

# Sequenzanalyse-Praktikum

Veranstalter: Roland Wittler, Tizian Schulz  
Tutor: Leonard Bohnenkämper

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2018winter/SequaPrak>  
[praktikum-seqan@CeBiTec.Uni-Bielefeld.DE](mailto:praktikum-seqan@CeBiTec.Uni-Bielefeld.DE)

# Inhalte und Ziele

- Modul Sequenzanalyse, 2. Teil
  - 5 LP (150 h, 60 h Kontaktzeit + 90 h Selbststudium)
- Voraussetzung: Bestandene Prüfung Sequenzanalyse + A&D oder ähnliches
- Vierstündige Übung
    - Zwei Stunden Theorie
    - Zwei Stunden Praxis
  - Vertiefung und Erweiterung der Vorlesung Sequenzanalyse
    - Datenformate, Datenbanksuche
    - Multiples Alignment in der Praxis
    - ...

# Zeiten und Räume

## Dienstags- und Mittwochskurs

- 10:15 – 11:45 in U10-146
- 12:30 – 14:00 in V6-116 (BRF Schulungsraum)

# Anforderungen

- Halten eines 30 bis 45-minütigen Vortrags
- Schriftliche Ausarbeitung des Vortrags in 5 bis 10 Seiten
- Bearbeitung der Übungsaufgaben
- Anfertigung von (akzeptablen) Protokollen über die Lösung der Aufgaben

# Ablauf

## Theorie:

- Besprechung des Protokolls letzter Woche
- Vortrag
- Diskussion und Fragen
- Besprechung der aktuellen Aufgabenstellungen

## Praxis:

- Bearbeitung der Praktikumsaufgaben
- Kurze Besprechung der Praktikumsaufgaben

# Das Protokoll

- Abgabefristen:
  - Dienstagsgruppe: Sonntag (24:00 Uhr)
  - Mittwochsgruppe: Montag (24:00 Uhr)
  - (Nachbesprechung im Seminar)
  - Finale Abgabe eine Woche später, selbe Zeiten
- Protokoll: PDF-Datei (< 3 MB)
- Quellcode: als Textdateien, nicht als Projekt-Ordner
- praktikum-seqan@CeBiTec.Uni-Bielefeld.DE
  
- Kurze Beantwortung aller Aufgaben
- Vollständige Sätze, Tabellenüberschriften, Achsenbeschriftungen etc.
- Siehe Beispiel auf Veranstaltungsseite
  
- Maximal zwei Personen pro Protokoll

# Das Protokoll – Fehlversuche

- Höchstens drei Fehlversuche erlaubt
- Fehlversuche bei:
  - unvollständiger vorläufiger Abgabe,
  - fehlerhafter finaler Abgabe,
  - verspäteter/keiner Abgabe (vorläufig sowie final).

# Der Vortrag

## Material und Hilfestellung:

- Literatur (zwei Wochen vor Vortrag ausgeteilt)
- Spezifische Fragestellungen
- Vorbesprechung (bis spätestens Freitag vor Vortragstermin)
  
- 30 bis 45-minütiger Vortrag
- Deutsch oder Englisch
- Latex, PowerPoint, OpenOffice, ...



