

# GEBA phase VI: the functional genomics and pangenomics of type strains



Michael Pester  
Head of the Dept. Microorganisms, DSMZ

Markus Göker  
GEBA project coordinator at DSMZ



<https://gold.jgi.doe.gov/gebaVI>

# Genomic Encyclopedia of Bacteria and Archaea

The GEBA project and its follow-up projects aim at systematically filling in genome gaps along the bacterial and archaeal branches of the tree of life.

Main contributors at DSMZ:

Hans-Peter Klenk, Markus Göker, Jan P. Meier-Kolthoff, Richard Hahnke,  
Brian Tindall, Stefan Spring, Rüdiger Pukall, Katharina Huber, Sabine  
Gronow, Elke Lang

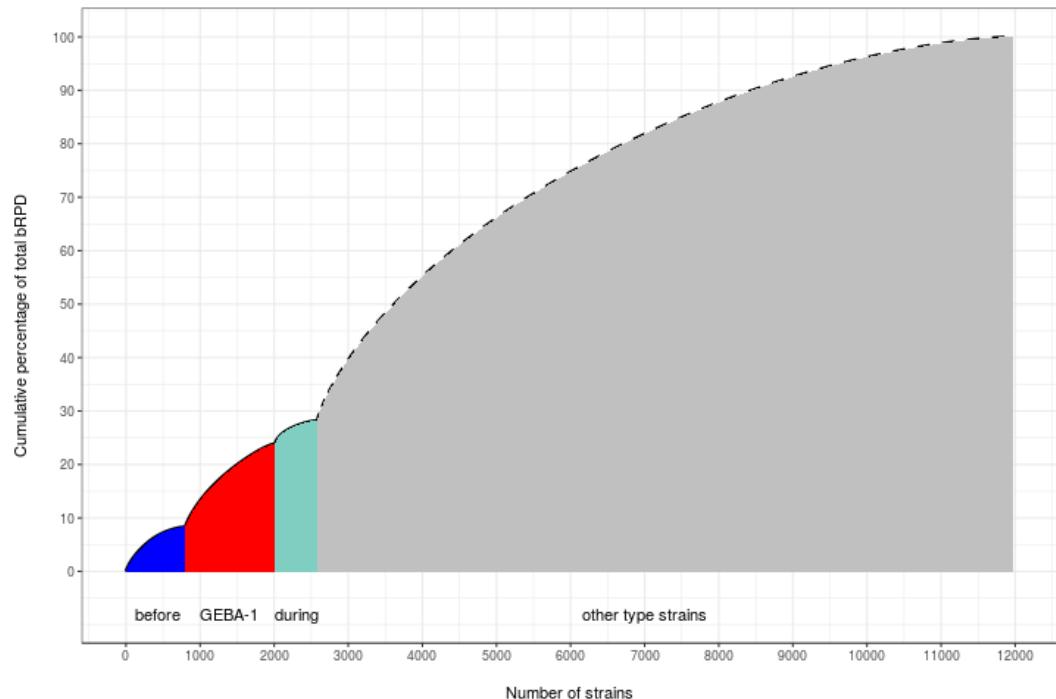
# Large-scale projects as sources of bacterial and archaeal type-strain genome sequences

- **GEBA**: finished, 250 type strains
- **KMG-1**: finished, 1000 type strains
- **KMG-2**: ongoing, 1000 type strains
- **ACTINO-1000**: ongoing, 1000 actinobacterial strains, PacBio sequencing
- **KMG-4**: ongoing, 1000 type strains of metagenomic interest
- **KMG-6**: started, 10,000 strains



