**Introduction**

The Ensemble Public FTP site provides biological information for many species in various formats. For this assignment, we will investigate files in the GFF3 format found at:

* [http://ftp.ensembl.org/pub/current\_gff3/Links to an external site.](http://ftp.ensembl.org/pub/current_gff3/)

The GFF3 format is described in (The ChatGPT can provide nice and short summaries):

* [https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.mdLinks to an external site.](https://github.com/The-Sequence-Ontology/Specifications/blob/master/gff3.md)

**Lecture 3**

Select an organism and download its corresponding GFF file.

Investigate this file with command line UNIX tools.

Find answers to the following questions:

* Tell us a bit about the organism. Hucho hucho, commonly known as the huchen or Danube salmon, is a large species of freshwater fish in the salmon family, Salmonidae. It is native to the Danube River basin in Central and Eastern Europe. The huchen is known for its size, with some individuals growing over 1.5 meters (about 5 feet) in length and weighing up to 50 kg (110 lbs).
* How many features does the file contain? 2449149
* How many sequence regions (chromosomes) does the file contain? 71639 - does the number line with sequence regions equate to the number of sequence regions?
* How many genes are listed for this organism? 50114
* What are the top-ten most annotated feature types (column 3) across the genome?
* 936981 exon
* 897978 CDS
* 91708 mRNA
* 80360 five\_prime\_UTR
* 71639 region
* 61879 three\_prime\_UTR
* 56505 biological\_region
* 50114 gene
* 2215 ncRNA\_gene
* 744 snoRNA
* Having analyzed this GFF file, does it seem like a complete and well-annotated organism? Yes !

Share any other insights you might note.

Create a text file that shows both how you downloaded the data and how you generated each of your results.  
  
Commit the file to your GitHub repository that you created for this course.  
  
Note that future assignments may ask someone else to repeat your findings. Build your report with repeatability/reproducibility in mind.

Submit the text file here to get graded.