Print Get PDF

Final Report 0 on the DZIF Project

Project number	TI 12.002
Title of the project	Bioinformatics Platform

Project partner and project duration

Project Coordination	**Start of Term**	**End of Term**
Helmholtz-Zentrum für Infektionsforschung GmbH	01.01.2021	28.02.2026

Project Partners	**Start of Term**	**End of Term**
Helmholtz-Zentrum für Infektionsforschung GmbH	01.01.2021	28.02.2026

Chapter 1: Short report (will be published)

a) Original task and scientific/technical status built on

Helmholtz-Zentrum für Infektionsforschung GmbH

Computational analyses of large biomedical and epidemiological data sets are of special significance for modern research on pathogens. In particular, the rapid development of omics technologies has led to an increased need for informatics in biological and clinical research. Bioinformatics is an essential component in the analysis of data within DZIF projects, and requires extensive expertise to either develop new methods, or identify the best available software tools for the task. Furthermore, educating DZIF researchers about computational data analysis methods is essential for translational infection research of the future.

Reflecting the needs of the DZIF partners, the TI Bioinformatics and Machine Learning offers the following services:

(1) Training: Organization and implementation of regular training opportunities for all DZIF scientists. In coordination with the DZIF Academy, the TI offers e.g. training schools on general computational data analyses (pipeline and R programming), bioinformatics software usage (e.g. microbiome and clinical pathogen data analyses, phylogenetics, resistances), and on machine learning.

(2) Consulting & Support: Best practice recommendations for data analysis for DZIF scientists and systematic benchmarking of state-of-the-art tools. This supports DZIF scientists in selecting a suitable method for a specific data analysis from the multitude of available computational methods.

(3) Infrastructure: Upon request, the TI offers a linux/ unix-based data processing environment and support for DZIF scientists in need of further compute resources. This enables to efficiently analyse modern sequence data, which can be seveal terabyte big, which often exceeds the ressources that are present locally.

b) Course of the project

Helmholtz-Zentrum für Infektionsforschung GmbH

1) Training

In coordination with the DZIF Academy, the unit offered two bioinformatics workshops per year. Topics were selected in consultation with DZIF scientists to address training needs at the time. Workshops lasted three to five days and were limited to 30 participants to ensure an intensive and interactive learning experience. Each workshop required several person-months for preparation and follow-up, as materials—including slides, datasets, and software—were revised based on prior evaluations or newly developed to meet DZIF-specific needs.

2) Consulting & Support

The unit supported ongoing and upcoming DZIF projects with bioinformatics expertise. It provided best-practice recommendations for data analysis, enabling scientists to choose suitable methods without requiring deep expertise or extensive time investment. Recommendations were based on the unit’s experience and a systematic evaluation of computational methods using an established benchmarking framework.

3) Infrastructure

The unit managed IT infrastructure, including software installation, virtual machine setup for workshop participants, and general system administration tasks. Following the InvestCall 2017, the Bioinformatics & Machine Learning unit was equipped with two high-memory servers, allowing it to deliver computationally intensive workshops, for example in metagenomics.

c) Main results and cooperation with other research institutions, if applicable

Helmholtz-Zentrum für Infektionsforschung GmbH

Over the past five years, the Bioinformatics and Machine Learning unit played a key role in supporting DZIF researchers through training, consulting, and tool development. It offered highly popular workshops on topics such as data analysis, statistics, and machine learning, helping researchers build practical skills and confidence. These courses were consistently in high demand and received very positive feedback.

In addition, the unit provided ongoing support for research projects by advising on data analysis, developing user-friendly tools and pipelines, and identifying best methods through systematic comparisons. It also created practical guidelines to improve data quality, ensure reliable results, and make research more transparent and reusable.

By combining training, expert support, and methodological development, the unit helped researchers better analyze complex infection and microbiome data. At the same time, it promoted high standards such as reproducibility and FAIR data management, and established itself as a central resource for bioinformatics expertise within DZIF.

Chapter 2: Detailed description (will be published)

a) Detailed presentation of the results achieved

Helmholtz-Zentrum für Infektionsforschung GmbH

Through high-demand workshops, targeted consulting, and development of robust computational pipelines, the TI significantly advanced the skills, knowledge, and capabilities of researchers, fostering a culture of methodological rigor and reproducibility. By supporting reproducible research, promoting FAIR data management, and aligning with international standards, the TI has established itself as a central resource for bioinformatics training and expertise, enabling DZIF researchers to effectively analyze and interpret complex infection and microbiome data.

