



Gene set characterization



In this part of the course, we will be working on analysing the gene sets that we have annotated on the assembled genome sequence. The work packages will have three main objectives

1. Assess the quality of the annotated gene set with respect to

- 1. possible contaminations.  We could have done this already during the assembly, e.g. using [BlobTools](#), but we didn't...
- 2. completeness in terms of gene loci
- 3. completeness of the gene structure

2. Trace evolutionary changes in the gene set.  We will be concentrating on the loss of genes in *C. parvum*¹⁾

Once you have completed this set of exercises, you should have an idea how to

- **run the individual analyses.**  Bear in mind, however, that input data is diverse with respect to data formats and data set size. Depending on your own data, you may have to adjust the commands that we provide in the individual exercises.
- **determine quality measures of your genome assembly on the level of gene sets.**  Watch out, *QUALITY* is not measured in absolute quantities. You will have to specify for yourself, and probably for each analysis separately, what data you consider excellent/very good/good/acceptable/poor
- differentiate between methodological artefacts and evolutionary changes

Follow the links below to

- [1\) Contamination check - taXaminer](#)
- [2\) Gene set completeness - BUSCO](#)
- [3\) Gene set completeness - fCAT](#)

Use the following links to navigate through the course

- [Main page of the Physalia course](#)
- [Previous work package: Genome annotation](#)
- [Move to the next package: Phylogenetic profiles - fDOG](#)

¹⁾

why is analysing gene gain harder??

From:

<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/> - **Teaching**

Permanent link:

<https://applbio.biologie.uni-frankfurt.de/teaching/wiki/doku.php?id=general:bioseqanalysis:genesetanalysis>

Last update: **2024/02/15 06:58**

