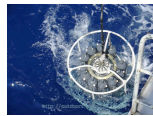


Aus einer Hand voll Erde – Metagenomik



Prof. Daniel Huson



Wie viele Arten gibt es?

Ungelöste Hauptfrage:



- Anzahl der verschiedenen Arten?
- Nicht mal bis auf einige Größenordnungen genau beantwortet
- Einige Schätzungen
 - 3-50 Millionen Arten von Arthropoden
 - 1-100 Millionen Arten von Fadenwürmer



Once the diversity of the microbial world is catalogued, it will make astronomy look like a pitiful science

- Julian Davies, Professor Emeritus, Microbiology and Immunology, UBC

Benannte Arten

~1.7 Million benannte Arten

- 287,655 plants, including:
 - 15,000 mosses
 - 13,025 ferns
 - 980 gymnosperms
 - 199,350 dicotyledons
 - 59,300 monocotyledons
- 74,000-120,000 fungi
- 10,000 lichens
- 5,700 prokaryotes
- ~1,250,000 animals, including:
 - 1,190,200 invertebrates:
 - 950,000 insects
 - 70,000 mollusks
 - 40,000 crustaceans
 - 130,200 others
 - 58,808 vertebrates:
 - 29,300 fish
 - 5,743 amphibians
 - 8,240 reptiles
 - 10,234 birds
 - 5,416 mammals

Quelle: <http://en.wikipedia.org/wiki/Biodiversity>

Sequenzen für ~200,000

Metagenomik

- “Das Studium der DNA von unkultivierten Organismen”
- > 99% aller Mikroben nicht kultivierbar
- **Genom:**
 - Gesamtheit der genetischen Informationen eines **einzelnen** Organismus
- **Metagenom:**
 - Gesamtheit der genetischen Informationen einer **Ansammlung** von Organismen
- **Meta-Transkriptom:** Gesamtheit aller RNA Sequenzen, u.s.w.

Daniel Huson 2013

Typische Quellen für Metagenome

- Bodenproben
- Meeresproben
- Bohrkerne aus dem Meeresgrund
- Luftproben
- Uralte Knochen
- Mikrobiota des Menschen

Daniel Huson 2013

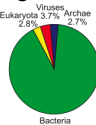
Global Ocean Sampling Expedition

- Craig Venter und Kollegen vom Venter Institut segeln um die Welt und sammeln Wasserproben

Daniel Huson 2013


Global Ocean Sampling Expedition

- DNA wird extrahiert und sequenziert
- 2007 (Sonderausgabe von PLoS Biology):
 - 41 Proben sequenziert
 - 7.7 Millionen Reads mittels Sanger Sequenzierung
 - Mittlere Länge der Reads: 882bp
 - Menge an Sequenz: 5.9 Gb



Sequenzierung 2013

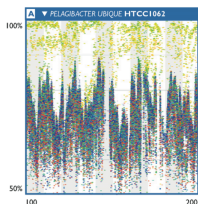
Alpha Proteobacteria	0.32
Unclassified Proteobacteria	0.153
Gamma Proteobacteria	0.132
Bacteroidetes	0.13
Cyanobacteria	0.079
Firmicutes	0.075
Actinobacteria	0.046
Marine Group A	0.022
Beta Proteobacteria	0.017
OP1	0.008
Unclassified Bacteria	0.008
Delta Proteobacteria	0.005
Planctomycetes	0.002
Epsilon Proteobacteria	0.001




Global Ocean Sampling Expedition

Zwei Hauptergebnisse:

- Sequenzen sehr inhomogen
- Ca. 6 Millionen neue Gene entdeckt, lineare Entdeckungsrate





Daniel Huson 2013


Metagenomik in der Medizin

- Gibt es eine Beziehung zwischen Krankheiten und der Darmflora bei Menschen?



