
YOU HAVE DIED OF DYSENTERY

Microbes and you

ON THE LATEST HUMAN MICROBIOME DISCOVERIES, COMPUTATIONAL QUESTIONS AND SOME SOLUTIONS

Elizabeth Tseng

Dept. of CSE, University of Washington
Johanna Lampe Lab, Fred Hutchinson Cancer Research Center

Bacteria = (deadly) pathogens

Pieter Bruegel's The Triumph of Death (c. 1562)

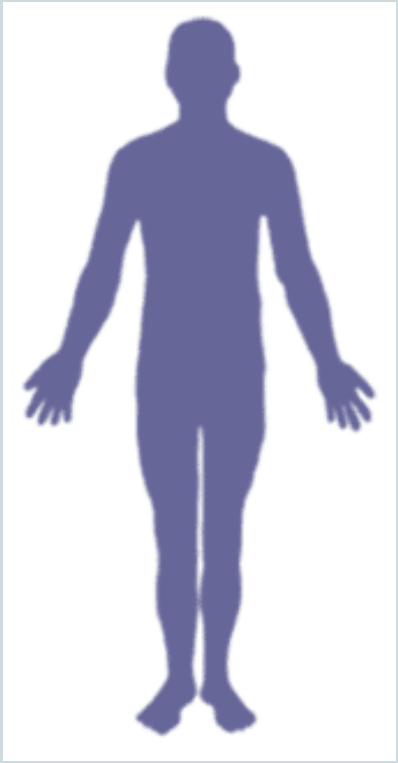
Yersinia pestis

- The plague – spread by rats
- Reduced world population 15-20% (or, half of Europe) in the 14th century
- Archaeological Y. pestis genome ~99% similar to the modern strains

But…

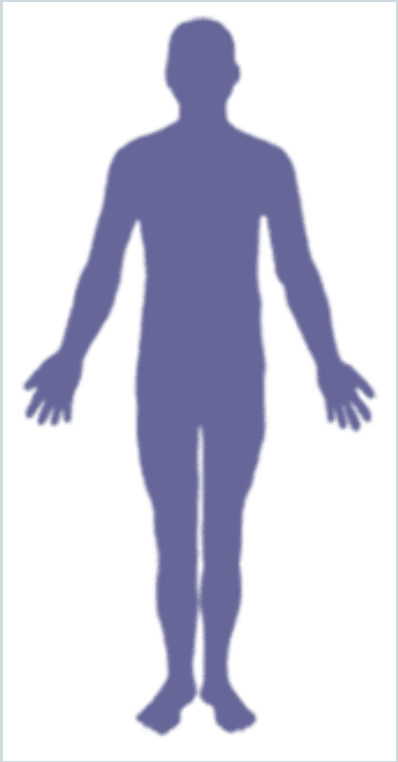
- A single human
  - 10 trillion cells
  - 20,000 genes

  - ? bacterial cells
  - ? bacterial genes



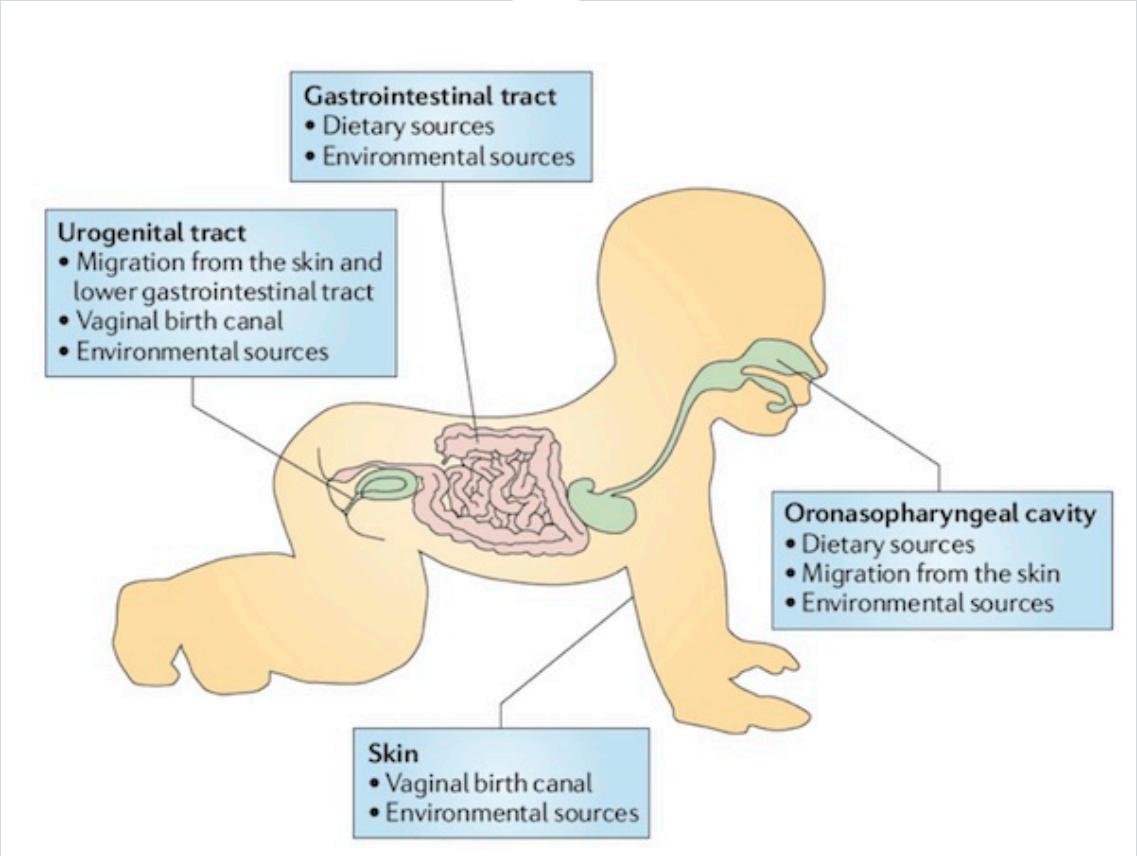
- A single human
  - 10 trillion cells
  - 20,000 genes

  - 100 trillion bacterial cells
  - 20 million bacterial genes



We are 90% bacteria!

