

Genetik, Genomik & Bioinformatik 551-1298-00L

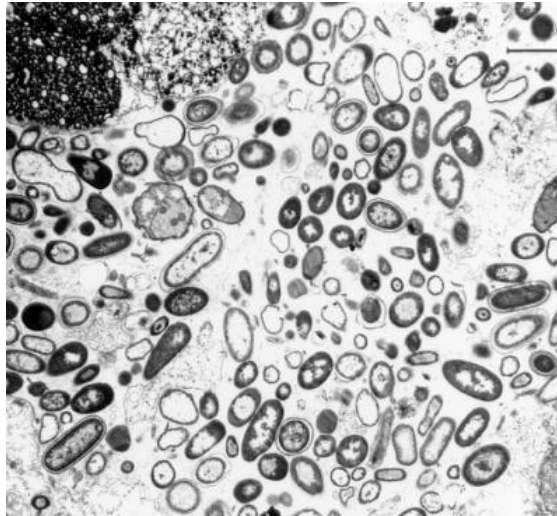
Metagenomik

Prof. Shinichi Sunagawa

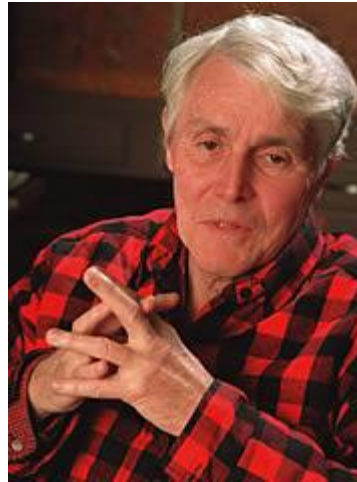
Bakterielle Diversität



Nucleotid-basierte Klassifizierung von Mikroorganismen



agtctcgctatgacgtcgtcgtcagactac
gtcgtacgtcgatatttctcgcgccggagc
gctagtacgtcgatatttctcgcgatcgtag
gagcctacgtcgtcgatagtgcgctagtgt



Carl Woese
he also found archaeas

Welche Nucleotidsequenz eignet sich als taxonomischer Marker?

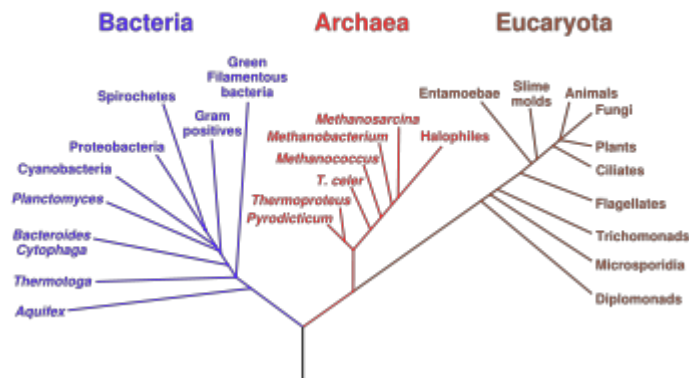
- enthalten in allen Organismen
- hoch konserviert
- nur selten horizontal transferiert
- mutiert in allen Organismen ähnlich schnell (“molecular clock“)

→ small subunit (SSU) rRNA
(Bakterien und Archeen: 16S;
Eukaryonten: 18S)

Nahe verwandte Bakterien haben nahe verwandte 16S rRNA-Sequenzen

Durch Vergleich können Verwandtschaftsbäume berechnet werden

Phylogenetic Tree of Life

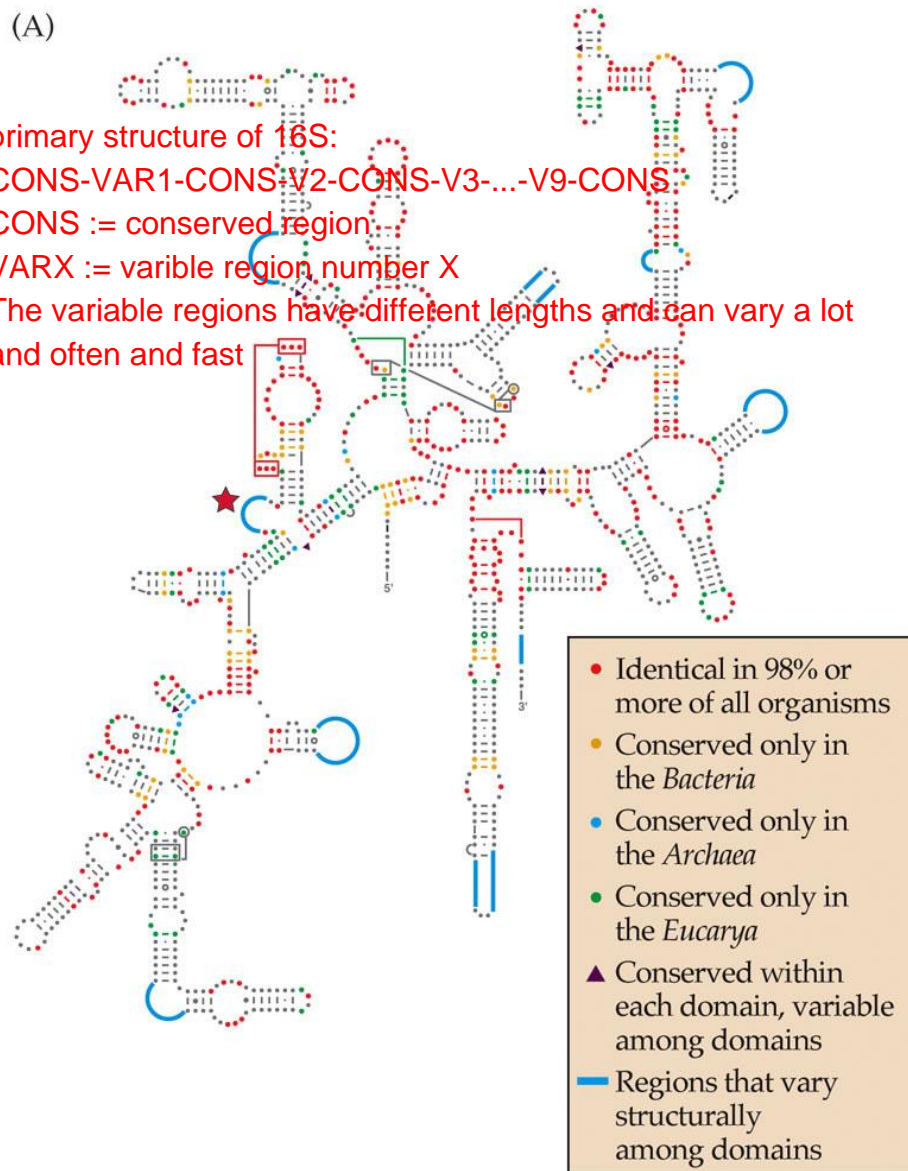


if a gene mutates too fast it diverges too quickly from the other organisms and it becomes very hard to track

Konservierung und Variation in der Small Subunit rRNA

(A)

primary structure of 16S:
 CONS-VAR1-CONS-V2-CONS-V3-...-V9-CONS
 CONS := conserved region
 VARX := variable region number X
 The variable regions have different lengths and can vary a lot
 and often and fast



Regionen sind konserviert bezüglich verschiedener taxonomischer Hierarchien. Klassifizierung durch Sequenzvergleich und Konstruktion phylogenetischer Bäume ist möglich.

