



Projektmanagement im Softwarebereich

AG Algorithmische Bioinformatik

Prof. Knut Reinert, Hannes Hauswedell, Rene Rahn

SeqAn Team

Projektmanagement im Softwarebereich

Qualifikationsziele:

Erwerb von allgemeinen Kenntnissen über die Anwendung von Software im beruflichen Alltag mit größeren Nutzergruppen, insbesondere praktische Erfahrungen mit typischen Problemen mit Software aus dem weiteren Umfeld der Bioinformatik und mit Lösungsansätzen zu deren Überwindung

Inhalte:

Verwendung von für den zu erwartenden Berufsalltag typischer Software für ein typisches Projekt.

Auswahl passender Software aus einer vorgegebenen Kollektion bzw. Anpassung oder Entwicklung fehlender Softwaremodule.

Erarbeitung von Lösungsstrategien im Team

Versuch einer Lösungsumsetzung.

Software libraries bridge gap

RNA-Seq

Structural variants

Metagenomics abundance

ChIP-Seq

Sequence assembly

Cancer genomics



FM-index

Multicore

Secondary memory

K-mer filter

Suffix arrays

Fast I/O

Hardware acceleration

Anforderungen

Biologie/Chemie

*

- Grundlegendes Wissen zu biologischen Sequenzen, DNA

Projektmanagement

**

- Ausarbeitung und Präsentation eines detaillierten Projektplans
- Umgang mit Werkzeugen zur kollaborativen Softwareerstellung
- Team-Arbeit, Arbeitsgruppe kennenlernen, Besuch des *SeqAn User Group Meetings*

Programmierarbeit

- Programmieren in C++, Templates, Einarbeiten in bestehende C++ Bibliotheken
- Testen und Dokumentieren der entwickelten Software

Es ist möglich unsere Projekte als BAen fortzusetzen!

Zeitplan

Datum	Termin/Dauer	Name	Ort
Di, 29.03.2016	10-18Uhr	Vorbesprechung, C++-Schnellkurs, Git, Compiler einrichten	??
Mi, 30.03.2016	09-18Uhr	SeqAn UGM Day1, Background + Introduction	??
Do, 31.03.2016	09-16Uhr	SeqAn UGM Day2, Tutorial Sessions	??
Do, 31.03.2016	16-19Uhr	SeqAn UGM Social Event (kostenlos, freiwillig)	??
Fr, 01.04.2016	09-16Uhr	SeqAn UGM Day3, Tutorial Sessions	??
Mo, 04.04.2016	10-18Uhr	Nachbesprechung SeqAn UGM, Weitere Tutorials, Aufteilung der Projekte	??
		selbstständige Arbeit	
Mo, 18.04.2016	10-13Uhr	Vorstellung der Projektpläne	??
		selbstständige Arbeit	
Mo, 23.05.2016	23:59Uhr	Abgabe des Abschlussberichts	(per Email)
Mo, 30.06.2016	10-18Uhr	Vorstellung der Ergebnisse; 12-14 Uhr Mittagspause	??