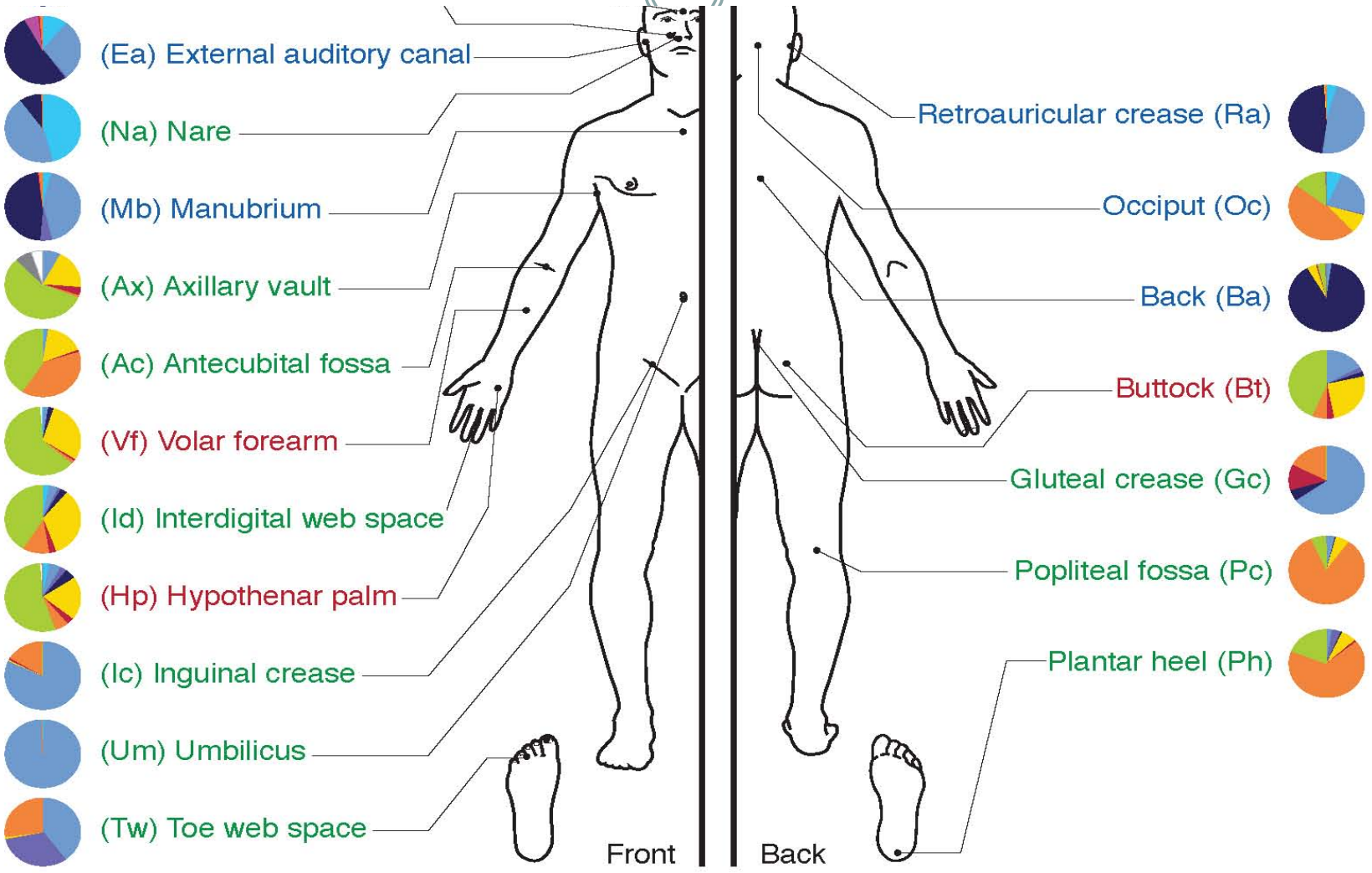
The Human Microbiome



Humans are born sterile, but quickly colonized and start to vary a lot…

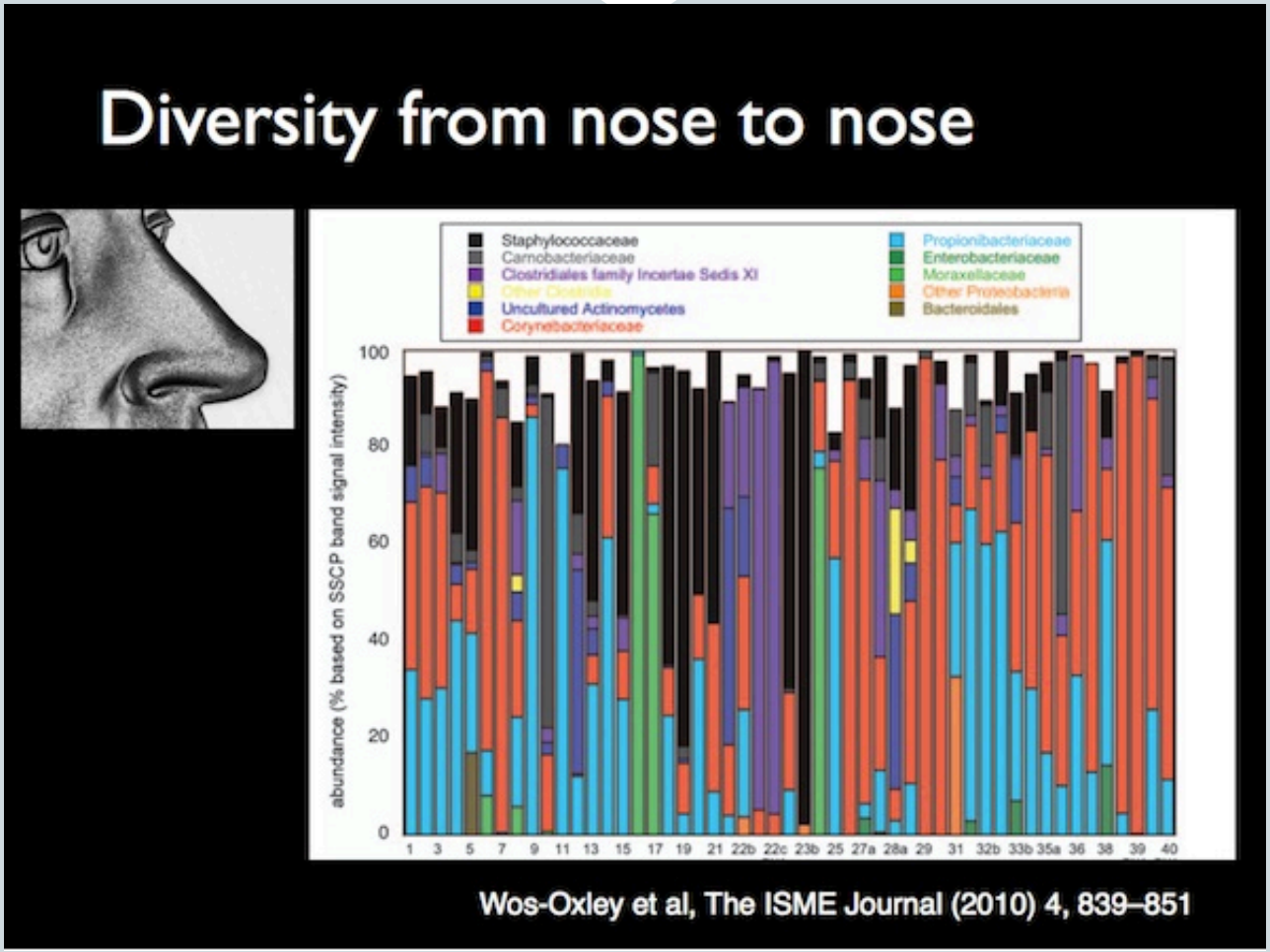
Carl Zimmer http://blogs.discovermagazine.com/loom/2011/03/31/the-human-lake/