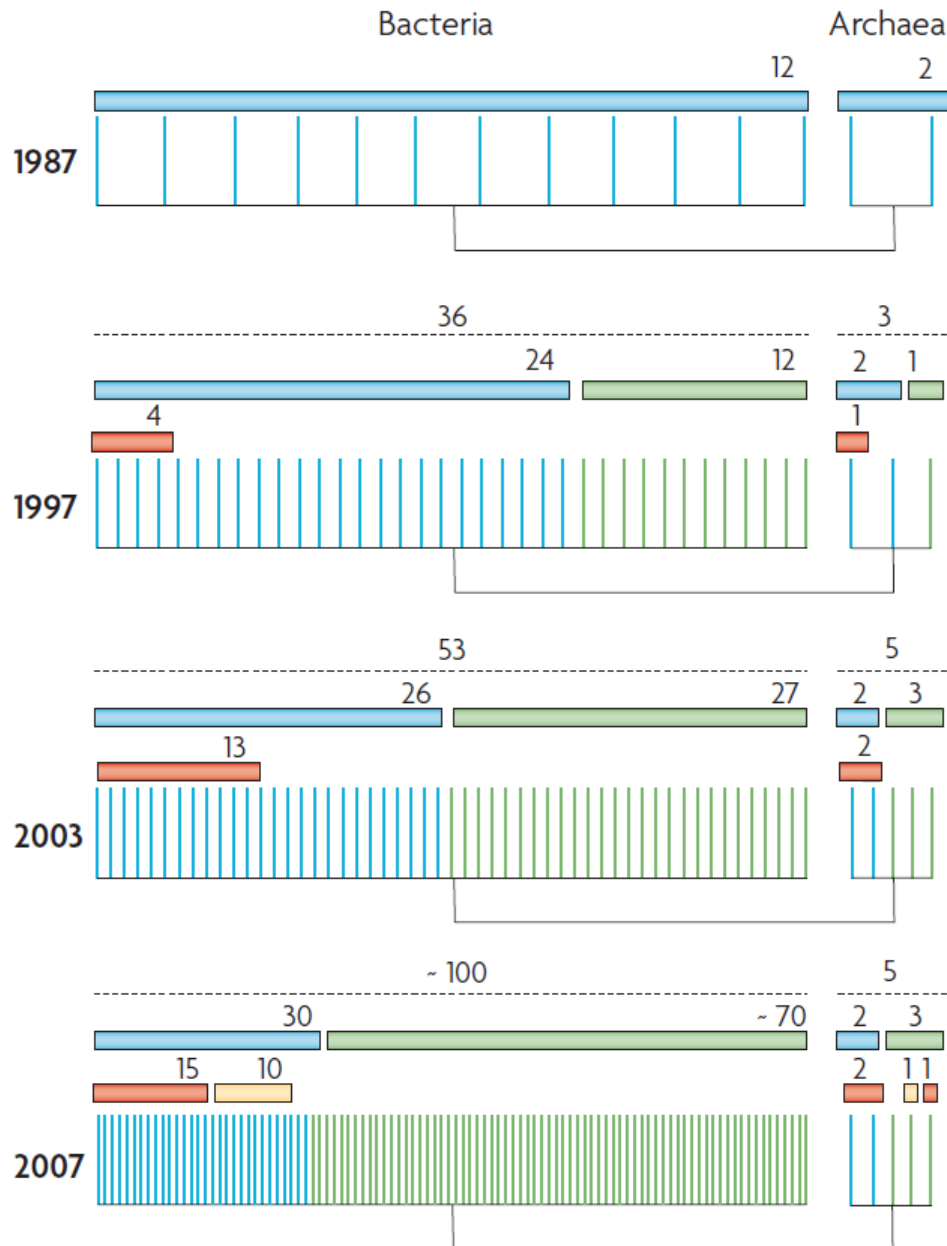
Table 17.1

Hierarchical classification of the bacterium *Spirochaeta plicatilis*

Taxon	Name
Domain	<i>Bacteria</i>
Phylum	<i>Spirochaetes</i> (vernacular name: spirochetes)
Class	<i>Spirochaetes</i>
Order	<i>Spirochaetales</i>
Family	<i>Spirochaetaceae</i>
Genus	<i>Spirochaeta</i>
Species	<i>plicatilis</i>

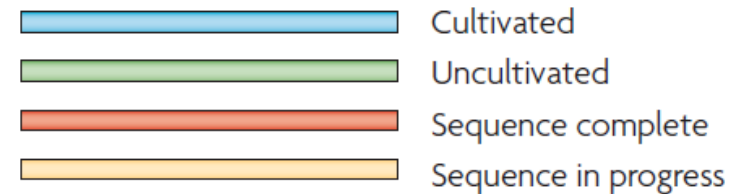
MICROBIAL LIFE, Table 17.1 © 2002 Sinauer Associates, Inc.

Die mikrobiologische Revolution



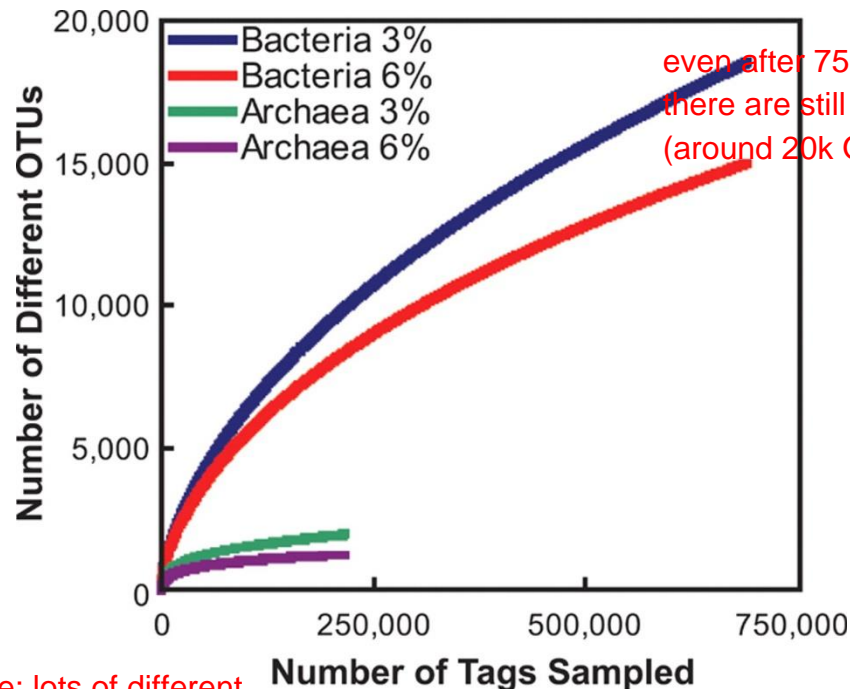
Entdeckung vieler bakterieller Phyla ohne kultivierte Vertreter (“Kandidatenphyla”)

