

Bioinformatics Services for Microbial Genome and BiGi Metagenome Research at Giessen University

S. Jaenicke, J. Blom, K. Brinkrolf, M. Dieckmann, F. Förster, B. Linke, R. Müller, O. Rupp, A. Goesmann.
Bioinformatics & Systems Biology, Justus-Liebig-University Giessen, Germany

Fkz 031A533B

Short description of the project

The dramatic drop in consumer costs for DNA sequencing is revolutionizing genomics research in general, and specifically in the field of biotechnology and medicine. DNA sequencing of microbial genomes and metagenomes of their habitats is generating data at a hitherto unforeseen pace. The real bottleneck that prevents realization of the full potential of the different omics technologies is not the data generation itself, but the subsequent data analysis.



The BiGi team
in Giessen

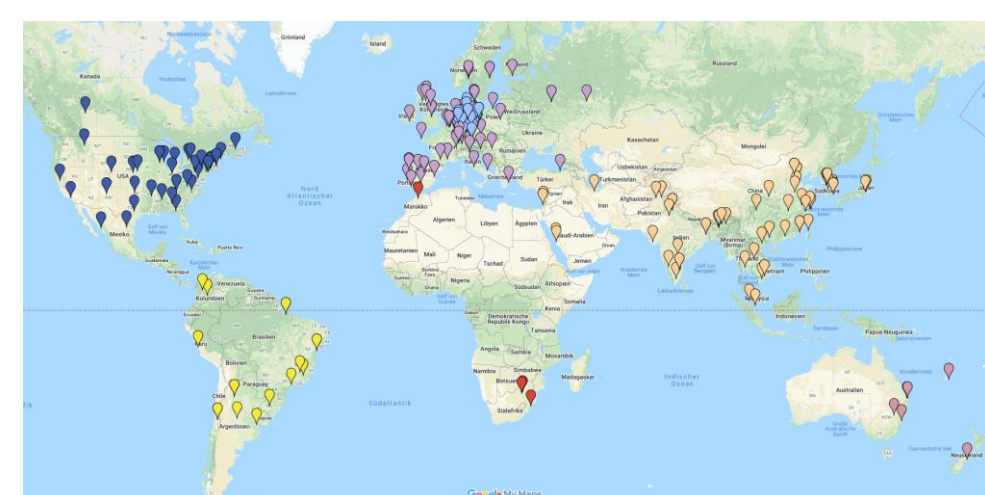
The BiGi competence center at Giessen University has a strong focus on high-performance computing services including an OpenStack infrastructure for cloud computing and a repository of reusable workflows suitable for high-throughput sequence data analysis. We provide bioinformatics software tools for manual genome annotation, high-throughput automatic genome annotation, large-scale comparative genomics and metagenomics, as well as RNA-Seq data analysis and visualization. As part of de.NBI, we contribute to standardization efforts for data interchangeability and software interoperability. Finally, we provide support and user training in the field of microbial bioinformatics to the life science community.