1. Training and Capacity Building (WP1)

Over the five-year period, the TI established a highly successful and continuously evolving training program in collaboration with the DZIF Academy. The training portfolio focused on essential bioinformatics skills for data-driven infection research, particularly in R programming, statistics, and machine learning, and included courses such as *Introduction to R*, *Deep Learning for Bioinformatics with R*, and, from 2023 onwards, *Basic Statistics with R*. Workshops typically lasted three to five days and were offered as online formats.

Demand for the courses was consistently high throughout all years, often significantly exceeding available places, with up to 100 applicants for 30 slots. At the same time, participant satisfaction remained at a very high level, ranging from 80% to over 95%. Participants frequently reported improved practical skills and confidence, and emphasized that they felt encouraged to actively engage and ask questions. The introduction of the statistics course in 2023 further broadened the training portfolio, with over 90% of participants confirming measurable learning gains.

In addition, the TI responded flexibly to emerging needs, for example by organizing a webinar in 2022 on bioinformatics analysis during the COVID-19 pandemic, which included the presentation of DZIF-funded tools such as CoVerage.

2. Consulting and Support (WP2)

In parallel to the training activities, the TI provided continuous and comprehensive bioinformatics consulting and support services to DZIF researchers. Consultation service was offered via the DZIF website (https://www.dzif.de/en/infrastructure/bioresources-biodata-and-digital-health), providing technical and strategic guidance on bioinformatics workflows, machine learning applications, and FAIR data management. These services supported a wide range of infection research projects and ensured that researchers could effectively design, analyze, and interpret complex datasets.

A key component of WP2 was the development and dissemination of best practices. The TI systematically evaluated and benchmarked computational methods, including metagenomic analysis software (Meyer *et al.*, 2022) and antimicrobial resistance phenotyping tools (Hu *et al.*, 2024). In cooperation with NFDI4Microbiota, the CAMI benchmarking portal was published, helping researchers to choose the best computation methods for metagenomics and promoting standards and best practices (https://cami-challenge.org/). Based on these analyses, recommendations were developed to support robust and reproducible research, including guidance on metadata submission (Cernava *et al.*, 2022), and the use of appropriate control samples, and contamination management in low-biomass studies (Kennedy *et al.*, 2023). In addition, the TI contributed to international discourse through a review articles on best practices for microbiome data sharing (Huttenhower *et al.*, 2023).

Methodological development was another central outcome of WP2. The TI developed and supported bioinformatics pipelines, such as genome assembly workflows for viral strains (Fritz *et al.*, 2021), and contributed to advanced tools including the GenomeNet-Architect for deep learning model selection (Gündüz *et al.*, 2024) and CoVerage (https://sarscoverage.org), an analytics platform that monitors the genetic and antigenic evolution of human SARS-CoV-2 viruses (Norwood *et al.,* 2025).

Through these combined activities, WP2 made a substantial contribution to strengthening the methodological rigor and analytical capabilities of DZIF researchers, while fostering the adoption of FAIR principles and internationally recognized standards. Furthermore, the TI maintained close collaboration with DZIF TTUs, which resulted in joint publications with DZIF partners (see Chapter 2g).

b) Milestones and Deliverables

Milestones

Helmholtz-Zentrum für Infektionsforschung GmbH

No.	Title	Work package	Status	Date according to the project description	Corrected date	Comment/reason
1	Bedarfsgerechte Workshops und Schulungen	1	Completed	31.03.2025		Annually by March 31
2	Best-Practice-Empfehlungen und Benchmarking	2	Completed	31.03.2025		Annually by March 31

