



The future of the SILVA database as part of the DSMZ Digital Diversity

Martinique Frentrop, Jan Gerken, Maria Chuvochina, Yeliz Sandikci, Robin Goldmann and Lorenz C. Reimer

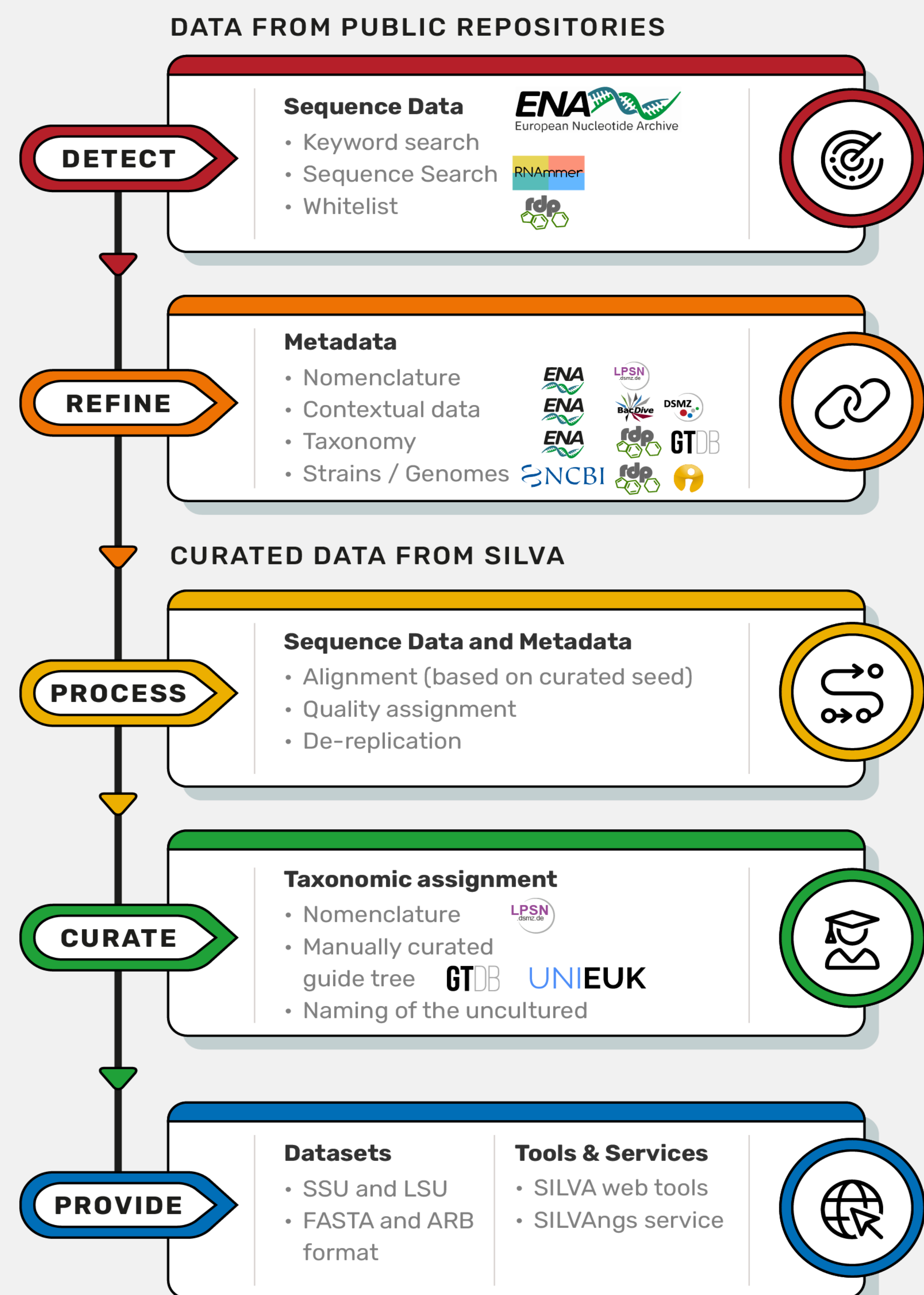
Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures

SILVA: The high-quality ribosomal RNA (rRNA) gene database

SILVA is a comprehensive resource of rRNA gene sequences spanning all three domains of life. Its primary purpose is to provide high-quality, standardized, and aligned reference datasets, a taxonomic framework, and tools for classifying microbial diversity across various ecosystems. SILVA is founding member of the German network for Bioinformatics (de.NBI) and ELIXIR Germany. In 2018, SILVA was recognized as ELIXIR Core Data Resource, followed by the acknowledgment as Global Core Biodata Resource in 2023.



