The general algorithm

We need nucleotide and amino acid sequences for all genes form two species, let’s call them human and mouse. We do blast of all human genes to mouse genes and reverse (using amino acid sequences). Than we pick those orthologous pairs for which human gene is the best hit for mouse gene and the mouse gene is best hit for human gene (it is not always the case because of gene duplications and gene lost in the genome). For orthologous pairs we take pairwise alignment of amino acid sequences and substitute corresponding nucleotide sequences instead of amino acids. Finally, we feed these nucleotide alignments into a software PAML 4. PALM counts the number of positions where non-synonymous (synonymous) substitution could happen and number of non-synonymous (synonymous) substitution which happened. Than:

dN=Number of non-syn substitutions/ Number of non-syn positions,

dS=Number of syn substitutions/ Number of syn positions

Some specific details

1. I took nucleotide sequences from here: <https://useast.ensembl.org/info/data/ftp/index.html> **CDS (FASTA)**This file has all isoforms for each gene. I use the longest of them.
2. Original blast is very slow. So, I use Diamond instead <http://www.diamondsearch.org/index.php>  
   It does the same but much faster. There could be some difficulties with it. There are several ways to install it, and one of them worked for me but the others didn’t.
3. We keep only those orthologous pairs which satisfy following requirments:

* Best bi-directional hit
* Length of alignment >30 aa
* Alignment cover >80% of both genes in a pair
* E-value is better than 10\*\*-6

1. Codeml from PAML is a software to calculate dN and dS. <http://abacus.gene.ucl.ac.uk/software/paml.html>

But I launch it via biopython. It I remember correctly you don’t need to install it independently.

1. You need to install biopython library to run my script. It makes it easier to work with sequence information.

I tried to keep and send to you the whole directory with all intermediate files. So, you can skip some steps. Actually, there are no need for you to redo blast. You can use any of the final files, for example, “blast\_clock\_short” and take orthologous pairs from it.