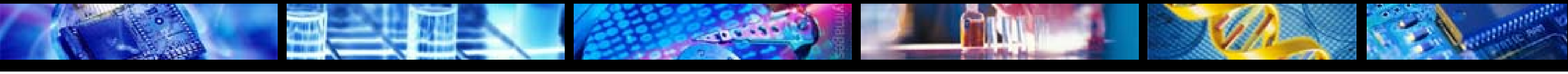
Computational Biology LU 2014

Microbiome Analysis

Bioinformatics Group
Institute for Knowledge Discovery
Graz University of Technology
http://genome.tugraz.at
Petersgasse 14, A-8010 Graz

Outline

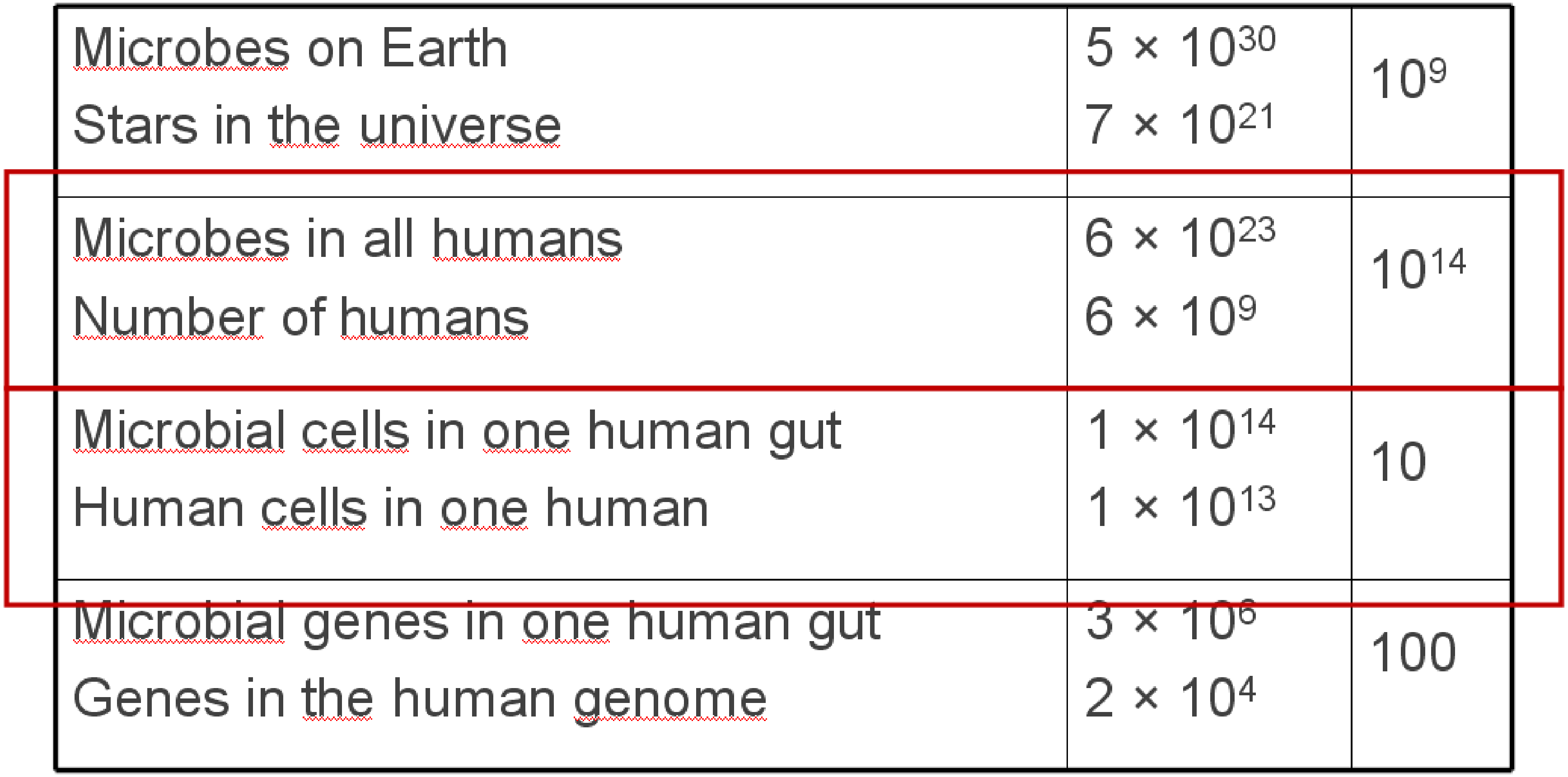
- Microbiome
  - Definition, facts, occurrence
  - Analysis
- Exercise