## KMG-1

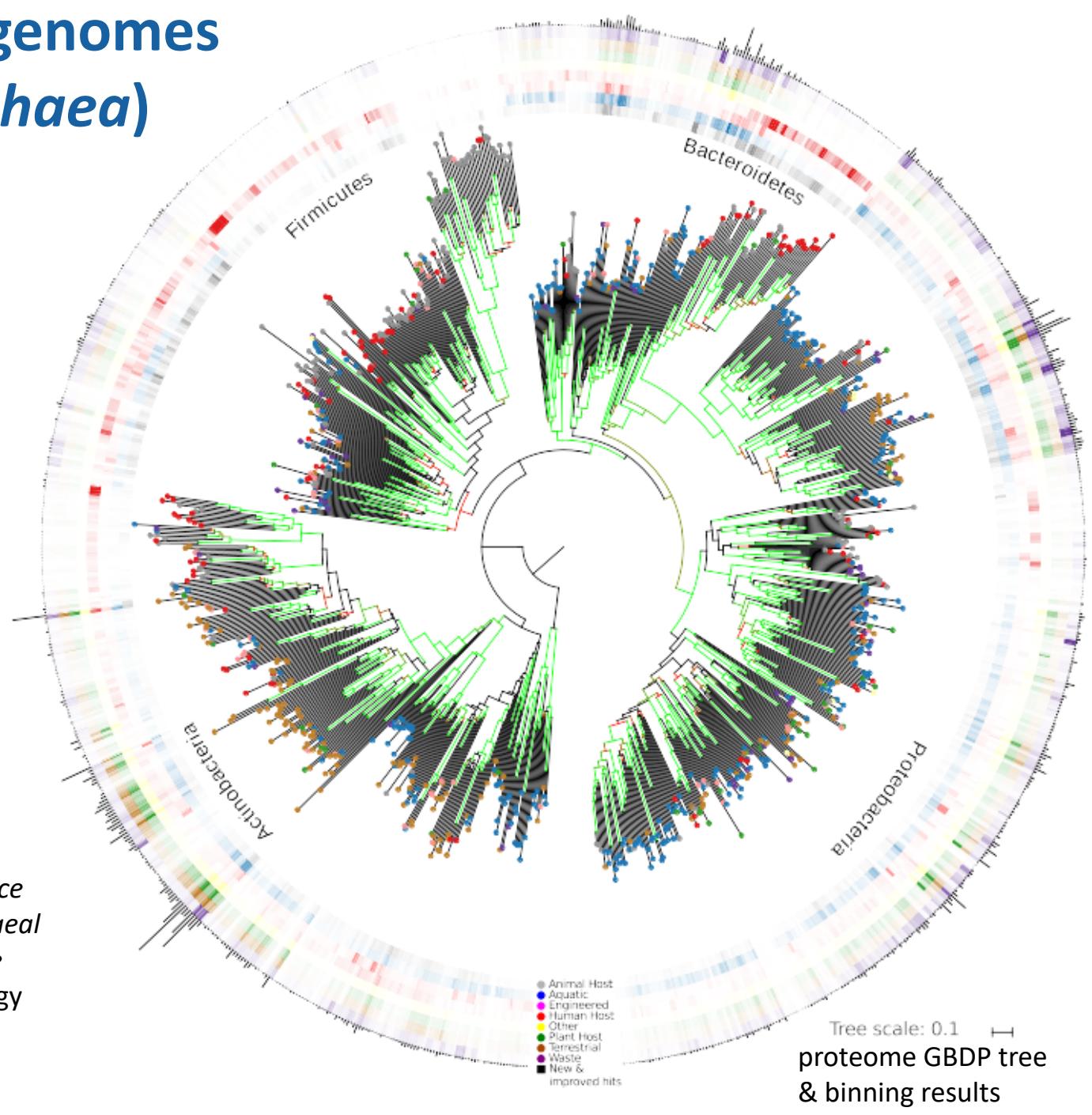
- *phylogeny-driven* genome-sequencing project
- increased phylogenetic coverage by genome sequences
- coverage increased considerably stronger than in previous projects



# 1000 KMG-1 genomes (*Bacteria, Archaea*)

- improved metagenomic binning
- detection of large number of biosynthetic clusters
- detection of new gene families
- highly resolved phylogenomic tree

Mukherjee et al., 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nature Biotechnology* 35: 676-683, 2017  
(doi:10.1038/nbt.3886)



# 1000 KMG-2 genomes

count	taxon
-----	
338	Bacteroidetes
192	Alphaproteobacteria
143	Actinobacteria
103	Firmicutes
45	Gammaproteobacteria
18	Deltaproteobacteria
18	Euryarchaeota
15	Betaproteobacteria
12	Tenericutes
10	Aquificae
8	Spirochaetes
3	Acidobacteria
3	Epsilonproteobacteria
3	Fusobacteria
2	Planctomycetes
2	Verrucomicrobia
1	Chloroflexi
1	Fibrobacteres
1	Thermotogae

## Genome-Based Taxonomic Classification of *Bacteroidetes*

Richard L. Hahnke<sup>1†</sup>, Jan P. Meier-Kolthoff<sup>1†</sup>, Marina García-López<sup>1</sup>, Supratim Mukherjee<sup>2</sup>, Marcel Huntemann<sup>2</sup>, Natalia N. Ivanova<sup>2</sup>, Tanja Woyke<sup>2</sup>, Nikos C. Kyripides<sup>2,3</sup>, Hans-Peter Klenk<sup>4</sup> and Markus Göker<sup>1\*</sup>

<sup>1</sup> Department of Microorganisms, Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, <sup>2</sup> Department of Energy Joint Genome Institute (DOE JGI), Walnut Creek, CA, USA, <sup>3</sup> Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, <sup>4</sup> School of Biology, Newcastle University, Newcastle upon Tyne, UK

## Genome-Scale Data Call for a Taxonomic Rearrangement of *Geodermatophilaceae* \*

Maria del Carmen Montero-Calasanz<sup>1,2\*</sup>, Jan P. Meier-Kolthoff<sup>2</sup>, Dao-Feng Zhang<sup>3</sup>, Adnan Yaramis<sup>1,4</sup>, Manfred Rohde<sup>5</sup>, Tanja Woyke<sup>6</sup>, Nikos C. Kyripides<sup>6</sup>, Peter Schumann<sup>2</sup>, Wen-Jun Li<sup>3\*</sup> and Markus Göker<sup>2\*</sup>

