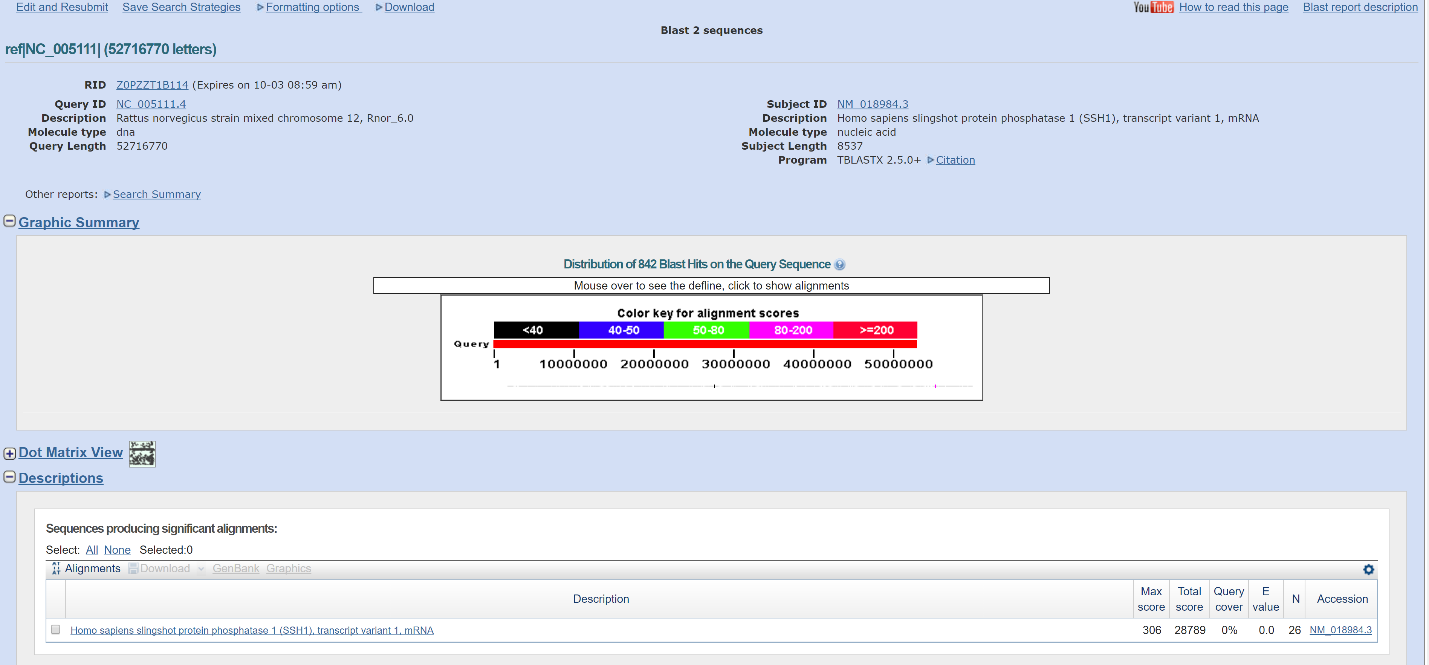
Exercicio 3:

BLASTN -

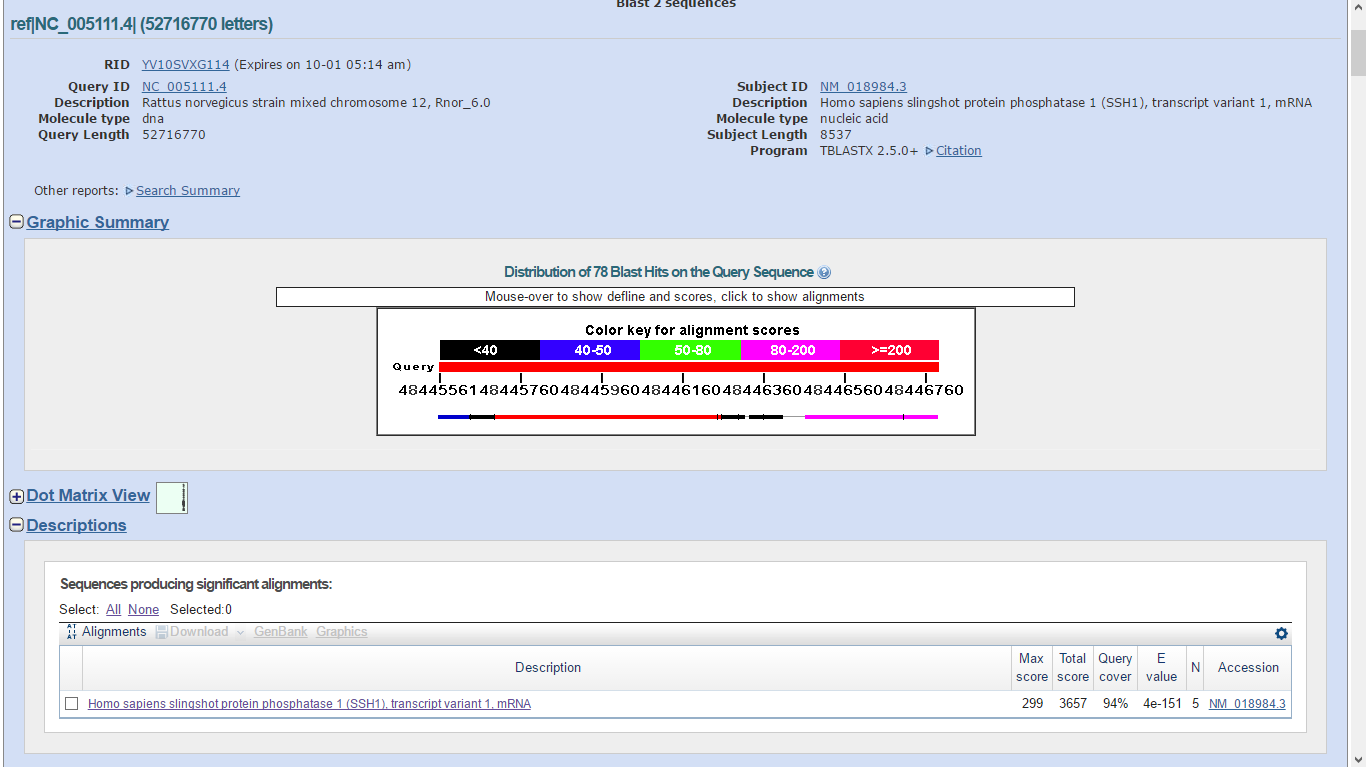




TBLASTX-



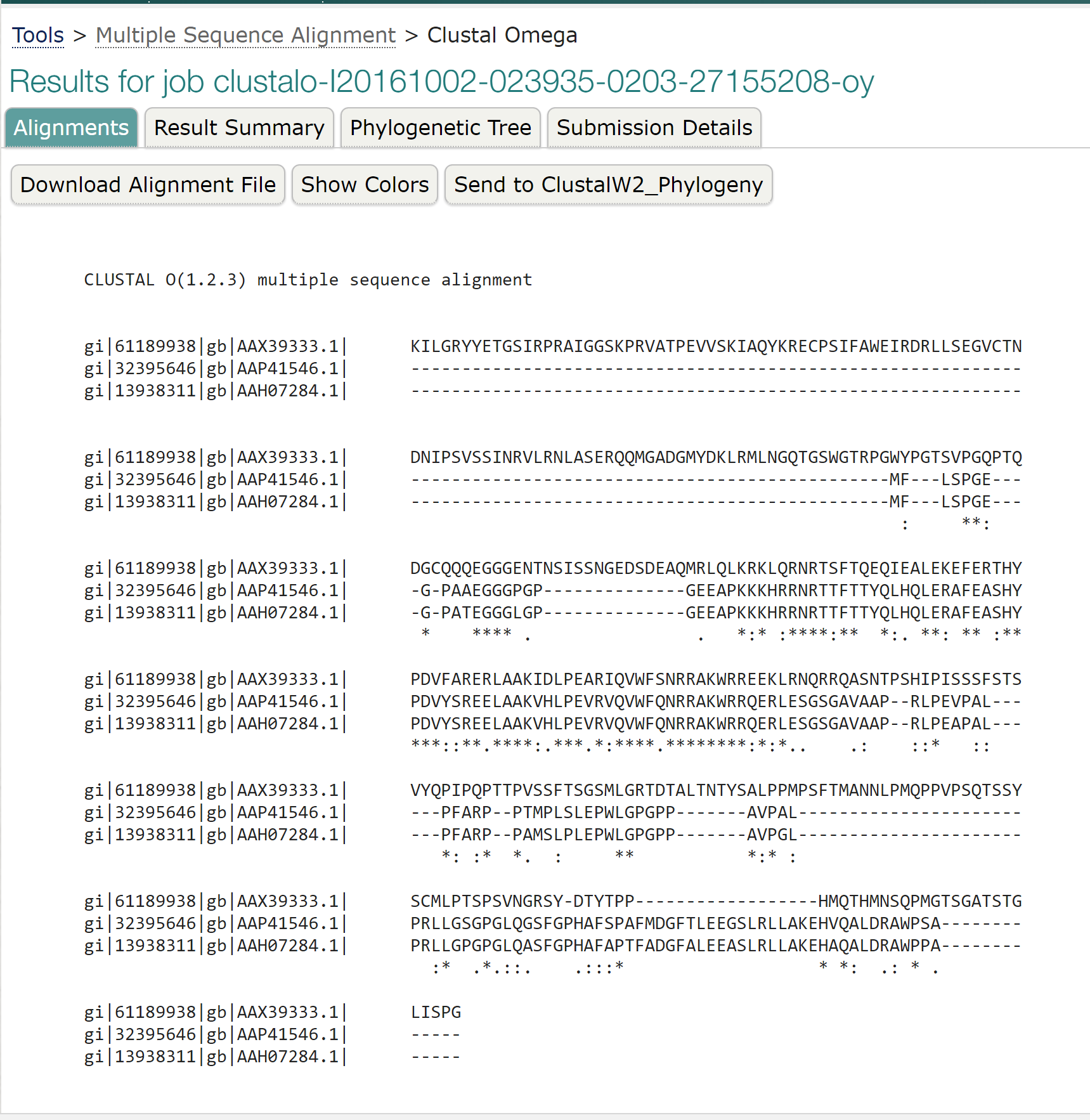
My results on the TBLASTX, are not right. Probably due to problems on the internet connection the online tool was not able to calculate the similarities with the genes. However, a class mate shared with me the screenshot of his result( since we ran the same genes). On his results you can observe that the Query Covers 94% on TBLASTX instead of 78% on BLASTN. Since TBLAST uses the BLOSUM, and it is more complex and precise than BLASTn.



So the difference between the BLASTn and TBBLASTx is because one analyses a more simple scoring function, while the TBLASTx uses BLOSUM. TBLASTx allows a “ overall alignment that reflects the biological similarity between the sequences more closely”. A more precise description is bellow taken from the book:

“Because nucleotides differ very little in biochemical functions, simple scoring functions are often used for DNA alignment. The simplest nucleotide matrix penalizes all changes equally; more complicated matrices can favor certain nucleotide substitutions over others (see Chapter 5). Amino acids, on the other hand, can be quite different from one another, and mismatches can be of varied effect depending on how similar or dissimilar amino acids are in their biochemical properties. Scores based on inferences about chemical or physical properties of proteins are possible and useful. It is well known that certain pairs of amino acids are much more likely to substitute for each other during evolution than others. This is likely due to certain physicochemical properties that they have in common, such as their hydrophobicity, size, or electrical charge. A good alignment should consider this and incorporate it into the scoring function so that the overall alignment reﬂects the biological similarity between sequences more closely. The two most common scoring functions that do this are based on observed substitution frequencies in proteins, and are called PAM and BLOSUM matrices (see Chapter 5 for more information). Whichever function we use, remember that the choice of scoring function can have a major impact on the resulting alignment, especially when choosing the “best” alignment, as deﬁned below.” (Cristianini,45)

Exercicio 4:



It is possible to observe that AAP41546.1 has a perfect alignment with AAH07284.1, meaning that the human and the cow’ protein perform the exact same function and did not suffer any evolutionary changes because both needed it for it survival. While the sheep’s protein sequence AAX39333.1 it is highly likely to be a protein that performs a totally different function since there is no similarities with the other two.