From toe to heel…



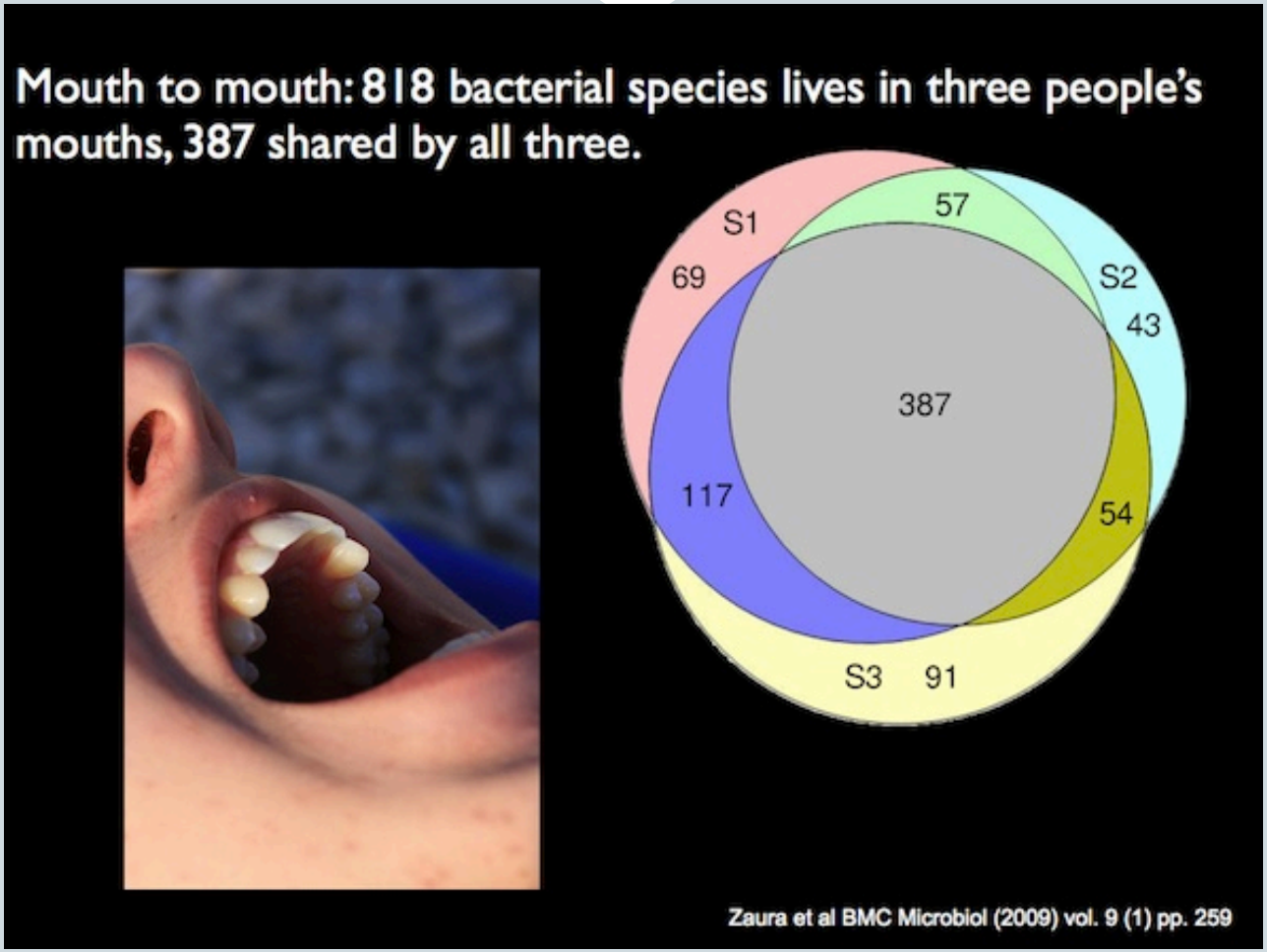
Darryl Leja, NHGRI, 2009

From nose to nose…



Carl Zimmer http://blogs.discovermagazine.com/loom/2011/03/31/the-human-lake/

And from mouth to mouth



Carl Zimmer http://blogs.discovermagazine.com/loom/2011/03/31/the-human-lake/

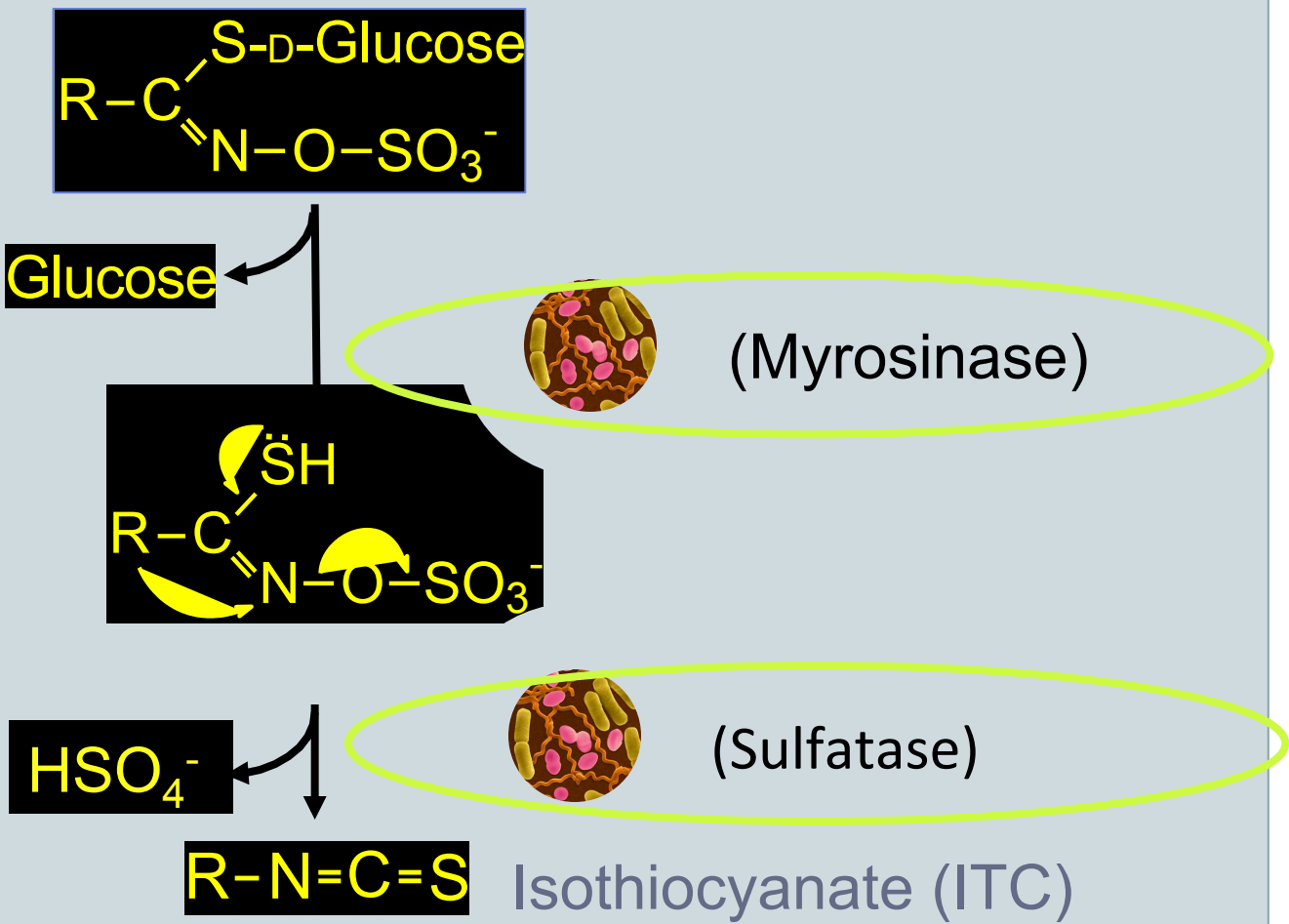
The human microbiome is associated with health risks

- Established causal relationship with
  - inflammatory bowels disease
  - bacterial vaginosis
  - cervical, liver, gastric cancer
- Shifts in community structure correlated with
  - HIV/AIDS
  - chronic wound inflammation
  - obesity
  - acne

But microbes can also be helpful

(to this baby)

Gut bacteria has enzymes humans lack for digesting particular nutrients



Yuesheng Zhang, Roswell Park Cancer Institute, Buffalo, NY
Slide courtesy of M. Hullar

Hot dogs + bacteria = carcinogens

(not so good)

Nitrate
NO_3^-

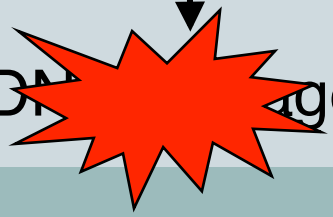
(Nitrate reductase)

Nitrite
NO_2^-

(Nitrite reductase)

***N*-nitroso compounds**

DNA adducts



DN ge

Slide courtesy of M. Hullar

The gut microbiome

- Clearly, microbes play a role in food metabolism

But,

- Do all animals/(humans) have the same gut microbes?
  - If not, what affects them?
- Do gut microbes influence nutrition intake from food?
  - Specifically, is there a “fat microbe” set vs “lean microbe” set?

Comparing the mammalian gut microbes

Evolution of Mammals and Their Gut Microbes

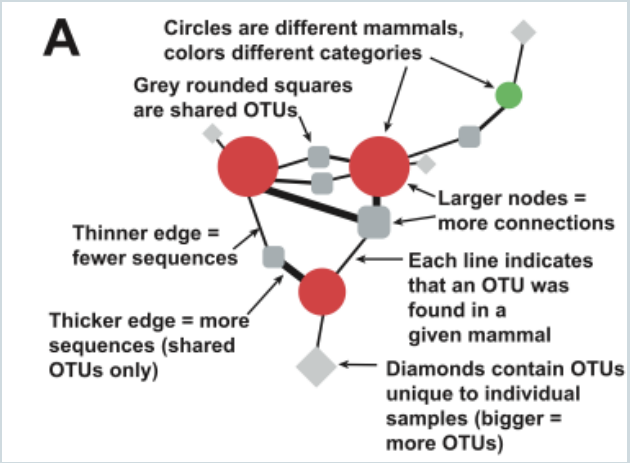
Ruth E. Ley,[1] Micah Hamady,[2] Catherine Lozupone,[1,3] Peter J. Turnbaugh,[1] Rob Roy Ramey,[4] J. Stephen Bircher,[5] Michael L. Schlegel,[6] Tammy A. Tucker,[6] Mark D. Schrenzel,[6] Rob Knight,[3] Jeffrey I. Gordon[1]*