Data processing workflow¹



Future developments

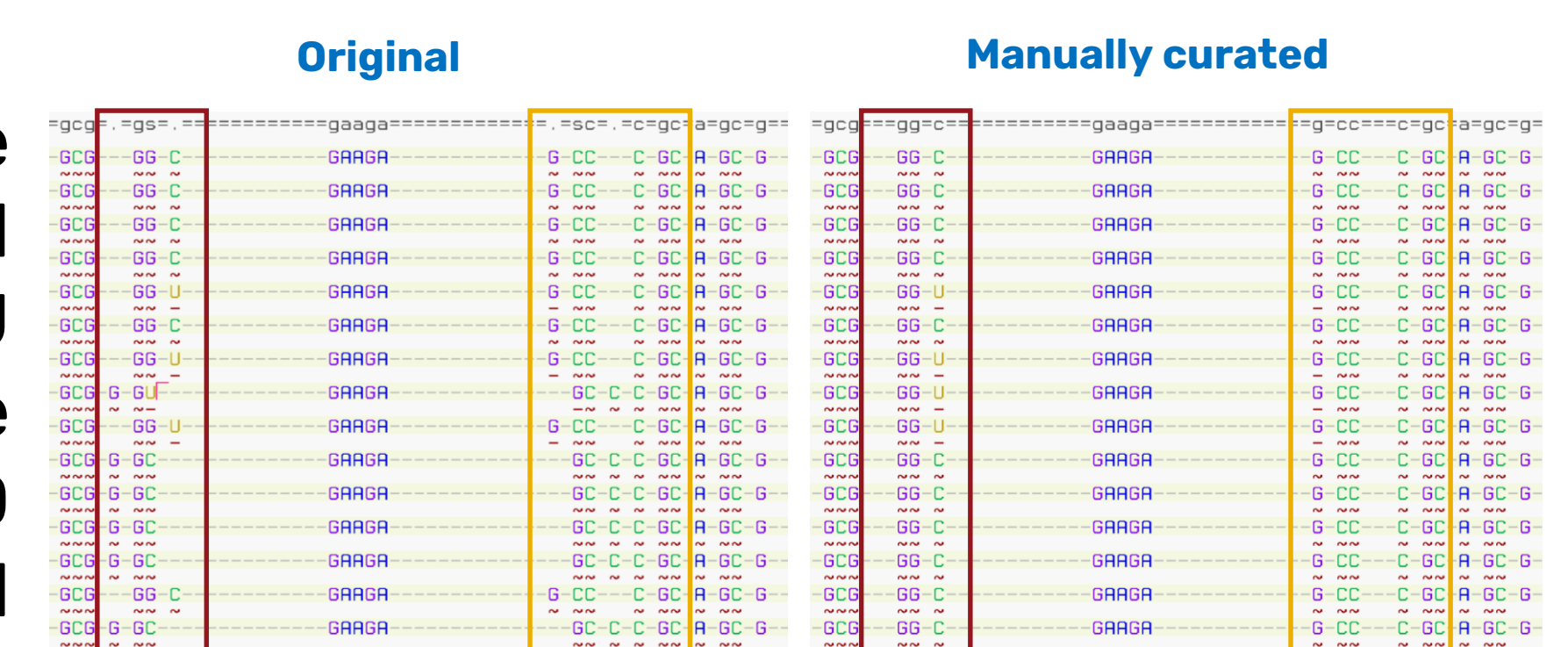
SILVA taxonomy

SILVA will continue to reconcile 16S rRNA gene and genome-based classifications by adopting the latest GTDB release. LPSN will serve be used as the primary source of nomenclature for prokaryotes, while UniEuk will become the official source of protist classification and nomenclature. Future developments will integrate de novo inferred phylogenies across all domains. To support user sequence classification, SILVA will provide classifiers for QIIME and other popular formats.



