



Projektmanagement im Softwarebereich

AG Algorithmische Bioinformatik

Prof. Knut Reinert, Manuel Holtgrewe

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** mittels der zusammengestellten Software und Dokumentation der Ergebnisse
- **Vortrag** zur Darstellung der Projektergebnisse

Probleme? Welche Probleme?

In theory, there is no difference between theory and practice.

But, in practice, there is.

- Jan L. A. van de Snepscheut

SeqAn – C++ Library for Sequence Analysis

- SeqAn ist eine **C++ Templatebibliothek** fuer Sequenzanalyse
- **Warum C++?**

Flexible Sprache, Multi-Paradigma

Erlaubt effiziente Programme (HiSeq2500 erzeugt etwa 100Gb/Tag).

“Managed languages were the hammer that made everything look like a nail during the last decade. Today, the focus is on high demanding applications that need performance. The “king” is Performance per Watt, per Cycle, per Transistor: Performance per Dollar.”

- Herb Sutter

- **Warum eine Bibliothek?**

“Without supporting libraries, most serious applications are unnecessarily hard in C++. With suitable libraries, most are reasonably easy.”

- Bjarne Stroustrup

```
seqan::Dna5String seq1 = "CGGAT";  
seqan::Dna5String seq2 = "CAGGAT";  
int score = globalAlignment(seq1, seq2, seqan::EditDistance());
```

SeqAn Funktionalität

Sequences

Strings
Modified Sequences
Gapped Sequences
Journaled Sequences, ...

Indices

Q-gram Index
Enhanced Suffix Array
Lazy Indices
...

Alignments

Global Alignment
Local Alignment
Heuristics
Seeds and Chaining, ...

String Search

Exact/Approximate
Search Heuristic
String Filters
Motif Search, ...

Graphs

Directed/Undirected
Alignment Graphs
Probabilistic Automata
Trees

Algorithms

SWIFT, T-Coffee,
Consensus, Re-Alignment,
Chaining, Alignment-Free
Comparison, ...

Biologicals

Dna, Dna+N, Rna, Rna+N
Amino Acids
String Filters
Motif Search, ...

Input / Output

Stream Abstraction
FASTA, FASTQ
SAM/BAM, BAM Index
GFF/GTF, ...

