Bioinformatik

Stefanie Hartmann

Wintersemester 2019 / 2020, Universität Potsdam

Metagenomics

Dec 13, 2019

Today's topics

- metagenomics
  - introduction
  - research questions & approaches
    - who is there?
      - phylotyping
      - OTU-based processing
    - what are they doing?
  - case studies: gut microbiome
    - inventory
    - relation to health
    - therapy
  - computer exercise
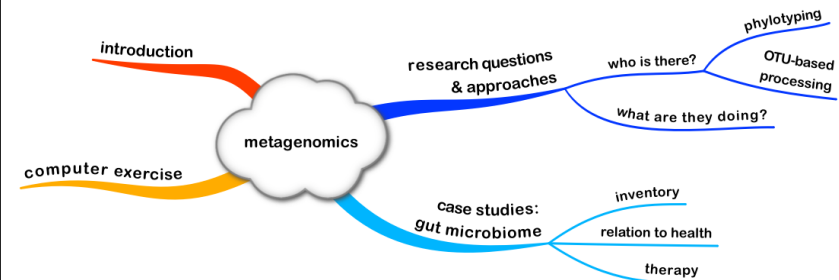
reading:
- **AD Tyler et al. Analyzing the human microbiome: A "how to" guide for physicians. Am J Gastroenterol. 2014**
  (Introduction; Library preparation; Sequence identification; The human microbiome -- lessons from community-level analysis)

Microbes run the world

microbes
- living things invisible to the human eye
- Bacteria, Archaea, some Eukaryotes (fungi, algae, protozoa)

  - convert key elements into accessible forms
  - some can live in extreme environments (temp., pH, etc)
  - are of commercial & medical relevance
  - interact with plants, fish, livestock, humans,…
  - make nutrients, metals, vitamins available to their hosts
  - help us to digest food, break down toxins

Studying microbes: culture-dependent

culturing
- grow microbes in isolation

analysis methods
- microscopy, biochemical tests, staining

challenges
- appearance → taxonomy?
- isolated microbes under lab conditions
- many uncultured, unculturable!

Cultures?

not free-living
- obligate intracellular symbionts of plants & animals
- dependence on host or co-symbionts

overwhelming abundance
- deep sampling to obtain less abundant microbes
- human gut microbes:
  millions of colonies would have to be picked!

growth conditions unknown or difficult
- environment? nutrients? competition?

Studying microbes: culture-independent

next-generation sequencing
- very sensitive, tiny amounts of DNA are sufficient
- allows to sequence the DNA in a sample
  - without culturing
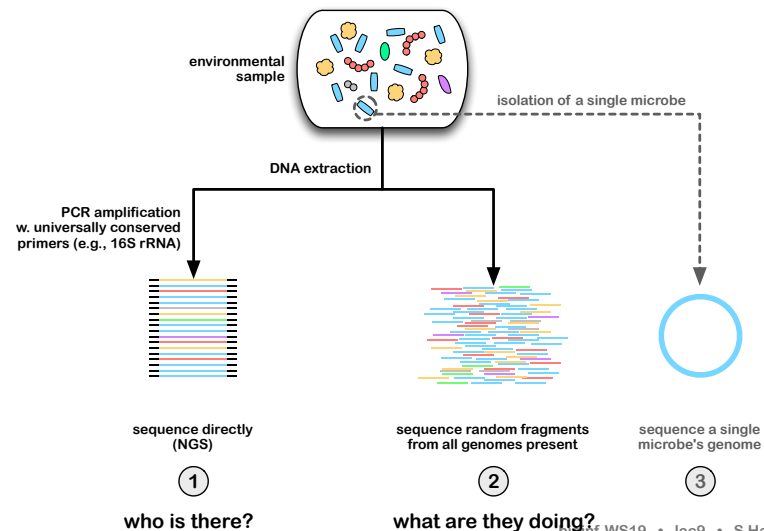 - without knowing who the DNA belongs to

microbiome
- all microbes in a given habitat

metagenome
- collection of all genomes of a microbiome

Studying microbes: culture-independent

environmental sample

isolation of a single microbe

DNA extraction

PCR amplification w. universally conserved primers (e.g., 16S rRNA)

sequence directly (NGS) ① who is there?