<http://en.wikipedia.org>

- Mausmodelle

Daniel Huson 2013

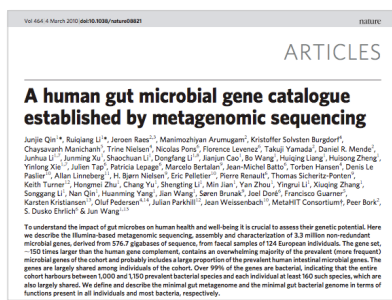
- Vergleich der Darmflora einer normalen Maus und einer obesen Maus (Turnbaugh et al, Nature, 2006)
 - Sanger Sequenzierung: 40 Mb
 - Roche(454) Sequenzierung: 160 Mb



Ergebnisse:

- Veränderung in der Zusammensetzung der Arten:
 - Firmicutes vs Bacteroidetes
- Veränderung in der Zusammensetzung der Gene:
 - Zunahme energieverarbeitender Gene

Daniel Huson 2013

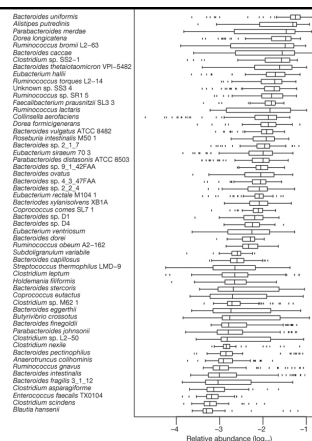


576 Gb DNA von 124 Individuen

Daniel Huson 2013

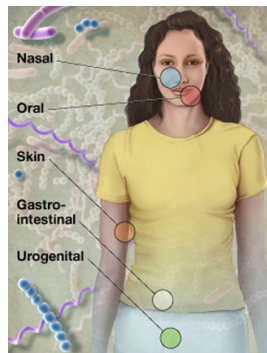
Darm Mikrobiom

- 57 Spezies vorhanden in $\geq 90\%$ der Individuen mit Anteil $>1\%$
- Hohe Variabilität
- Bacteroidetes und Firmicutes die häufigsten Bakterien



Human Microbiome Project

- **Ziel des HMP:** Charakterisierung der Mikrobiota, die auf dem menschlichen Körper leben.
- **Hauptfrage:** Welche Rolle spielen diese Mikrobiota für die menschliche Gesundheit und für Krankheiten?



<http://commonfund.nih.gov/hmp/>

Daniel Huson 2013

Human Microbiome Project

Volume 486 Number 7402 pp157-286 14 June 2012

14 June 2012

Structure, function and diversity of the healthy human microbiome **OPEN**

The Human Microbiome Project Consortium

The Human Microbiome Project Consortium reports the first results of their analysis of microbial communities from distinct, clinically relevant body habitats in a human cohort; the insights into the microbial communities of a healthy population lay foundations for future exploration of the epidemiology, ecology and translational applications of the human microbiome.

See also [News & Views by Reimar](#)

A framework for human microbiome research **OPEN**

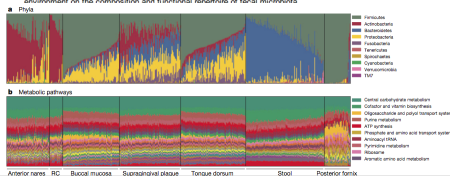
The Human Microbiome Project Consortium

The Human Microbiome Project Consortium has established a population-scale framework to study a variety of microbial communities that exist throughout the human body, enabling the generation of a range of quality-controlled data as well as community resources.

Human gut microbiome viewed across age and geography

Tanya Yatsunenko, Federico E. Rey, Mark J. Manary, Indi Trehan, Maria Gloria Dominguez-Bello **+ et al.**

The human gut microbiome from a large cohort of more than 500 individuals living on three continents with three distinct cultures is analysed, emphasizing the effect of host age, diet and environment on the composition and functional repertoire of faecal microbiota



Daniel Huson 2013

Pansen Mikrobiom

Science 28 January 2011:
Vol. 331 no. 6016 pp. 463–467
DOI: 10.1126/science.1200387

[< Prev](#) | [Table of Contents](#) | [Next >](#)