In vielen Habitaten sind 0.1-1% aller Bakterien bisher nicht kultivierbar (“the great plate count anomaly”)



Artenvielfalt in natürlichen Habitaten

Analyse einer Tiefsee-Bodenprobe:



sep increase: lots of different species

Huber et al., *Science* **318**, 97 (2007)

Terrestrische Bodenproben:



Studie A: 10,000 Arten

Torsvik et al., *Environ. Microbiol.* 56, 782 (1990)

Studie B: >1,000,000 Arten

Ca. 99.9% der Arten sind selten

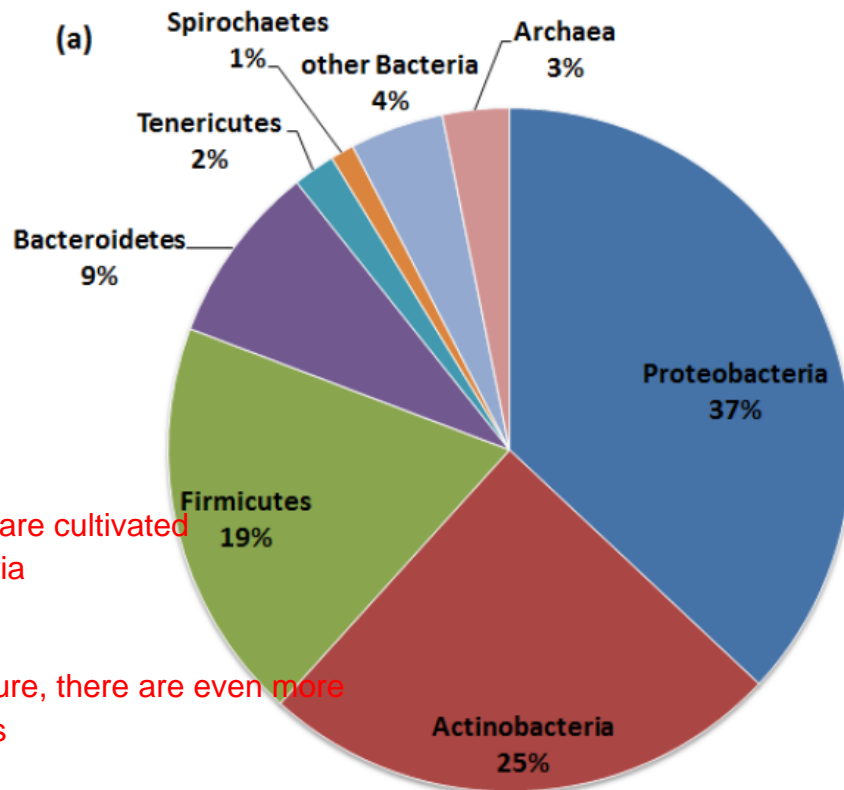
Gans et al., *Science* 309, 1387 (2005)

Einfache Diversitätsabschätzungen durch **Rarefaktions-Analyse**:

Probenanzahl gegen die Zahl der ermittelten “Arten” auftragen (für nichtkultivierte Bakterien verwendet man den Begriff **OTU**, **operational taxonomic unit**),

dann mathematische Ermittlung der Asymptote. Beispiel: MOTHUR software

Datenverzerrung durch Isolierung: Nur vier Phyla repräsentieren 88% aller isolierten Bakterien

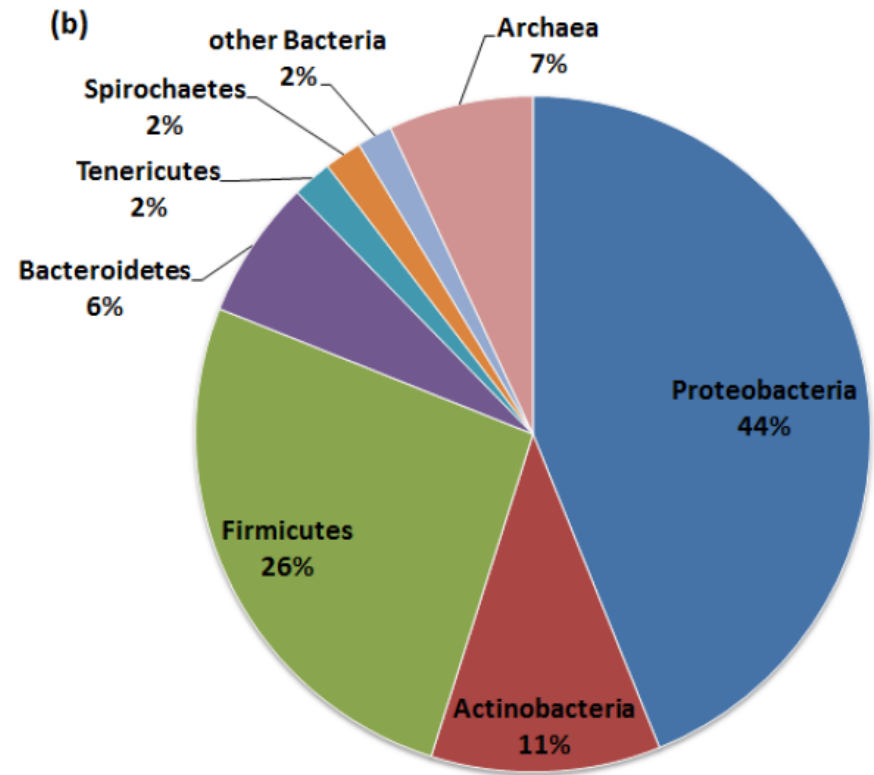


those are cultivated
bacteria

in nature, there are even more
phylas

there also ways to reconstruct non-cultivated phylas

Taxonomische Zusammensetzung
kultivierter Bakterien in der DSMZ
(Deutsche Sammlung von
Mikroorganismen und Zellkulture)



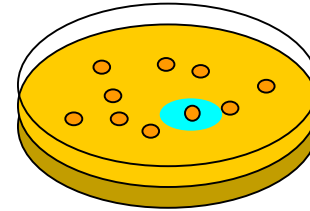
Taxonomische Zusammensetzung
sequenzierter Bakterien