<sup>1</sup> School of Biology, Newcastle University, Newcastle upon Tyne, United Kingdom, <sup>2</sup> Leibniz Institute, German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, <sup>3</sup> State Key Laboratory of Biocontrol and Guangdong Provincial Key Laboratory of Plant Resources, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, <sup>4</sup> Department of Biotechnology, Middle East Technical University, Ankara, Turkey, <sup>5</sup> Central Facility for Microscopy, Helmholtz Centre for Infection Research, Braunschweig, Germany, <sup>6</sup> Department of Energy, Joint Genome Institute, Walnut Creek, CA, United States

## Genome-based classification of micromonosporae with a focus on their biotechnological and ecological potential \*

Lorena Carro<sup>1</sup>, Imen Nouiou<sup>1</sup>, Vartul Sangal<sup>2</sup>, Jan P. Meier-Kolthoff<sup>1,3</sup>, Martha E. Trujillo<sup>4</sup>, Maria del Carmen Montero-Calasanz<sup>1</sup>, Nevzat Sahin<sup>1,5</sup>, Darren Lee Smith<sup>2</sup>, Kristi E. Kim<sup>6</sup>, Paul Peluso<sup>6</sup>, Shweta Deshpande<sup>7</sup>, Tanja Woyke<sup>8</sup>, Nicole Shapiro<sup>7</sup>, Nikos C. Kyripides<sup>7</sup>, Hans-Peter Klenk<sup>1</sup>, Markus Göker<sup>1,3</sup> & Michael Goodfellow<sup>1</sup>

## Genome-Based Taxonomic Classification of the Phylum *Actinobacteria* \*

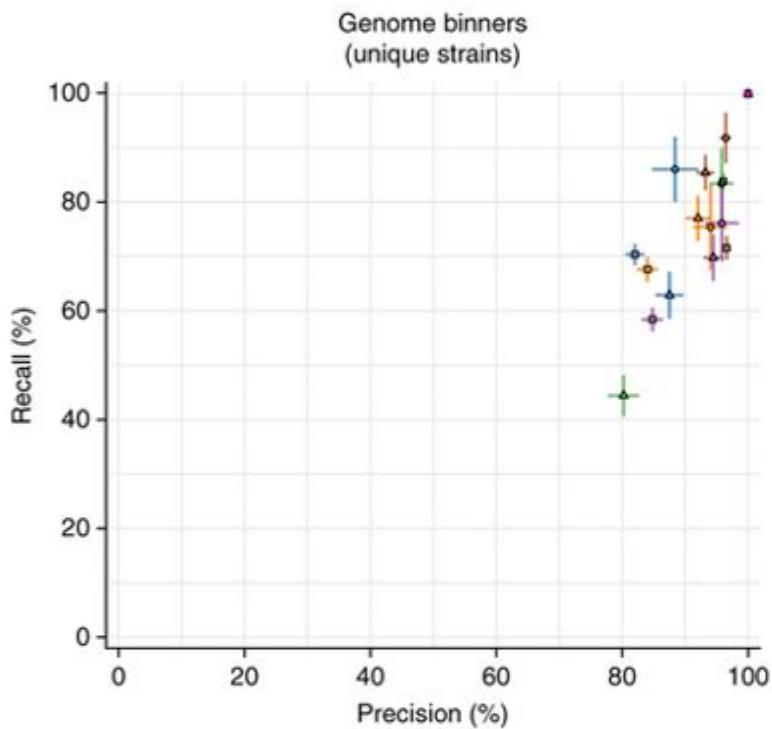
Imen Nouiou<sup>1†</sup>, Lorena Carro<sup>1†</sup>, Marina García-López<sup>2†</sup>, Jan P. Meier-Kolthoff<sup>2</sup>, Tanja Woyke<sup>3</sup>, Nikos C. Kyripides<sup>3</sup>, Rüdiger Pukall<sup>2</sup>, Hans-Peter Klenk<sup>1</sup>, Michael Goodfellow<sup>1</sup> and Markus Göker<sup>2\*</sup>

<sup>1</sup> School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, <sup>2</sup> Department of Microorganisms, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, <sup>3</sup> Department of Energy, Joint Genome Institute, Walnut Creek, CA, United States