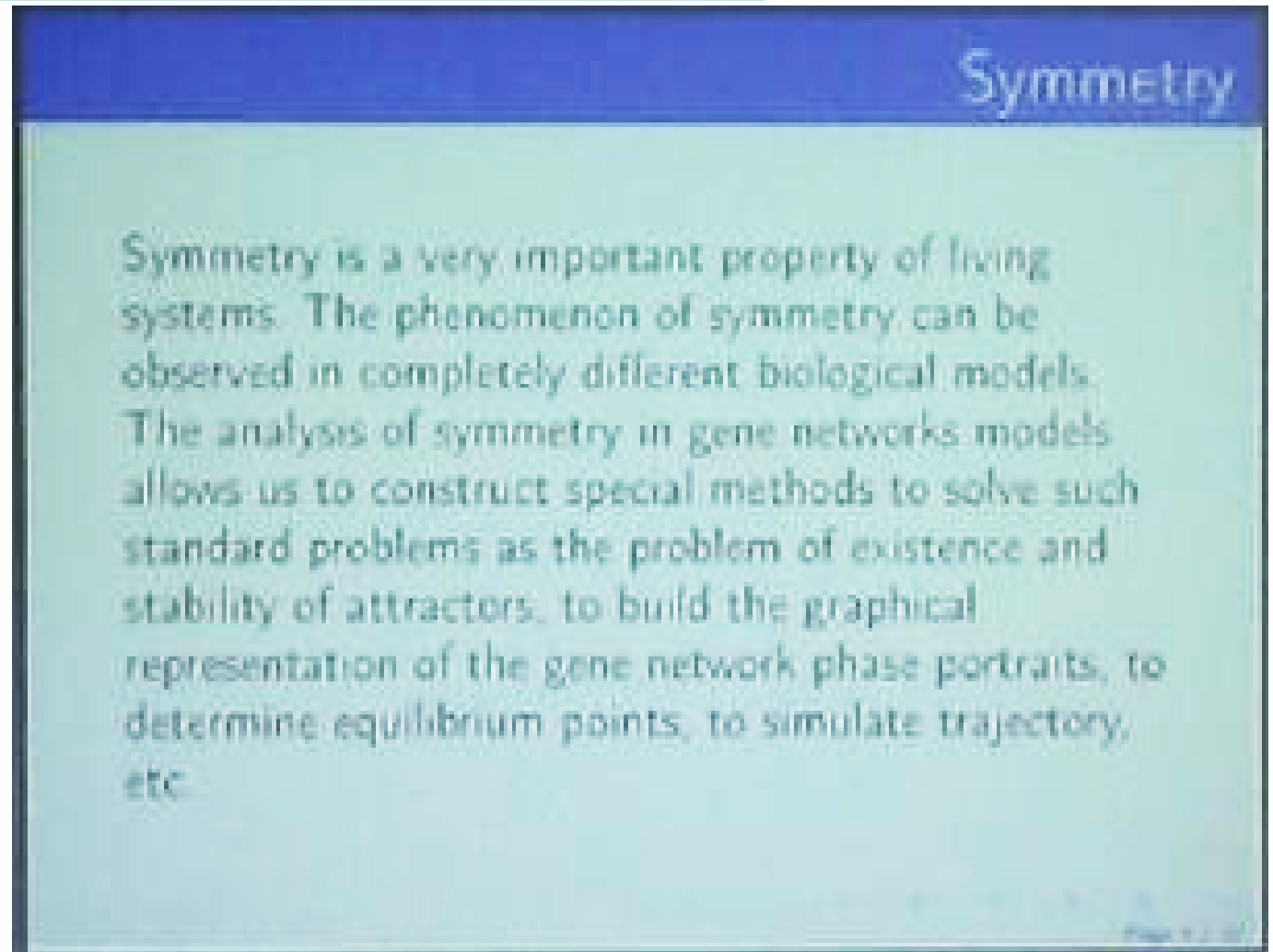
# Der Vortrag – Folien

## Einige Hinweise:

- Nicht zu viel Text
- Abbildungen verwenden (und zitieren!)
- Text groß genug
- Nicht zu viele Schriftgrößen/arten
- Gut zu erkennende Schriftfarbe
- ...

# Folien – ein schlechtes Beispiel

So  
nicht:



# Die Ausarbeitung

Abgabe zwei Wochen nach eurem Vortrag.

Hilfestellung:

- Inhaltliche Rückmeldung zum Vortrag
- Zweimaliges Einreichen, Korrektur spätestens zwei Wochen nach Rückgabe
  
- 5–10 Seiten
- Deutsch oder Englisch
- PDF

# Die Ausarbeitung - Aufbau

- Inhaltsverzeichnis
- Motivation
- Ausarbeitung des Vortrags in (kurzen) Kapiteln
- Zusammenfassung
- Quellen

# Die Ausarbeitung – Abbildungen und Tabellen

- Abbildungen: Unterschriften
- Tabellen: Überschriften
  
- Nummer, (Titel)
- Beschreibung
- Quellenangabe (falls nicht selbst erstellt)
  
- Schriftgröße in Grafiken = Schriftgröße im Text
  
- Referenzierung im Text: „ ... as visualized in Figure 3.“  
„ ... in Tabelle 2 gegenübergestellt.“

# Tabellen – ein Beispiel

**Table 6.1:** Listing of *Corynebacterium* [80] and *Mycobacterium* [125] genomes used in the study. For each strain, its proper identifier, its length, its G+C content and the number of coding regions is given. For the latter value, also putative, automatically predicted coding regions are considered.

Genome	Length (in base pairs)	G+C Content (in %)	Number of Coding Regions
<i>C. glutamicum</i> ATCC 13032	3,282,708	53.84	3,058
<i>C. glutamicum</i> R	3,314,179	54.13	3,052
<i>C. efficiens</i> YS-314	3,147,090	63.14	2,950
<i>C. diphtheriae</i> NCTC 13129	2,488,635	53.48	2,264
<i>C. ulcerans</i> 809	2,483,825	53.31	2,228
<i>C. aurumicosum</i> ATCC 700975	2,790,189	60.63	2,539
<i>C. jeikeium</i> K411	2,462,499	61.40	2,105
<i>C. urealyticum</i> DSM 7109	2,369,219	64.19	2,024
<i>C. urealyticum</i> DSM 7111	2,269,359	64.40	1,890
<i>C. kroppenstedtii</i> DSM 44385	2,446,804	57.46	2,037
<i>M. tuberculosis</i> KZN 1435	4,398,250	65.61	4,059