Die Ökologie, Physiologie und das biotechnologische Potential nichtkultivierter (=fast aller) Bakterien ist so gut wie unbekannt und von grossem Interesse für die Forschung

Meta-omische Methoden zur Analyse und biotechnologischen Nutzung nicht kultivierter Bakterien

functional based metagenomics

1. Konstruktion und Screening metagenomischer DNA-Bibliotheken



sequence based metagenomics

2. Metagenomische DNA-Sequenzierung, Metatranskriptomik, Metaproteomik

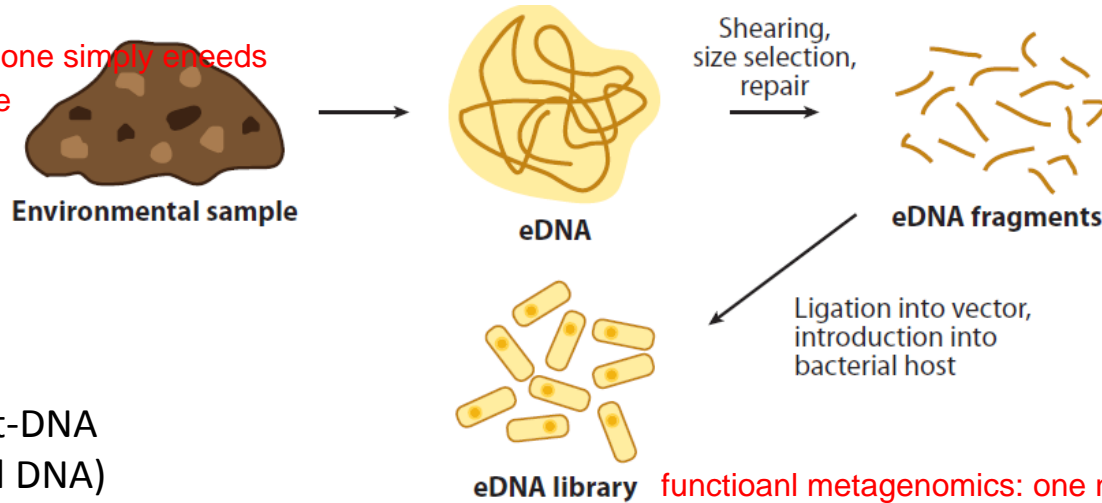
no DNA libraries needed

what kind of microorganisms occur, what genes, what genomes, what proteins are present (what genes expressed)?



Konstruktion und Screening Metagenomischer DNA-Bibliotheken

sample can be anything, one simply needs enough DNA to sequence

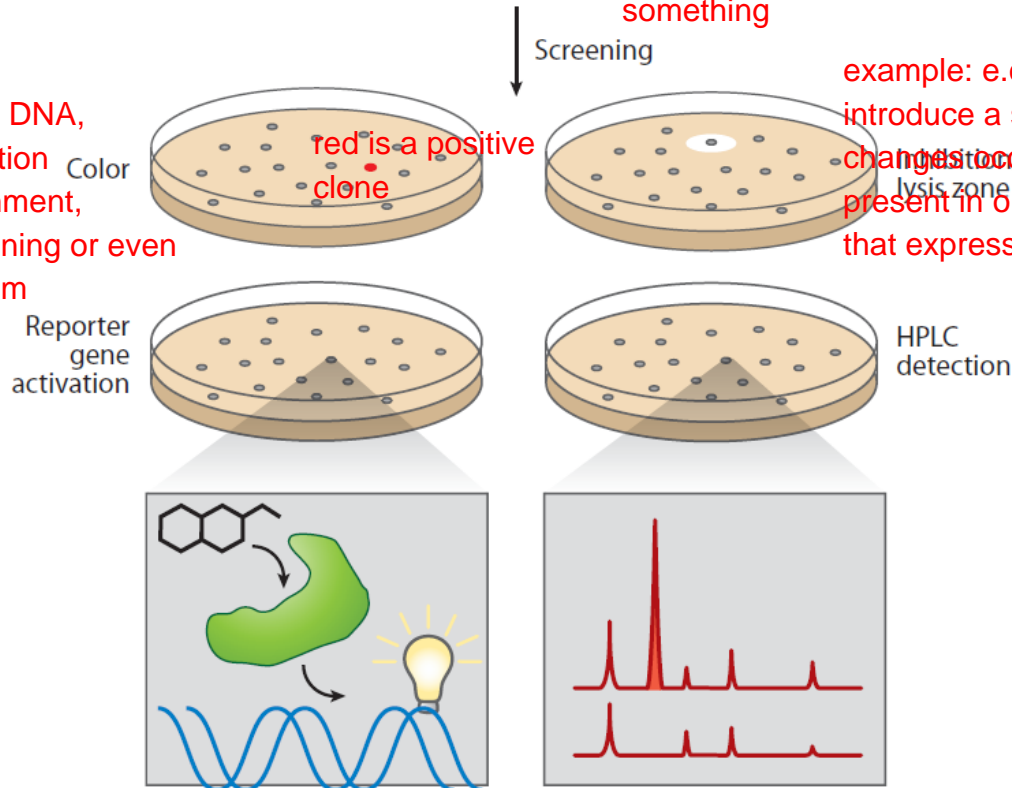


eDNA: Umwelt-DNA
(environmental DNA)

def. eDNA := environmental DNA, because that's DNA information that comes from the environment, without specifically mentioning or even knowing from which organism

functional metagenomics: one needs an assay to screen for something

example: e.coli grow on agar plates. introduce a substance such that color changes occur when a metabolic pathway is present in order to screen for the right e.coli's that express the eDNA



Wahl der Vektoren

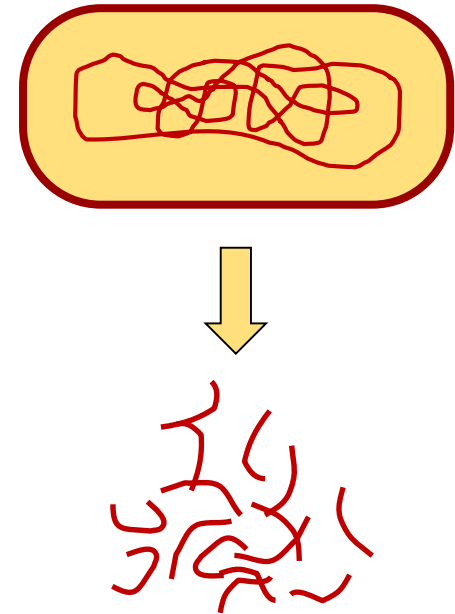
Genomgröße *E. coli*: 4.600.000 bp

Bibliothek aus 3 kb-Genomfragmenten:

>1500 Klone benötigt für 1x Abdeckung des Genoms

Bibliothek aus 40 kb-Genomfragmenten :