de.NBI services

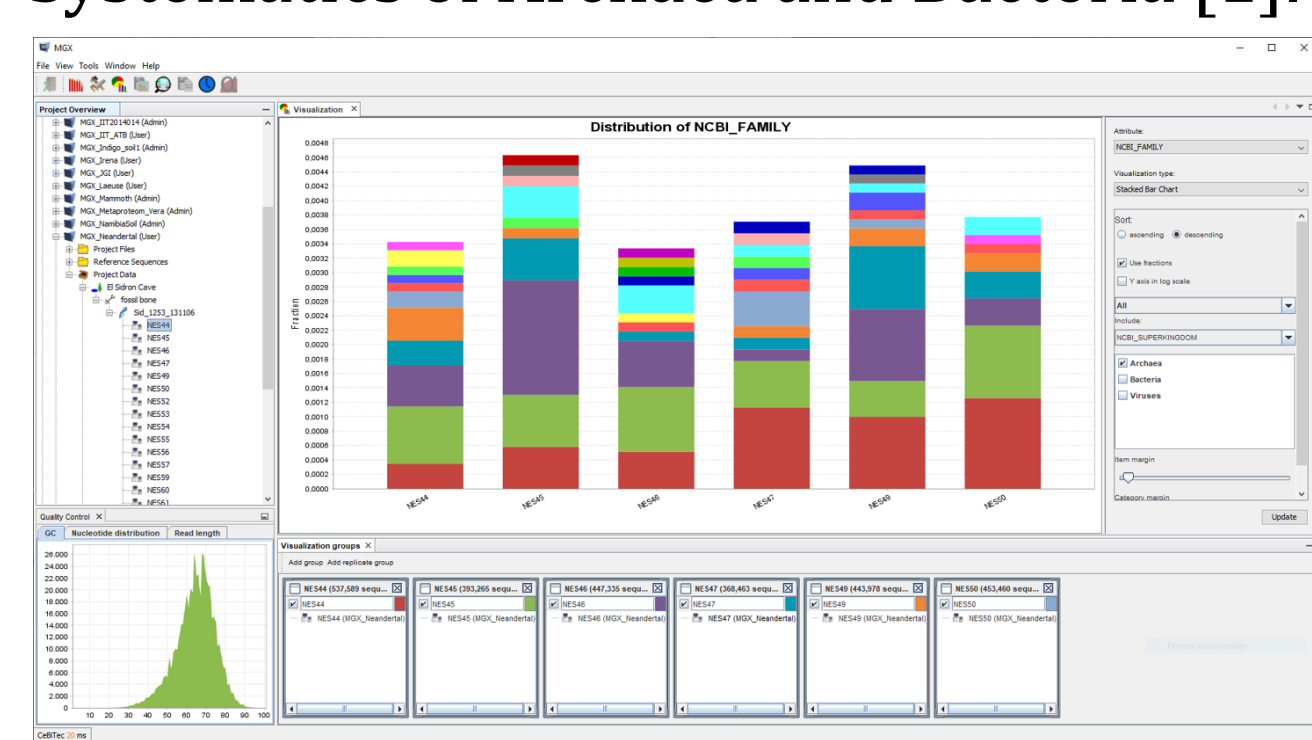
The BiGi competence center at Justus-Liebig-University Giessen (JLU) provides a broad range of bioinformatics services for the analysis of microbial sequence data. This includes downloadable as well as web server-based solutions, for example the renowned **GenDB genome annotation system**. Furthermore, infrastructure as a service (IaaS) is offered, e.g. by hosting of applications on de.NBI hardware.

EDGAR: Comparative genomics

The EDGAR platform for comparative genome analysis is the most heavily used de.NBI service at JLU. Since 2017, **675 new EDGAR projects** have been created, and almost **75,000 genomes** have been processed in 1,660 project updates. Currently,



1,382 EDGAR projects are hosted in total for scientific users worldwide. Recently, EDGAR has been presented as a tool for phylogenomics in Bergey's Manual of Systematics of Archaea and Bacteria [1].



The MGX framework for flexible metagenome analysis

The MGX platform for metagenome analysis allows convenient and high-throughput processing of environmental community datasets. As one of the BiGi flagship services, **over 80 projects** are currently hosted on servers in Bielefeld

and Giessen with individual project sizes of up to **123 Gbp**. In 2018, the MGX framework was successfully published in BMC Microbiome [2].

ASA³P: Scalable assembly, annotation and analysis of bacteria

A recent addition to the BiGi portfolio is ASA³P [3], featuring the complete workflow from raw WGS data of all major sequencing platforms to annotated genomes and in-depth genome characterizations in a fully automated manner. It can process thousands of genomes per day using cloud resources. So far, **more than 6,700 genomes** have successfully been analyzed with ASA³P.

Galaxy: Bioinformatics analysis pipelines

A Galaxy server hosted in Giessen provides tailored analysis pipelines for microbial genome and transcriptome data. The server has **over 300 registered users** and is constantly used in student education as well as BiGi training activities.



