



## CONTENT

### GOAL OF RETREAT

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. When did the project 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 detailed overview of the retreat program. The tutorials, the applications, the code.

### MORE INFO:

[www.seqan.de](http://www.seqan.de)

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

Contact: Anja Kasseckert  
([Anja.Kasseckert@fu-berlin.de](mailto:Anja.Kasseckert@fu-berlin.de))  
for signing up and information.

Freie Universität Berlin

## Algorithmic Bioinformatics SeqAn Retreat 2010, 04.10.-08.10.

### GOAL OF RETREAT

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 retreat.

Software Highly accessed Open Access

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

Andreas Döring<sup>1</sup> ✉, David Weese<sup>1</sup> ✉, Tobias Rausch<sup>1,2</sup> ✉ and Knut Reinert ✉

<sup>1</sup> Algorithmische Bioinformatik, Institut für Informatik, Takustr. 9, 14195 Berlin, Germany  
<sup>2</sup> 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 20-30 people working in NGS

The retreat fee is 50 EUR for including all drinks and food.

### Keywords:

- Support for **NGS data analysis** such as read mapping, variant analysis, RNA-Seq support, assembly support
- **Advanced tutorials** for **bioinformaticians** and **algorithm engineers**.
- Discussion about the **future content**. Algorithmic components, multicore support, data structures.

## 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

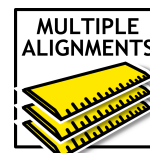
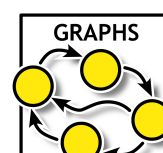
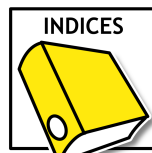
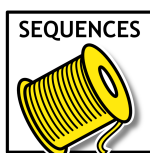
most of all (having the tremendous data volumes in genomics in mind) be efficiently implemented.

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. Although this was work-intensive, it led to a lot of insights and finally to the current SeqAn design which in our eyes fulfills our initial goals.

In 2006, the PhD students David Weese and Tobias Rausch joined the SeqAn team and by the end of 2007, the first version of SeqAn was released. In 2008, Anne-Katrin Emde and in 2009, Birte Kehr and Manuel Holtgrewe joined the team as PhD students.

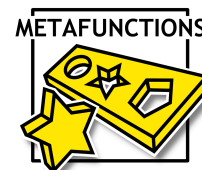
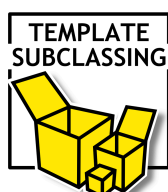
Their help in augmenting the functionality of SeqAn, in implementing algorithms, data structures, and providing documentation and tutorials was indispensable in making SeqAn a great product.

## SEQAN CONTENT



SeqAn contains now about 20 man 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 your 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 DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

## THE RETREAT

We will meet at the Freie Universität Berlin, one of Germany's elite universities. The Math/CS department is located in Dahlem, the green, southwestern part of Berlin (see photo on the right).



### Preliminary Schedule

This year, there will be a „basic“ and an „advanced“ 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.

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

04.10.	Introduction, Overview, Applications	
Morning	Welcome and Intro (Knut Reinert)	
Morning	SeqAn: Basic Concepts	SeqAn: Novelties, Advanced Usages
Afternoon	Participants Talks, Usage in academia and the industry	
Evening	Get Together	

05.10.	„Hands On“	
Morning	SeqAn Setup, Design Principles	SeqAn Advanced Tutorial, User Problem Reports
Afternoon	Basic Tutorials and next Steps	
Evening	Social Event	

06.10.	Contents / Code Review	
Morning I	Breakout Sessions, Wish Lists	
Morning II	Discussion of Breakout Sessions, Proposals, TODOs for Developers	
Afternoon	Code Review, Problem Resolution	

07.10.	Code Review	
08.10.	Code Review	
	This time is allocated as time of focused development and problem resolution in the SeqAn library. However, all participants are welcome to stay.	

**We invite you to come to the retreat!**

**So far a number of people have confirmed from:**

Illumina Ltd (USA/UK)

MDC, Berlin

FU Berlin, Germany

Uni Dortmund, Germany

MPI, Molecular Genetics, Berlin

Uni Tübingen, Germany

CRG/Uni Barcelona, Spain

University of Queensland

..and you ?