115 Klone benötigt



→ Für Metagenombibliotheken werden Vektorensysteme für grosse Inserts verwendet

Fosmid- und **Cosmidvektoren**: Insertgröße 35-45 kb

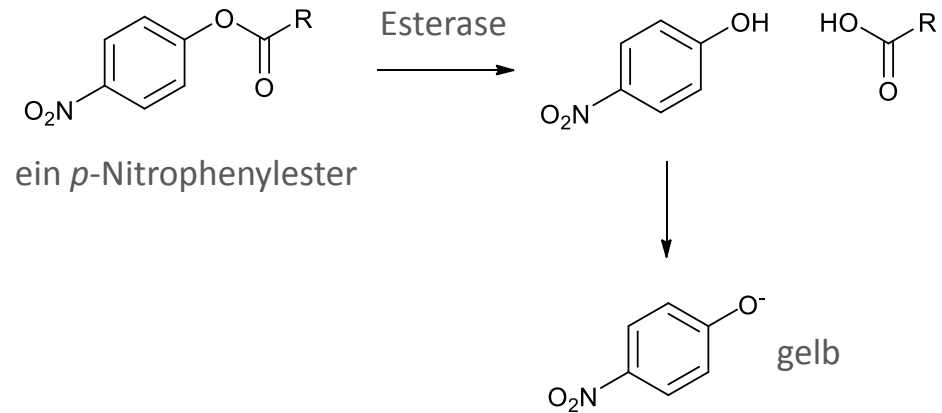
Fosmide haben eine Kopiezahl von 1 und sind daher oft stabiler als Cosmide mit höherer Kopiezahl.

BAC-(bacterial artificial chromosome) Vektoren: bis 350 kb möglich, aber Konstruktion grosser Bibliotheken schwieriger.

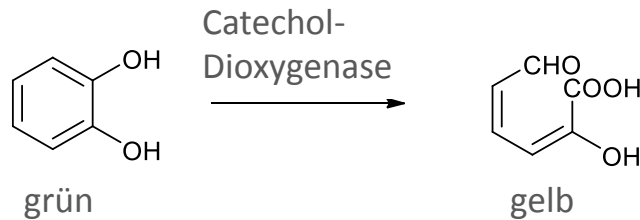
BAC for operons or lots of genes so that a specific function can be reached

Neue Enzyme durch Aktivitäts-Screening

Esterasen und Lipasen

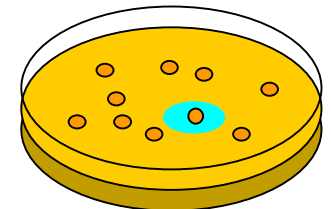


Catechol-Dioxygenasen



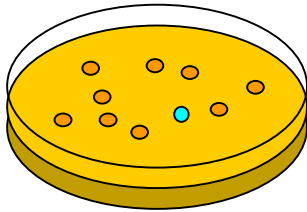
Einige weitere Enzyme aus Aktivitätsscreens:

Agarasen
Amidasen
Amylasen
Xylanasen
Hydroxybutyrat-Dehydrogenasen
Alcohol-Oxidoreductasen
Pectatlyasen, etc.

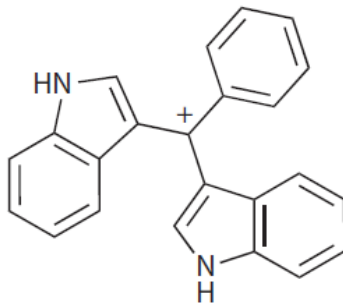
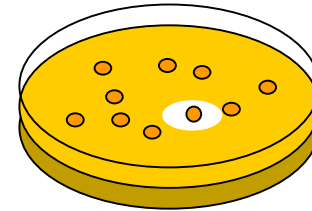


Neue Antibiotika durch Farb- oder Aktivitäts-Screening

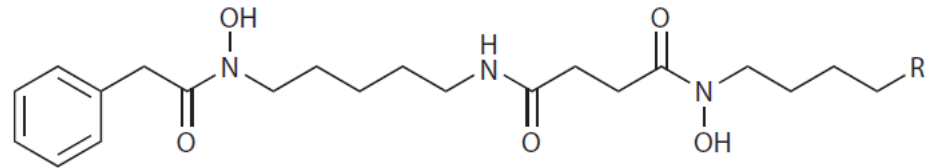
(chemische Strukturen sind nicht Examens-relevant)



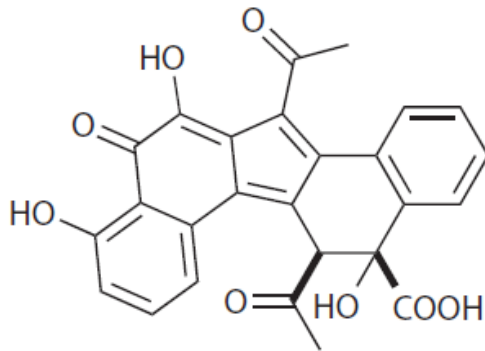
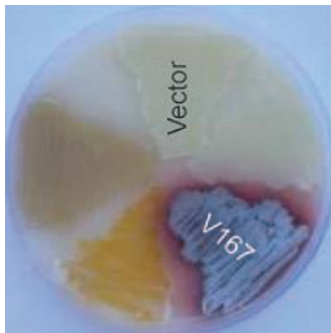
DNA-Quelle: Erde



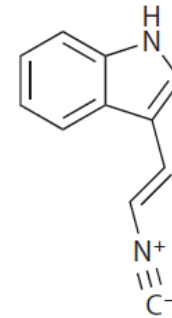
Turbomycin B
rot



Terragin A



Erdacin
rot



ein Isonitril

Nachteil phänotypischer Screens

Hits bei Screens sind viel seltener als mathematisch erwartet.

Gründe:

DNA stammt oft aus „exotischen“ Bakterien, die nur entfernt zu *E. coli* verwandt sind

Promotoren und Regulatoren werden nicht erkannt

Codon-Gebrauch ist anders als bei *E. coli* (seltene Codons)

