Explanations

The nature of the Arabidopsis proteome is composed of nucleotides sequences whereas the S. pombe one is composed of amino acids sequences.

To decide on "sensible" BLAST parameters, I did several research. First, to select the right BLAST option, I was on this site: https://www.arabidopsis.org/help/helppages/BLAST_help.jsp.

- I used **blastx** to compare a nucleotide query sequence translated in all reading frames against a protein sequence database.
- And I used **tblastn** to compare a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.

Then I considered that the result of a blast was significant, so that we could therefore consider two homologous sequences, if the **E-value < 10e-2**. I took this decision thanks to this website: https://www.metagenomics.wiki/tools/blast/evalue. They said *"Blast hits with an E-value smaller than 1e-50 includes database matches of very high quality. Blast hits with E-value smaller than 0.01 can still be considered as good hit for homology matches."*

Bonus: 1%

Reciprocal-best-BLAST is only the first step in demonstrating that two genes are orthologous.

Indeed, by definition, orthologs are genes derived from a common ancestor by speciation and so defined on a phylogenetic tree. Then, finding reciprocal best hits is not the proper way of inferring orthology. So it could be a good idea to verify the orthologs found by realize a phylogenetic tree.

I also read an article about an other method called the reciprocal smallest distance algorithm (RSD) on https://wall-lab.stanford.edu/docs/dpwall_2007_1.pdf.

"This approach improves upon the common procedure of taking reciprocal best Basic Local Alignment Search Tool hits (RBH) in the identification of orthologs by using global sequence alignment and maximum likelihood estimation of evolutionary distances to detect orthologs between two genomes. RSD finds many putative orthologs missed by RBH because it is less likely to be misled by the presence of close paralogs in genomes."