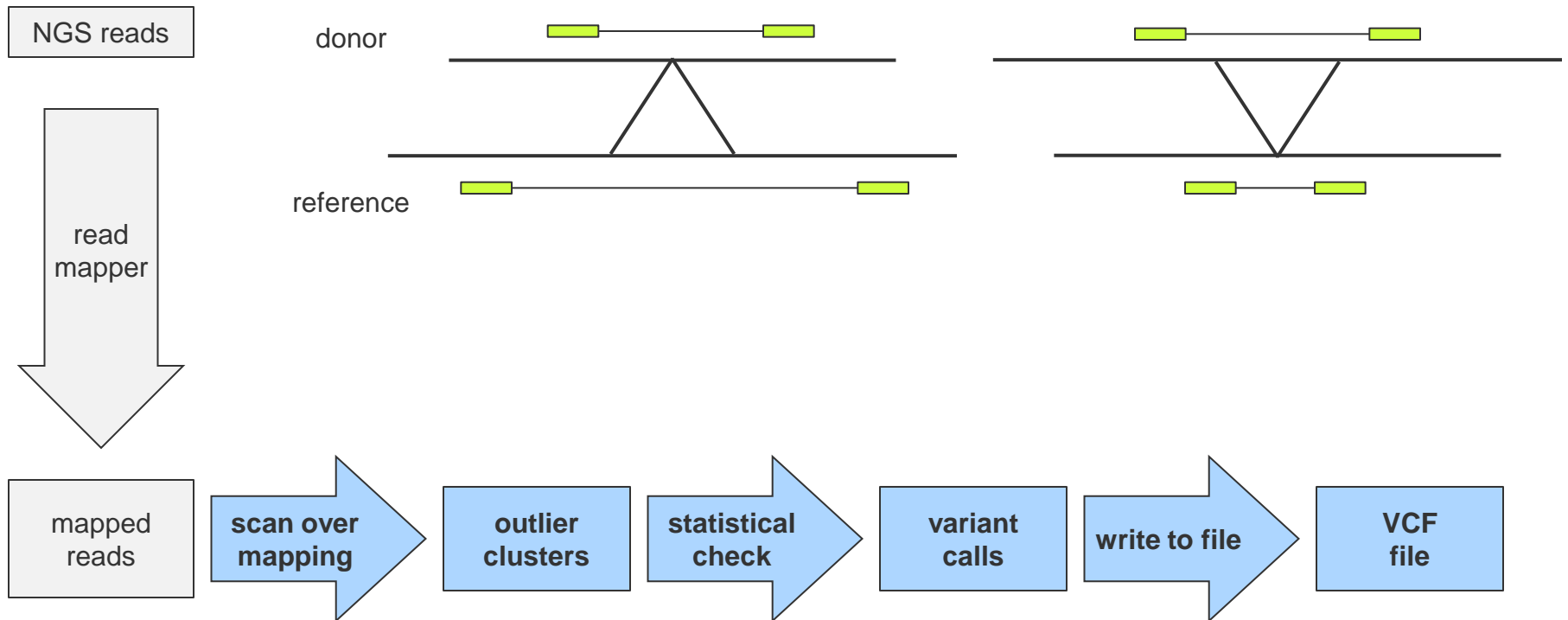
Miscellaneous

Argument Parsing,
File Parsing Helpers,
OS Abstractions,
Statistics, ...

Projekte

- **Mini FastQC**
Analysiere FASTQ Datei auf Qualitätsverteilung, Biases wie C+G Content usw.
- **Mini PEMer**
Suche in Paired-End Read Daten nach Strukturvarianten
- **Waterman-Eggert**
Effiziente Nachimplementierung des Waterman-Eggert Algorithmus fuer lokale Alignments
- **LAGAN**
Implementiere LAGAN Algorithmus zum effizienten Alignment langer Sequenzen

Mini PEMer



Anforderungen

Biologie/Chemie

*

- Grundlegendes Wissen zu biologischen Sequenzen, DNA

Projektmanagement

**

- Ausarbeitung und Präsentation eines detaillierten Projektplans
- Umgang mit Werkzeugen zur kollaborativen Softwareerstellung (SVN)
- Zusammenarbeit in kleineren Projektteams und Entwurf von Schnittstellen

Programmierarbeit

- Programmieren in C++, Templates, Einarbeiten in bestehende C++ Bibliotheken
- Testen und Dokumentieren der entwickelten Module
- **Empfohlene Teilnahme am Kurs „C++ für Fortgeschrittene“ (25.3.-2.4.)**

Zeitplan

Siehe auch: [ABI Wiki](#) -> [Lecture Wiki](#) -> [PMSB 2013](#)

Block 1: Vorbesprechung (2.4.)

- Vorstellung der Teilprojekte, Einführung in Teilgebiete, Literatur

Block 2: SeqAn Tutorial (8.4.-12.4.)

- Hilfe bei der Softwareinstallation (CMake, Compiler usw.)
- Vorträge zu SeqAn, Durcharbeiten von Tutorial Aufgaben
- Verteilen der Teilprojekte

Block 3: Recherche und Präsentation des Projektplans (12.4.-24.4.)

- Vorbereiten des Vortrags
- Welche Datentypen/Algorithmen brauche ich? Welche Schnittstellen brauche ich? Was gibt es davon schon in SeqAn?

Block 4: Umsetzung des Projektplans, Implementierung (24.4.-24.5.)

- Selbstständig, regelmäßige Treffen mit Betreuer/in

Vorstellung der Ergebnisse (29.5.)

The End – Fragen?

Siehe auch: [ABI Wiki](#) -> [Lecture Wiki](#) -> [PMSB 2013](#)