- Look at the gut microbe composition

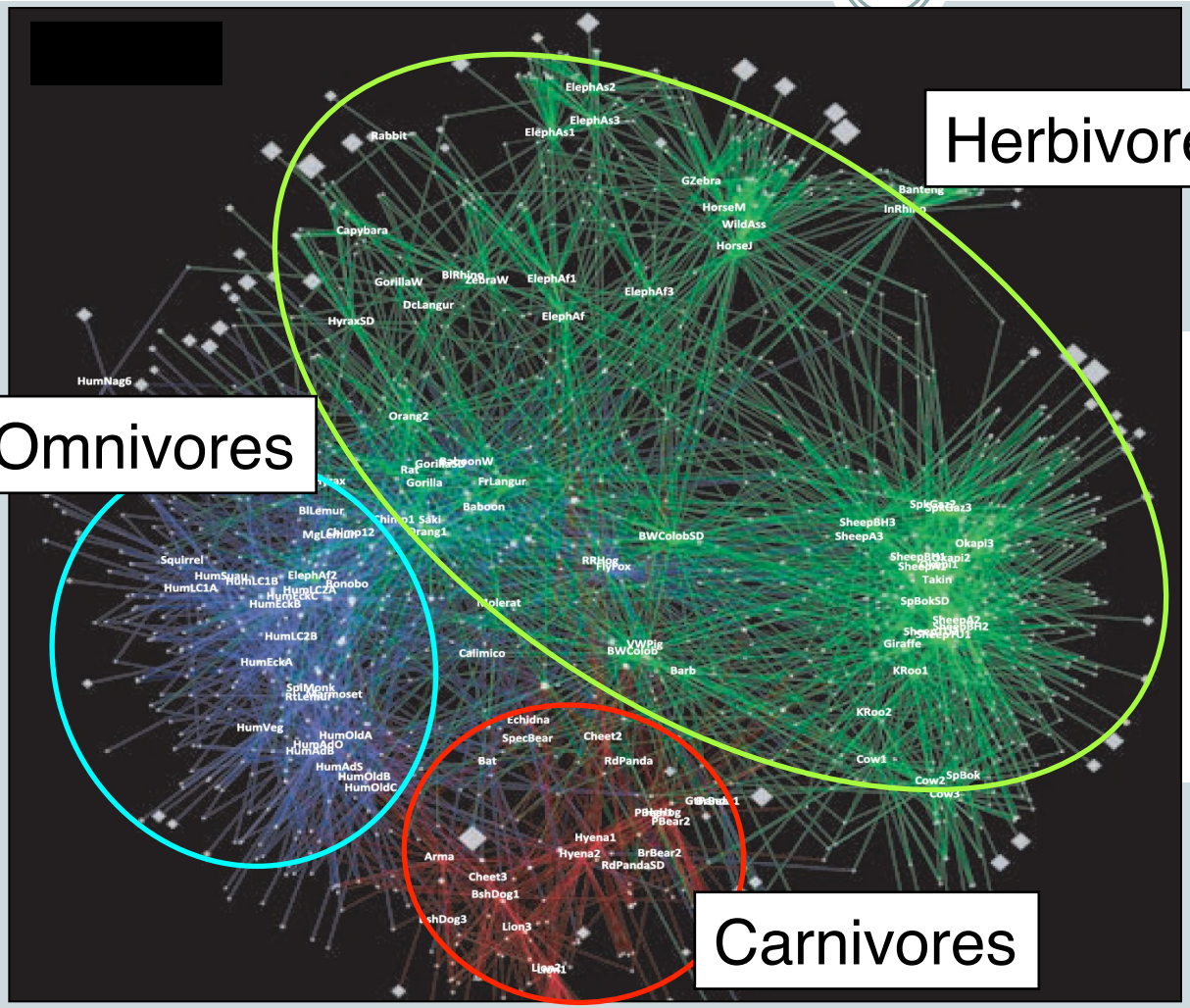
of 106 different mammals covering 60 species

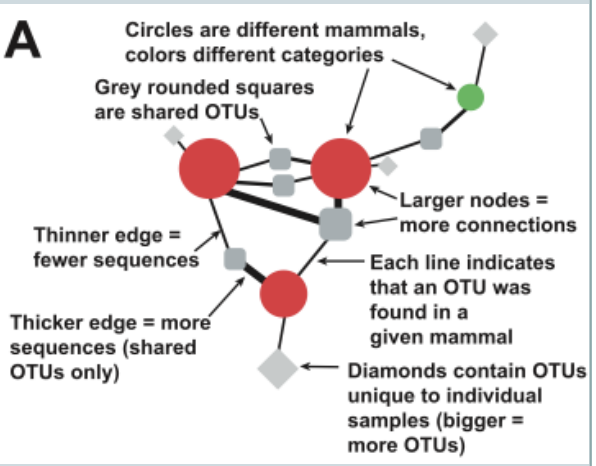
- circle = mammal
- square = OTU/species
- edge (between circle – square) if this mammal has this species in its gut

OTU = Operational Taxonomic Unit
(think of them as species for now)

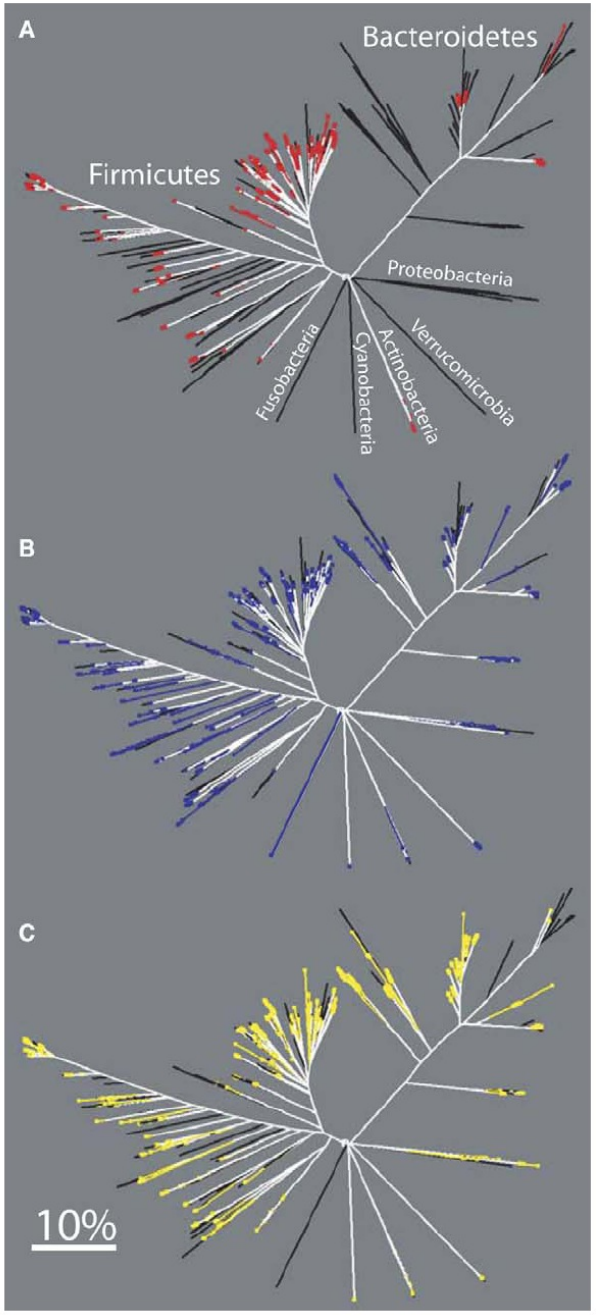


You are what you eat





Ley, RE et al., *Science*, 2008



Inter-individual variation

*shared (white)

*unique to an individual
(red, blue, or yellow)

*absent
(black)

20% culturable

Eckberg et al.,2005,Science 308:1635-1638

A study of obese and germ-free mice

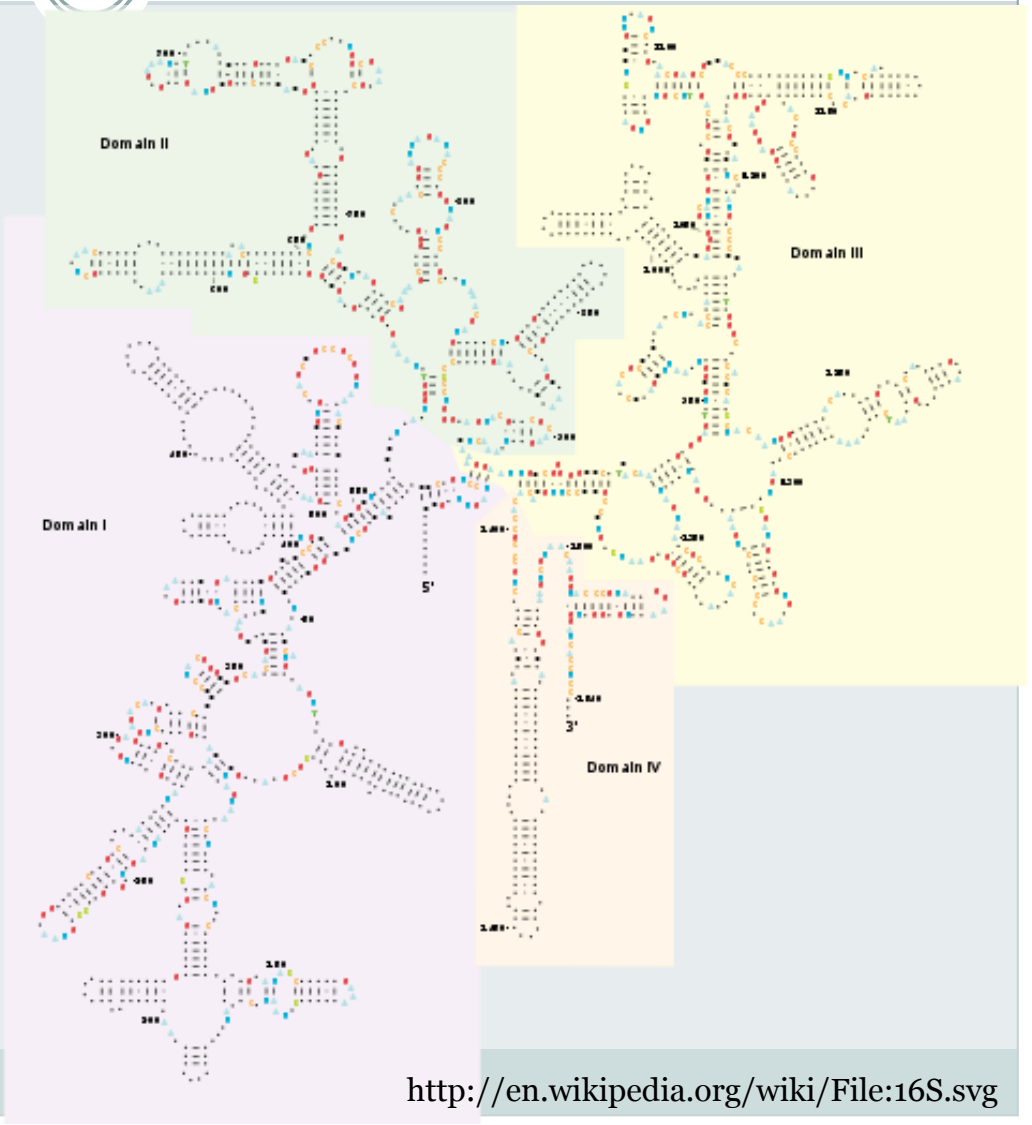
- Obese mice vs germ-free mice
  - Increased capacity fermenting polysaccharides (complex carbs)
  - Increased adiposity (fat cells)
- Transplanting ob mice gut microbe into germ-free mice
  - More efficient energy uptake

Gordon et al, *Nature*, 2006

How do we know what bacteria is there?