sequence random fragments from all genomes present ② what are they doing?

sequence a single microbe's genome ③

Prokaryotic taxonomy

prokaryotic species definition
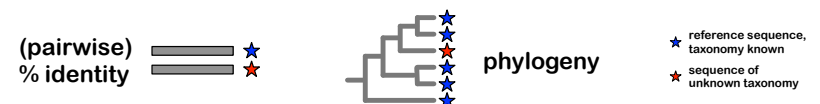- based on phenotype, chemotaxonomy, genetic data

genetic data
- genome-wide average nucleotide identity (> ~95%)
- 16S rRNA sequence identity (>= 97%)

species identification: compare to reference (phylotyping)
- 16S rRNA sequence identity
- 16S rRNA phylogentic analysis

(pairwise) % identity

phylogeny

reference sequence, taxonomy known
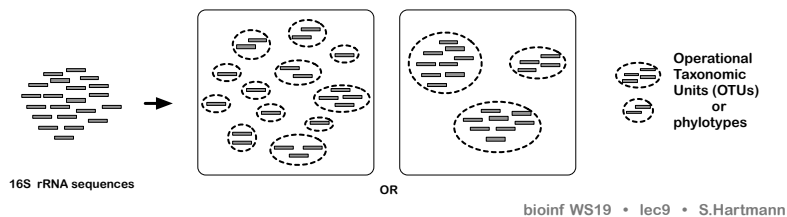
sequence of unknown taxonomy

Prokaryotic taxonomy

available genetic data

- type strains in culture collections: ~11,000
- prokaryotic genome projects: ~25,000
- non-redundant 16S rRNA sequences: ~500,000
- number of bacterial species: tens of millions?

"species" identification independent of reference data

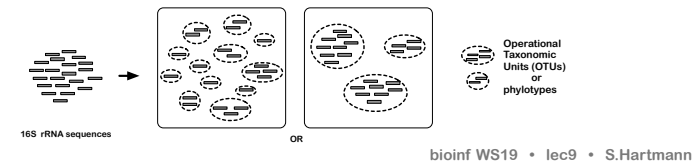- grouping of data based on sequence identity

16S rRNA sequences → Operational Taxonomic Units (OTUs) or phylotypes

OR

---

① who is there?

A. classify: compare to reference data (phylotyping)

(pairwise) % identity

phylogeny

★ reference sequence, taxonomy known
★ sequence of unknown taxonomy

B. cluster: compare to sequences in the same data set (OTU-based processing)

16S rRNA sequences → Operational Taxonomic Units (OTUs) or phylotypes

OR

---

② what are they doing?

- mixture of DNA from different taxa; different abundance, relatedness
- composite genome for each strain
- fewer (even single) reads per region
- shorter assembled fragments
- huge data volumes
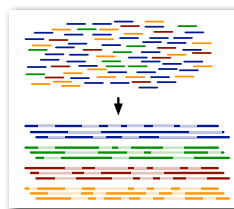
for each (composite) genome (or the entire data set)

- predict location & sequence of genes
- predict gene function
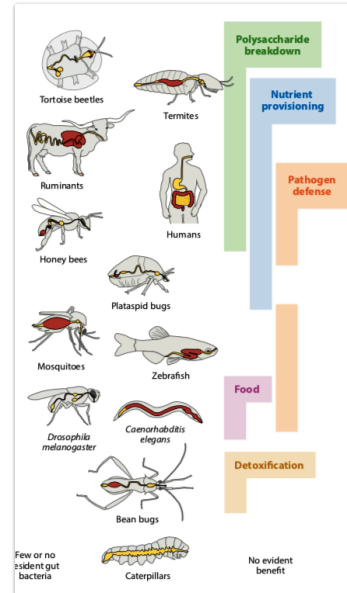- predict metabolic capabilities

