

Center for Integrative Bioinformatics (CIBI Berlin)

Software library and tools for biological sequence analysis

Fkz 31A535B

René Rahn, Marcel Ehrhardt, Svenja Mehringer, Knut Reinert
Algorithmic Bioinformatics, Freie Universität Berlin, Germany

Short description of the project

C++ library for biological sequence analysis

- Open Source “BSD 3-Clause License”
- Support for Linux, Windows and OS X
- Support for GCC, CLANG, Visual Studio, ICPC (Intel)

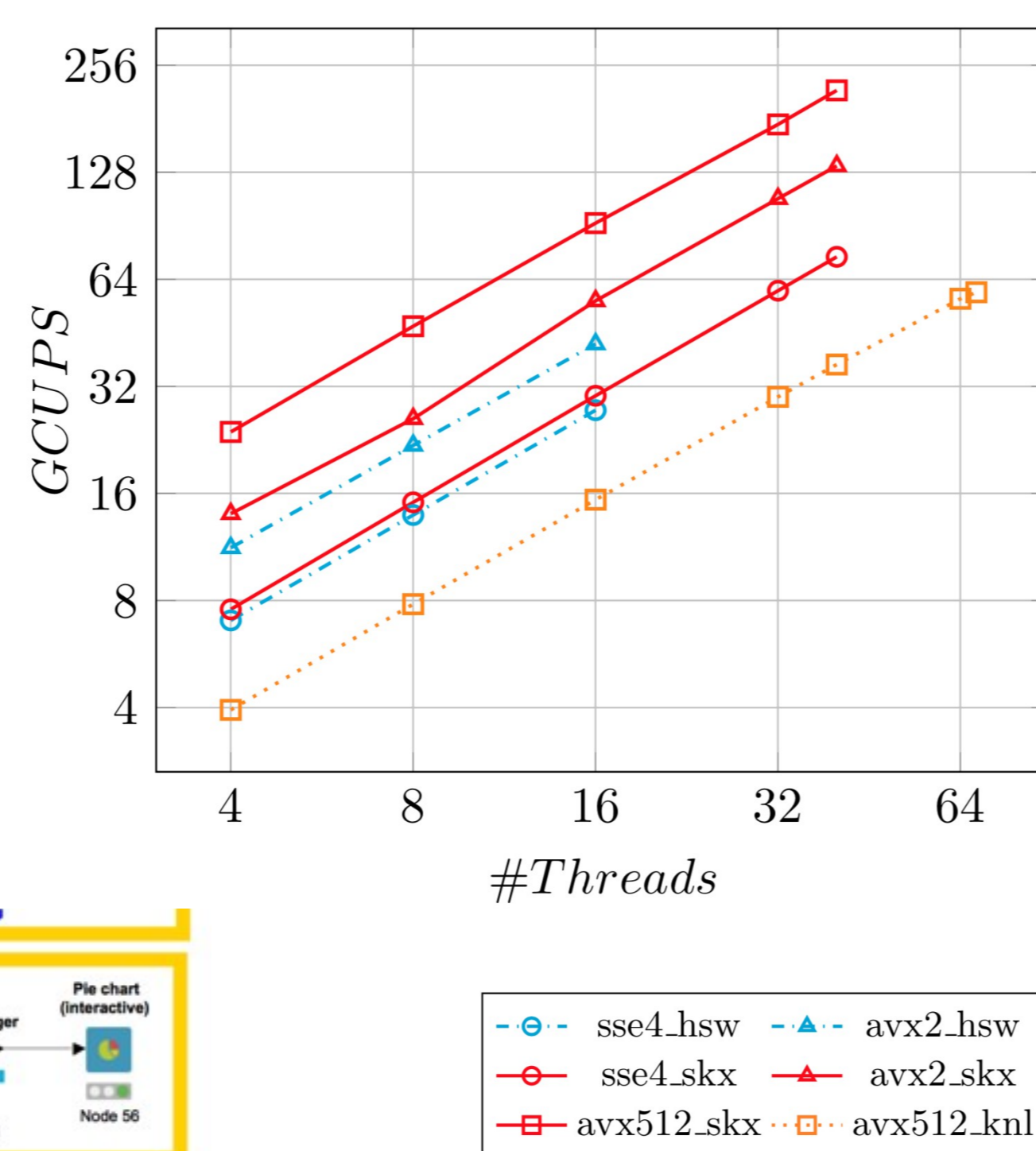
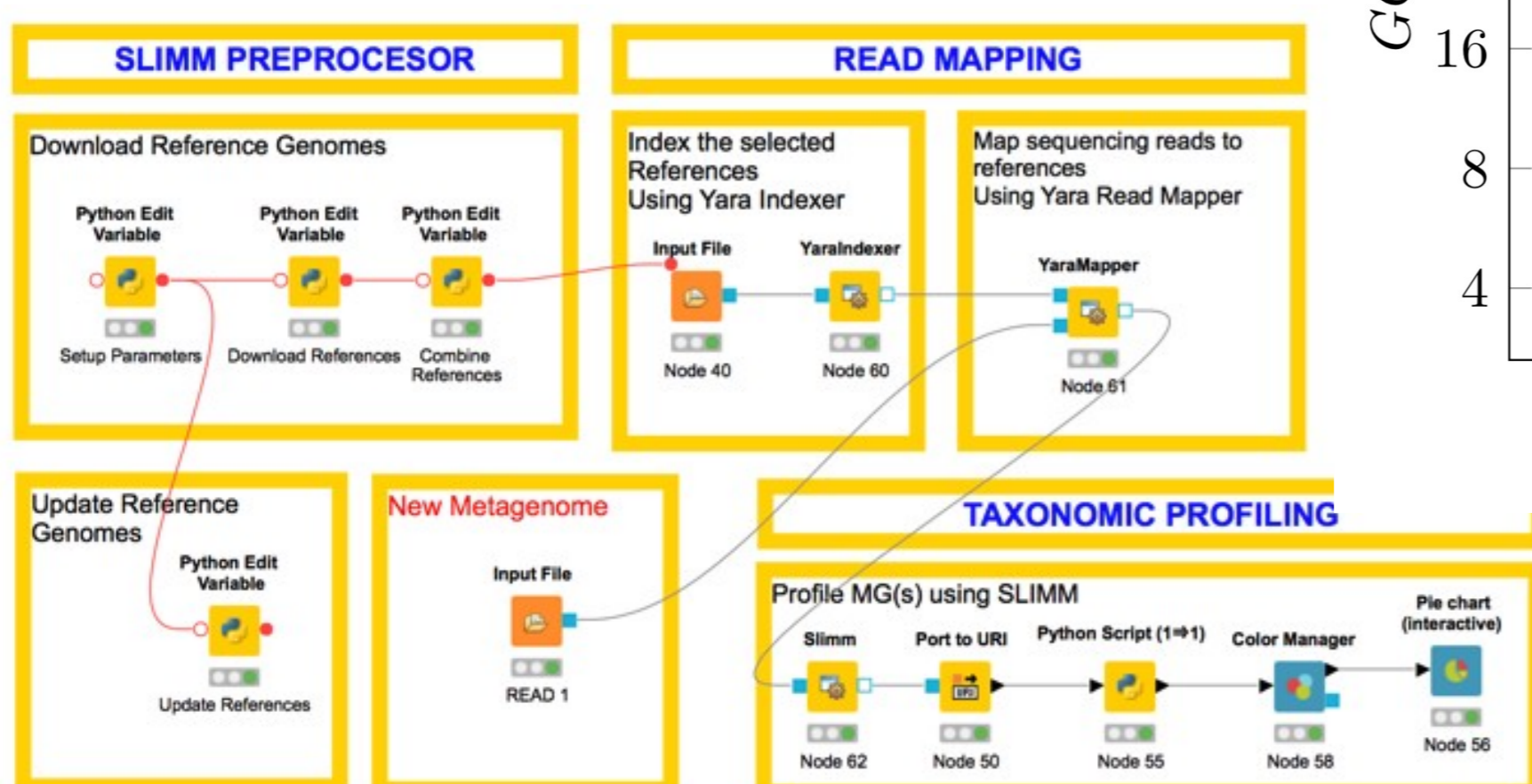
Key library functionalities