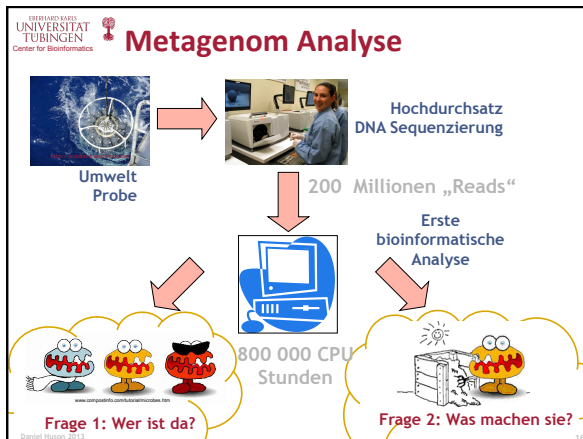
REPORT

Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen

Matthias Hess^{1,2,*}, Alexander Sczyrba^{1,2,*}, Rob Egan^{1,2}, Tae-Wan Kim³, Harshal Chokhawala³, Gary Schroetz⁴, Shujun Luo⁴, Douglas S. Clark^{3,5}, Feng Chen^{3,5}, Tao Zhang^{3,5}, Roderick I. Mackie⁶, Len A. Pennacchio^{3,5}, Susannah G. Tringie^{1,2}, Axel Visel^{1,2}, Tanja Woyke^{1,2}, Zhong Wang^{1,2}, and Edward M. Rubin^{1,2,3}



- Ziel: Entdeckung von Genen zur Verdauung von Zellulose (Anwendungen in Biotreibstoffen)
- 268 Gb Sequenz
- Ergebnis: 27 000 neue Gene zur Verdauung von Biomasse





Taxonomische Analyse

Taxonomische Analyse – Who is out there?

- Um dieses zu beantworten, werden die Reads mit einer Referenzdatenbank verglichen:

Probe → ACTGTGCACGTTGACGTAAGTTT...CGTGT → NCBI

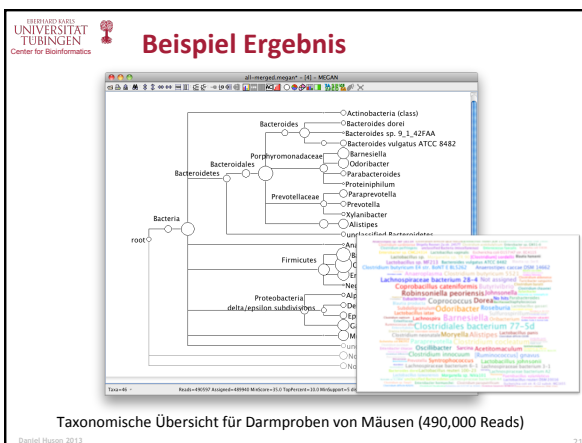
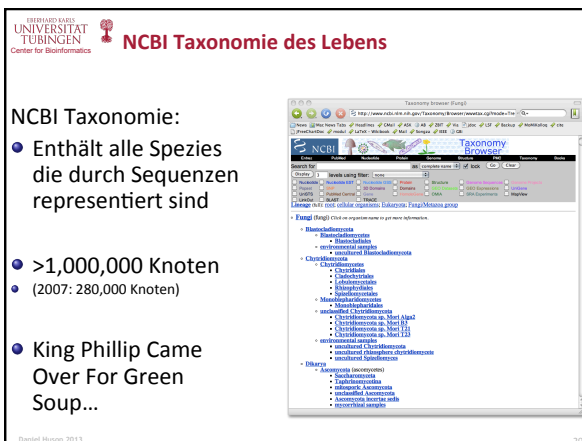
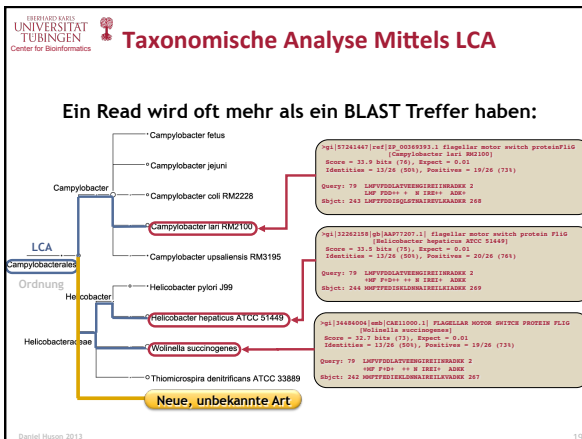
- Ergebnis sind BLAST Treffer:

```
>gi|57241447|ref|ZP_00369393.1| flagellar motor switch protein
[Campylobacter lari RM2100]
Score = 33.9 bits (76), Expect =0.01
Identities = 13/26 (50%), Positives = 19/26 (73%)

Query: 79 LMFVFDLATVEENGIREIINRAKK 2
      LMF FDD++ + N IRE++ ADK+
Sbjct: 243 LMFDFDISQLSTNAIREVLKADKR 268
```

