

Vorlesung Einführung in die Bioinformatik SoSe 2014

Von der DNA zur Datenbank: Sequenzierung & Assemblierung



Prof. Daniel Huson

ERBERHARD-KARLS
UNIVERSITÄT
TÜBINGEN

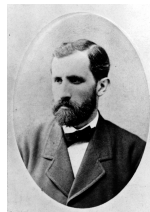


Singapore Center on Environmental
Life Sciences Engineering (SCELESE)

ZBIT Center for
Bioinformatics

ERBERHARD-KARLS
UNIVERSITÄT
TÜBINGEN
Center for Bioinformatics

Entdeckung der DNA



Friedrich Miescher
(1844-1895)

1869 Miescher entdeckte DNA in der
Küche des Schlosses Hohentübingen

Daniel Huson 2014

2

ERBERHARD-KARLS
UNIVERSITÄT
TÜBINGEN
Center for Bioinformatics

Biomolekulare Grundlage des Lebens



...
A-T
C-G
G-C
T-A
A-T
A-T
...

1953 Watson und Crick

- Struktur der DNA ist eine Doppelhelix
- Die **Reihenfolge** der Basen enthält Informationen, die in der Zelle verarbeitet werden

Daniel Huson 2014

3

FRIEDRICH-SCHILLER-UNIVERSITÄT
 TUBINGEN
 Center for Bioinformatics

DNA und Proteine

Daniel Huson 2014
4

FRIEDRICH-SCHILLER-UNIVERSITÄT
 TUBINGEN
 Center for Bioinformatics

Protein Sequenzierung

1955 Biochemiker Frederick Sanger bestimmt die erste vollständige Aminosäuresequenz eines Proteins, Insulin.

```

1 malwmrlpl lalllalwgd paaafvnghl cgshlvealy lvcgergffy tpktrreadd
61 lqvgqvelgg gpgagslqpl alegslqkrg iveqcctsic slyqlenycn

```

Daniel Huson 2014
5

FRIEDRICH-SCHILLER-UNIVERSITÄT
 TUBINGEN
 Center for Bioinformatics

DNA Sequenzierung

1975 Frederick Sanger entwickelt die Kettenabbruchmethode zur Sequenzierung von DNA

Daniel Huson 2014
6

UNIVERSITÄT
TÜBINGEN
Center for Bioinformatics


Genomics

- Genomics ist das Studium der Genomsequenzen einzelner Organismen



- Sequenzierung und Assemblierung
- Annotation und Analyse
- Genomvergleich

UNIVERSITÄT
TÜBINGEN
Center for Bioinformatics


Genomics

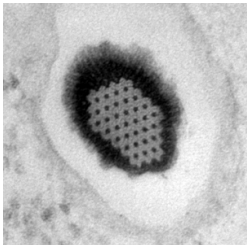
- Genomgrößen:
 - Virus 5-10 Kilobasen (kb)
 - Bakterien: 1-10 Millionen Basen (Mb)
 - *Arabidopsis thaliana*: 120 Mb
 - Fruchtfliege: 140 Mb
 - Mensch: 3 Milliarden Basen (Gb)

“Genetics is a way of thinking.
 Genomics is a set of tools.”
 Mary-Claire King (Science, Feb 2011)

UNIVERSITÄT
TÜBINGEN
Center for Bioinformatics

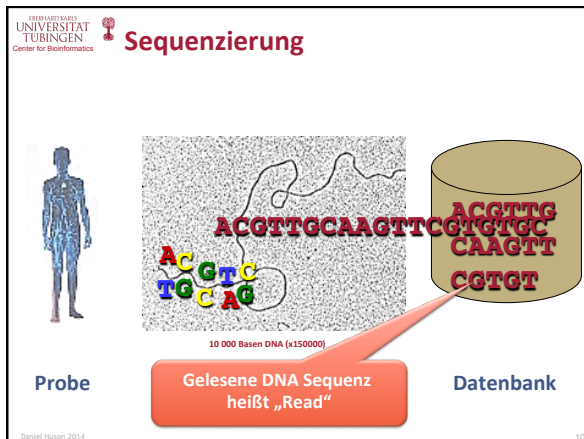

Mega Virus *Pithovirus sibericum*

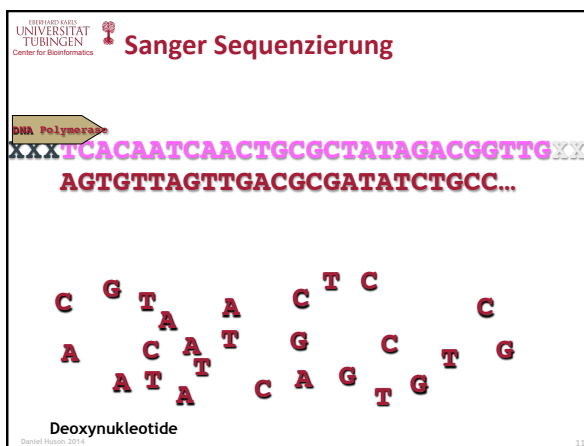
- 2013 publiziert
- 30,000 Jahre geforen
- 2.9 Mb Genom
- Größe eines kleines Bakteriums