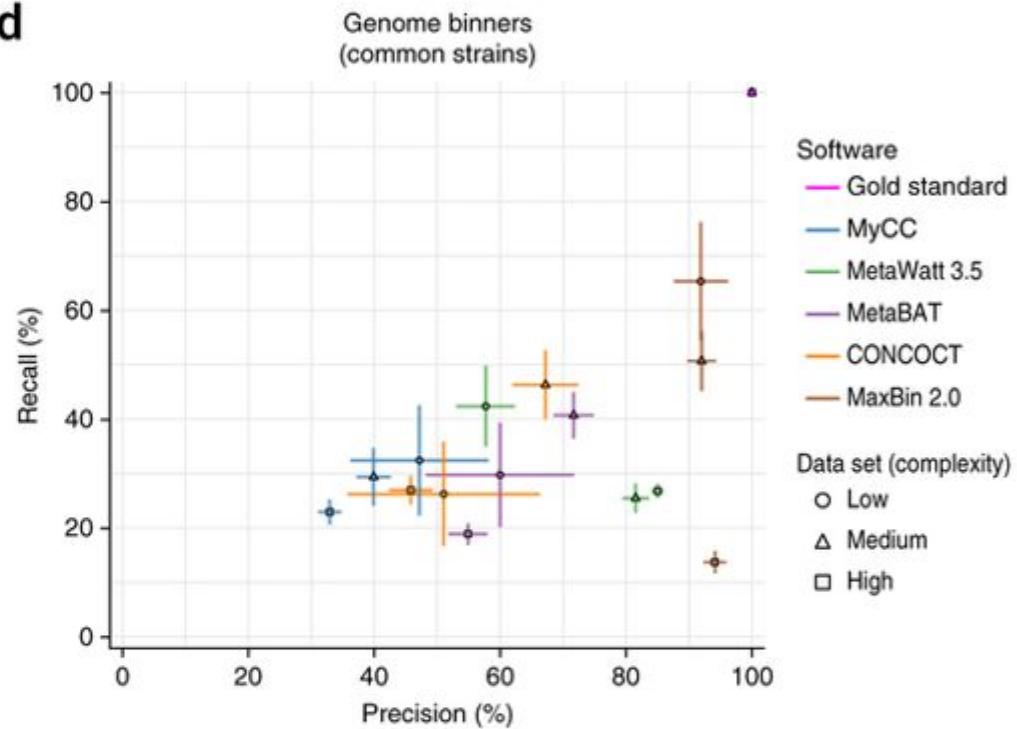
\*) genome sequences also from other projects, e.g. ACTINO-1000