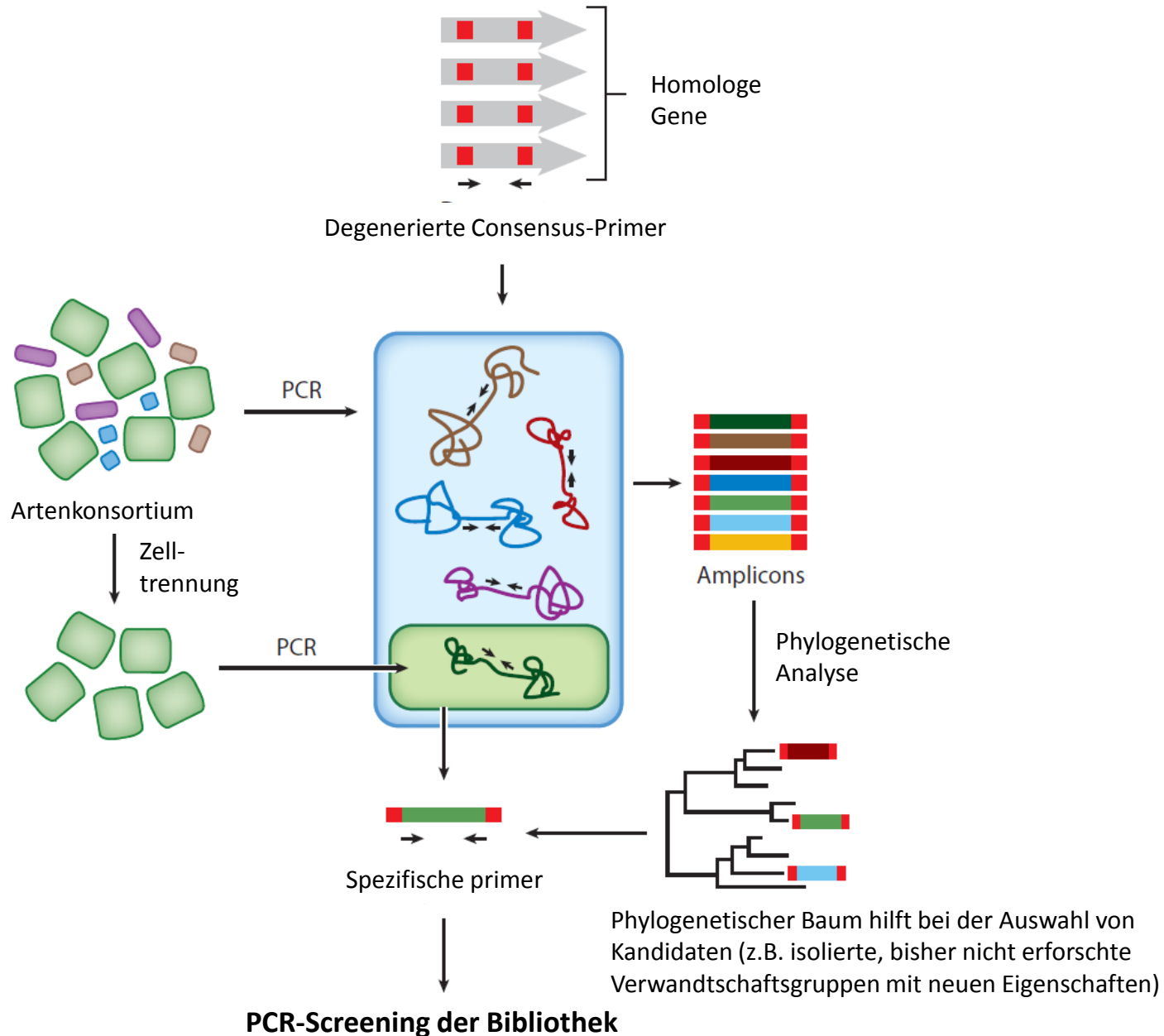
E. coli kann die Enzyme nicht falten oder posttranslational aktivieren

Mögliche Lösung:

Andere Wirtsbakterien verwenden (aber: hier oft niedrigere Klonzahlen erreichbar)

DNA-basierte statt funktionale Screens, dann gezielte Expression

Arbeitsablauf in Sequenz-basierten Screens



Vor- und Nachteile Sequenz-basierter Screens

(+) Gensequenz als Output

(+) Gene können gefunden werden, die

die nicht im bakteriellen Wirt exprimiert werden

für die kein Assay im Bibliotheksformat existiert

die neue Funktionen haben

(-) Zeitaufwändig

(-) Höhere Kosten

(-) Sequenz liefert oft keine Information über die Funktion (z.B. des Enzyms oder eine biosynthetischen Route)

2. Sequenzierung von Metagenomen

Environmental Genome Shotgun Sequencing of the Sargasso Sea

J. Craig Venter,^{1*} Karin Remington,¹ John F. Heidelberg,³
Aaron L. Halpern,² Doug Rusch,² Jonathan A. Eisen,³
Dongying Wu,³ Ian Paulsen,³ Karen E. Nelson,³ William Nelson,³
Derrick E. Fouts,³ Samuel Levy,² Anthony H. Knap,⁶
Michael W. Lomas,⁶ Ken Nealson,⁵ Owen White,³
Jeremy Peterson,³ Jeff Hoffman,¹ Rachel Parsons,⁶
Holly Baden-Tillson,¹ Cynthia Pfannkoch,¹ Yu-Hui Rogers,⁴
Hamilton O. Smith¹

Science 304 (2005) 66-74

1 Gb sequenziert

>1.2 Millionen Gene

148 neue "Arten"



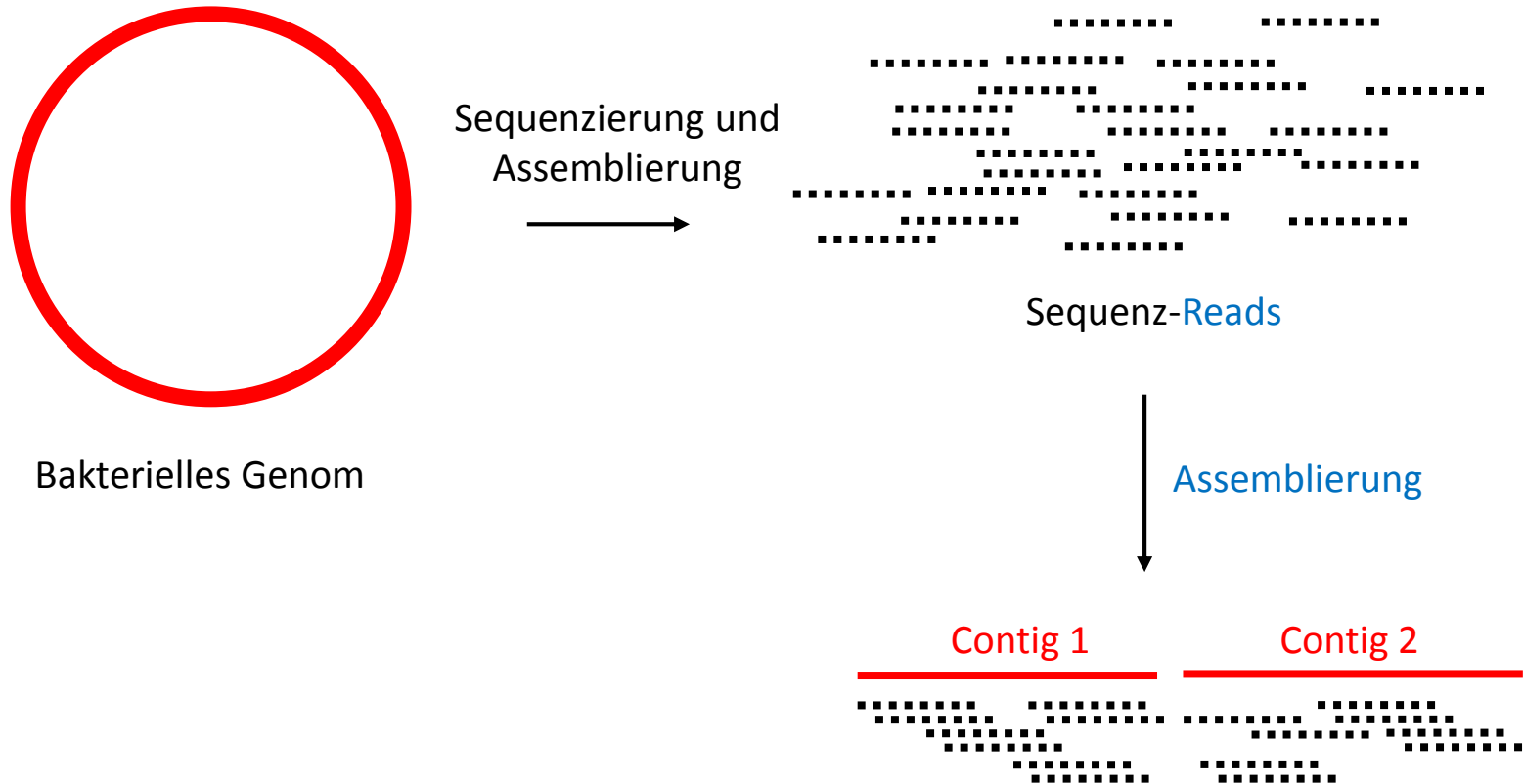
PLoS Biol. 5, e16 (2007)