Sequencing the 16S ribosomal RNA

- A non-coding RNA
  - Transcribed but not translated
- Part of the ribosome
- ~1,500 bp
- Common species identifier
  - Present in all bacteria
  - Evolves 25X slower than other genes
  - Conserved regions
  - Variable regions

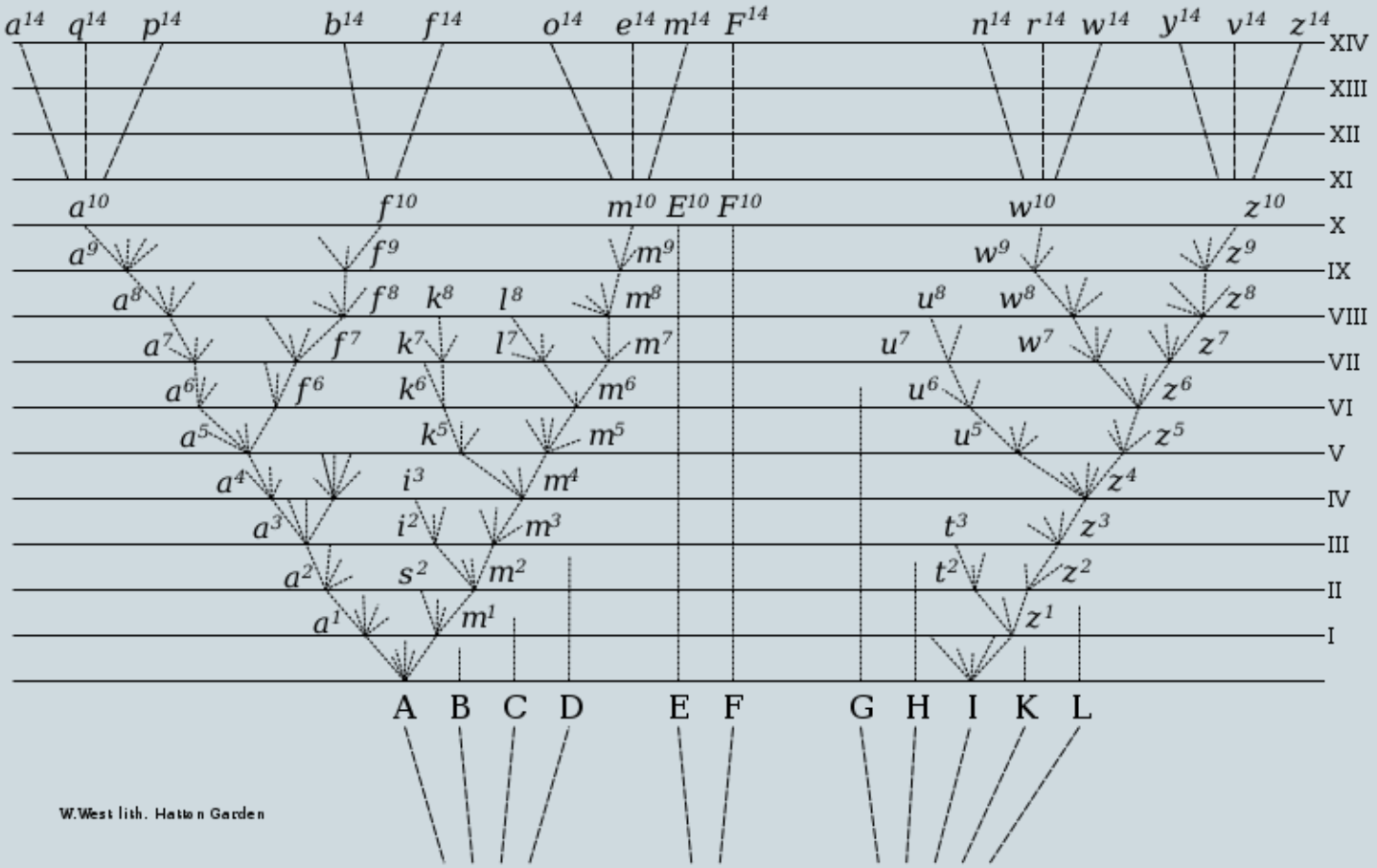


http://en.wikipedia.org/wiki/File:16S.svg

Species differences are reflected in 16S

```
bacteria 1   GTCATGGCGAATAAGCCACGGCGAACTACGTG
bacteria 2   GTCATCGCGAATAAGCCACGGCTAACTACGTG
bacteria 3   GTCATCGCGAACAAGCCACGGCAAACTACGTG
bacteria 4   GTCATGGCGAATTAGCCACGGCTAACTACGTG
```

Bacteria species trees are often defined by 16S rRNA phylogeny



Charles Darwin, *On the Origin of Species*, 1859

Recruit study participants

Collect samples

Extract microbial DNA from samples

These are the steps that lead to finding out what microbes are in your gut

```
@SOLEXA-1GA-1:4:63:1674:855#GCCAAT/1
GAATATTGGTCNATGGCCGAGAGGCTGAACCAGCCAAGTCGCGTGAGGGA
+SOLEXA-1GA-1:4:63:1674:855#GCCAAT/1
@@@A@?8A=@)%:BBBBB@A>ABBBB@B??ABBBACCCBCA@CCBCBBCC
@SOLEXA-1GA-1:3:24:1487:98#GCCAAT/1
TTCTATCAGCAGGGAAGATAGTGACGGTACCTGACTAAGAAGCCCCGGCC
+SOLEXA-1GA-1:3:24:1487:98#GCCAAT/1
BBCBBBA3@CACBC;5CAB@C?BBBCB38?;(>651'855=13:A=/33&
@SOLEXA-1GA-1:2:8:845:25#GCTAAT/1
GGAGGAAACTCTGACCCAGCACCGCCGCGTGGAGGAAGAAG
+SOLEXA-1GA-1:2:8:845:25#GCTAAT/1
```

Analyze DNA sequences

Run DNA sequencing machine

What species does this 16S fragment come from?

>300292::HABQLFT02C43GG rank=0001964 x=1171.0 y=302.0 length=501
AGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGGGAAACCCTGATGCAG
CGACGCCGCGTGGAGGAAGAAGGTCTTCGGATTGTAAACTCCTGTTGTTGAGGAAGATA
ATGACGGTACTCAACAAGGAAGTGACGGCTAACTACGTGC

Species identification

- Smith-Waterman
  - BLAST
- Length-*k* substring matches (k-mer)
  - RDP classifier
- HMMs
- Phylogenetic trees
  - pplacer
  - ARB

RDP classifier: a naïve bayes approach

- Training set: N genus set, G_1, G_2, ...G_N each having n_i species sequences.
- If a sequence belongs to G_i, what is the probability that it contains a particular k-mer word w?

$$P(w \mid G_i)^* = \frac{\#\text{ of seqs in } G_i \text{ containing } w}{n_i}$$

- Given a sequence x from G_i, assuming k-mer independence, the joint probability is:

$$P(x \mid G_i) = \prod_{j=1,2,\ldots} P(w_j \mid G_i)$$

RDP classifier: a naïve bayes approach

- Given a sequence x, the posterior probability that it came from genus G_i is:

Can ignore if assume uniform prior

$$P(G_i \mid x) = \frac{P(x \mid G_i)P(G_i)}{P(x)}$$

Constant. Ignore.

$$P(x \mid G_i) = \prod_{j=1,2,\ldots} P(w_j \mid G_i)$$

- Output: report G_i that gives the highest $P(x \mid G_i)$

HMM approach

Genomics. 2012 Feb 3. [Epub ahead of print]

C16S - A Hidden Markov Model based algorithm for taxonomic classification of 16S rRNA gene sequences.