Campylobacter lari RM2100

Daniel Huson 2013



Funktionale Analyse – What are they doing?

- Auch diese Frage wird mittels eines Vergleiches mit einer Referenzdatenbank beantwortet:

>gi|57241447|ref|ZP_00369393.1|flagellar motor switch protein
 [Campylobacter lari NZL100]
 Score = 33.9 bits (76), Expect = 0.01
 Identities = 13/26 (50%), Positives = 19/26 (73%)
 Query: 79 LMFVDDLATVEENGREINRAKK 2
 LMF FDD++ + N RDK
 Sbjct: 243 LMFVDDISQLSTNAIREVLKAKDKR 268

flagellar motor switch protein

Daniel Huson 2013

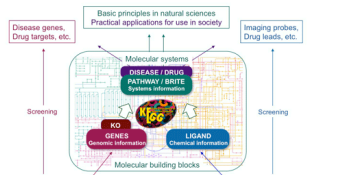
- [KEGG Home](#)
 - [Release notes](#)
 - [Current statistics](#)
 - [Plea from KEGG](#)
- [KEGG Database](#)
 - [KEGG overview](#)
 - [Searching KEGG](#)
 - [KEGG mapping](#)
 - [Color codes](#)
- [KEGG Objects](#)
 - [Pathway maps](#)
 - [Brite hierarchies](#)
- [KEGG Software](#)
 - [KegTools](#)
 - [KEGG API](#)
 - [KGML](#)
- [KEGG FTP](#)
 - [Subscription](#)

- [GenomeNet](#)
- [DBGET/LinkDB](#)
- [Feedback](#)

KEGG Overview

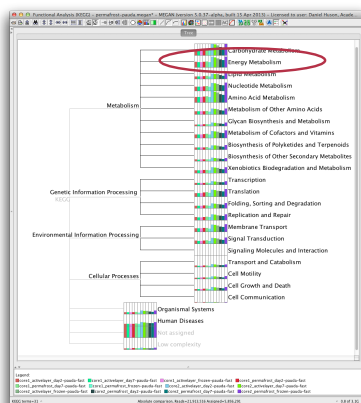
1. Genomes to Biological System

KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from genomic and molecular-level information. It is a computer representation of the biological system, consisting of molecular building blocks of genes and proteins (genomic information) and chemical substances (chemical information) that are integrated with the knowledge on molecular wiring diagrams of interaction, reaction and relation networks (systems information).



Daniel Huson 2013

Funktionaler Vergleich verschiedener Proben

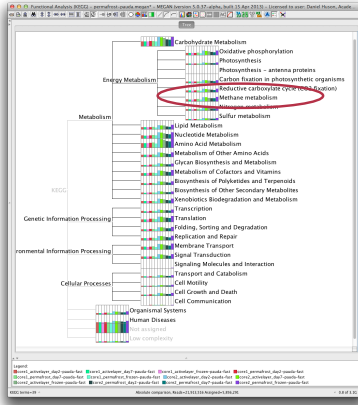


10



KEGG
Analyse

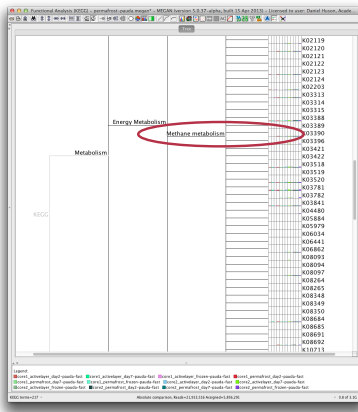
Funktionaler
Vergleich
verschiedener
Proben





