



CONTENT

GOAL OF WORKSHOP

Find out, whether it is worthwhile for you to spend a couple of days in Berlin at one of Germany's top universities.

ABOUT SEQAN

More Details about the content and intentions of SeqAn and the BMBF BioStore project. When did the projects start, what are the plans for the future?

VENUE AND SCHEDULE

The Freie Universität Berlin is one of Germany's elite universities. Located in Dahlem, the home of Max Planck and Einstein. [Here](#) you also find a overview of the retreat program, the tutorials, the applications, the code.

MORE INFO:

www.seqan.de

[BioStore homepage](#)

Knut Reinert
Algorithmic Bioinformatics
Freie Universität Berlin
Takustrasse 9, 14195 Berlin

Contact: Sabrina Krakau
(krakau@mi.fu-berlin.de) for signing up and information.

Freie Universität Berlin

BioStore - SeqAn Workshop 2011, 12.09.-14.09.

GOAL OF WORKSHOP

SeqAn is by now the most comprehensive C++ library for biological sequence analysis. The number of people working with SeqAn and its large code base

data analysis, data mining, or other sequence related topics. If you plan to use SeqAn as a bioinformatician or make it your platform for algorithmic development, you should sign up for this workshop.

Software

Highly accessed Open Access

SeqAn An efficient, generic C++ library for sequence analysis

Andreas Döring¹ ✉, David Weese¹ ✉, Tobias Rausch^{1,2} ✉ and Knut Reinert ✉

¹ Algorithmische Bioinformatik, Institut für Informatik, Takustr. 9, 14195 Berlin, Germany

² International Max Planck Research School for Computational Biology and Scientific Computing, Ihnestr. 63 - 73, 14195 Berlin, Germany

✉ author email ✉ corresponding author email

BMC Bioinformatics 2008, **9**:11 doi:10.1186/1471-2105-9-11

The electronic version of this article is the complete one and can be found online at:
<http://www.biomedcentral.com/1471-2105/9/11>

make it worthwhile to take a closer look at the latest release and to identify :

- further applications for which algorithmic components should be built.
- potential areas of improvement or enhancement for the current code base.

In addition to maintenance and discussing new features we will offer advanced tutorials for people who are not yet SeqAn experts but want to make use of its generic design and large, efficient code base. We expect 30-40 people working on NGS

The workshop is free of charge.

Content:

- Support for **NGS data analysis** such as read mapping, variant analysis, RNA-Seq support, assembly support
- **Advanced tutorials** for **bioinformaticians** and **algorithm engineers**.
- Presentation of **BioStore** concept and workflow engine presentation by the **KNIME** group.
- Presentation of planned **CONVEY** FPGA support
- Discussion about the **future content**.

ABOUT SEQAN

In 2003, the Algorithmic Bioinformatics group initiated the SeqAn project at the Freie Universität Berlin. Before setting up his group at the university, Knut Reinert had been working for years at Celera Genomics where he participated at large genome assembly projects.

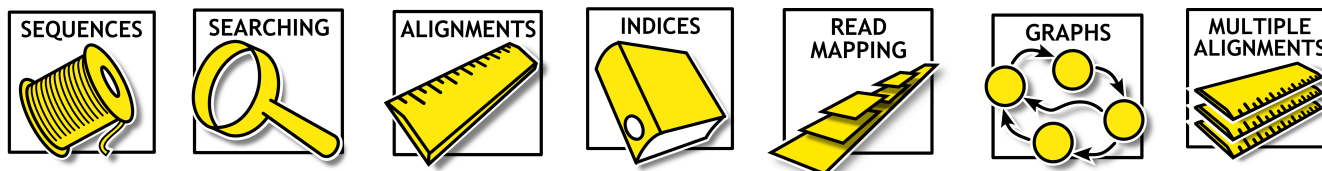
A central part of these projects was the development of large software packages containing assembly and analysis algorithms developed by the Informatics Research team at Celera. Although successful, the endeavor clearly showed the lack of available implementations in sequence analysis, even for standard tasks. Needed algorithmic components were either simply not available, or hard to access in third-party, monolithic software products.

The idea for SeqAn was that the library should be comprehensive for the field of sequence analysis, it should be easy to use, and

most of all (having the tremendous data volumes in genomics in mind) be efficiently implemented.

In 2003, Andreas Döring started the project. In the course of 18 months, the evaluation and discussion of goals, software designs and possible contents led to various prototypes and then to the final design. From 2006 on many PhD students joined the project. Their help in augmenting the functionality of SeqAn, in implementing algorithms, data structures, and providing documentation and tutorials was indispensable in making SeqAn a great product. From summer 2011 SeqAn is supported by the BMBF VIP program. The goal is to establish an SDK, a shop modul and workflow engine integration to make the application of tools developed in SeqAn easy and robust and help to analyse the growing amount of NGS data.

SEQAN CONTENT



SeqAn contains about 24 men years of work and is now a comprehensive software library for biological sequence analysis. It contains data structures for sequences, (multiple) alignments, q-gram indices and enhanced suffix arrays, directed and undirected graphs (as well as specialized automata). Its algorithmic components encompass algorithms for fast (approximate) string matching, exact and heuristic alignments, read mapping components, x-drop extension, gapped q-gram indices, motif searching algorithms and various graph algorithms just to name a few. The retreat will make you acquainted with the possibilities of SeqAn and future additions.

SEQAN DESIGN



Adding you own algorithmic components to SeqAn is no black magic. We will introduce you to the main design concepts used. **Template subclassing** for introducing polymorphism, **global interfaces** as means for easy integration, and **metafunctions** for typing.

THE HARNACK HOUSE

We will meet at the Max-Planck owned Harnack house, in Berlin-Dahlem (see photo). Formely a meeting point of Berlin's science elite (Einstein, Planck, Heisenberg, Haber, Hahn, etc.) it is now a modern conference venue. More detailed information will follow later this summer.



Preliminary Schedule

This year, there will be a „basic“ and an „advanced“ tutorial track. The basic track caters to those new to SeqAn, the advanced track is for more experienced users that are interested in advanced techniques and bringing specific problems for discussion with SeqAn developers and other users. In addition we will present you the progress in generic parallelization and the planned workflow solution of BioStore.

Please note that this schedule is preliminary and can be changed if the participants of the workshop want to.

12.09.	Introduction, Overview, Applications	
Morning	Welcome and Intro to SeqAn and BioStore (Knut Reinert)	
Morning	SeqAn: Basic Concepts, Novelties, Advanced Usages: Talks by SeqAn team	
Afternoon	SeqAn Parallelization: Talks by FU Berlin, Convey	
	participant talk	
	participant talk	
	participant talk	
	participant talk	
Evening	Get Together (Harnack Haus)	

13.09.	Hands On	
Morning	SeqAn Setup, Design Principles	SeqAn Advanced Tutorial, User Problem Reports
Afternoon	Basic Tutorials	
Afternoon	Knime tutorial (2 hrs hands-on)	
Evening	Get together (Harnack Haus)	

14.09	Discussion	
Morning I	Breakout Sessions, Wish Lists	
Morning II	Discussion of Breakout Sessions, Proposals, Knime for Developers (2 hrs)	
Afternoon	Conclusion of meeting	