# KMG-4 directly targets phylogenetic regions difficult for metagenomic binning

c



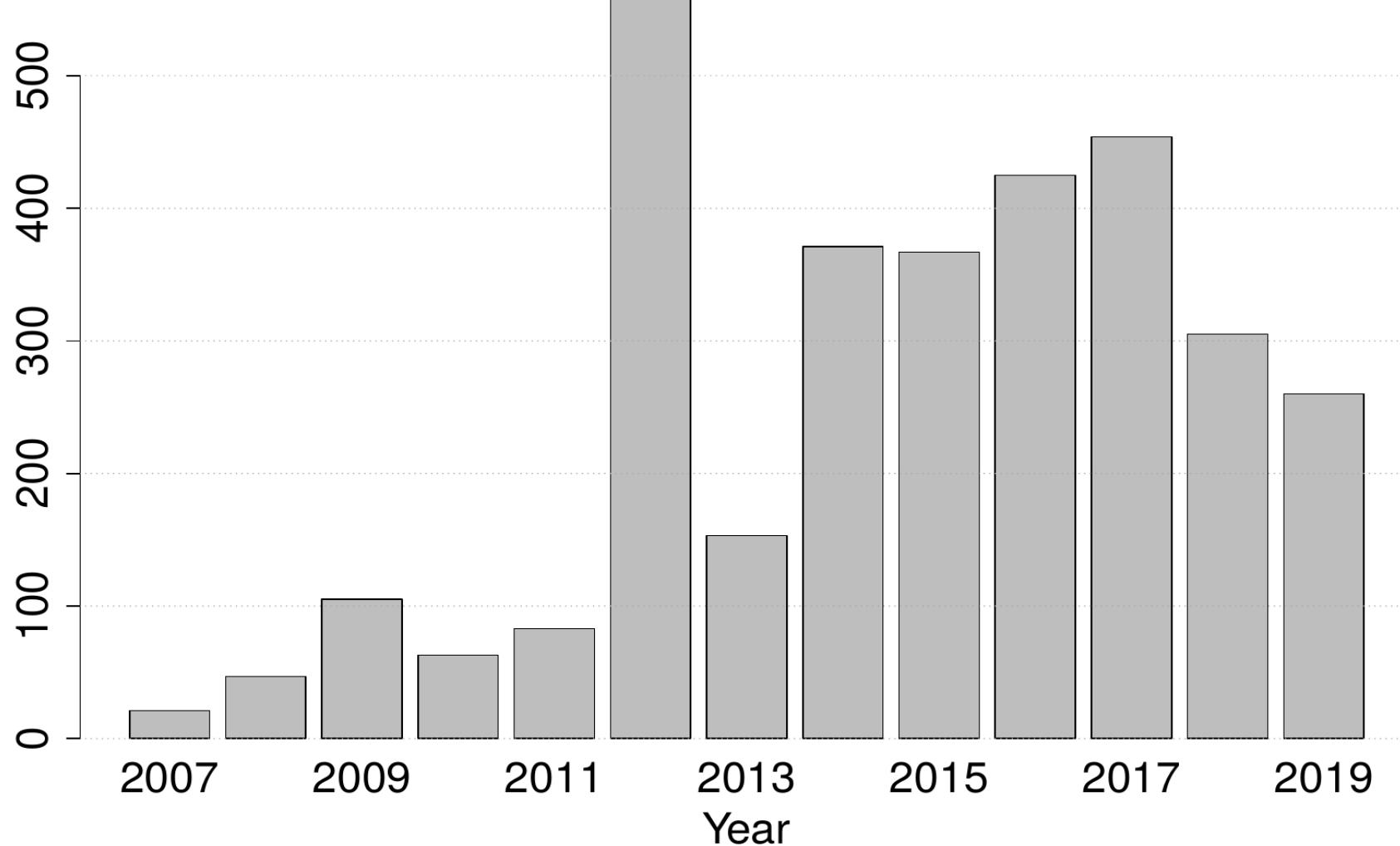
d



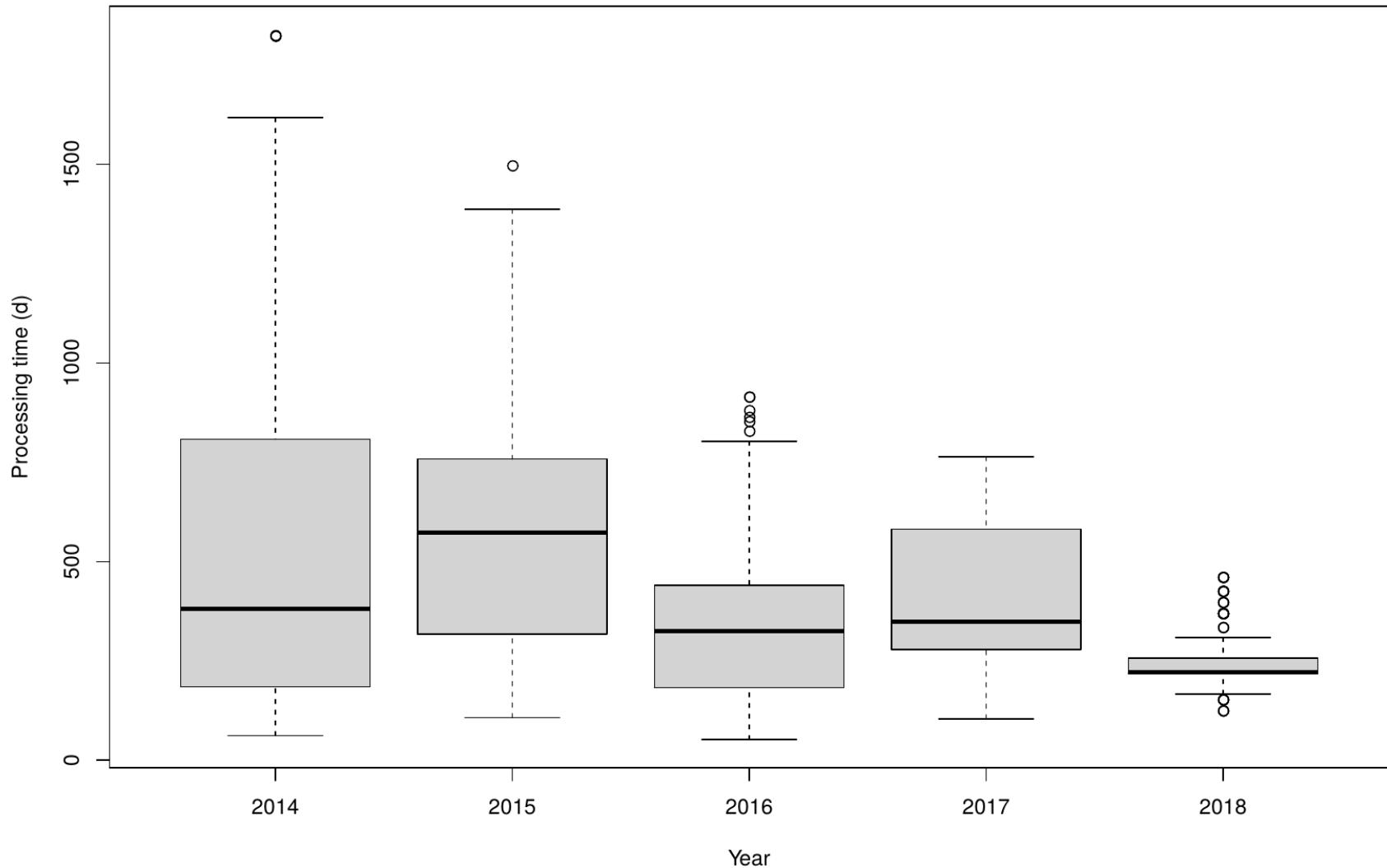
Sczyrba et al., *Critical Assessment of Metagenome Interpretation – a benchmark of computational metagenomics software*. Nature Methods, 2017 (doi:10.1038/nmeth.4458).