KEGG
Analyse

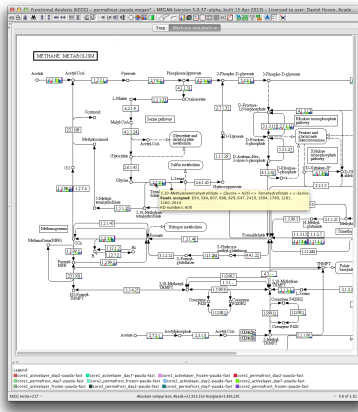
Funktionaler
Vergleich
verschiedener
Proben

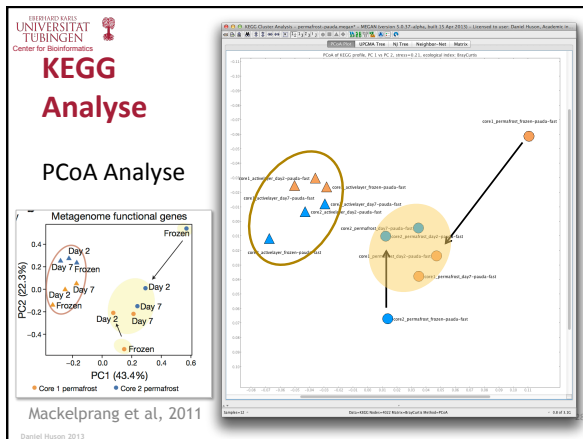




KEGG
Analyse

Funktionaler
Vergleich
verschiedener
Proben





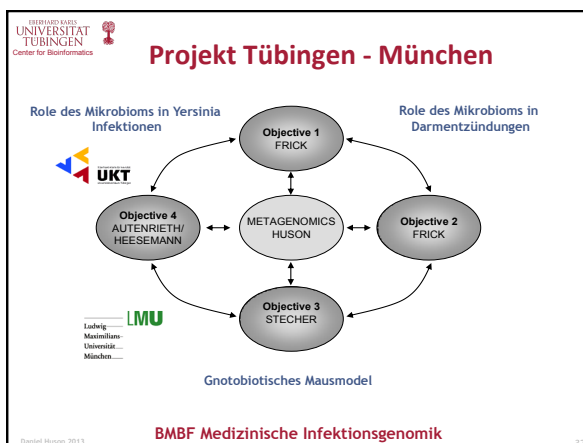
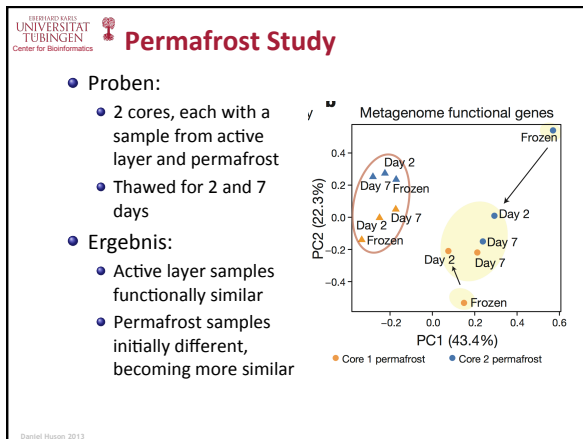
Funktionale Analyse mit SEED

- Die SEED Klassifikation bildet Gene auf **functional roles** in **subsystems** ab
- Ein **subsystem** ist eine Menge von **functional roles** die ein spezifisches biologisches System ergeben
- Grundlage der RAST und MG-RAST Dienste:
Rapid Annotation using Subsystem Technology
- Graph hat $\approx 13,000$ Knoten
- www.theSEED.org