https://www.mi.fu-berlin.de/w/ABI/PMSB_Seqan_2016

SeqAn Demo in Javascript / im Browser

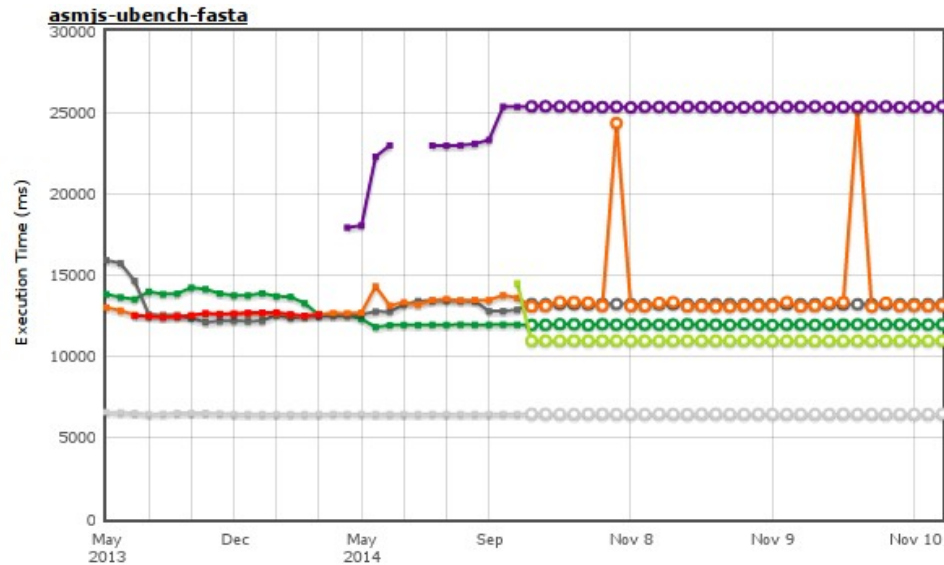


emscripten

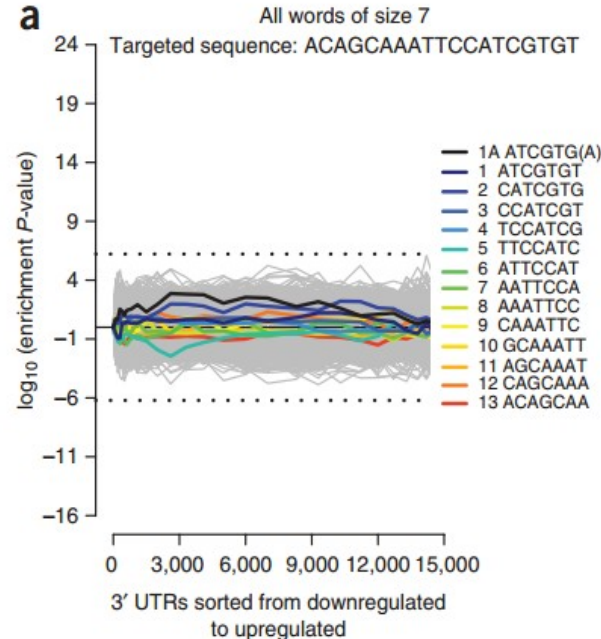
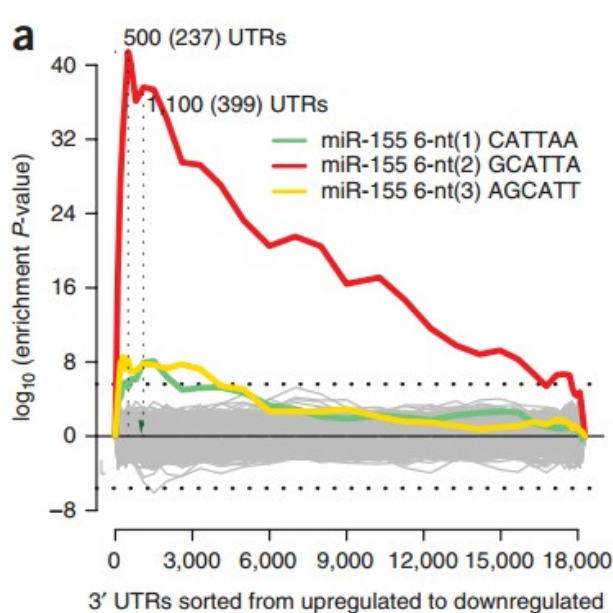
Emscripten is an LLVM-based project that compiles C and C++ into highly-optimizable JavaScript in asm.js format. This lets you run C and C++ on the web at near-native speed, without plugins.