Performance of metagenomic binning dramatically decreases for lower taxonomic ranks

# Increasing contribution of DSMZ to GEBA



# Streamlined processing times



# The new GEBA VI phase – open to everybody!

## The goal

Improve our understanding of gene function for a wide variety of prokaryotes utilizing the tools of genomics, pangenomics, transcriptomics and exometabolomics.

## The approach

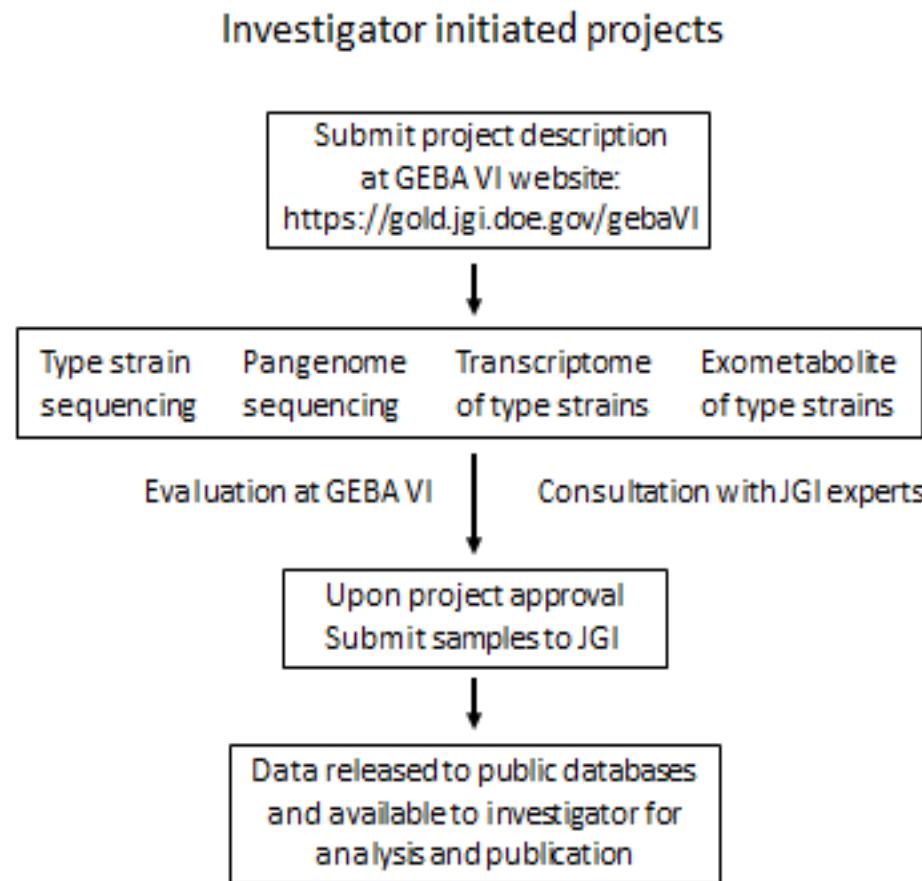
- Two experimental approaches
- **comparative genomics of type strains**, including surveys of genome sequences, transcriptomes and exometabolomes
- in depth studies of a **single species**, including characterizations of the **pangenome** or **transcriptomes** and **exometabolomes** under different growth conditions

## Available resources

- Illumina sequencing of the genomes of 10,000 prokaryotes
- These genomes will comprise 5,000 type strains
- The remainder will be 5-10 reference strains for determination of the pangenomes for 500-1000 species
- transcriptome and exometabolites can be determined for up to 1,000 type strains

