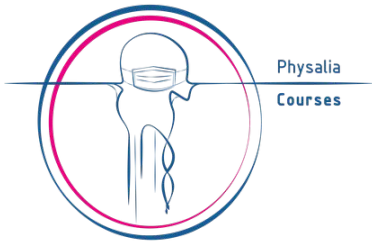
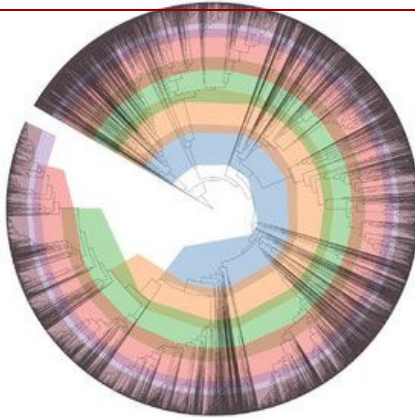


ENVIRONMENTAL METAGENOMICS

Physalia course, online, 11-15 November 2024

Gene catalog approach

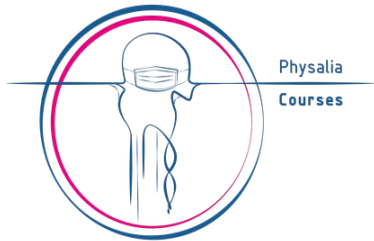
Nikolay Oskolkov, Lund University, NBIS SciLifeLab
Luis Pedro Coelho, Queensland University of Technology

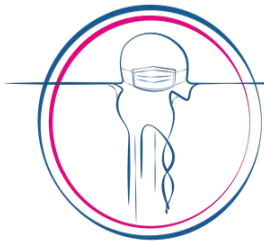


NB: original course material courtesy:
Dr. Antti Karkman, University of Helsinki
Dr. Igor Pessi, Finnish Environment
Institute (SYKE)


Gene catalogue approach

1. Take *contigs* (binned or unbinned)
2. Predict genes (ORFs)
3. Cluster all the genes together
 - a. 95% nucleotide identity
4. This is your **gene catalogue**
5. Annotate it with eggnog-mapper/RGI/...
6. Align reads to the catalog





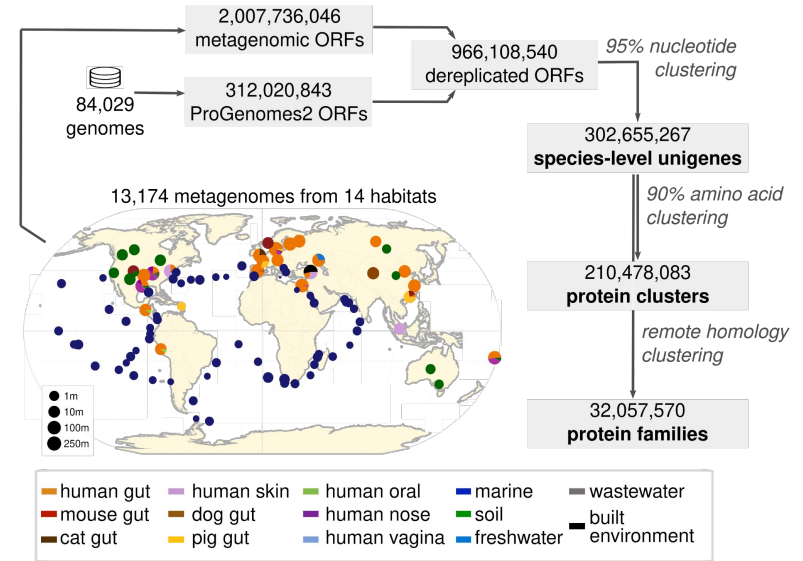
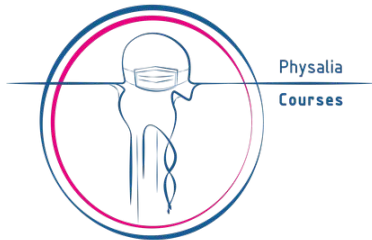
Existing gene catalogues

 **Global Microbial Gene Catalog v1.0**

The Global Microbial Gene Catalog is an integrated, consistently-processed, gene catalog of the microbial world, combining metagenomics and high-quality sequenced isolates. A total of 2.3 billion ORFs from 13,174 metagenomes (covering 14 habitats) and the complete [ProGenomes2](#) database were clustered together at 95% nucleotide identity to build a catalog of 302,655,267 unigenes. [Read more...](#)

Query by sequence or identifier	Find a unigene family (eggNOG orthologs)	Map a (meta)genome to the GMGC	Explore GMGBs (genomic bins)	Antibiotics resistance data
---------------------------------	--	--------------------------------	------------------------------	-----------------------------

Find homologues by sequence (BLAST-like) or search by identifier



From [\(Coelho et al., Nature, 2022\)](#)