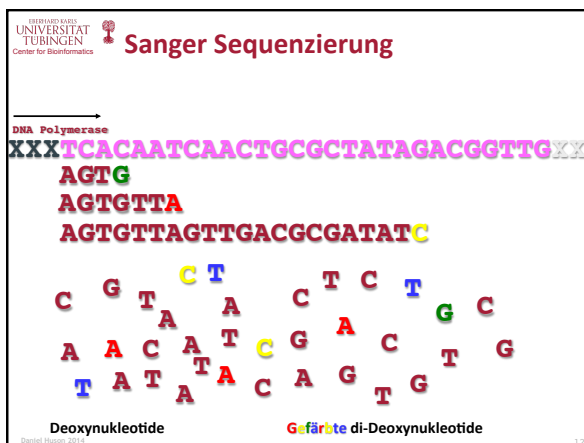


Ed Yong, Nature News, April 2014

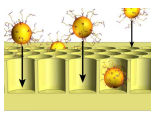
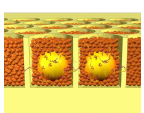
3

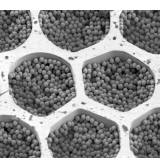


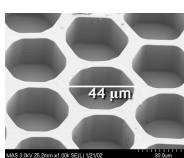
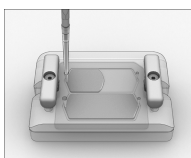





Kugeln verteilen auf einer PicoTiterPlate™

Kugeln werden in Mikrowells verteilt →
 
 Enzyme hinzugefügt →
 

Zentrifugation
 

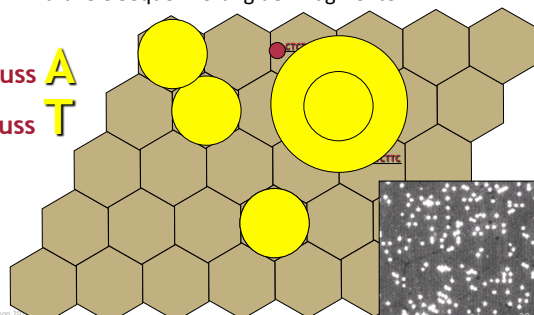



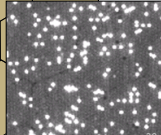
Source: 454


Pyrosequenzierung

• Parallele Sequenzierung der Fragmente:

Fluss **A**
 Fluss **T**

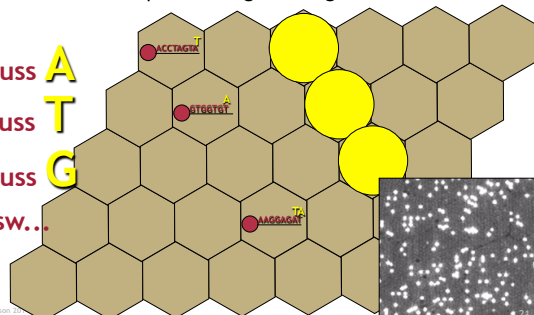


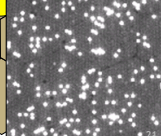
Daniel Huson 2014
 


Pyrosequenzierung

• Parallele Sequenzierung der Fragmente:

Fluss **A**
 Fluss **T**
 Fluss **G**
 usw...



Daniel Huson 2014
 

Pyrosequenzierung

- Basenbestimmung durch Analyse einer Serie von Bildern

Sequencing and Basecalling Results for 101base Read

TCAGCGTAAG...

Source: 454

Preisentwicklung DNA Sequenzierung

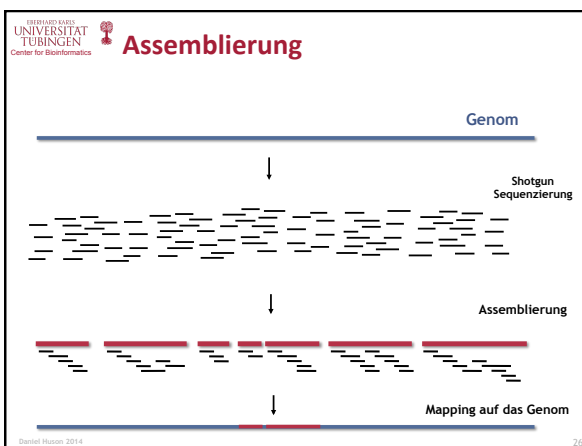
Quelle: <http://spectrum.ieee.org/biomedical/devices/the-gene-machine-and-me>

Sequenzierungsprojekte

Genome Sequencing Projects on GOLD
September 2008, 5643 projects

GOLD: Genomes online database
www.genomesonline.org





UNIVERSITÄT TUBINGEN
Center for Bioinformatics

Assemblierung

- Sequenzierung produziert Millionen von kurzen DNA Reads
- Bioinformatik: Suche überlappende Sequenzen und füge sie zusammen

AGGTTTAG

TTTAGGAT

AGGTTTAGGAT

Daniel Huson 2014

UNIVERSITÄT TUBINGEN
Center for Bioinformatics

NCBI Datenbank

- Die Hauptresource für Sequenzdaten ist NCBI:

© 2014

UNIVERSITÄT TUBINGEN
Center for Bioinformatics

NCBI Datenbank

© 2014

UNIVERSITÄT TUBINGEN
Center for Bioinformatics

NCBI Datenbank

Hund-Genom?

© 2014

[illegible]