

Übungen zum Sequenzanalyse-Praktikum

Universität Bielefeld, SoSe 2013
Prof. Dr. Jens Stoye · M.Sc. Stefan Jannsen · B.Sc. Linda Sundermann
<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013summer/SequaPraktikum-seqan@CeBiTec.Uni-Bielefeld.DE>

Übungsblatt 2 vom 19.04.2013
Abgabe am Mittwoch, den 24.04.2013

Aufgabe 1 (Kompression mit bzip2)

Erstelle 6 Textdateien mit je 100.000 Zeichen, die folgendermaßen aufgebaut sein sollen:

1. alle Zeichen sind identisch,
2. Zufallszeichen aus einem 4-Buchstabenalphabet,
3. DNA-Sequenz,
4. Protein-Sequenzen,
5. natürlichsprachlicher deutscher Text,
6. natürlichsprachlicher Text einer anderen Sprache deiner Wahl.

Komprimiere diese Dateien mit `bzip2`. Beschreibe deine Beobachtungen und versuche sie zu erklären.

Aufgabe 2 (Bowtie 2)

Installiere auf deinem System die aktuelle Version von Bowtie 2, die du hier findest: <http://bowtie-bio.sourceforge.net/bowtie2>. Verwende entweder eines der vorkompilierten Pakete oder übersetze die Quellversion.

1. Sieh dir das MANUAL an. Keine Angst, du musst nicht alles im Detail lesen. Verschaffe dir lediglich einen Überblick.
2. Gegen Ende des MANUALs gibt es einen Abschnitt „Getting started with Bowtie 2: Lambda phage example“. Dieser besteht aus fünf Unterabschnitten mit Beispielaufrufen von Bowtie 2 und `samtools/bcftools`. Führe diese nacheinander aus und schau dir die Ergebnisse an. (Für die Beispiele des letzten Unterabschnitts „Using SAMtools/BCFtools downstream“ musst du auch Zugriff auf `samtools` und `bcftools` haben, die du hier findest: <http://sourceforge.net/projects/samtools>.) Beschreibe in deinem Protokoll die einzelnen Aufrufe und was sie bewirken.