

[00:00] Welcome to Bioinformatics for Biologists, I'm Stevie Bain, a researcher from the University of Edinburgh.

[00:08] In this video we are going to introduce the NCBI database, that's the National Center for Biotechnology Information, and explain how to do a keyword search for a gene or protein of interest. You can access the website at the address below: ncbi.nlm.nih.gov

[00:30] The database has a search bar that allows the user to search using keywords, similar to the way that you would use a web search engine. At the left-hand side there is a drop down menu that lets you choose which specific database you would like to search, for example nucleotide or protein. Alternatively, you can search across all databases. Let's run through an example, imagine we are interested in the enzyme catalase, we type 'catalase' into the search box and hit search.

[01:00] After a few moments, we should see a results page that looks like this. As you can see we have our search results for catalase categorised by database. We have some results in literature which include books and scientific journals. We have results in genes, we have some results in proteins that can be divided into conserved domains and clusters. We have some results in genomes and some results in genetics. We also have results in chemicals. The blue boxes next to each database tell us how many results we have in each. If we look at proteins, we can see that we have just under 400,000 search results.

[01:44] If we click on protein, we are taken to a page that shows us the search results for catalase in the protein database. As we can see, there are just under 400,000 results which is around 20,000 pages. Our first result is catalase from *Drosophila melanogaster*. If we look underneath, we can also see it has 506 amino acids. We can click on this description to find out more.

[02:14] When we click on this description, we are taken to a page that gives us some more information about our protein search result. We can see the NCBI reference sequence which contains the accession number the unique id for this sequence. We can also find out more info such as the source of the protein which we know is *Drosophila melanogaster*, but we can also find out the common name and a bit more about the organism's taxonomy.

[02:44] Here we can also find scientific literature related to the result. If we scroll back to the top, we can see FASTA. If we click on this, it takes us to the FASTA sequence of the protein. This first line is the defline - it contains information about the protein name and the species it comes from. Underneath we have the amino acid sequence, each amino acid is represented by one letter for example M for Methionine.

[03:14] We can also conduct more specific searches. Say, for example, we wanted to search for the human catalase protein. We would select protein from the dropdown menu and type catalase in the search bar; but this time we would use the Boolean operator capital AND followed by the species name, in this case *Homo sapiens*. We would then type square brackets ORGN.

[03:44] This time our results page is specifically showing matches in the protein database for catalase in Homo sapiens. As you can see, we have much fewer results - only 91 compared to around 400,000 in the last search. Each of these results has the name catalase followed by Homo sapiens in square brackets. Our first result has 527 amino acids.

[04:13] When we click on the description of our first result, we are taken once again to a page that gives us some more information about this protein. This includes the accession number and some more information about the source including taxonomical information. Once again, we also find scientific literature related to this result. Let's click on FASTA and take a look at the amino acid sequence.

[04:44] Here we have the amino acid sequence of catalase in FASTA format with the defline giving a description at the top and the amino acid sequence underneath.

[04:57] We hope you found this overview of the NCBI database and instructional video on how to do keyword searches useful. If you would like some more information about our project, please visit our website at 4273pi.org, you can also follow us on Twitter @4273pi.