<https://gold.jgi.doe.gov/gebaVI>

# The Potential contribution of DSMZ to GEBA VI

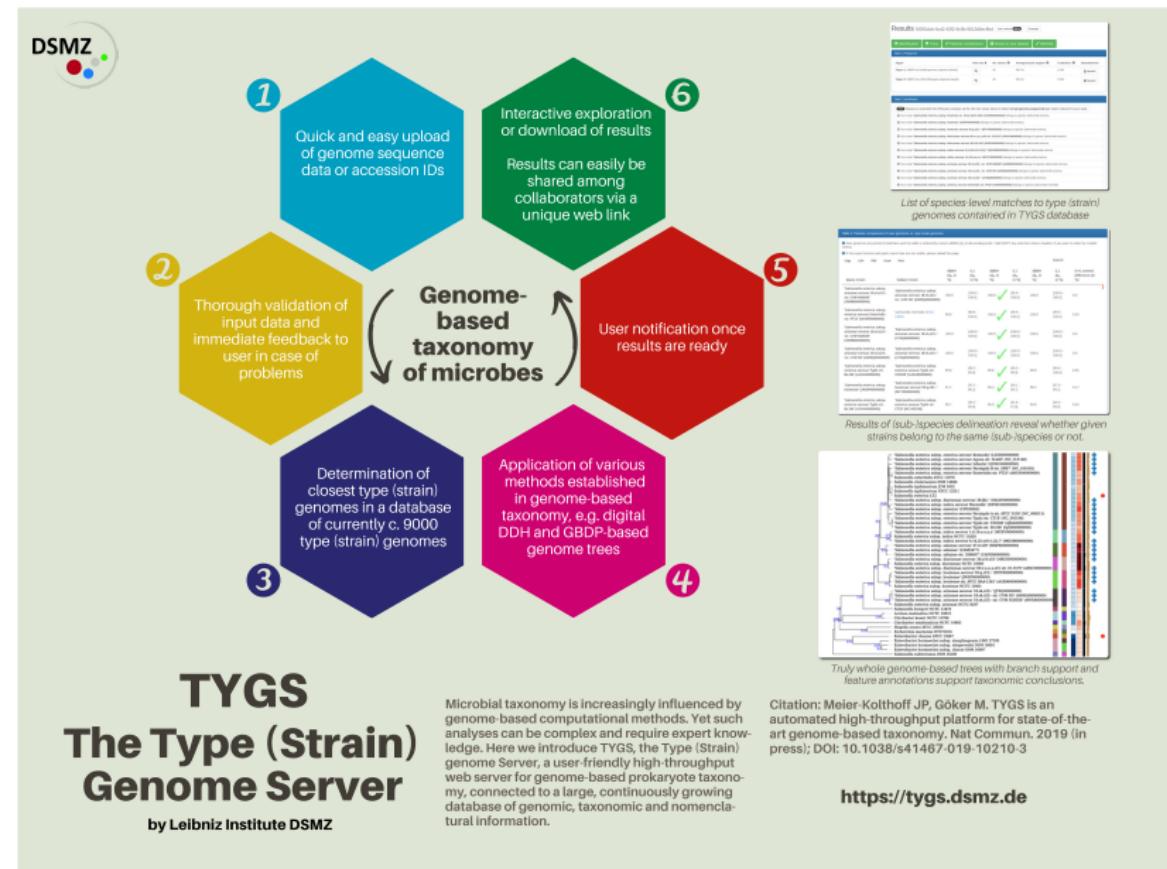


## Available DSMZ-strains

- > 4,000 unsequenced **type strains**
- > 11,000 unsequenced non-type strains

# Type (Strain) Genome Server

- Large collection (>9000) of verified type-strain genomes
- Linked to information on nomenclature and taxonomic literature
- Species and subspecies assignment of user-defined genomes
- Automated detection of closest genomes
- Genome-based phylogenetic tree