SILVA reference alignment (SEED)

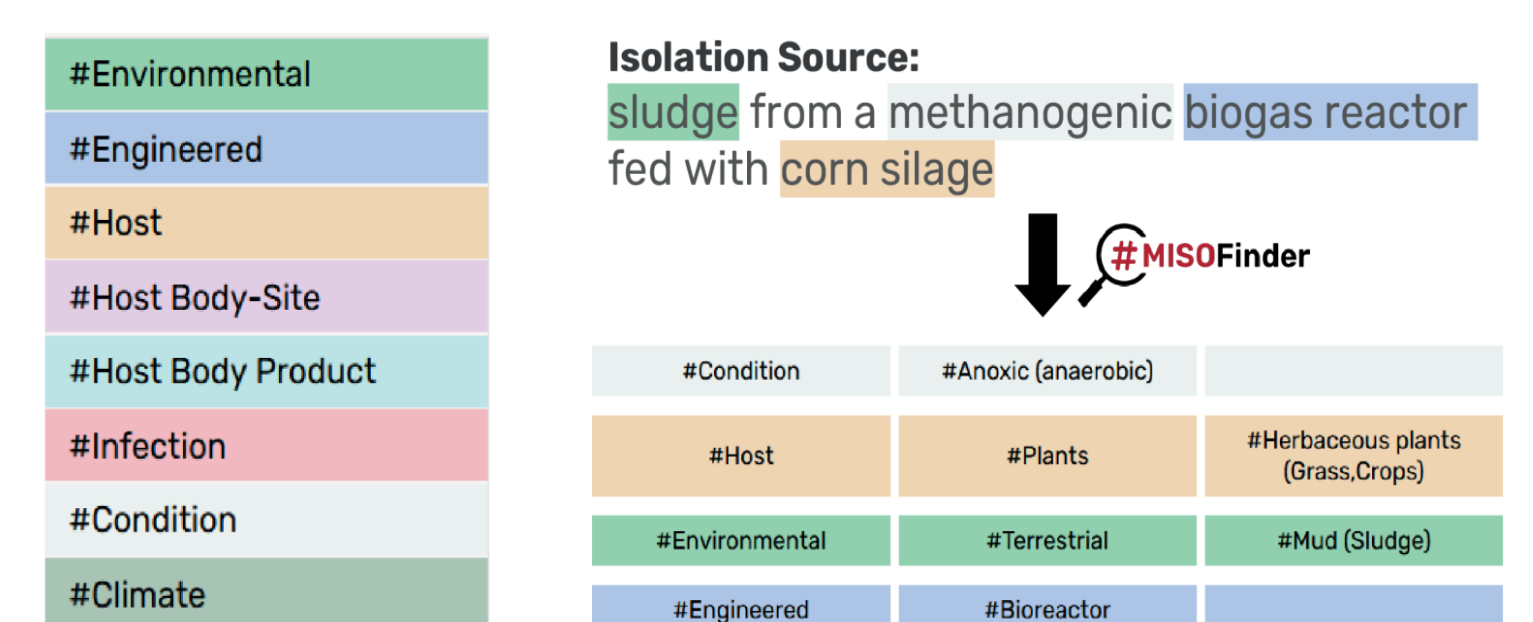
Due to the high diversity of sequences in the SEED, the alignment has artificially added gaps or incorrectly shifted bases. The LSU SEED was manually re-curated to improve alignment accuracy in SINA. The SSU SEED will be improved through manual curation and programmatic approaches.



The LSU SEED includes 4,508 rRNA sequences that represent phylogenetic branches of the SILVA LSU Ref NR 99. The boxes show typical misalignments and their post-manual curation improvements.

Standardizing metadata with #MISO Finder

Isolation sources are often provided in various formats. The Microbial Isolation Source Ontology (#MISO) Finder provides tags to supplement original isolation source entries with tags to standardize them and will be integrated into SILVA.

