

## Cellosaurus newsletter 7 of April 2019

### 1) What changed from release 26 to release 29

Since our last newsletter in May 2018, we have deployed three new Cellosaurus releases; the last one is release 29 of February 25, 2019 which contains information on 110,948 cell lines from 609 species. During this one year period we added more than 3,000 new entries, but also updated the information regarding many existing entries. The most noteworthy update efforts were the following: the addition of sequence variation information to about 3,500 entries (now in 9,504 cell lines with over 14,000 somatic and genomic mutations), of population doubling time to ~1,000 entries (now available in 5,033 entries), and the retrofit of 7,000 additional cell lines with the age of the cell line donor at the time of sampling (currently 88,232 entries, 79% of all entries). We also continued to collect STR profile information and we now have 6,130 cell lines with an STR profile.

In release 28 we introduced information on the HLA alleles that are present in human cell lines. This data is stored in a new CC line termed "HLA typing". Here is an example of this new line:

```
CC   HLA typing: A*01:01:01,03:02; B*08:01; C*07; DRB1*04:08; DRB4*01:03:01:01 (IPD-  
IMGT/HLA; 25999).
```

Thanks a collaboration with the [IPD-IMGT/HLA](http://www.ebi.ac.uk/ipd/imgt/hla/) database the number of cell lines with HLA typing information will more than double from the current release (807 entries) to the next one (>2,200).

The Cellosaurus is now cross-referenced to 83 distinct resources as we have added cross-references to three more companies distributing cell lines, namely Applied Biological Materials (ABM), FujiFilm Cellular Dynamics, Inc (FCDI) and KYinno as well as three additional resources: the AntiBodies Chemically Defined (ABCD) database, the IARC TP53 database and the Sanger Cell Model Passports.

### 2) What is planned for release 30

The next release of the Cellosaurus which is planned for May will introduce two important changes.

#### 2.1 Genome ancestry

Julie Dutil from the Ponce Health Sciences University has accepted to make available in the Cellosaurus the data from the ECLA (Estimated Cell Line Ancestry) tool (<http://ecla.moffitt.org/>) developed by her group. Currently ECLA estimates the ancestry for ~1,400 cancer cell lines. This data will be stored in a new CC line termed "Genome ancestry". Here is an example of this new line:

```
CC   Genome ancestry: African=0.21%; Native American=0.92%; East Asian, North=1.76%;  
East Asian, South=0.01%; South Asian=1.21%; European, North=67.1%; European,  
South=28.8% (PubMed=30894373).
```

Of course this will look much nicer on the ExpASy web version of the Cellosaurus!

#### 2.2 Cell line entry 'history'

We will introduce a new line type ("DT") which provides information on when a Cellosaurus entry was created, when it was last updated and which version of the entry is currently available. Here is an example of how this line will look like in the text version of the Cellosaurus:

```
DT   Created: 06-06-12; Last updated: 07-09-18; Version: 5
```

### 3) The Cellosaurus on ExPASy

The traffic toward the Cellosaurus on ExPASy is continuing to increase. Since it was made available in May 2015, it has been visited 1.1 million times by 644,000 distinct users that have browsed 3.5 million pages. Two important changes were carried out on the web version of the Cellosaurus.

#### 3.1 New STR similarity search tool

Thibault Robin, a PhD student in our group, has developed a very powerful STR similarity search tool that was made publicly available a few weeks ago: <https://web.expasy.org/cellosaurus-str-search/>

With this tool you can search for similarities between your cell line samples and the human cell lines STR profiles stored in the Cellosaurus. Highlights of this tool are the following:

- It accepts all STR markers that have been used for authentication (31 different markers + amelogenin).
- You can select between 3 algorithms: [Tanabe](#) and two versions of the [Masters](#) algorithm.
- You can define search modes: deciding if amelogenin is taken into account or not and how the program treats empty markers in the query sample or in the stored STR data.
- The results are displayed in a user-friendly tabular format that can be exported as a file (CSV, JSON or PDF formats).
- If you wish to search for similarity to multiple cell line samples you can do so by using the "batch mode", i.e. by specifying a file that contains the STR marker information for the cell lines.
- It can be accessed programmatically through a RESTful API. The documentation for the API is found in the [help](#) page.

In addition to being available on the Cellosaurus home page, the tool is accessible from any human cell line entry that contains a STR profile and will search for similarity to that profile.

#### 3.2 Text version of a cell line entry

It is now possible to directly obtain from the server a text version of a cell line entry. As these pages have their own URLs (example: [https://web.expasy.org/cellosaurus/CVCL\\_0178.txt](https://web.expasy.org/cellosaurus/CVCL_0178.txt)) this should allow users that want to programmatically download a specific entry to do in RESTful manner.

### 4) Other news

We have added two new articles in the ExPASy [Invitromaticists spotlight](#) section: one on Jorgen Fogh and the other on Robert Ian Freshney. Both were written by Amanda Capes-Davis.

The Cellosaurus [paper](#) in the Journal of Biomolecular Techniques has been published and is referenced by DOI: 10.7171/jbt.18-2902-002; PMID: 29805321. Very recently, we participated in a eLIFE publication by Babic et al untitled [Incidences of problematic cell lines are lower in papers that use RRIDs to identify cell lines](#).

PS: Subscribe to our Twitter page (<https://twitter.com/Cellosaurus>) for tweets about new developments regarding the Cellosaurus and the universe of cell lines.