- Full-text indexing
- Reference based alignment
- Formatted Input / Output
- Parallelized data structures and algorithms



Tools for high-throughput sequencing

- Flexbar – Read quality control
- Mason – Read simulation
- Yara – Read mapping
- Lambda – Protein BLAST
- And many more...



de.NBI services

- Maintenance of build and test infrastructure for developers
- Integration of external code, data repositories or software contributions into SeqAn and SeqAn3 (325 issues solved, 676 pull requests integrated in 2019)
- Tool and algorithm development (358 citations of SeqAn publications in 2019)
- Consulting, dev. training and software support on daily basis (e.g. Bioinformatics Solution Center)
- Improvements to software documentation and teaching material
- Deployment of our tools (8,000+ unique users worldwide)
 - Source and binary releases (3663 downloads on github)
 - KNIME plugins to construct workflows for sequence analysis data processing and downstream processing
 - Available on Ubuntu, Debian, Brew, Bioconda (11.4k downloads) and Biocontainer
 - Ready for cloud computing (deNBI Cloud)

General information on the project

- 1.5 full position (FTE)
- 1.0 student assistant (FTE)
- All positions are currently filled

Progress report

Infrastructure development

- Released SeqAn 3.0.0 and 3.0.1
- Started Github Actions for CI and CD
- More Architecture Support (e.g. Arm64, PPC, x64)
- Setup an Application Infrastructure
- Packaged on Major Distributions (Debian, BioConda, Brew, etc..)