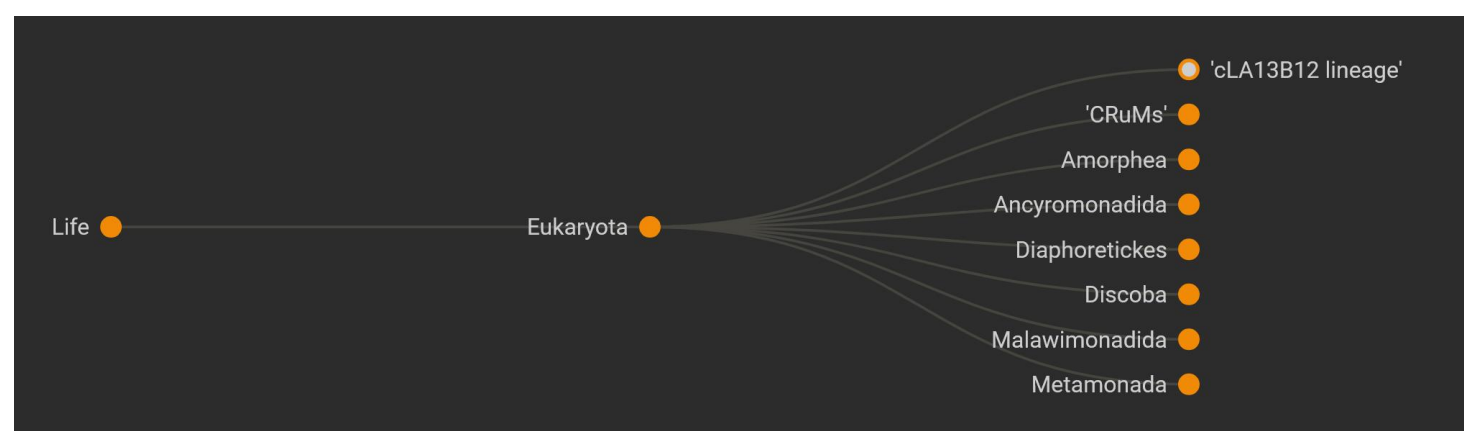


The first level of #MISO consists of 8 different terms. Isolation sources are mostly textual descriptions of the source from which the organism was obtained. In the future, the #MISO Finder will be developed and applied to transform the unstructured text into the standardized ontology.

SILVA's integration within DSMZ Digital Diversity

Adapting EukMap as TaxMap for LPSN

As part of the UniEuk project, SILVA developed a taxonomic curation platform. This platform will be adapted as TaxMap to visualize the taxonomic hierarchy of LPSN. The development aims to facilitate SILVA curation and improve users' access to LPSN-derived taxonomy.