1700 neue Proteinfamilien ohne bekannte Funktion

Grosses biotechnologisches Potential, aber es ist schwierig, Informationen über einzelne Organismen zu bekommen

Herausforderung Sequenzanalyse

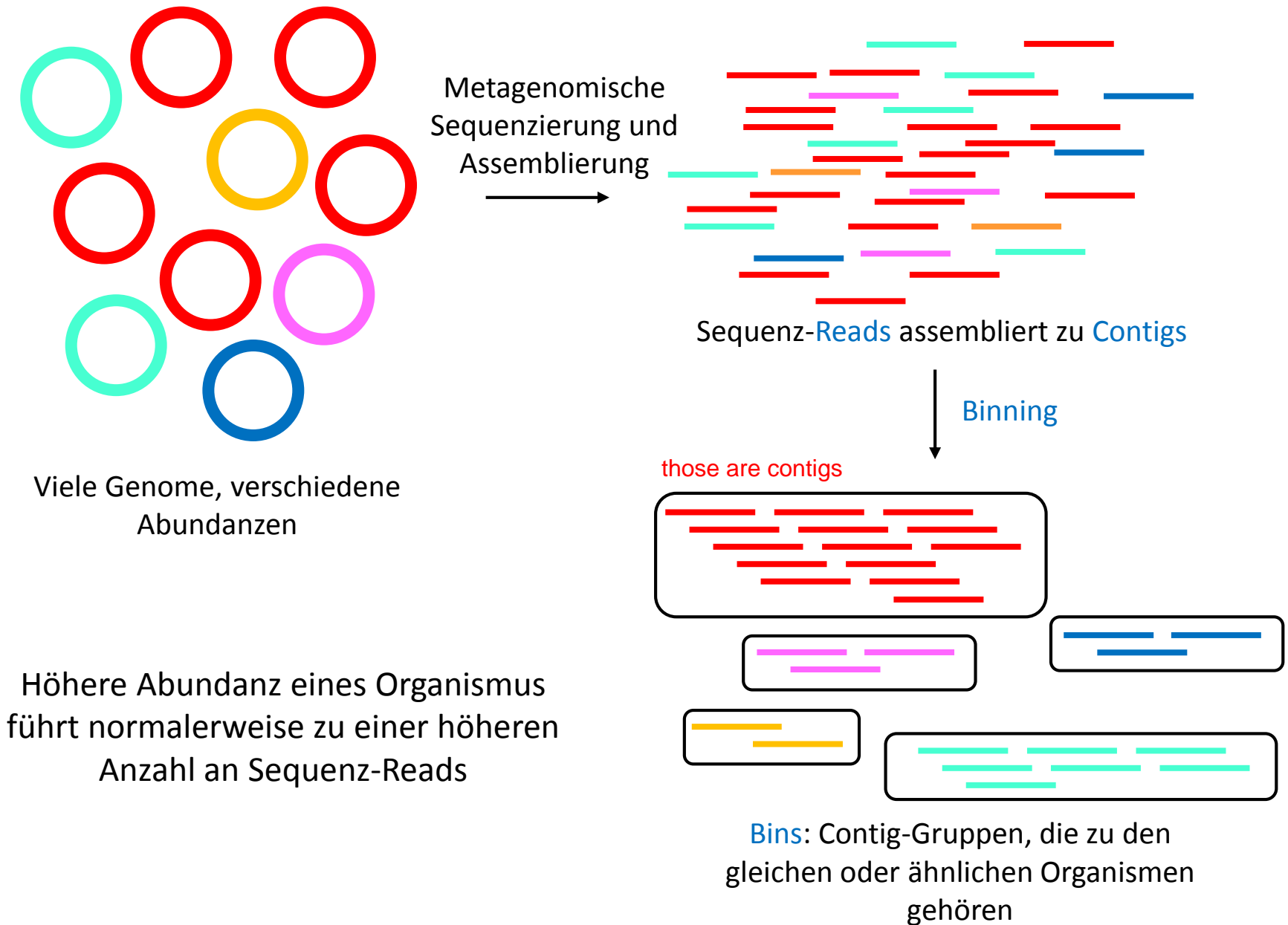
Sequenzierung einzelner Genome:



In der Metagenomik werden je nach Komplexität der Probe nur wenige abundante Genome relativ vollständig abgedeckt

Zuordnung der Contigs zu Organismen ist oft schwierig

Wie werden die Sequenzen Organismen zugeordnet?



Binningmethoden

Taxonomie-abhängig (auf Basis von Referenzgenomen):

Taxon der nächsten Sequenzhomologen
oder des letzten gemeinsamen Vorfahren

Kompositions-abhängig (ohne Referenzgenom):

Sequenzabdeckung

G+C-Gehalt

Häufigkeiten von Tetranucleotiden (z.B. GGAG vs. GGAC) oder längeren Nucleotidfolgen

u.a.