Microbiome

- A microbiome is the totality of microbes, their genetic elements (genomes), and environmental interactions in a defined environment.
- Microbes: viruses, bacteria, fungi, algae
- Environment: human inter toe space, soil probe, table, your hand, …

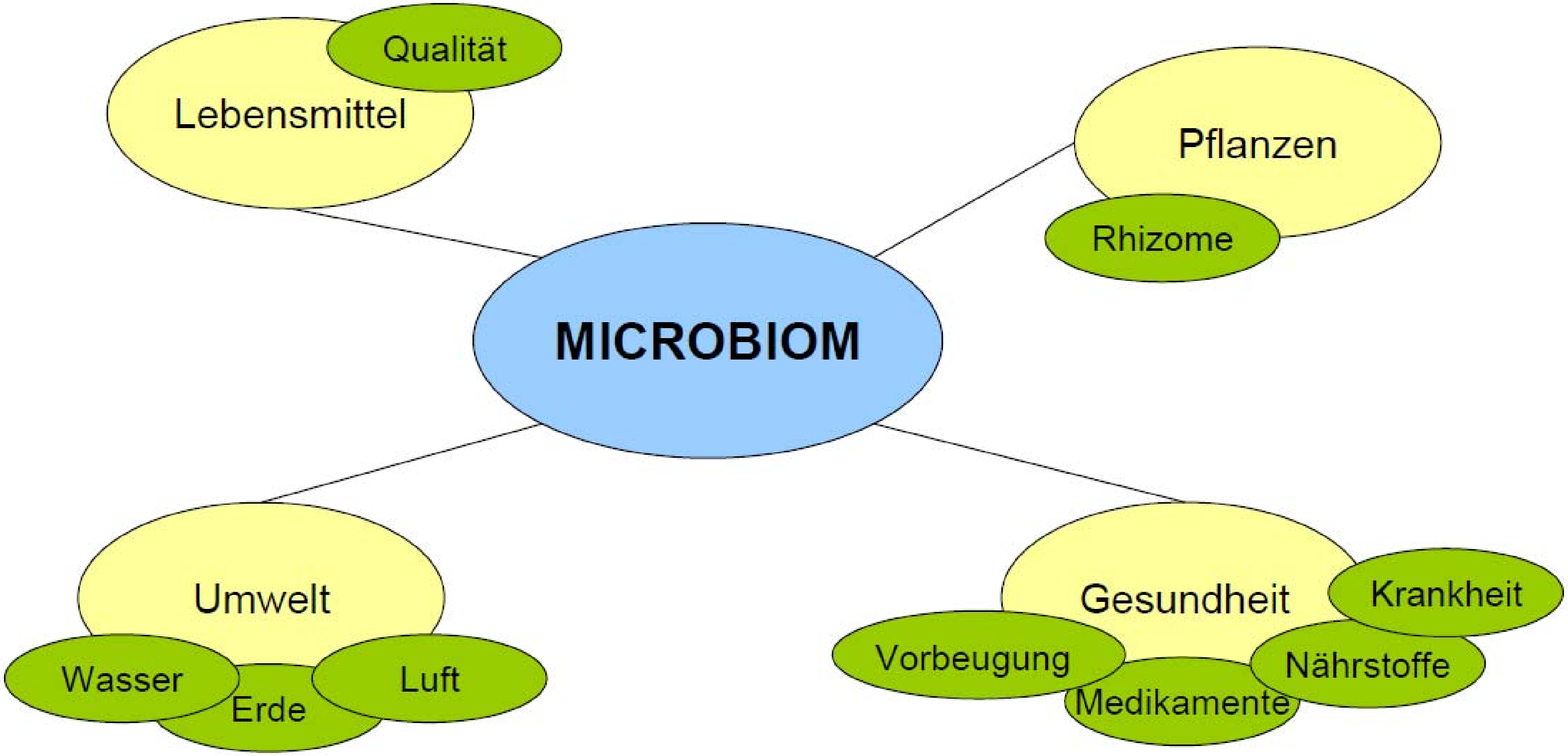
Microbiome – interesting facts



Microbes on Earth	5×10^{30}	10^{9}
Stars in the universe	7×10^{21}	
Microbes in all humans	6×10^{23}	10^{14}
Number of humans	6×10^{9}	
Microbial cells in one human gut	1×10^{14}	10
Human cells in one human	1×10^{13}	
Microbial genes in one human gut	3×10^{6}	100
Genes in the human genome	2×10^{4}	