- Chrome (v8)
- Safari (jsc)
- Firefox (Ion)
- Native C++ (Clang)
- Firefox (no asmjs)
- Firefox (Ion, GGC old)
- Chrome (Full turbofan, no crankshaft)

Machines >
Breakdown >
About >



Assessment of off-target effects of non-coding RNAs



Questions?

j.kim@fu-berlin.de

- You will develop a software that can evaluate off-target effects of siRNA using SeqAn.
- The entire procedure will be based on **Stijn van Dongen et al., (Nature Method, 2008)** – figures.
- The students who seek for “more realistic” biological questions are particularly encouraged to apply. The only requirement is the basic understanding of C++. (I can help you for other parts)
- Unfortunately, the mentor can't speak German. If you need a help in German, you should consider other options.

SeqAn project: short exact read mapper

- build index on reference sequence
- search single-end / paired-end reads
- find best position for every read (Hamming / edit distance)
- output hits in SAM format



Add BAM support to SeqAn SequenceFile

A SAM file (test.sam)

```

@HD VN:1.4 SO:unsorted
@PG ID:Yara PN:Yara VN:0.9.5
@SQ SN:gi|9632547|ref|NC_002077.1| LN:4718
seq1 4 * 0 0 * * 0 0 CGATCGATAAT IXXXXXXXXX RG:Z:none
seq2 4 * 0 0 * * 0 0 CCTCTCTCTCCCT IXXXXXXXXX RG:Z:none
seq3 4 * 0 0 * * 0 0 CCCCCCCC IXXXXXXX RG:Z:none
    
```

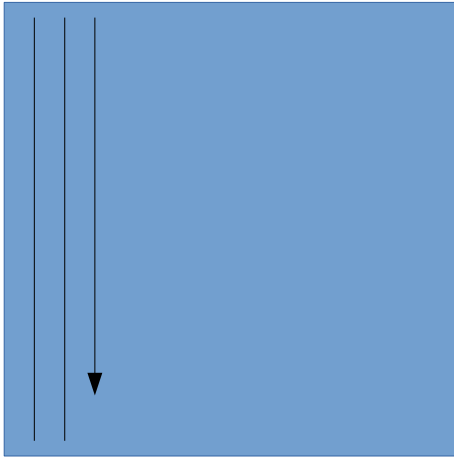
A FASTQ file (test.fq)

```

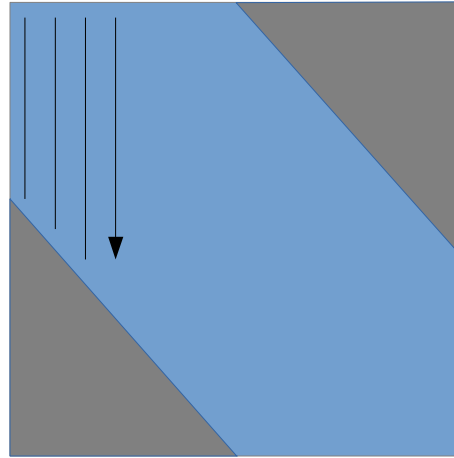
@seq1
CGATCGATAAT
+
IXXXXXXXXX
@seq2
CCTCTCTCTCCCT
+
IXXXXXXXXX
@seq3
CCCCCCCC
+
IXXXXXXX
    
```

- BAM Format – compressed SAM
- Take less space
- Can hold all the information in a FASTQ file
- In SeqAn it's used only as alignment

