

## Cellosaurus newsletter of March 2016

Last week a new version of the Cellosaurus (<http://web.expasy.org/cellosaurus/>) was released containing information on more than 60'000 cell lines from 535 species.

While human (48'274), mouse (5'433) and rat (1'441) cell lines represent just over 90% of the entries, there is a growing wealth of entries regarding other species including insects, ticks, fish and birds. In this context it is useful to point out that the "Group" topic in the Cellosaurus allows users to search for cell lines belonging to such taxonomic categories. For example the query:

[http://web.expasy.org/cgi-bin/cellosaurus/search?input="fish cell line"](http://web.expasy.org/cgi-bin/cellosaurus/search?input=)

retrieves the list of fish cell lines. A "group" also exists for "insect", "tick", "amphibian", "reptilian", "bird", "bat", "cetacean" and "non-human primate".

The Cellosaurus is growing not only in terms of the number of cell lines it contains but also in terms of the richness of the captured information. In the last months two new important information items were added:

Information on the toxic chemical compound to which a resistant cell line has been generated was added starting in release 15. Today such information is provided for over 1'000 cell lines (most of them human cancer cells). There are now 135 different compounds listed, the most frequent being cisplatin and doxorubicin (adriamycin). The compound information is cross-referenced to ChEBI which we thank for creating the chemical entities that were not yet available in that resource.

On the subject of supporting ontologies we also thank the NCI Thesaurus team for creating many new terms that are used to represent the disease "status" of human and mammalian model organism cell lines. The Cellosaurus now contains over 25'000 cell lines linking to over 1'000 diseases in the NCI Thesaurus.

In release 16 we started indicating which biological, chemical or physical (radiation) "transformant" was used to create a transformed cell line or a model organism cancer cell. Such information is now available for 27'340 cell lines and we plan to complete this task in the coming months. By far the most represented transformant is EBV (more than 25'000 cell lines) due to the high number of lymphoblast cell lines (LC) distributed by the Coriell Biorepository, but SV40 (>1000 cell lines) and various adenoviruses and papillomaviruses (HPV) are also well represented.

Our next challenge is to add the STR (short tandem repeat) DNA profile for human cell lines when such information is available. The STR profile is critical to the important process of cell line authentication. We will start to introduce this information in the next release of the Cellosaurus (release 17) in April 2016.

Last but not least, we wish to thank the individual researchers, the representatives of cell line collections and companies and Amanda Capes-Davies of the International Cell Line Authentication Committee (ICLAC) which have answered our numerous e-mail inquiries which all aimed at providing the scientific community with the most accurate representation of cell line information in the Cellosaurus.