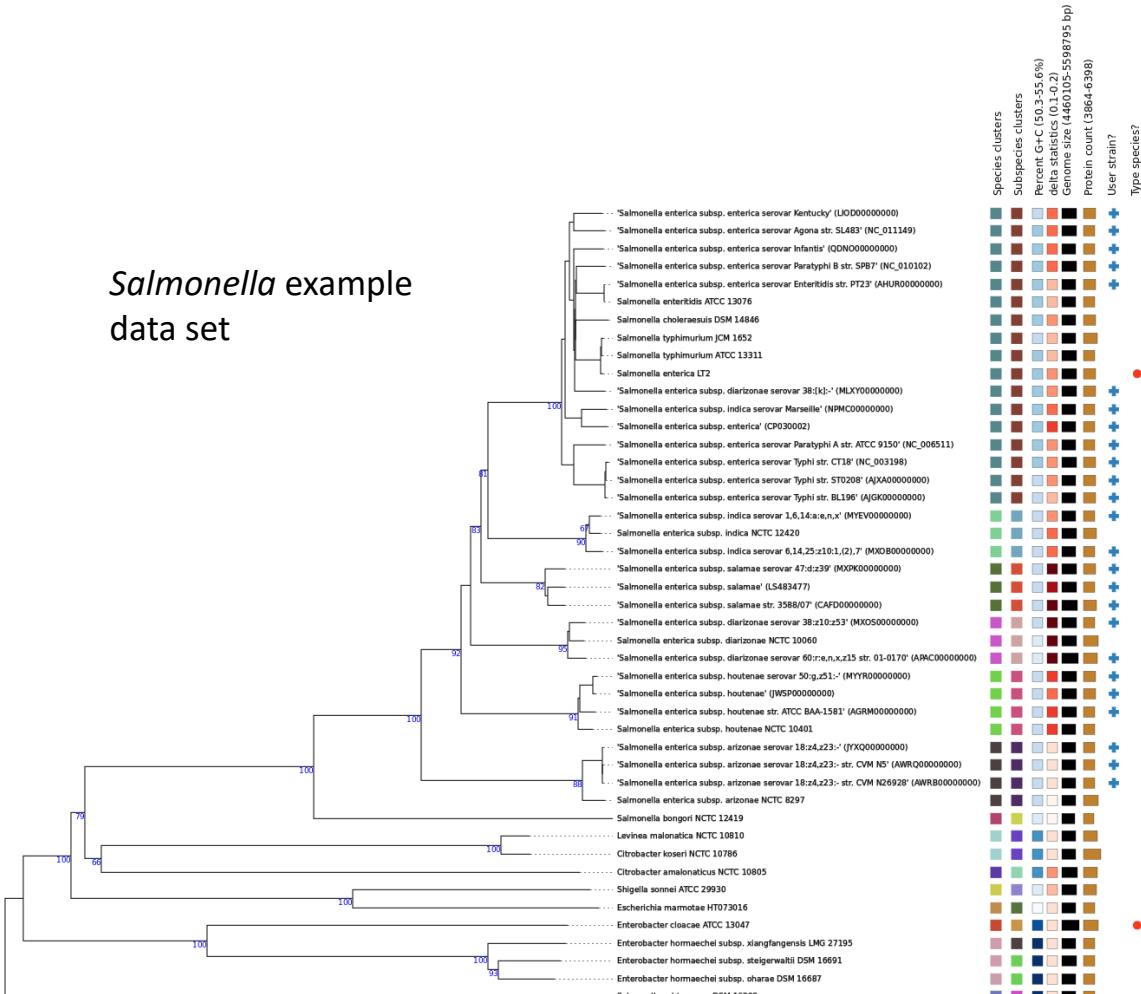


Meier-Kolthoff J.P., Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **10**:2182, 2019.

<https://tygs.dsmz.de/>

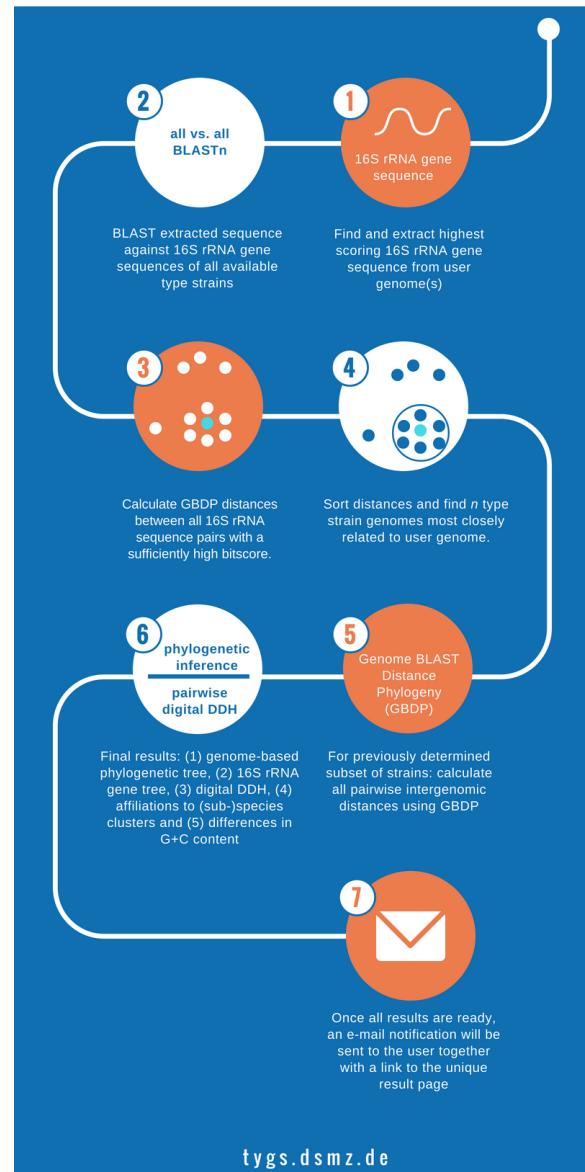
# Type (Strain) Genome Server

*Salmonella* example  
data set



<https://tygs.dsmz.de/>

# Basic TYGS STEPS



**By the way.... Anyone out there interested in working at the DSMZ?**

**In August, we will be hiring for an extramural project related to collections and the Nagoya Protocol (start date December 2019)!**

**Talk to Amber Scholz: [amber.h.scholz@dsmz.de](mailto:amber.h.scholz@dsmz.de)**

<https://gold.jgi.doe.gov/gebaVI>

Thank you

<https://tygs.dsmz.de/>



Leibniz-Institut • DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH

# GEBA genomes in IMG

## DNA sent to JGI

in_img	project	sent	project
860	KMG-1	915	KMG-2
838	KMG-2	863	KMG-4
319	KMG-4	860	KMG-1
242	GEBA	591	ACTINO-1000
239	ACTINO-1000	242	GEBA