---

② what are they doing?
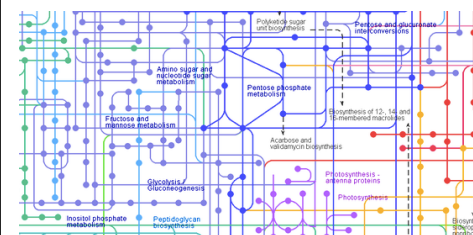
information about metabolic pathways of (prokaryotic) organisms

- sequenced metagenomes
  - genes for which enzymes are present?
  - which metabolic pathways are possible?

ocean water, soil, permafrost, homes, university campus, ...

ATCCGGACTAGC

Polysaccharide breakdown

Nutrient provisioning

Pathogen defense

Food

Detoxification

Tortoise beetles
Termites
Ruminants
Honey bees
Humans
Plataspid bugs
Mosquitoes
Zebrafish
Drosophila melanogaster
Caenorhabditis elegans
Bean bugs
Caterpillars

Few or no resident gut bacteria

No evident benefit

The human microbiome

Nasal
Oral
Skin
Gastro-intestinal
Urogenital

Human Microbiome Project
http://commonfund.nih.gov/hmp/

- the surface and insides of our bodies are colonized by bacteria
- colon > skin > rest
- some bacteria cause infections
- many bacteria benefit us
  - help digest our food
  - produce vitamins
  - regulate our immune system
  - protect against disease-causing bacteria

Gastrointestinal (GI, gut) microbiome

Esophagus
Jejunum
Ileum
Stomach
Duodenum
Colon
Colon
Appendix
Rectum

stomach

colon

epithelium   feces

- increase in abundance & diversity of microbes
- up to ~10^{11} cells per gram

---
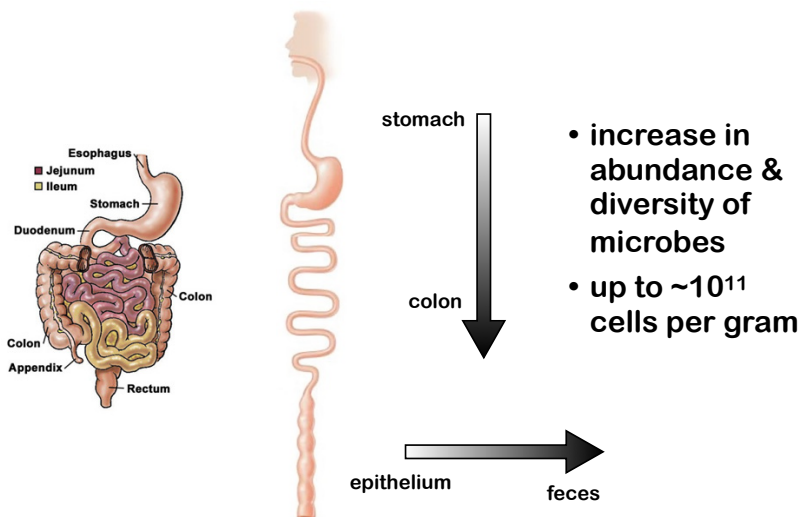
Gastrointestinal (gut) microbiome

- fecal samples or endoscopic biopsies
- high-throughput sequencing
  - 16S rRNA and/or shotgun approach
- questions
  - who is there, and what are they doing?
  - how is that related to disease & health?
  - microbiome therapeutics?
- highly interdisciplinary work
  - microbial ecologists, clinicians, immunologists, physiologists, molecular biologists, nutritionists, bioinformaticists, and many more