Ghosh TS, Gajjalla P, Mohammed MH, Mande SS.

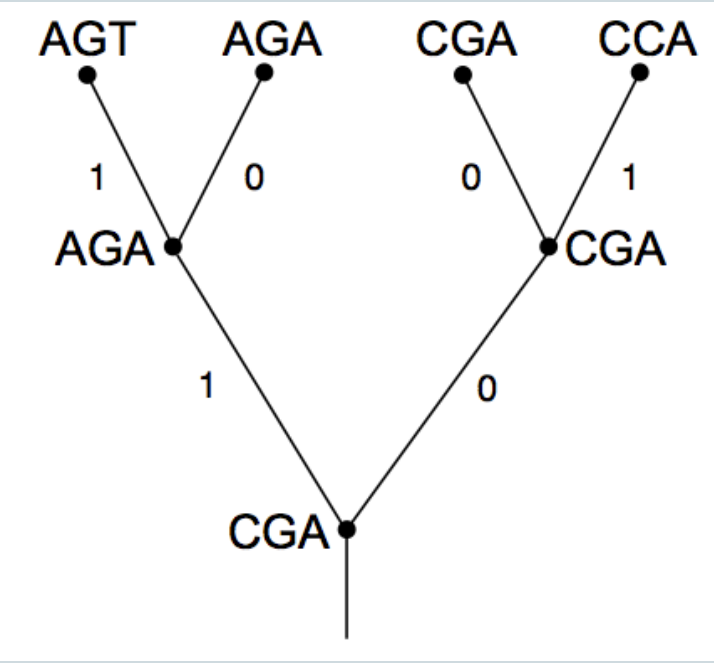
Abstract

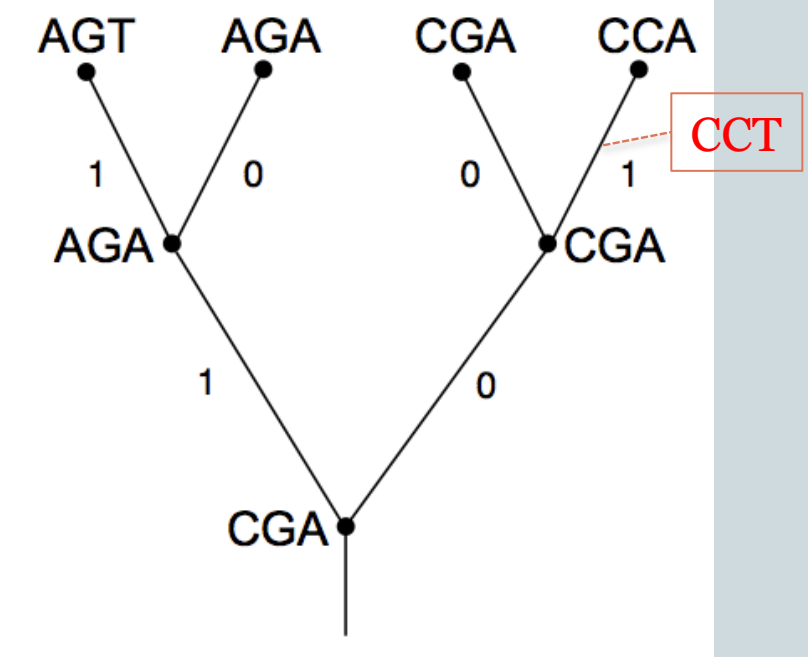
Recent advances in high throughput sequencing technologies and concurrent refinements in 16S rDNA isolation techniques have facilitated the rapid extraction and sequencing of 16S rDNA content of microbial communities. The taxonomic affiliation of these 16S rDNA fragments is subsequently obtained using either BLAST-based or word frequency based approaches. However, the classification accuracy of such methods is observed to be limited in typical metagenomic scenarios, wherein a majority of organisms are hitherto unknown. In this study, we present a 16S rDNA classification algorithm, called C16S, that uses genus-specific Hidden Markov Models for taxonomic classification of 16S rDNA sequences. Results obtained using C16S have been compared with the widely used RDP classifier. The performance of C16S algorithm was observed to be consistently higher than the RDP classifier. In some scenarios, this increase in accuracy is as high as 34%. A web-server for the C16S algorithm is available at http://metagenomics.atc.tcs.com/C16S/.

PMID: 22326741 [PubMed - as supplied by publisher]

Phylogenetic approach

- Given: reference phylogenetic tree *T*
- Input: unknown sequence *x* = *'CCT'*
- Output: *T* with *x* inserted into one of the branches





- Now move to HGT

Horizontal Gene Transfer: an example

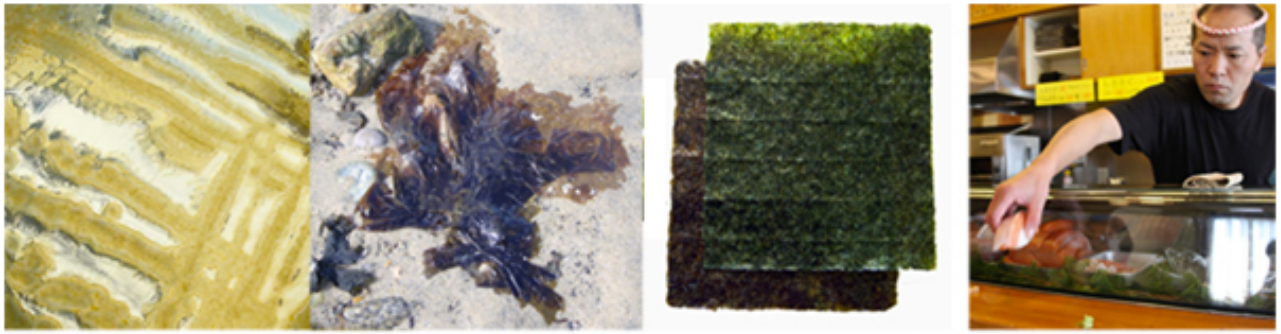
nature Vol 464|8 April 2010|doi:10.1038/nature08937

LETTERS

Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota

Jan-Hendrik Hehemann[1,2]†, Gaëlle Correc[1,2], Tristan Barbeyron[1,2], William Helbert[1,2], Mirjam Czjzek[1,2] & Gurvan Michel[1,2]

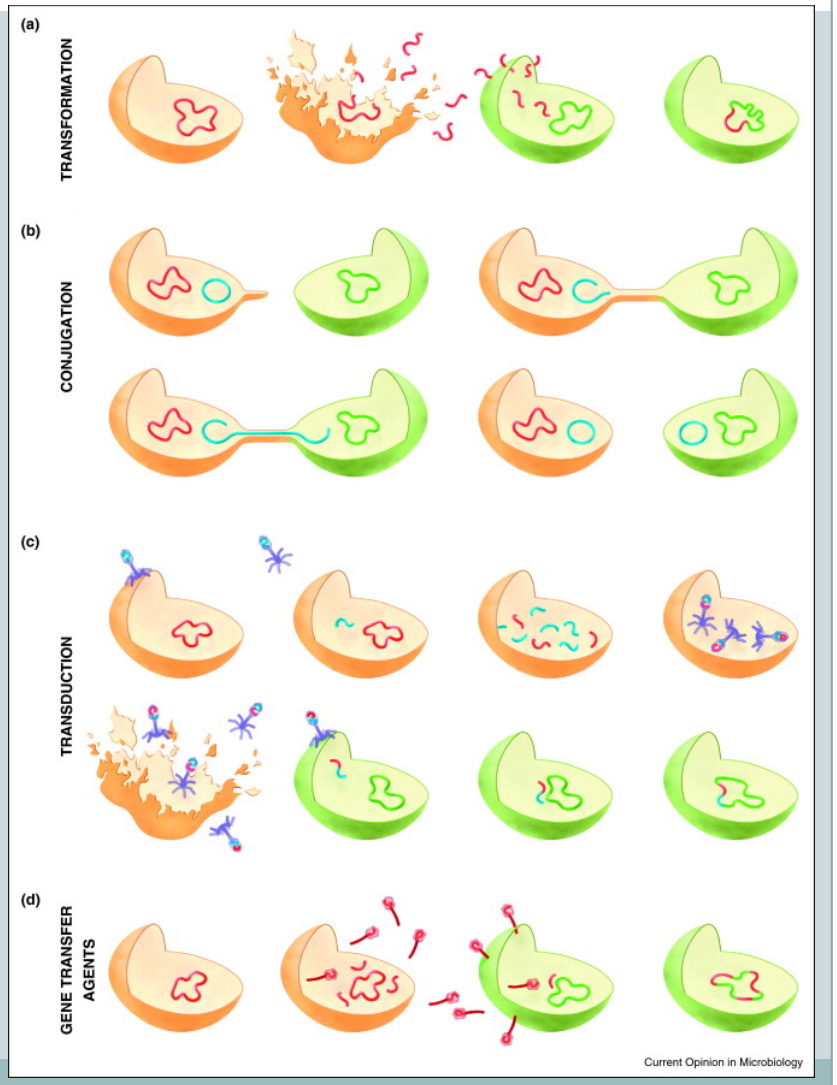
From oceans to bowels



Genes from *Zobellia* lurk on this seaweed linger on this food and end up in his gut