Deliverables

Helmholtz-Zentrum für Infektionsforschung GmbH

No.	Title	Work package	Status	Date according to the project description	Corrected date	Comment/reason
1	2 Bioinformatik-Workshops pro Jahr	1	Completed	31.12.2025		Annually by December 31
2	Benchmarking von state-of-the-art Methoden	2	Completed	31.12.2025		Annually by December 31

c) Most important items of the financial report

Personnel cost

Helmholtz-Zentrum für Infektionsforschung GmbH

employment of software developer, E13/level 2-3: 100%, 46.5 months (02/2021-12/2024)
employment of postdoc 1, E14/level 3: 20%, one year (01-12/2024)
employment of postdoc 2, E14/level 4: 60%, one month (07/2024)
employment of postdoc 3, E14/level 3-4: 50%, 14 months (05/2021-06/2022), and 35%, 2 months (05-07/2025)

employment of postdoc 4, E13/level 4: 100%, 14 months (01/2025-02/2026)
employment of postdoc 5, E14/level 4: 67%, 4 months (09-12/2025), and 50%, 2 months (01-02/2026)

Consumables (also contracts)

Helmholtz-Zentrum für Infektionsforschung GmbH

Not applicable

Investment funds

Helmholtz-Zentrum für Infektionsforschung GmbH

Not applicable.

Travel expenses

Helmholtz-Zentrum für Infektionsforschung GmbH

meetings with project partners and scientific conferences

d) Necessity and adequacy of the work done

Helmholtz-Zentrum für Infektionsforschung GmbH

The use of resources was necessary and appropriate to achieve the objectives and the results achieved.

e) Prospective benefit of the project, in particular the usability of the results according to the updated exploitation plan

Helmholtz-Zentrum für Infektionsforschung GmbH

The activities of the TI have generated substantial and sustainable benefits that extend well beyond the project duration. Through its integrated approach combining high-demand training, targeted consulting, and the development of robust computational tools, the TI has created a strong foundation for continued capacity building, methodological standardization, and innovation in infection research.

A key prospective benefit lies in the long-term usability and scalability of the established training program. The continuously refined workshop portfolio, developed in collaboration with the DZIF Academy, addresses core competencies in bioinformatics, statistics, and machine learning. Given the persistently high demand and demonstrated learning success, these training formats will be maintained and further expanded in future initiatives. Early-career researchers, clinicians, and data scientists within and beyond DZIF will continue to benefit directly from these offerings.

The consulting and support structures developed within the TI also provide lasting value. The established advisory services and best-practice guidelines enable researchers to design and implement robust, reproducible analyses of complex biological data. These services are expected to remain an essential component of DZIF’s research infrastructure and can be further expanded in future projects.

Another major prospective benefit is the sustained usability of the computational tools, pipelines, and benchmarking resources developed within the TI. Platforms such as CoVerage, as well as established workflows for genome assembly and metagenomic analysis, provide readily applicable solutions for ongoing and future research projects. These tools are designed to be reusable, extensible, and aligned with international standards, making them highly suitable for integration into follow-up research efforts, including large-scale collaborative projects and infrastructure initiatives. They will be particularly valuable for researchers working with high-throughput sequencing data, antimicrobial resistance analysis, and pathogen surveillance.

Furthermore, the TI's strong emphasis on FAIR data management and reproducibility ensures that its outputs remain accessible, interoperable, and reusable in the long term. The developed recommendations, benchmarking studies, and review contributions will continue to inform best practices and support harmonization efforts at both national and international levels. Stakeholders who benefit from these results include not only DZIF researchers but also the broader bioinformatics and microbiome research communities, public health institutions, and data infrastructure providers. Overall, the TI has established a sustainable ecosystem of training, tools, and expertise that will continue to support high-quality, data-driven infection research and foster innovation in the years to come.

f) Progress/Advances in the field of the project by third parties during implementation of the project

Helmholtz-Zentrum für Infektionsforschung GmbH

Not applicable.

g) Successful or planned publication of the results according to No. 5 of the "Nebenbestimmungen für Zuwendungen" (NABF/NKBF 2017)

Helmholtz-Zentrum für Infektionsforschung GmbH

Bankwitz D, Bahai A, Labuhn M, Doepke M, Ginkel C, Khera T, et al. Hepatitis C reference viruses highlight potent antibody responses and diverse viral functional interactions with neutralising antibodies. *Gut*. 2021;70(9):1734–1745. doi:10.1136/gutjnl-2020-321190.

Deng ZL, Dhingra A, Fritz A, Götting J, Münch PC, Steinbrück L, et al. Evaluating assembly and variant calling software for strain-resolved analysis of large DNA viruses. *Brief Bioinform*. 2021;22(3):bbaa123. doi:10.1093/bib/bbaa123.