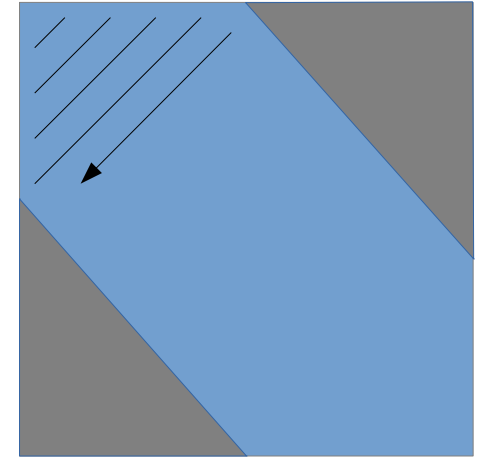
Needleman-Wunsch mit Vektorisierung



Regulär. DP



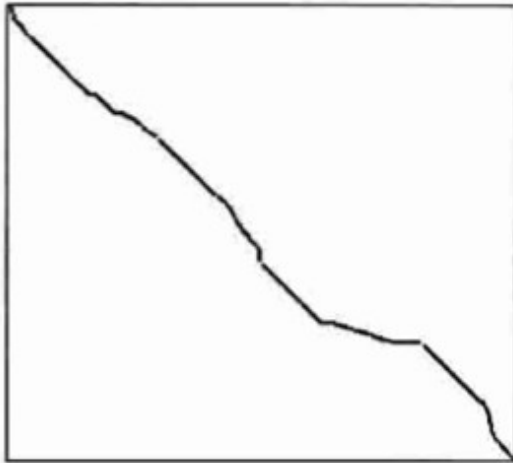
Mit Band



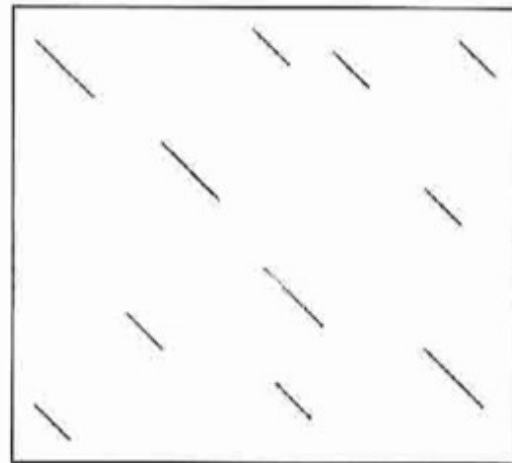
Anti-Diagonalen

Anti-diagonale Berechnung ermöglicht das Nutzen von SSE, AVX um mehrere Berechnungen gleichzeitig in einem CPU-Takt zu machen → sehr hohe Parallelisierung!

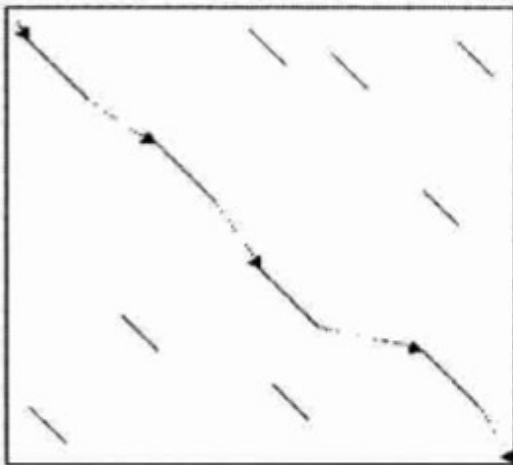
Lagan (genome size alignment)



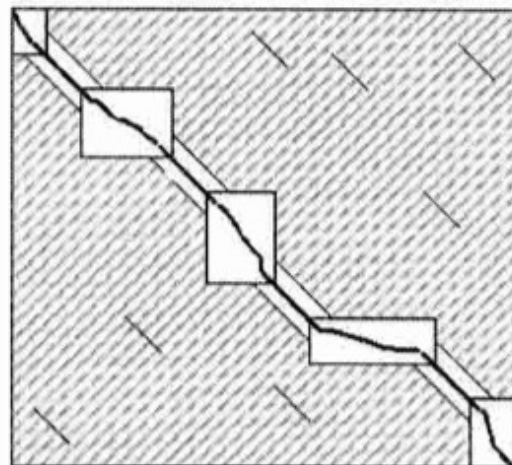
A



B



C



D