Mechanisms of horizontal gene transfer (HGT)

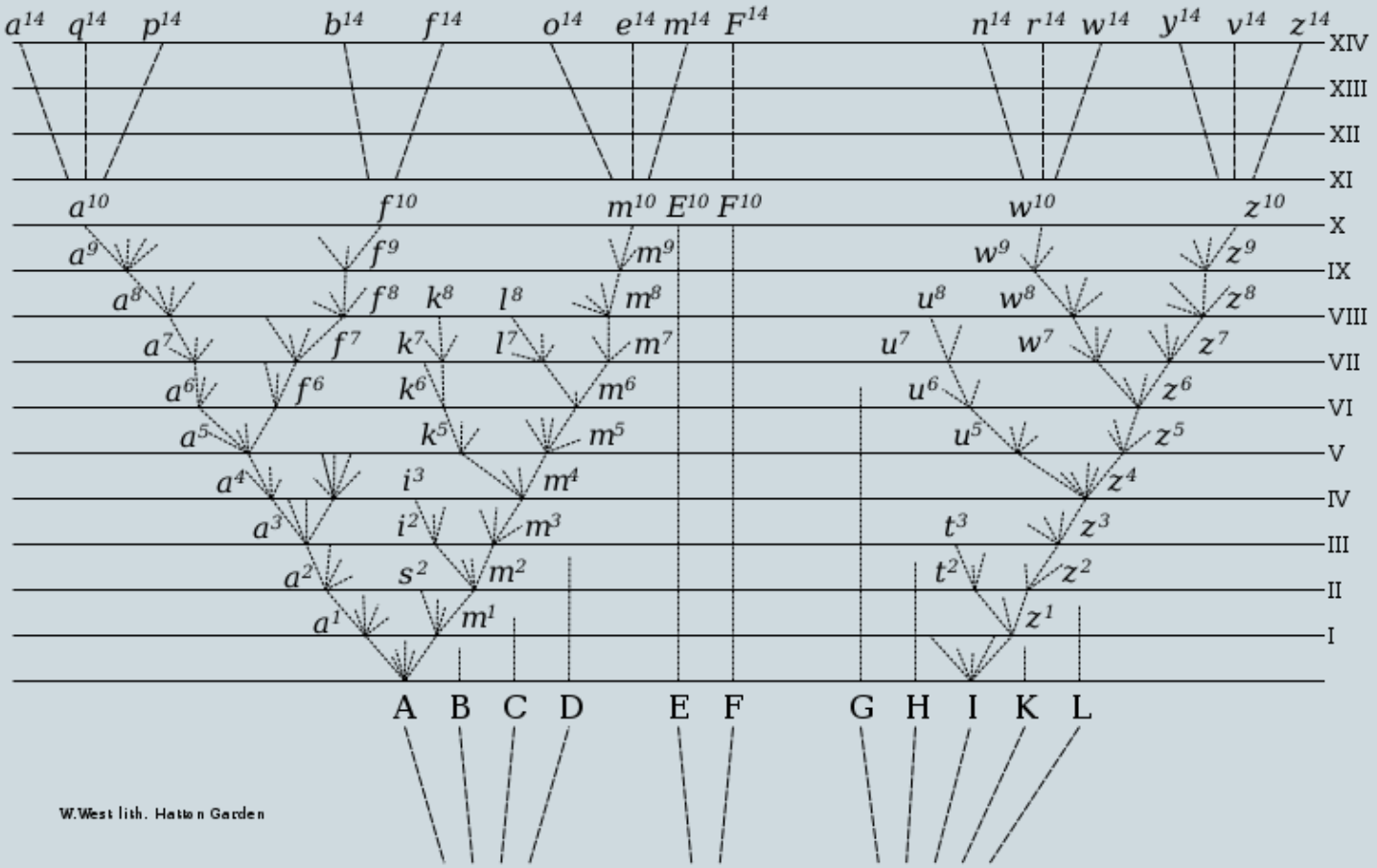
- Transfer of genetic material between organisms
- Not between ancestor – child (vertical transfer)
- Mechanisms of HGT:
  - Transformation
  - Conjugation (plasmid)
  - Transduction (phage)
  - Gene transfer agents



Driving forces of HGT

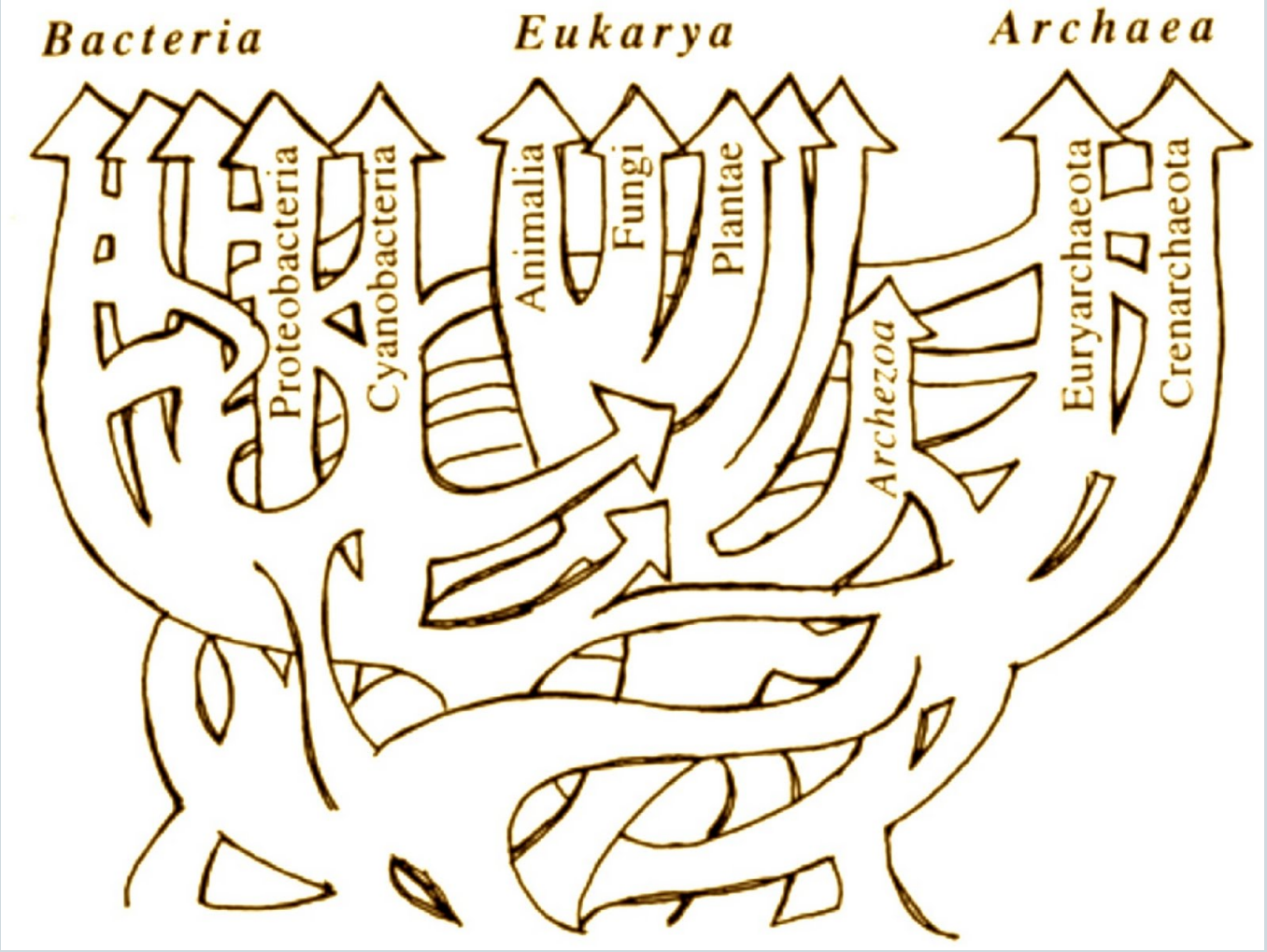
- Phylogeny – compatibility
  - Similar mechanisms for transfer
  - Similar species, similar functions, new gene more likely to survive
- Geography – limitation of dispersal
  - Separated by hard geographical limits (ex: continents)
- Ecology – adaption
  - Increasing fitness (ex: drug resistance, metabolism)

Tree of life?