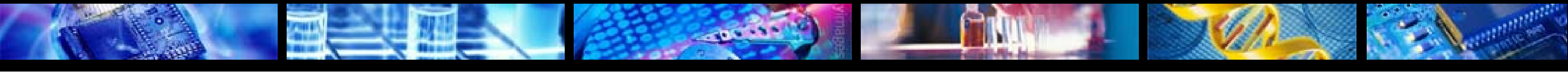
Kyrpides NC. *Nat Biotechnol.* 2009; 27(7): 627-632

Where can it be found



Microbiome - Analysis

- Phylotype generation and classification
- Alpha and beta diversity
- Differential abundance (taxa)

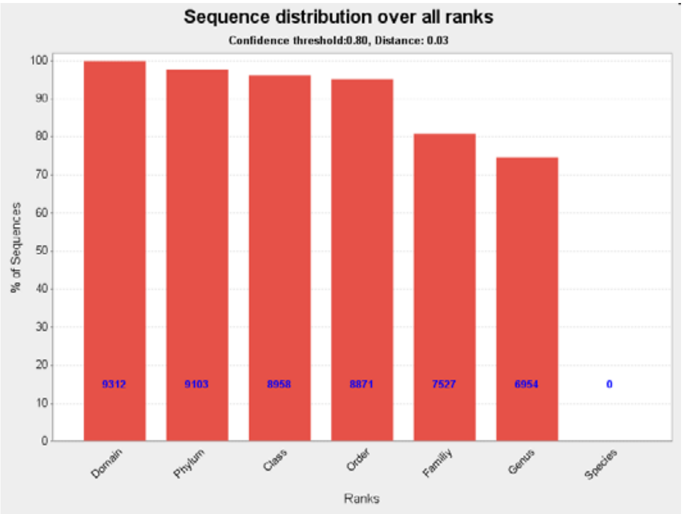


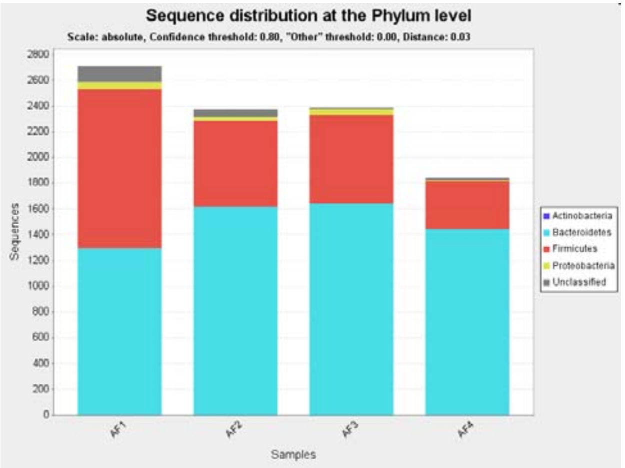
Phlyotype generation

- Multiple sequence alignment
- Distance calculation
- Clustering at a certain similarity (97% similarity ≡ 3% maximum sequence distance)
  For 16S sequences this represents a species (by convention)
  This is called a Phylotype or OTU (Operational taxonomic Unit)
- Extraction of representative sequences