Snapshot: adults from Europe, Japan, US
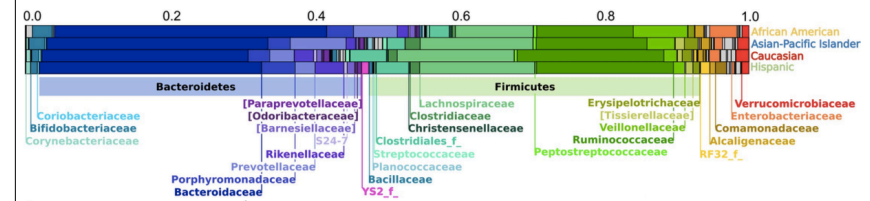
- dominant phyla: Bacteroidetes, Firmicutes, Actinobacteria, Proteobacteria, Verrucomicrobia
- great variation between individuals
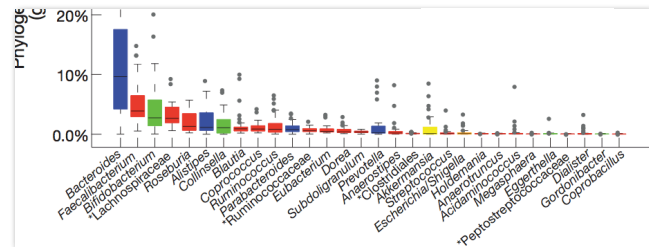
genus abundance variation box plot for the 30 most abundant genera;
35 adults from Europe, Japan, US; Arumugam et al., Nature 2011

Diversity across ethnicities in the US

AW Brooks, PLOS. 2018

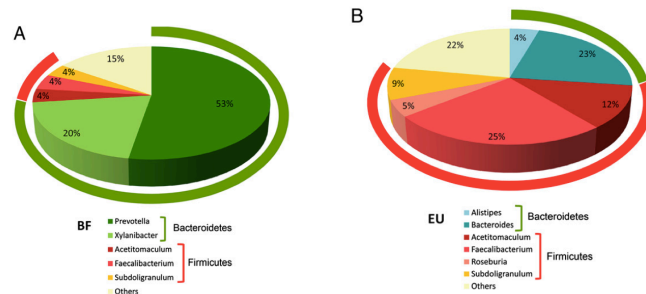- 1,673 individuals, self-reported ethnicity
- socioeconomic, cultural, geographic, dietary, genetic diversity
- difference in microbiota composition: related to difference metabolism & health?
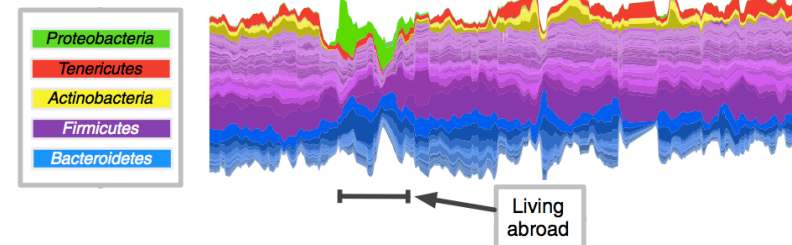
Snapshot: impact of diet

predominantly vegetarian diet: starch, fiber, plant polysaccharides (rural African village)

typical western diet: high in animal protein, sugar, starch, fat; low in fiber (Florence, Italy)

- data from 29 healthy children, aged 1-6 years
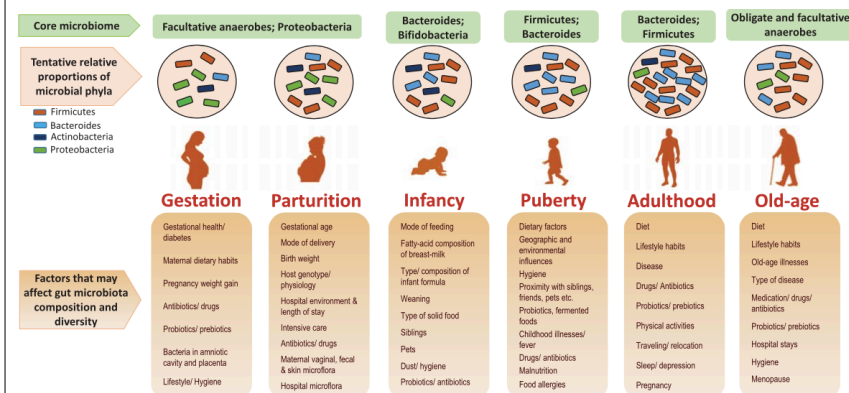- CD Filippo et al., PNAS 2010