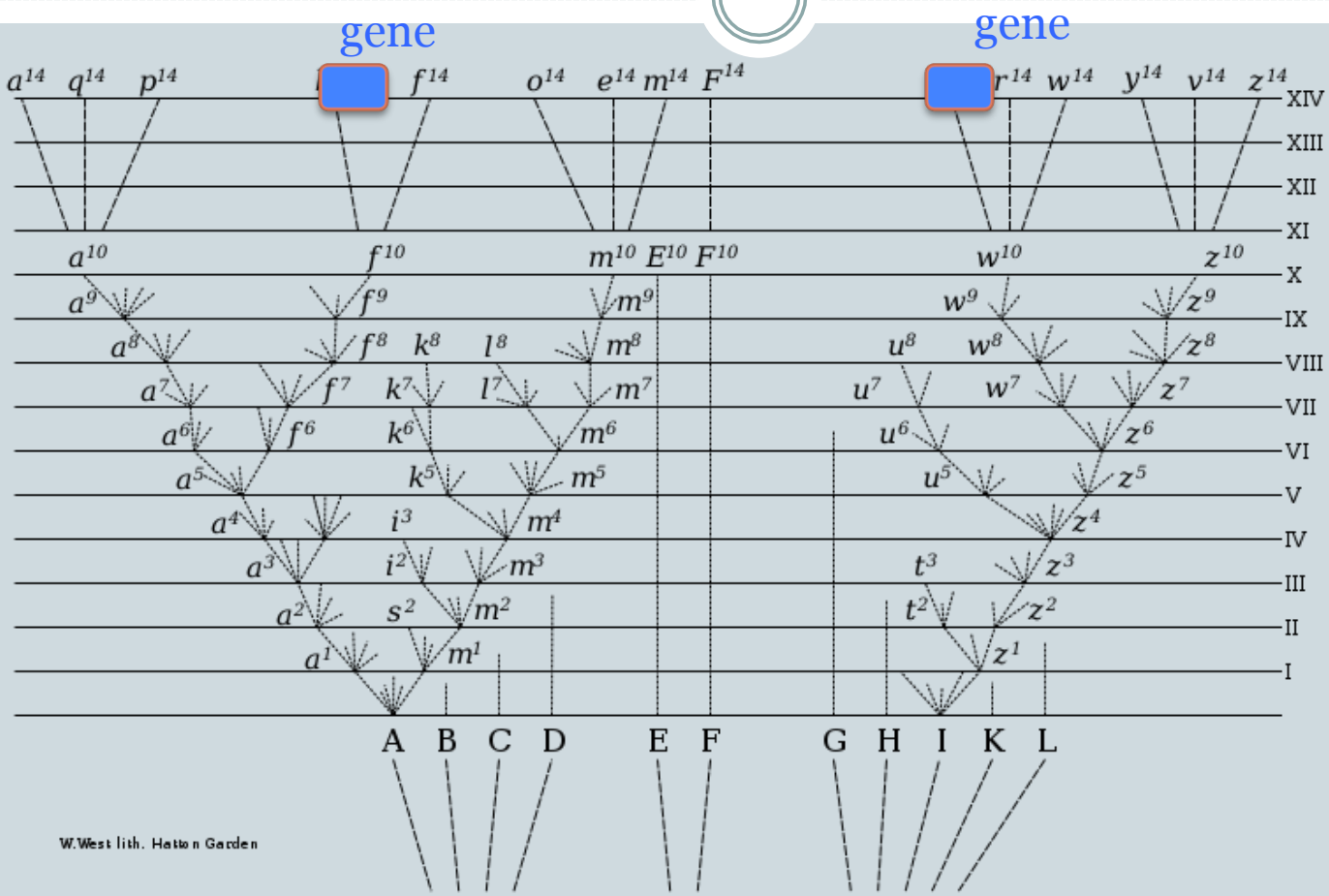
Charles Darwin, *On the Origin of Species*, 1859

Tree of life?



Doolittle & Bapteste, PNAS, 2006

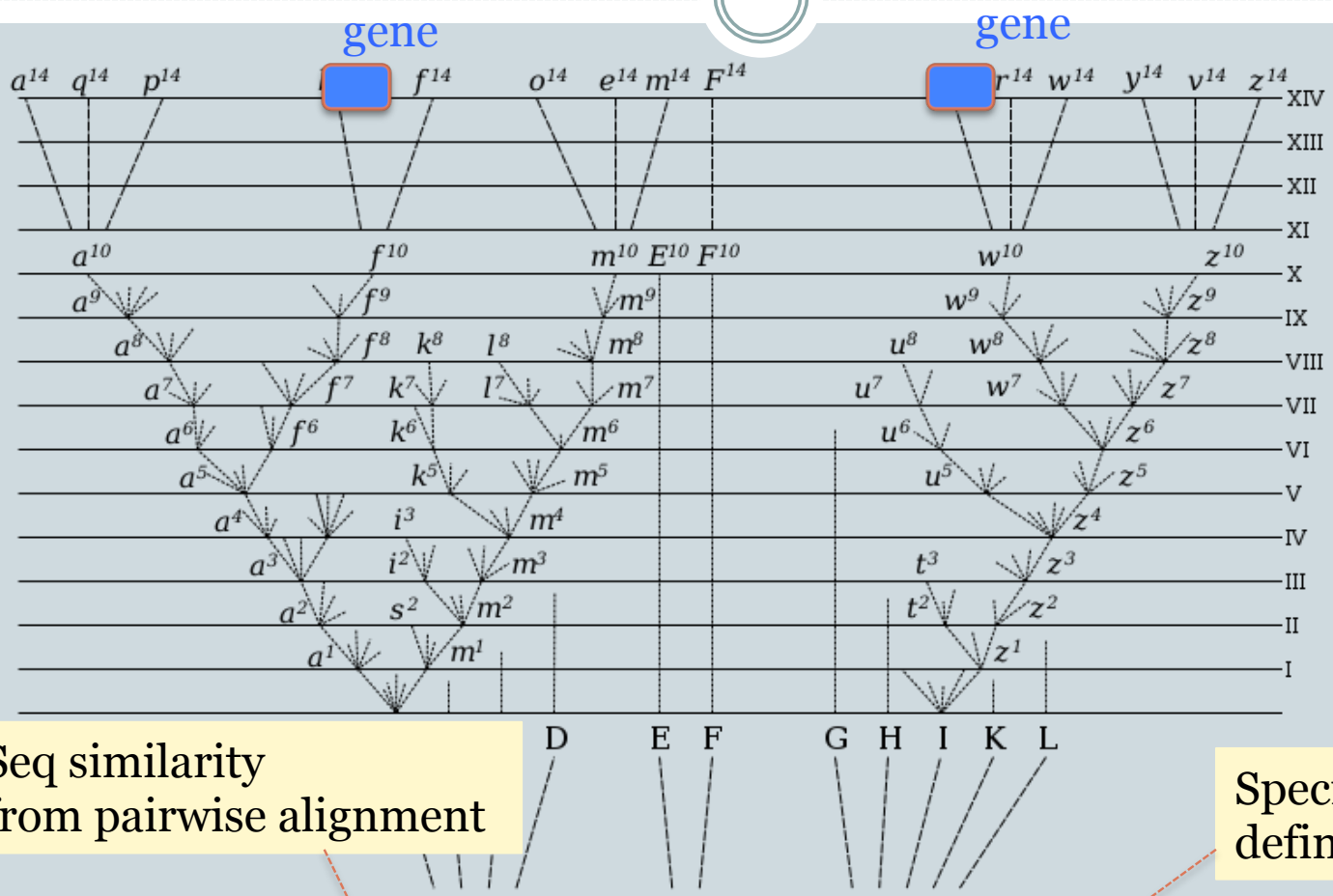
How do we find HGTs?



Unusually highly similar genes between distant species

➔ Likely HGT

How do we find HGTs?



Seq similarity
from pairwise alignment

Species distance
defined by 16S

Unusually highly similar genes between distant species
➔ Likely HGT

doi:10.1038/nature10571

Ecology drives a global network of gene exchange connecting the human microbiome

Chris S. Smillie[1]*, Mark B. Smith[2]*, Jonathan Friedman[1], Otto X. Cordero[3], Lawrence A. David[4] & Eric J. Alm[3,5,6]

- 2,235 bacterial genomes
- From human/animal body sites, natural environments, etc
- Find blocks of **nearly identical DNA (> 99% id, > 500 bp)** in **distant species (< 97% 16S rRNA id)**

Output:

- 16,954 likely HGTs
- ~98% protein coding
- Independent validation confirmed 99% are HGTs

doi:10.1038/nature10571

Ecology drives a global network of gene exchange connecting the human microbiome

Chris S. Smillie[1]*, Mark B. Smith[2]*, Jonathan Friedman[1], Otto X. Cordero[3], Lawrence A. David[4] & Eric J. Alm[3,5,6]

Results from the study:

- HGT was more frequent among bacteria in the same human body site than between body sites
  - Why? Because need to adapt?
- Antibiotic-resistant genes have been frequently transferred between human, farm animals, and food
- These recent HGTs genes could be potential drug targets