Fritz A, Bremges A, Deng ZL, Lesker TR, Götting J, Ganzenmüller T, et al. Haploflow: strain-resolved de novo assembly of viral genomes. *Genome Biol*. 2021;22(1):212. doi:10.1186/s13059-021-02426-8.

Münch PC, Franzosa EA, Stecher B, McHardy AC, Huttenhower C. Identification of natural CRISPR systems and targets in the human microbiome. *Cell Host Microbe*. 2021;29(1):94–106.e4. doi:10.1016/j.chom.2020.10.010.

Cernava T, Rybakova D, Buscot F, Clavel T, McHardy AC, Meyer F, et al. Metadata harmonization—standards are the key for better usage of omics data for integrative microbiome analysis. *Environ Microbiome*. 2022;17(1):33. doi:10.1186/s40793-022-00425-1.

Hotop SK, Reimering S, Shekhar A, Asgari E, Beutling U, Dahlke C, et al. Peptide microarrays coupled to machine learning reveal individual epitopes from human antibody responses with neutralizing capabilities against SARS-CoV-2. *Emerg Microbes Infect*. 2022;11(1):1037–1048. doi:10.1080/22221751.2022.2057874.

Meyer F, Fritz A, Deng ZL, Koslicki D, Lesker TR, Gurevich A, et al. Critical assessment of metagenome interpretation: the second round of challenges. *Nat Methods*. 2022;19(4):429–440. doi:10.1038/s41592-022-01431-4.

Gündüz HA, Binder M, To XY, Mreches R, Bischl B, McHardy AC, et al. A self-supervised deep learning method for data-efficient training in genomics. *Commun Biol*. 2023;6:928. doi:10.1038/s42003-023-05310-2.

Münch PC, Eberl C, Woelfel S, Ring D, Fritz A, Herp S, et al. Pulsed antibiotic treatments of gnotobiotic mice manifest in complex bacterial community dynamics and resistance effects. *Cell Host*

Microbe. 2023;31:1007–1020.e4. doi:10.1016/j.chom.2023.05.013.

Stanelle-Bertram S, Beck S, Mounogou NK, Schaumburg B, Stoll F, Al Jawazneh A, et al. CYP19A1 mediates severe SARS-CoV-2 disease outcome in males. *Cell Rep Med*. 2023;4:101152. doi:10.1016/j.xcrm.2023.101152.

Kennedy KM, de Goffau MC, Perez-Muñoz ME, Arrieta MC, Bäckhed F, Bork P, et al. Questioning the fetal microbiome illustrates pitfalls of low-biomass microbial studies. *Nature*. 2023;613:639–649. doi:10.1038/s41586-022-05546-8.

Huttenhower C, Finn RD, McHardy AC. Challenges and opportunities in sharing microbiome data and analyses. *Nat Microbiol*. 2023;8:1960–1970. doi:10.1038/s41564-023-01484-x.

Hu K, Meyer F, Deng ZL, Asgari E, Kuo TH, Münch PC, et al. Assessing computational predictions of antimicrobial resistance phenotypes from microbial genomes. *Brief Bioinform*. 2024;25:bbae206. doi:10.1093/bib/bbae206.

Goliaei S, Foroughmand-Araabi MH, Roddy A, Weber A, Översti S, Kühnert D, et al. Importations of SARS-CoV-2 lineages decline after nonpharmaceutical interventions in phylogeographic analyses. *Nat Commun*. 2024;15:5267. doi:10.1038/s41467-024-48641-2.

Gündüz HA, Mreches R, Moosbauer J, Robertson G, To XY, Franzosa EA, et al. Optimized model architectures for deep learning on genomic data. *Commun Biol*. 2024;7:516. doi:10.1038/s42003-024-06161-1.

Norwood K, Deng ZL, Reimering S, Robertson G, Foroughmand-Araabi MH, Goliaei S, et al. In silico genomic surveillance by CoVerage predicts and characterizes SARS-CoV-2 variants of interest. 2025.

Deng ZL, Safaei N, Schütte SL, Ohlendorf V, Maasoumy B, McHardy AC. High prevalence and local dissemination of daptomycin-resistance mutations for *Enterococcus faecium* in cirrhotic patients. *Gastroenterology*. 2026;170(1):228–230. doi:10.1053/j.gastro.2025.08.046.