Overbeek et al., Nucleic Acids Res 33(17), 2005

Permafrost Studie

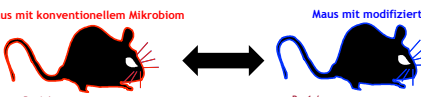
- Welche Folgen hat das Schmelzen der Permafrost?
- 12 Proben, 250 Millionen Illumina reads
- 800,000 CPU Stunden für KEGG Analyse
- Ein Hauptergebnis: Permafrost Proben werden sich beim schmelzen funktional ähnlicher



Darm Mikrobiom

Ziel 1: Charakterisierung der Darmflora bei einer Darmentzündung (Kolitis)

- **Taxonomischer Vergleich:**

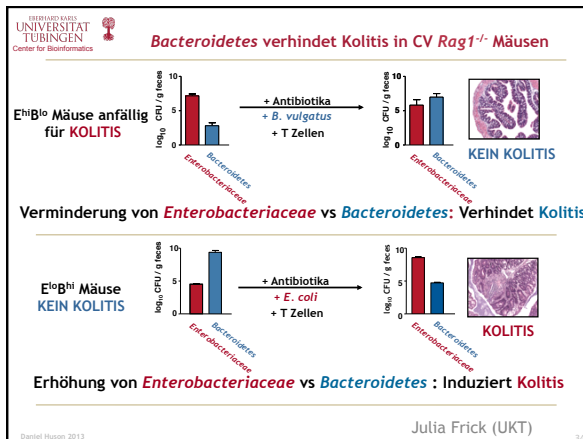


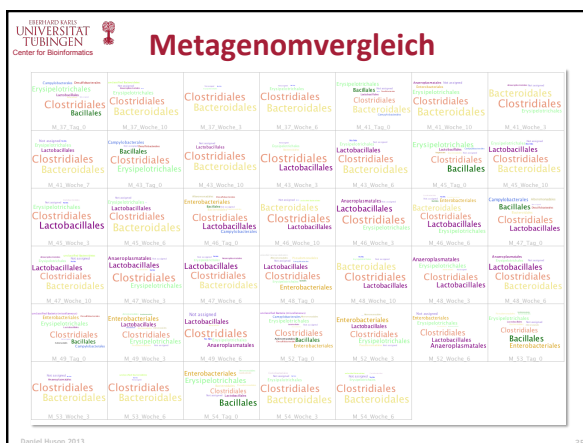
Maus mit konventionellem Mikrobiom $\longleftrightarrow$ **Maus mit modifiziertem Mikrobiom**

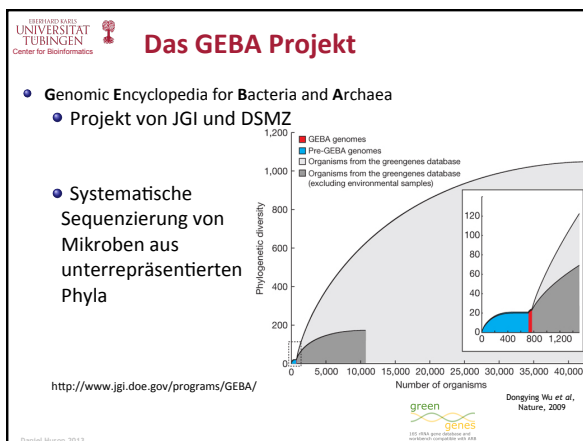
Rag1^{-/-} **Rag1^{-/-}**

- **Welche Bakterien verursachen eine Darmentzündung (IBD)?**

Julia Frick (UKT)







Zusammenfassung

- Ein Metagenom ist die vollständige DNA einer Ansammlung von Organismen
- Das "Human Microbiome Project" untersucht Metagenome von Mensch-assoziierten Mikrobiota
- Die drei grundlegenden bioinformatischen Fragen sind:
 - Taxonomische Analyse (Who is out there - Wer ist da?)
 - Funktionale Analyse (What are they doing? Was machen sie?)
 - Vergleichende Analyse (How do they compare - Welche Unterschiede gibt es?)
