Method 4: Use a yeast protein sequence for identifying the corresponding protein in another organism (with sequence info compiled from research paper):

Another lab has compiled sequence info for some of our proteins of interest. So this method involves extracting the relevant sequences from their data & using them for the alignments.

1. Download the fasta sequence files from this site:

<https://www.embopress.org/doi/full/10.15252/embr.201744102#supplementary-materials>

The zip file you want is DATASET EV1

1. Remove the files for the following proteins (they don’t have an equivalent in yeast): ARHGEF17, Astrin, Borealin, BugZ, CenpF, CenpM, CenpR, Cep57, p31comet, Rod, Ska1, Ska2, Ska3, SKAP, Spindly, Zwilch
2. From the remainder files, extract sequences for the organisms listed below. Each file has a protein name as the file name & contains all the sequences for that protein from the different organisms. Note that some organisms will not be present in all files.

Homo sapiens, Mus musculus, Xenopus tropicalis, Takifugu rubripes, Danio rerio, Ciona intestinalis, Branchiostoma floridae, Saccoglossus kowalevskii, Drosophila melanogaster, Anopheles gambiae, Caenorhabditis elegans, Brugia malayi, Schistosoma mansoni, Nematostella vectensis, Tichoplax adhaerens, Amphimedon queenslandica, Mnemiopsis leidyi, Monosiga brevicollis, Salpingoeca rosetta, Capsaspora owczarzaki, Saccharomyces cerevisiae, Candida glabrata, Kluyveromyces lactis, Debaryomyces hansenii, Yarrowia lipolytica, Neurospora crassa, Schizosaccharomyces pombe, Cryptococcus neoformans, Ustilago maydis, Mucor circinelloides, Phycomyces blakesleeanus, Mortierella verticillata, Mortierella elongata, Conidiobolus coronatus, Coemansia reversa, Allomyces macrogynus, Catenaria anguillulae, Batrachochytrium dendrobatidis, Spizellomyces punctatus, Encephalitozoon intestinalis, Vavraia culicis, Edhazardia aedis

1. For each protein (file name above), use all of the sequences extracted above in step #3 for the alignment