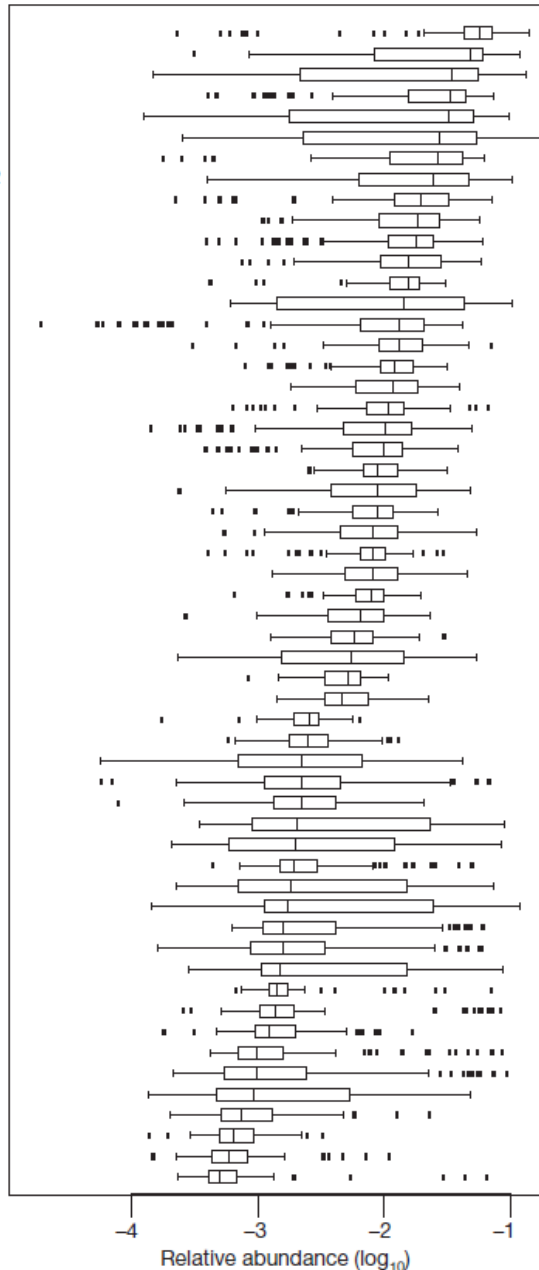
Visualisierung des Darm-Mikrobioms eines Neugeborenen



Unbinned					
				12, 13	<i>Actinomyces urogenitalis</i>
<i>Clostridium butyricum</i>	1			12	<i>Actinomyces urogenitalis</i> - Phage
<i>Leuconostoc</i> sp.	2			14	<i>Escherichia coli</i> - Strain B
<i>Veillonella</i> sp. - Species A	3, 16			14	<i>Enterococcus faecalis</i> - Plasmid
<i>Enterococcus faecalis</i>	4			14	<i>Escherichia coli</i> - Strain A/B - Plasmid
<i>Escherichia coli</i> - Strain A	5, 21			15	<i>Propionibacterium</i> sp.
<i>Escherichia coli</i> - Strain A - Phage	5			16	<i>Veillonella</i> sp. - Species A - Plasmid
<i>Staphylococcus</i> sp.	6, 22			22	<i>Veillonella</i> sp. - Species A - Phage
<i>Staphylococcus</i> sp. - Plasmid A	unbinned			17, 25	<i>Negativicoccus succinicivorans</i>
<i>Staphylococcus</i> sp. - Plasmid B	unbinned			18, 19	<i>Veillonella</i> sp. - Species B
<i>Staphylococcus</i> sp. - Phage	unbinned			23	<i>Veillonella</i> sp. - Species B - Phage
<i>Streptococcus anginosus</i>	7			20	<i>Varibaculum cambriense</i>
<i>Clostridium bartlettii</i>	8, 9			23, 24	<i>Streptococcus parasanguinis</i>
<i>Streptococcus</i> sp.	10			23	<i>Streptococcus parasanguinis</i> - Plasmid
<i>Veillonella dispar</i>	11				

Beispiel für grosse metagenomische Datensätze

Bacteroides uniformis
Alistipes putredinis
Parabacteroides merdae
Dorea longicatena
Ruminococcus bromii L2-63
Bacteroides caccae
Clostridium sp. SS2-1
Bacteroides thetaiotaomicron VPI-5482
Eubacterium hallii
Ruminococcus torques L2-14
Unknown sp. SS3 4
Ruminococcus sp. SR1 5
Faecalibacterium prausnitzii SL3 3
Ruminococcus lactaris
Collinsella aerofaciens
Dorea formicigenerans
Bacteroides vulgatus ATCC 8482
Roseburia intestinalis M50 1
Bacteroides sp. 2_1_7
Eubacterium siraeum 70 3
Parabacteroides distasonis ATCC 8503
Bacteroides sp. 9_1_42FAA
Bacteroides ovatus
Bacteroides sp. 4_3_47FAA
Bacteroides sp. 2_2_4
Eubacterium rectale M104 1
Bacterioides xylanisolvens XB1A
Coprococcus comes SL7 1
Bacteroides sp. D1
Bacteroides sp. D4
Eubacterium ventriosum
Bacteroides dorei
Ruminococcus obeum A2-162
Subdoligranulum variable
Bacteroides capillosus
Streptococcus thermophilus LMD-9
Clostridium leptum
Holdemania filiformis
Bacteroides stercoris
Coprococcus eutactus
Clostridium sp. M62 1
Bacteroides eggertii
Butyrivibrio crossotus
Bacteroides finegoldii
Parabacteroides johnsonii
Clostridium sp. L2-50
Clostridium nexile
Bacteroides pectinophilus
Anaerotruncus colihominis
Ruminococcus gnavus
Bacteroides intestinalis
Bacteroides fragilis 3_1_12
Clostridium asparagiforme
Enterococcus faecalis TX0104
Clostridium scindens
Blautia hansenii



Human gut microbiomes of 124 individuals

577 Gb sequenced

>1000 total bacterial species, but only at least 160 species per individual (per human)

150x more genes than in the human complement

How to interpret such data? The mere presence of a gene does not mean that it is of significance.

Importance to conduct additional studies on expression ([metatranscriptomics](#), [metaproteomics](#)) and function (biochemical assays)

Zusammenfassung

Wir wissen nur sehr wenig über Bakterien, da nur ein Bruchteil der Organismen bisher kultiviert wurde.

Die 16S rRNA Gen-Analyse liefert Informationen über die taxonomische Diversität in Umweltproben. Die Diversität kann durch Rarefaktionsanalyse abgeschätzt werden.

Metagenomische Methoden sind Library-basiert oder beruhen auf der Sequenzierung der Gesamt-DNA.

Phänotypisches Screening von Libraries liefert schnell funktionale Daten. Hier gilt: you get what you search for.

Die metagenomische Sequenzierung liefert eine Fülle von Sequenzdaten, aber keine funktionalen Informationen (nur Hypothesen). Sequenzen sind normalerweise hoch fragmentiert.

Contigs können Organismen durch Binning zugeordnet werden.