Classification

Representative sequences of a phylotype are classified taxonomically. A full classification covers the following ranks: Domain, Phylum, Class, Order, Family, Genus, Species

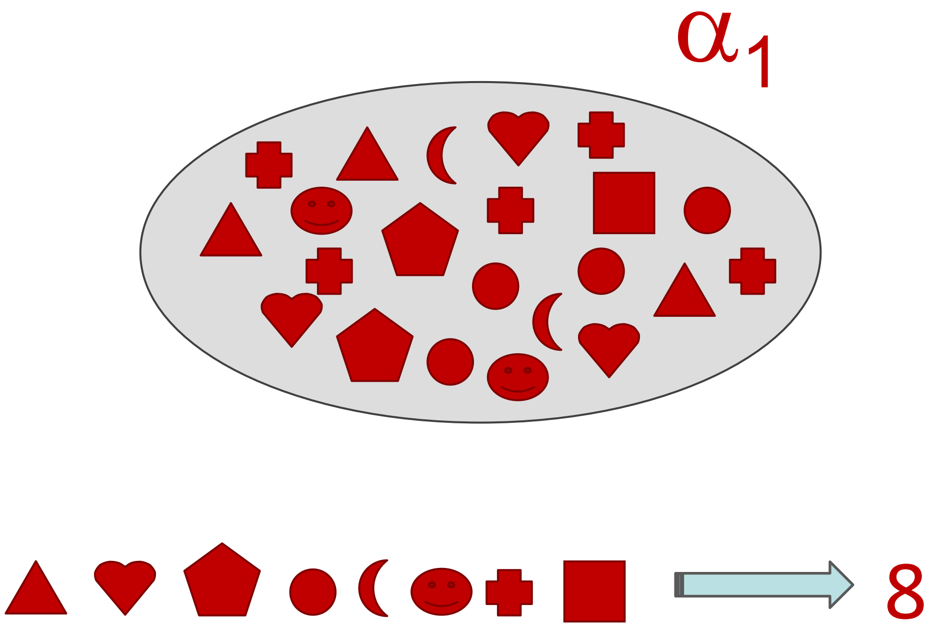




Absolute view of sequence distribution at the rank Phylum of sample A at different time points.

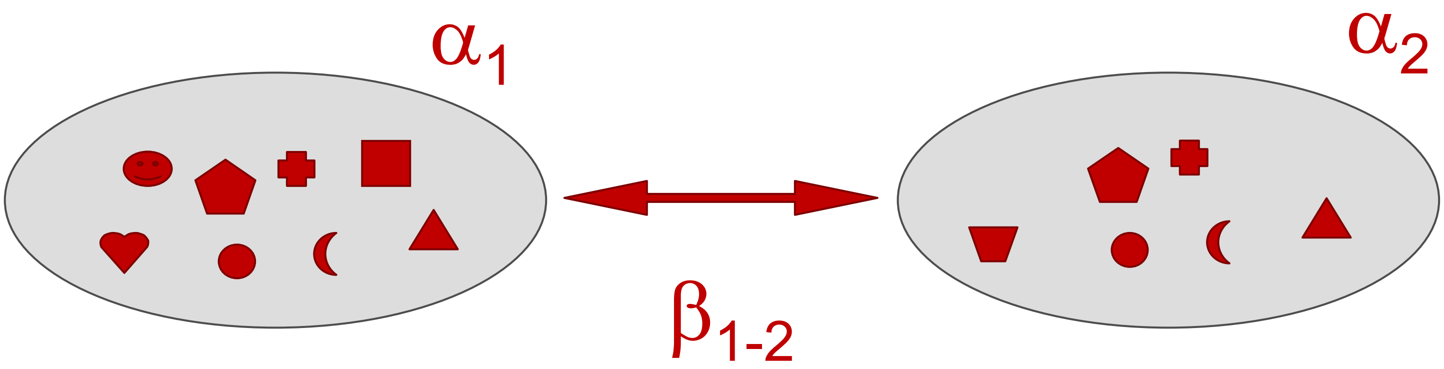
Alpha and beta diversity

- α-diversity
  - species richness
  - characterizes the number of species at a particular time point, area ...



