

1. How is PSI-BLAST different from blastp?

PSI-BLAST builds up a multiple alignment above the highest score pairs of a normal BLASTp program. By that, it takes the e-value into account. Then, PSI-BLAST produces a position-specific score matrix (PSSM) from the resulting multiple alignment. The score is dependent on the conservation status. In the following step the PSSM is updated with a second or more rounds of searches in the database to find other sequences with matching conservation score to add up to the matrix until desired or until convergence is reached. PSI-BLAST is therefore better to detect more distant sequences with similarity, since it connects conservation information of a range of related sequence into a single score matrix.

2. What is a PSSM? Explain what kind of information it contains and how it is used.

PSSM - position specific score matrix. PSIBLAST generates a multiple alignment of the pairs with the highest score from a previous BLASTp run dependent on a certain e-value threshold. It then produces the position specific score matrix, which shows the conservation pattern. The PSSM can be used as a new substitution matrix which can be updated in order to improve the results.

3. What does it mean when a PSI-BLAST search converges?

It means, that after convergence no new sequences are detected with the specific threshold.

4. Which sequence is most similar to P08487 when searching the Swissprot database according to PSI-BLAST first round?

P19174.1

5. Run PSI-BLAST iteration 2. Which one of the sequences that was found after iteration 2 but not in iteration 1 is most similar to your query protein P08487?

Q95M30.3

6. Explain briefly why you would use a multiple sequence alignment to study a group of related proteins.

It becomes easier to show evolutionary relationships, linkages and ancestors. With the resulting sequence homology one can do a phylogenetic analysis to assess their evolutionary origins. The MSA can be used to identify point mutations, insertion or deletion mutations, sequence conservation status of protein domains and tertiary or secondary structures.

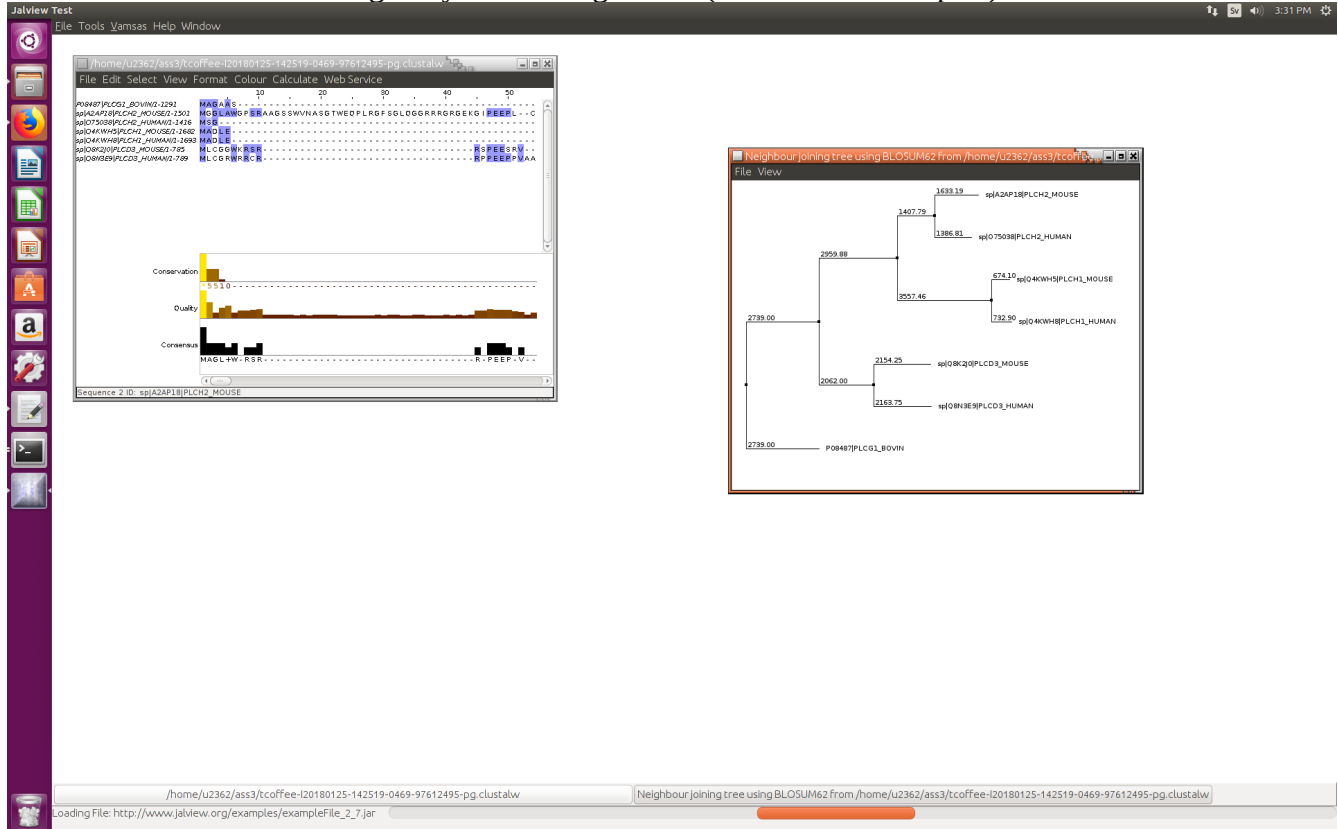
7. What does it mean that a site is conserved?

It means, that the site of amino acid has been preserved during evolution, which means that it was not changed, and it probably has an important function.

8. Color the alignment according to BLOSUM62. Explain the coloring scheme.

The gaps are colored white. It is colored dark blue, when a residue matches the consensus residue at that position. If there is a positive BLOSUM62 score, but the residues do not match, the color is light blue.

9. Calculate a phylogenetic tree using Neighbour joining tree BLOSUM 62. Include the image of the obtained tree and an image of your full alignment. (Hint: Save As/Export)



10. Explain (briefly) the ClustalX coloring scheme.

The residues in the alignment are assigned with a color, dependent whether a residue meets a minimum criteria, which is specific for the residue type, for that specific position in the alignment.

11. Include again an image of the alignment and the phylogenetic tree obtained.