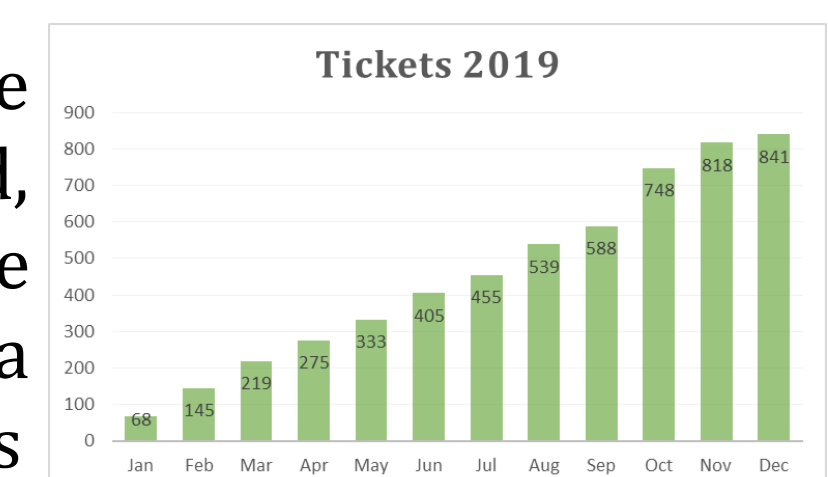
General information on the project

Currently, five scientists in Giessen are directly funded by de.NBI. These employees provide project work to the extent of 4.65 full time equivalents (FTE). They are involved in service, training, cloud operations as well as ELIXIR coordination. A further 7 persons contribute to de.NBI related tasks. To support these efforts, Giessen University created one new permanent position in September 2018.

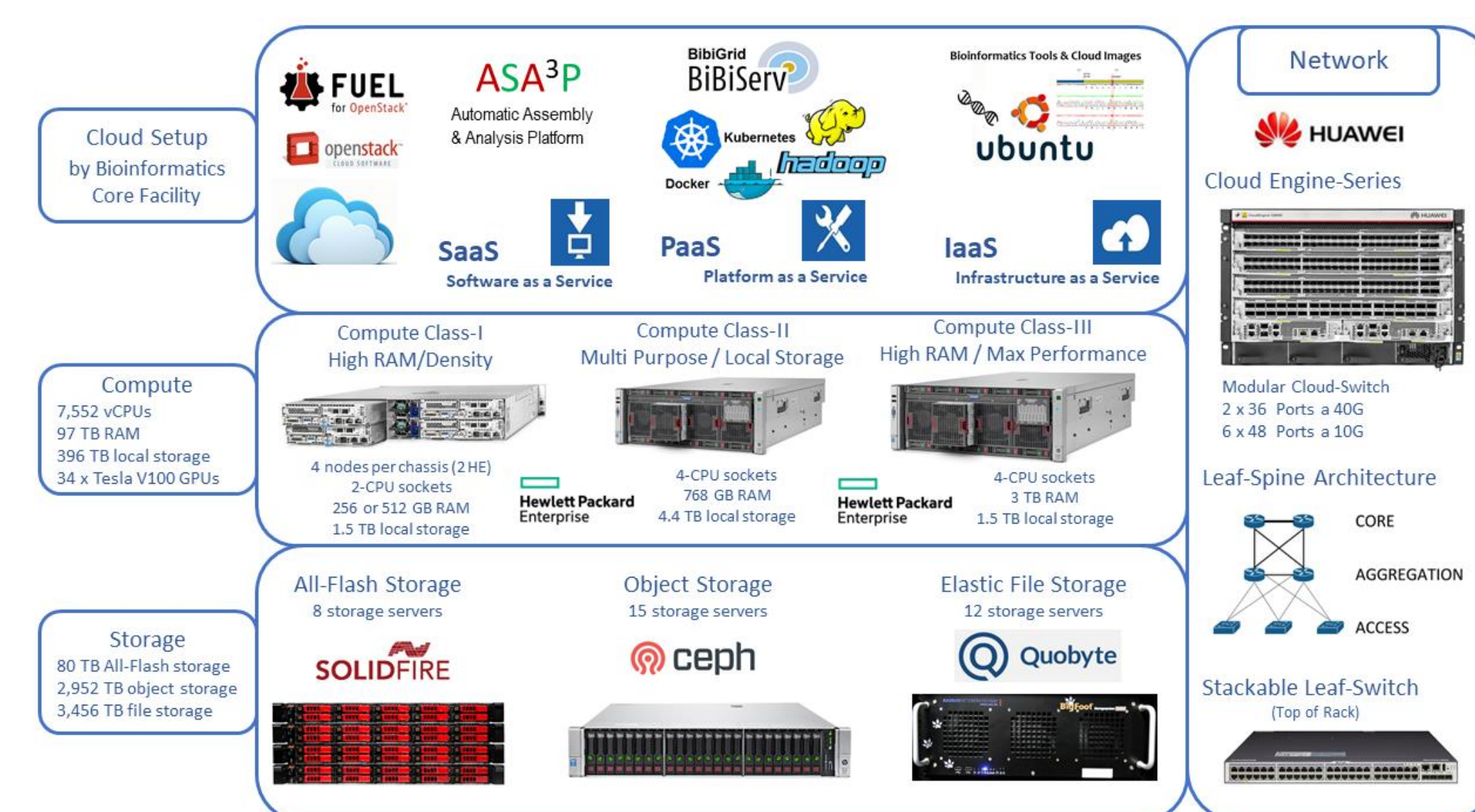
	Personnel funded by de.NBI	Contributors to de.NBI projects, not funded by de.NBI
Total number of employees:	4.65	7
- Scientific personnel (FTE):	4.65	3
- Technical staff (FTE):	--	1

