Frequently Asked Questions

Conrad Schoch[1]

Created: May 10, 2011; Updated: February 21, 2020.

How can I tell when new species appear in GenBank?

You can register with MyNCBI to get 'What's New' email updates for any Entrez query. For example, if you are interested in amphibian species you can execute the Entrez query:

Amphibia[subtree] AND species[rank] AND specified[prop]

Click the 'Save search' hotlink above the Entrez search box to receive email updates whenever this query retrieves something new. There is documentation here –

http://www.ncbi.nlm.nih.gov/books/NBK3843

and some tutorials here –

http://www.nlm.nih.gov/bsd/disted/pubmed.html

http://www.nlm.nih.gov/bsd/viewlet/myncbi/saving_searches.html

How can I track the status of a list of taxa in the GenBank taxonomy?

If you have a list of taxa that are of interest to you, you can upload it (or copy and paste) to the Name/ID Status page – this will tell you which taxa are listed in our database, and whether they are primary or secondary names (synonyms, misspellings, etc.) You can track changes in the GenBank taxonomy by saving a copy of the status report, and comparing it with the status report for the same list of taxa at a later date.

Are taxids stable?

Taxids are assigned sequentially as taxa are added to the taxonomy database, and are never reused for a different taxon once they have been assigned. When two taxa are determined to be synonymous they are merged, and the taxid that disappears becomes a 'secondary' taxid for the one that remains. Taxa are occasionally removed from the taxonomy database (particularly internal nodes, during a taxonomic revision) – these taxids are deleted, and are not reused. Lists of merged and deleted taxids are included in the taxonomy dump files on the ftp site.

The public web pages at the NCBI only show taxa that are linked to public sequence entries. If a taxid is linked to a public sequence entry that is later removed, or assigned to a different taxon, the corresponding taxid may disappear from the NCBI web pages without appearing in either the merged or deleted taxid list.

Is there a regular taxonomy ftp update cycle, or are files updated as needed?

Curators do not make updates based on a specific schedule. Entrez Taxonomy is reindexed and updated weekly, the Taxonomy Browser is updated in real time (as the taxonomy database is edited by the taxonomy

Author Affiliation: 1 NCBI; Email: schoch2@ncbi.nlm.nih.gov.

group) and the Taxonomy FTP dump files are updated hourly. An archive of monthly historical taxdumps is available: ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/taxdump_archive/

Unpublished taxonomic names are not allowed in NCBI records. Consequently, placeholder names are substituted until such time as associated proposed novel names have been published. It is made clear to submitters that it is their responsibility to notify NCBI when a new name associated with their sequence submission(s) has been published. Unfortunately, most submitters do not do so. As a result, updates usually occur only when NCBI staff happen to come across a publication with a new taxonomic name or when database users, such as yourself, bring it to our attention. NCBI taxonomy staff do make an effort to find such names, but this is not their primary responsibility. It occurs only when they are able to find time to do so. There is no specific time cycle associated with such updates.

What does the Taxonomy Common Tree represent?

Found here: https://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi . The Common Tree display shows a hierarchical view of the relationships among the taxa and their lineages. This is NOT a phylogenetic tree but only a representation of our classification, to visualize the subset of the NCBI taxonomy that spans a set of named TaxNodes. The classification itself should resemble a combination of several different phylogenetic concepts - this is not consistently applied, includes uncertainly placed lineages and relies on whatever is broadly accepted in current taxonomic literature.