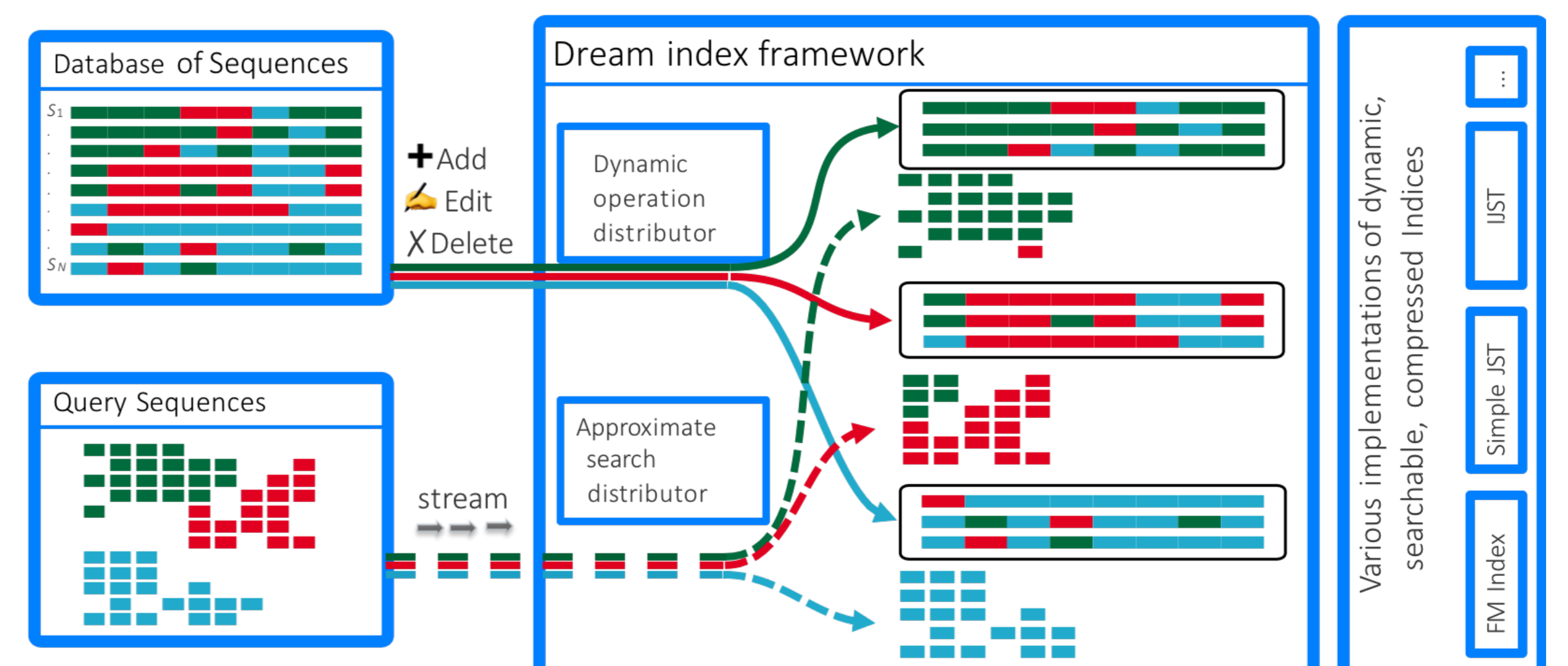
External Contribution

- Ranges Library: Bug Reports + Upstream Fixes
- SDSL Library: Became a Maintainer
- Reported Compiler Defects (e.g. Internal Compiler Errors, clang, gcc)
- Participated in C++ Standard Committee Meeting + Paper
- Collaboration with HD-HUB for Python Bindings

Training and education

- Added new training materials
- Developer, user meetings and conferences
- Online lectures and tutorials
- Participated in Train a Trainer Workshop