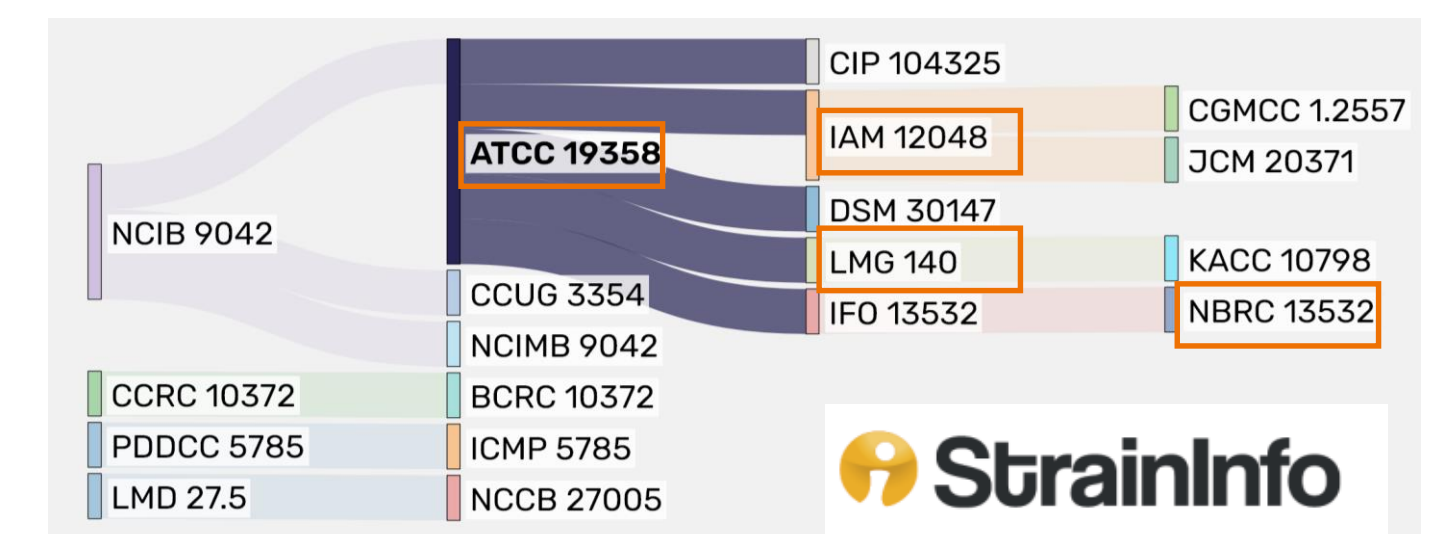


Screenshot from <https://eukmap.unieuk.net/taxonomy>



Integrating StrainInfo Identifiers

In the past, SILVA used Straininfo.net to indicate sequences derived from type strains or cultured isolates. We will continue this feature based on the new StrainInfo service. Future integration will also be realized via the DSMZ Digital Diversity Knowledge Graph, allowing users to establish connections between strain and sequence data.



Strain history for *Agrobacterium radiobacter* (<https://straininfo.dsmz.de/strain/311029>)

Meet the team



Jan Gerken
Technical Lead



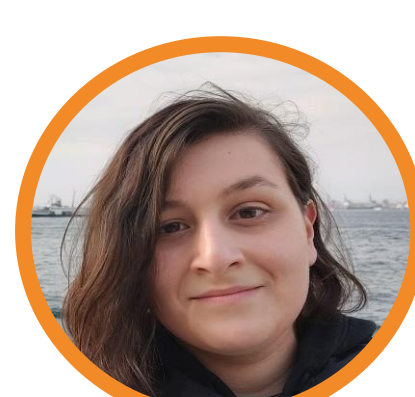
Maria Chuvochina
Scientific Lead



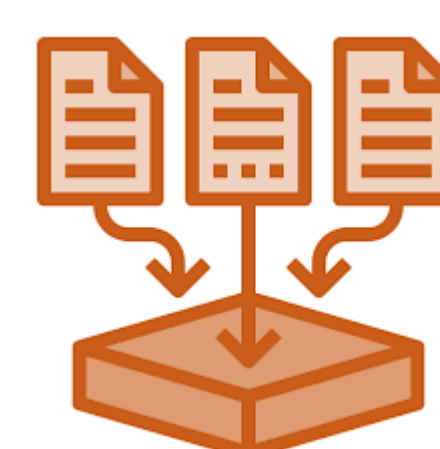
Martinique Frentrop
Metadata Curator



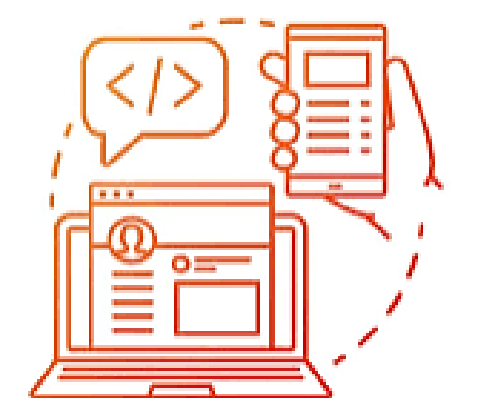
Robin Goldmann
Pipeline Developer



Yeliz Sandikci
Backend Developer



Interested in Frontend Development or Database Curation work?



Join us and get in contact!

1. Koblitz, J., et. al. (2024). DSMZ Digital Diversity Report. p.25–29. 10.5281/zenodo.14203734