

DSMZ Digital Diversity: Building a global biodata infrastructure

Lorenz Christian Reimer, Julia Koblitz, Joaquim Sardá Carbasse, Antje Jäde, Julia Hauenstein, Lisa Jeske, Christian-Alexander Dudek, Jan Gerken, Robin Goldmann, Martinique Frentrup, Markus Göker, Jan P. Meier-Kolthoff, Heike M. Freese, Patrick Nguyen, Adam Podstawka, Boyke Bunk, Jörg Overmann

¹*Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen*

For over a decade the DSMZ has been developing and providing biodata webservices for the wider microbial research community, such as the Bacterial Diversity Metadatabase BacDive or the “Prokaryotic Names Up-to-date” (PNU) database, which was combined with the List of Prokaryotic Names with Standing in Nomenclature (LPSN) in 2019. More recent developments are the Type Strain Genome Server (TYGS), a high-throughput database-driven platform for state-of-the-art genome-based taxonomy of prokaryotes and MediaDive, the largest database for standardized cultivation media for microorganisms.

Here we present the development of the new IT platform DSMZ Digital Diversity, that integrates the two renowned databases BRENDA and SILVA into the DSMZ database portfolio, with the aim to build a global biodata infrastructure. BRENDA is the world’s most comprehensive enzyme database, providing enzyme and enzyme-ligand related data. SILVA is a sequence database specialized on providing high-quality ribosomal RNA sequence data for identification of microorganisms. Together this database consortium covers the fields taxonomy, nomenclature, metabolism, phenotypic and environmental data, as well as high quality ribosomal and genomic sequence data. The goal is to establish an all-encompassing platform for standardized and integrated data and state-of-the-art analysis tools. The DSMZ Digital Diversity (<https://stb.dsmz.de>) will serve researchers as a one-stop-shop for connecting, retrieving and analyzing data and thereby will pave the way for large scale high-throughput analyses.