Time series: 1 person, 1 year

- healthy male, daily sample collection for one year
- also recorded: diet, exercise, sleep, illness, etc
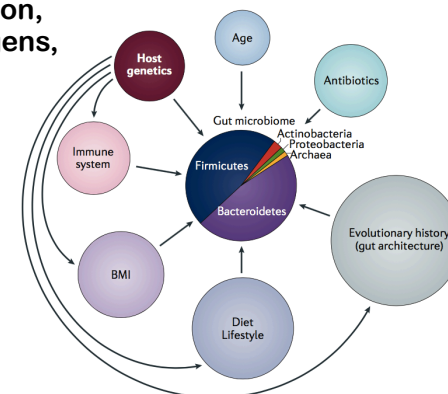- LA David et al., Genome Biology 2014

Age-related changes, factors

Major conclusions so far

- great variation between taxonomic composition
- a core microbiome is essential for our health
- affected by sanitation, exposure to pathogens, antibiotics, other medications, environmental factors, host genetics, diet, …

Functional analysis of sequence data

- required housekeeping functions: information processing, main metabolic pathways
- biodegradation & fermentation of polysaccharides indigestible by the host
- synthesis of essential vitamins
- metabolism of foreign compounds (drugs, food supplements)
- pathogen protection (stimulate immune system and/or kill or inhibit invading pathogenic species)

➤ • involved in host intestinal function, metabolism, physiology, immune system, organ development, cell proliferation, bone mass, adiposity, behavior, …

Examples of gut bacteria, metabolites, functions

Metabolites	Related bacteria	Biological functions
Short-chain fatty acids (e.g. acetate, propionate, butyrate)	*Firmicutes,* including species of *Eubacterium, Roseburia, Faecalibacterium,* and *Coprococcus*	inhibit the growth of pathogens; provide energy to the colonic epithelial cells, implicated in human obesity, insulin resistance and type 2 diabetes, colorectal cancer.
Bile acids	*Lactobacillus, Bifidobacteria, Enterobacter, Bacteroides, Clostridium*	absorb dietary fats, facilitate lipid absorption maintain intestinal barrier function, signal systemic endocrine functions to regulate triglycerides, cholesterol, glucose and energy homeostasis.
Vitamins (K, B12, biotin, folate, thiamine, riboflavin)	*Bifidobacterium*	provide complementary endogenous sources of vitamins, strengthen immune function, exert epigenetic effects to regulate cell proliferation.

excerpt from JK Nicholson et al., Science 2012

Alterations in gut microbiome (dysbiose)

diseases associated with GI-tract

- irritable bowel syndrome (IBD)
(e.g., Crohn's disease, ulcerative colitis)
- antibiotic-associated diarrhea

diseases not associated with the GI-tract

- allergy
- obesity
- diabetes
- rheumatoid arthritis
- colorectal cancer
- heart disease
- depression
- and many more

Positive effects on microbiome & health?

recolonization

- appendix
  - a safe-house for "good" bacteria;
    re-inoculation after gut infection & purging (cholera)
- bacteriotheraphy
  - transfaunation / faecal microbiota transplant to
    re-establish normal intestinal flora
  - common practice in veterinary medicine, successful
    in humans after *Clostridium difficile* infection

diet

- probiotics
- prebiotics

Prebiotics, probiotics

prebiotics: food ingredients that

- are resistant to gastric acidity,
  hydrolysis by host enzymes, or
  absorption by the upper
  gastrointestine
- are fermented by the gut microbiota
- stimulate growth or activity of of
  microbial species beneficial to the
  host's health