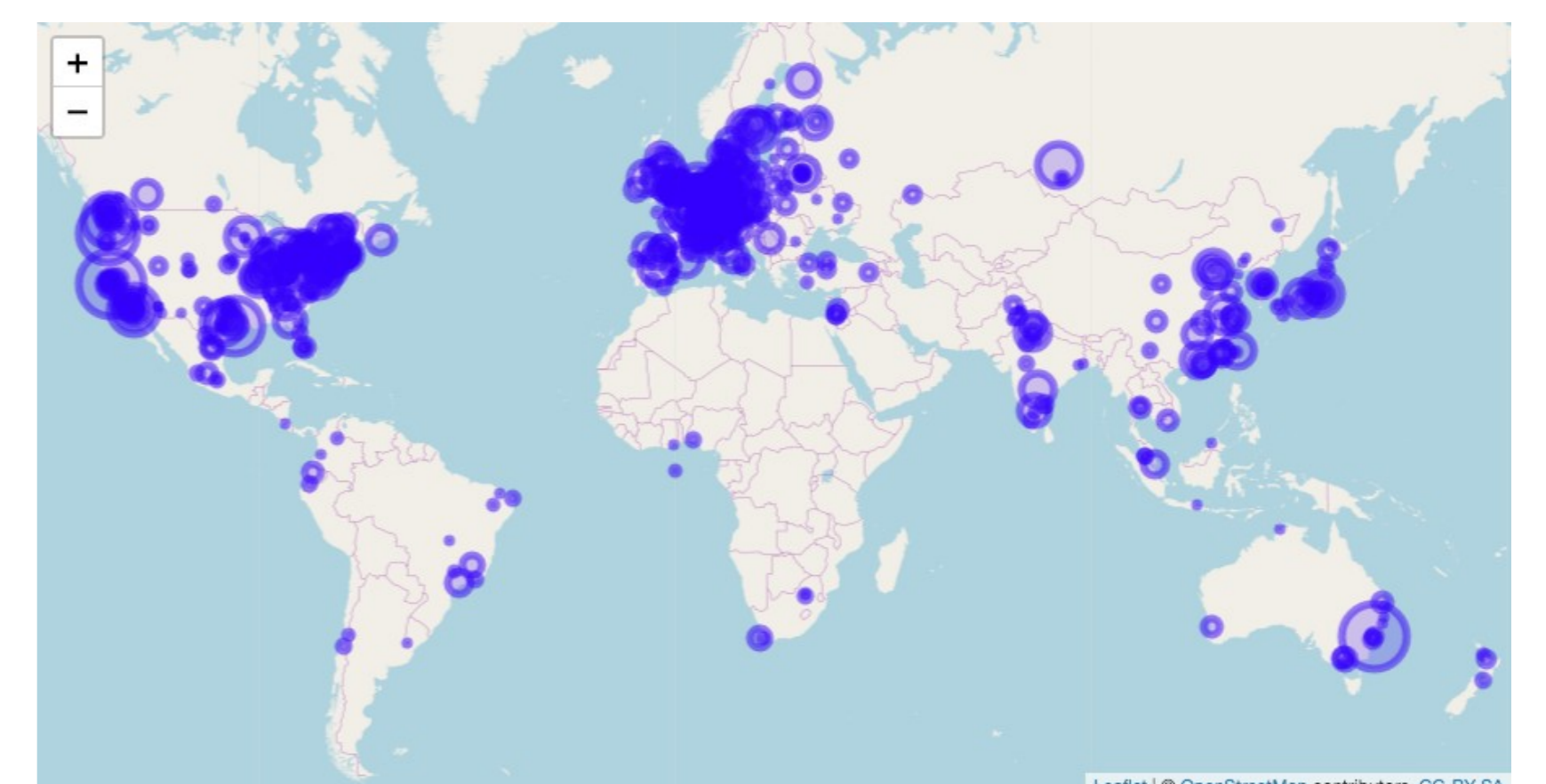
DREAM = Dynamic seaRchABLE pARallel coMpressed index



de.NBI Training and education

Training events at user/developer meetings and Conferences

- 18 Mar 2019 (Berlin): SeqAn & OpenMS (CIBI) Integration Workshop at the KNIME Spring Summit
- 22 Mar 2019 (Escal): SeqAn3 Developer Retreat 2019
- 24 Feb 2020 (Malta): BioStec (Accepted)
- 25 Mar 2020 (Italy): SeqAn3 Developer Meeting
- 26 Jul 2020 (Montreal): ISMB (Accepted)



Publications

- The SeqAn C++ template library for efficient sequence analysis: A resource for programmers. K Reinert, R Rahn, S Mehringer et al. - *Journal of Biotechnology*.
- DREAM-Yara: An exact read mapper for very large databases with short update time T. H. Dadi, E. Siragusa, V. Piro, A. Andrusch, E. Seiler, B. Renard, & K. Reinert - *Bioinformatics*
- Generic accelerated sequence alignment in SeqAn using vectorization and multi-threading R. Rahn, S. Budach, P. Costanza, M. Ehrhardt, J. Hancox, and K. Reinert, - *Bioinformatics*
- Vaquita: Fast and Accurate Identification of Structural Variation Using Combined Evidence J. Kim and K. Reinert - *LIPIcs – Leibniz International Proceedings in Informatics*