Bioinformatics, computer lab XII

University of Potsdam WS 2019/2020, S. Hartmann

Networks: the KEGG database

1 KEGG

The Kyoto Encyclopedia of Genes and Genomes is a collection of databases. It contains information about biochemical reactions, the gene products that are involved in these reactions, as well as entire metabolic pathways. Today you will only explore a specific pathway map at KEGG.

☞ Start at `http://www.genome.jp/kegg/`. Navigate to the KEGG PATHWAY database as discussed in class.

☞ Under "1. Metabolism", you find the heading "1.1 Carbohydrate metabolism". Under "1.1 Carbohydrate metabolism", follow the link to "Ascorbate and aldarate metabolism".

This will lead you to a reference map of the corresponding pathways that was manually drawn. (With the help of a software, of course, but nevertheless manually done.) This reference map is a union of what is known from multiple species: not all the pathways are present and functional in all species, and we'll get to this aspect in a moment.

☞ Click on the Button "Help" on the upper right-hand corner of the page.

- What do boxes, open circles and boxes with rounded corners refer to?
  (The numbers in the boxes refer to the EC (Enzyme Commission) IDs for the gene product that carries out a given reaction.)
- What do the arrows with open and filled arrow tips refer to?

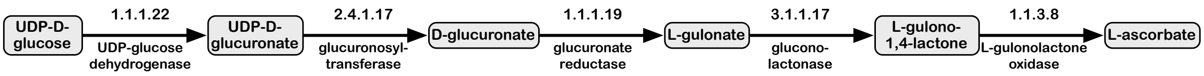
☞ Approximately in the center of the network there is a compound with the ID C00072 and the name "L-Ascorbate", with several arrows pointing to and from it. Do you know the common name for this molecule? If not, click on the circle that represents the L-Ascorbate node in the network to find out. You will be directed to an entry for this compound.

☞ Go back to the pathway. As mentioned, this is a generic reference pathway. If you want to see which of the involved genes have been identified in the genome of specific organisms, you can use the pulldown menu on the upper left-hand corner of the page to select an organism, where it currently says "Reference pathway":

- Select "Mus musculus" from the pulldown menu and click the "Go" button next to it. The enzymes/genes in green boxes have been identified in the mouse genome.
- You can click on any of the green boxes, and you will be directed to the corresponding entry for the gene: you'll find information about enzymatic activity, DNA and protein sequences, cross-links to other databases, and more.

- Go back to the pathway, with identified mouse-genes colored green.
- Do you expect the mouse to be able to synthesize vitamin C?
- To find out if the mouse can synthesize vitamin C, you need to check if all genes that are required for the biosynthesis pathway have been identified.

Below is a graph for the biosynthesis of Vitamin C in mouse, starting from UDP-D-glucose:



- In the graph above, what do the nodes/vertices represent, what do the edges represent?
- Is this graph directed or undirected? Weighted or unweighted? Cyclic or acyclic?
- The gene encoding glucuronate reductase seems to be missing in the mouse, which would mean that it can't synthesize vitamin C. However, there is another enzyme that can carry out this function: in a new browser tab, look up the alcohol dehydrogenase protein with the ID "AK1A1_MOUSE" in the Uniprot database. Under the heading "Function", "GO - Molecular function", the third activity listed is glucuronate reductase – which is exactly what is needed for the synthesis of Vitamin C.
- To summarize: can a mouse synthesize its own vitamin C, or is it dependent on dietary vitamin C uptake?

☞ Look up the vitamin C synthesis pathway of the following animals to determine if they are able to synthesize vitamin C. If the node for the enzyme glucuronate reductase isn't colored green, you can assume that alcohol dehydrogenase can substitute and carry out this function: cat (*Felis catus*), horse (*Equus caballus*), elephant (*Loxodonta africana*).

☞ Look up the vitamin C synthesis pathway of human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), and gorilla (*Gorilla gorilla*). Which genes/enzymes are missing? Note that there is no other enzyme that has L-gulonolactone oxidase (GULO) activity, so if the box for this enzyme (E.C. number 1.1.3.8) isn't colored green, this gene/enzyme/functionality is not present in the animal.

It seems that these three primates cannot synthesize their own vitamin C – although vitamin C is an indispensable compound! It acts as antioxidant, is involved in collagen synthesis, and has a number of other cellular functions. This means that humans, chimpanzees, and gorillas can only survive if their diet contains sufficient quantities of vitamin C, which it fortunately does!

☞ Can you draw the species tree of the animals you investigated today? Where on this tree could the ability to synthesize vitamin C have been lost?

☞ Without additional information, you don't know if the gene encoding L-gulonolactone oxidase was lost or mutated in the common ancestor of humans, chimps, and gorillas – or if it was lost three times independently within these lineages. Can you think of a sequence analysis that would, in theory, allow you to distinguish between these two options? (You don't have to do this analysis, but which approach would you suggest?)