probiotics

- live microorganisms that, when
  administered in adequate amounts,
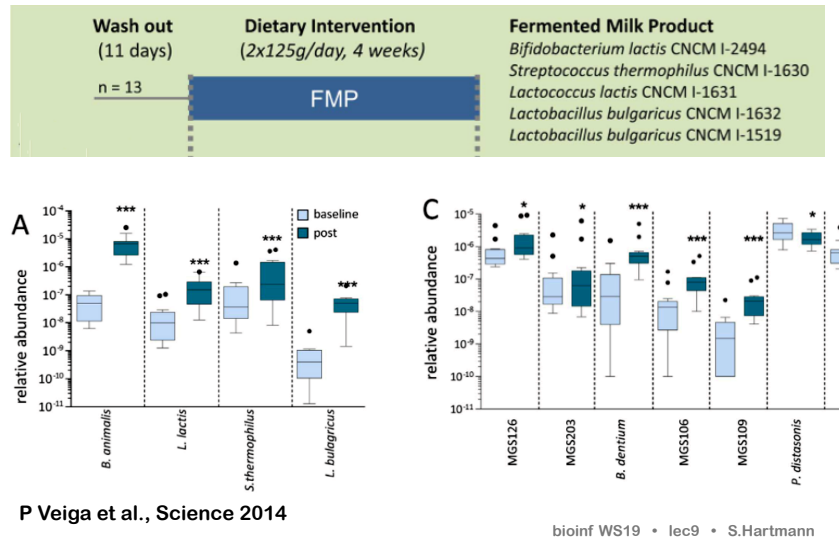 confer a health benefit on the
  consumer

Probiotics

fermented foods

- sauerkraut, kimchi, kefir, dry fermented sausage,
  yogurt, cheese, kombucha, miso, …
- if not further processed (heat, acid, etc),
  might contain 10^6-10^9 viable cells per g/ml
- many survive passage through the GI tract
  - composition of microbiota in GI tract changed
  - transient effect
  - magnitude?
  - importance?

Prebiotics, probiotics

Wash out (11 days)	Dietary Intervention (2x125g/day, 4 weeks)	Fermented Milk Product
n = 13	FMP	*Bifidobacterium lactis* CNCM I-2494 *Streptococcus thermophilus* CNCM I-1630 *Lactococcus lactis* CNCM I-1631 *Lactobacillus bulgaricus* CNCM I-1632 *Lactobacillus bulgaricus* CNCM I-1519

P Veiga et al., Science 2014

Microbiome therapeutics

Irritable Bowel Syndrome (IBD)

- functional disorder of the intestine
- host genetic factors implicated
- associated with altered gut microbiome
  - ↓ *Firmicutes*, *Bacteroidetes*
  - ↓ beneficial (butryrate, vitamin producing) microbes
  - ↑ proteobacteria (pathogens)
  - ↑ fungi
  - - other specific changes

Key concepts and terms

- approaches: taxonomic inventory, functional capabilities
- phylotyping vs. OTU-based processing
- species definition for prokaryotes?
- how are metagenomic projects different from single-genome sequencing projects? why is it so difficult to assemble genomes from metagenomics projects?
- main findings: gastrointestinal human microbiome (variation? affected by? medical implications?)
- terms: metagenome, microbiome
- bioinformatics component of a metagenomics project?

Today's exercise: online BLAST

Slide 1

Today's exercise: online BLAST

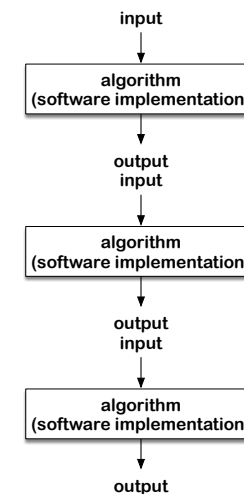**1. decide which version of BLAST to use**

Slide 2

2. paste the "query sequence" into the online form

3. decide which database to search

4. accept all default parameters, start the search

Slide 3

Slide 4

Computer exercise

input

↓

algorithm
(software implementation)

↓

output
input

↓

algorithm
(software implementation)

↓

output
input

↓

algorithm
(software implementation)

↓

output