



Extending the database system of BACTOME

Dr. Sarah Pohl

Institute for Molecular Bacteriology, TWINCORE, Hannover

Damianos P. Melidis

L3S Research Center, Leibniz Universität Hannover

March 26, 2019

1 Motivation

After the first human genome was sequenced, many genome databases and online browsers were created to study each species' genomic and transcriptomic landscape. With the same goal, researchers have developed BACTOME [Hornischer et al., 2018], an interactive reference database for the opportunistic pathogen *Pseudomonas aeruginosa*.

Currently, the database is designed and implemented using a relational database system (MySQL). With the goal to extend the current design we are going to review existing works for the application of databases in genomics such as Dorok et al. [2017], Kozanitis et al. [2013], Schulz et al. [2016]. Then we aim to redesign and implement an efficient database system for BACTOME. After implementation we will evaluate its memory and time requirements. Finally, we will test the database system offline and online in order to ensure its applicability.

2 Goal/Deliverables

Upon successful completion, the master student should:

- Deliver a redesigned and efficient database system for BACTOME
- Implement at least 1-2 database architectures
- Design, implement and test code
- Conduct experiments for (time and memory) performance comparison
- Explain her/his work in a scientific report

A supervisor will always help with major questions, testing of the code and writing of the report.

3 Prerequisites

1. Motivation & interest in the topic
2. Self-contained work in a multi-disciplinary team
3. MySQL, PostgreSQL
4. Python, JavaScript, html, CSS
5. Experience in genomics and transcriptomics is a plus

4 Contact

Dr. Sarah Pohl: Sarah.Pohl@helmholtz-hzi.de.
Damianos Melidis: melidis@l3s.de.

References

- Klaus Hornischer, Ariane Khaledi, Sarah Pohl, Monika Schniederjans, Lorena Pezoldt, Fiordiligie Casilag, Uthayakumar Muthukumarasamy, Sebastian Bruchmann, Janne Thöming, Adrian Kordes, et al. Bactome—a reference database to explore the sequence-and gene expression-variation landscape of pseudomonas aeruginosa clinical isolates. *Nucleic acids research*, 47(D1):D716–D720, 2018.
- Sebastian Dorok, Sebastian Breß, Jens Teubner, Horstfried Lapple, Gunter Saake, and Volker Markl. Efficient storage and analysis of genome data in databases. *Datenbanksysteme für Business, Technologie und Web (BTW 2017)*, 2017.
- Christos Kozanitis, Andrew Heiberg, George Varghese, and Vineet Bafna. Using genome query language to uncover genetic variation. *Bioinformatics*, 30(1):1–8, 2013.
- Wade L Schulz, Brent G Nelson, Donn K Felker, Thomas JS Durant, and Richard Torres. Evaluation of relational and nosql database architectures to manage genomic annotations. *Journal of biomedical informatics*, 64:288–295, 2016.