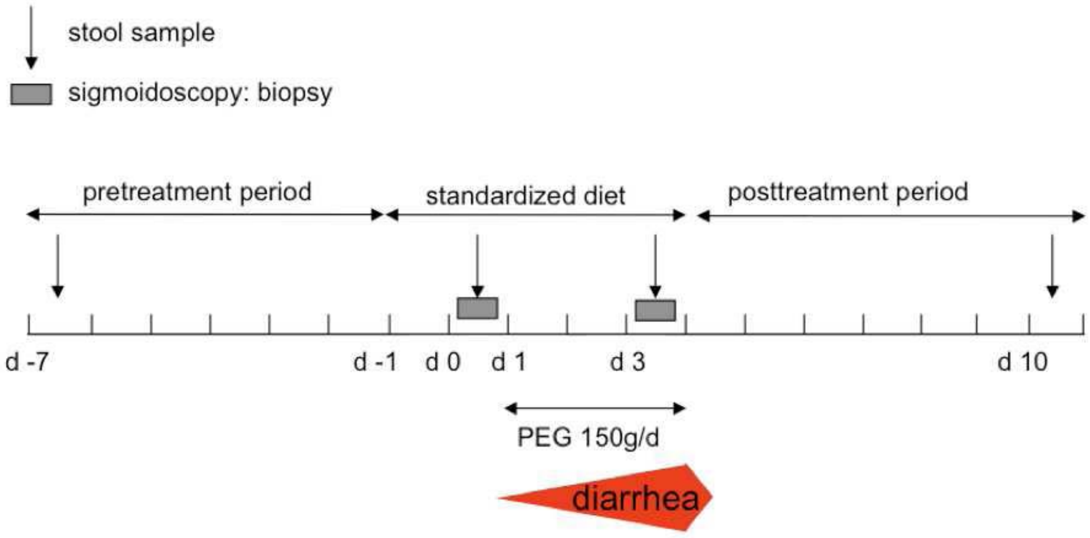
Alpha and beta diversity

- β-diversity
  - describes similarity or dissimilarity between two samples (species turn-over)
  - calculated index (Anderberg coefficient, Bray-Curtis index)



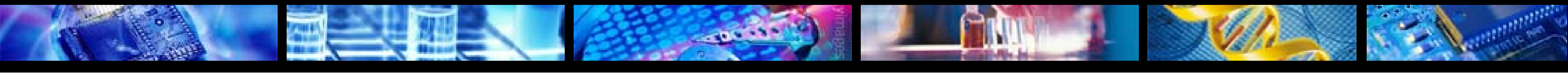
Exercise - Microbiome Analysis

- Data from an induced diarrhea study.
- Probes taken from four patients at four different time points



Exercise - Microbiome Analysis

- DNA was extracted, 16S rDNA variable region 1-2 has been amplified and sequenced with Roche/454 technology
- Basic analysis of 16S sequence data was performed with SnoWMAn (http://SnoWMAn.genome.tugraz.at): Sequences were clustered with 97% similarity and the resulting OTUs (operational taxonomic units) are taxonomically classified
- Result is a data matrix consisting of rows representing OTUs. Columns contain read counts per sample and taxonomic classifications
(described in `SnoWMAn Summary.txt`)



Exercise – R hints

- Install packages from bioconductor:
  edgeR
  Use: `biocLite()`
- Install CRAN packages:
  vegan, biom, BiodiversityR
  Use: `install.packages()`
- Do not forget to load the packages before working with them: library("…")

Exercise – useful functions

- read.table(), write.table() – read/write tab del. files
- as.matrix() – converts data to matrix form
- barchart() – creates are barchart
- aggregate() – calculates sums on subsets
- estimateR(), diversityresult() – calculate alpha diversity
- DGEList(), calcNormFactors(), estimateCommonDisp(), estimateTagwiseDisp(), exactTest(), topTags(et) – for calculation of differential abundance.