# Abbildung – ein Beispiel

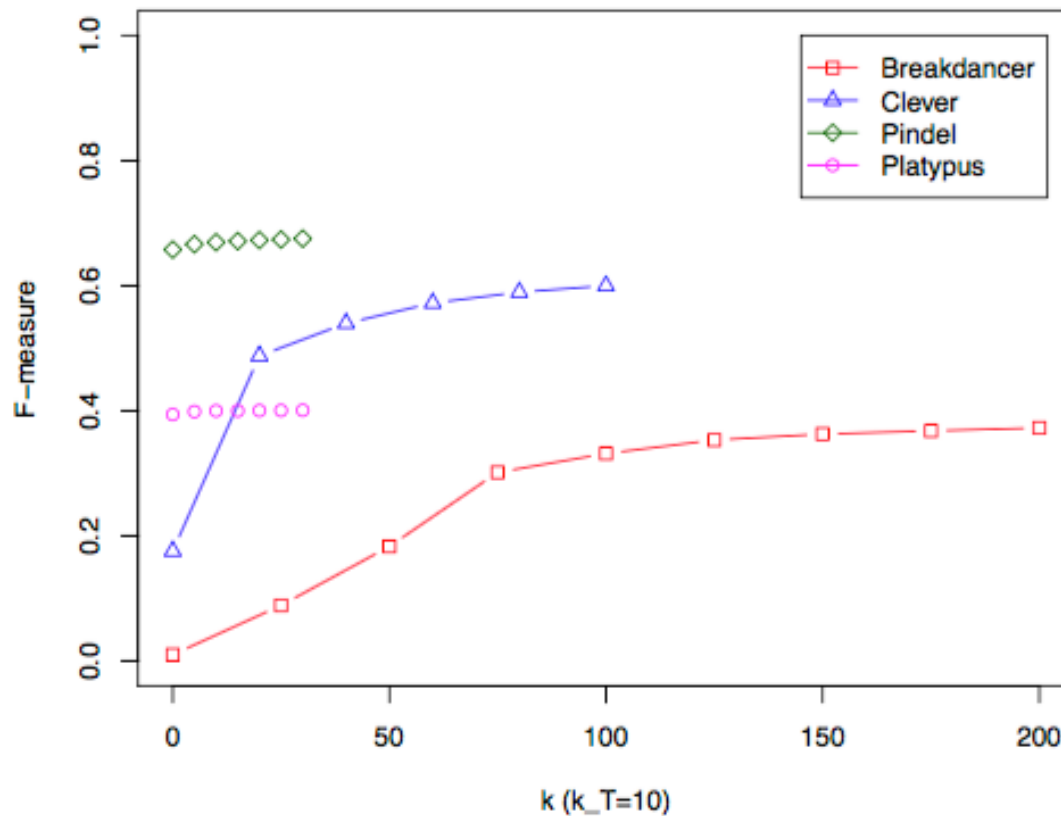


Fig. 4: F-measure for callsets of different tools compared to truth (Venter dataset). Predictions have been compared with varying neighborhood sizes to the truth with a fixed neighborhood size of  $k_T = 10$ . Precision and recall have been determined by the matching approach described in Section 2.2 and are shown in Section D in the Supplement.

# Die Ausarbeitung - Zitieren

- Letztes Kapitel „Literaturverzeichnis“ mit allen verwendeten Quellen
- Angabe im Text, in Abbildungen und Tabellen
  - Beispiele: „Laquer entwickelte eine Approximation [2].“  
„Durch die Erforschung von verwandtschaftlichen Beziehungen, ... umfassende Informationen über deren Evolution [15, Kapitel 2].“  
„Abbildung 3. Schema der Datenbank. ... Grafik entnommen aus [4].“
- Einheitliches Format
  - Buch: Autoren, Titel, Verlag, Erscheinungsjahr
  - Paper: Autoren, Titel, Journal, Volume, Seitenzahl, Erscheinungsjahr
  - Internetseite: Autoren (wenn erkennbar), Seitentitel, URL, Zugriffsdatum



# Literaturverzeichnis – ein Beispiel

## Literaturverzeichnis

- [1] R. D. M. Page und E. C. Holmes. *Molecular Evolution – A Phylogenetic Approach*. Blackwell Science Ltd, 1998.
- [3] D. A. Benson, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Rapp und D. L. Wheeler. GenBank. *Nucleic Acids Res.*, 28(1):Seiten 15–18, 2000.
- [15] J. Bischoff, M. Domrachev, S. Federhen, C. Hottel, D. Leipe, V. Soussov, R. Sternberg und S. Turner. The NCBI Entrez Taxonomy Homepage. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Taxonomy>. Stand: 06.10.2006.

# Plagiatismus

- Verwenden von Texten aus Papern/Büchern/Internetseiten ohne Kenntlichmachung als Zitat mit Quellenabgabe
- Übersetzen einer Quelle ins Deutsche ohne Kenntlichmachung als Zitat mit Quellenabgabe
- Umformulieren (Synonyme, Satzumstellung, etc.) von Texten
- Eine Ausarbeitung darf nicht nur aus Zitaten bestehen – auch nicht, wenn sie kenntlich gemacht werden.
- Verwenden von Texten anderer aus Vorjahren

... = Abschreiben

= Nicht bestanden!

# Die nächsten Wochen...

<b>Datum</b>	<b>Thema</b>	<b>Vortragender</b>
09/10.10.	Anzahl Alignments	Roland
16/17.10.	BWT in der Praxis	Roland
23/24.10.	Q-Gramme und de Bruijn-Graphen	?

... *siehe Internetseite.*