Progress report

The success of the JLU branch of BiGi is reflected by the huge number of **tickets** accumulated in the reporting period, **exceeding 800**. This reflects the vast amount of service tasks, including general support, feature requests, and data processing issues, that have been completed within this project. Based on user feedback, the de.NBI services offered are constantly extended and enhanced. Due to the increasing number of large-scale differential gene expression (DGE) RNA-Seq experiments, a high-throughput pipeline generator and executor to create reproducible DGE analyses is under development. The system supports various mapping and analysis tools and can be executed in parallel on a compute cluster. Furthermore, it can visualize DGE results in various interactive charts.



de.NBI cloud activities



- Over 100 external users in over 45 projects

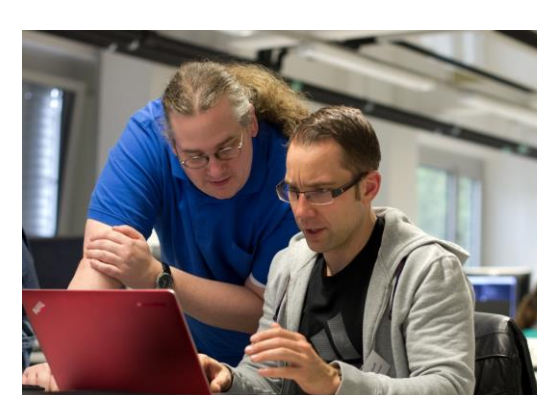
- Over 5.4 million vCPU hours (01/2019 – 12/2019)

ELIXIR collaboration



The BiGi node Giessen is participating in two ELIXIR work programs, EOSC-Life and Cloud and AAI focusing on implementing the standards set by the Global Alliance for Global Health (GA4GH) initiative. Also, participation in the Microbial Biotechnology Community was recently approved. The BiGi node Giessen provides cloud architecture expertise and training in both programs to make optimal use of state of the art infrastructures.

de.NBI Training and education



de.NBI Summer School on Cloud Computing for Bioinformatics

The de.NBI Summer School 2017 was conducted at Justus-Liebig-University Giessen from 25. June until 1. July with 24 participants.

Annual Genomics Training Course

The topic of this workshop is sequence analysis with a focus on short-read data, genome annotation and comparative genomics. Participants get hands-on experience in the usage of BiGi tools such as GenDB, ASA³P, and EDGAR. The latest workshop was held in May 2019; in the last five years, over 120 participants took part in this course.

Workflow management systems in the cloud

This workshop held in 2018 covered the usage of workflow management system for NGS data analysis in the de.NBI cloud. It introduced three workflow management systems, CWL, Nextflow, and Snakemake and was attended by 15 participants.

de.NBI Summer School on Metagenomics

In 2020, Justus-Liebig-University Giessen will host a Summer School targeting the analysis of microbial community datasets. The Summer School is scheduled for September 2020 and will be announced via de.NBI and ELIXIR channels.



Publications

- Blom et al. (2019) EDGAR: a versatile tool for phylogenomics. *Bergey's Manual of Systematics of Archaea and Bacteria*.
- Jaenicke et al. (2018) Flexible metagenome analysis using the MGX framework. *Microbiome* 6(1): 76
- Schwengers et al. (2020) ASA³P: An automatic and scalable pipeline for the assembly, annotation and higher-level analysis of closely related bacterial isolates. *ACCEPTED*.

More than 150 (co)-authored peer reviewed publications, more than 50 acknowledgements