PR12_Mikrobna genomika

Insights on biology and evolution from microbial genomes

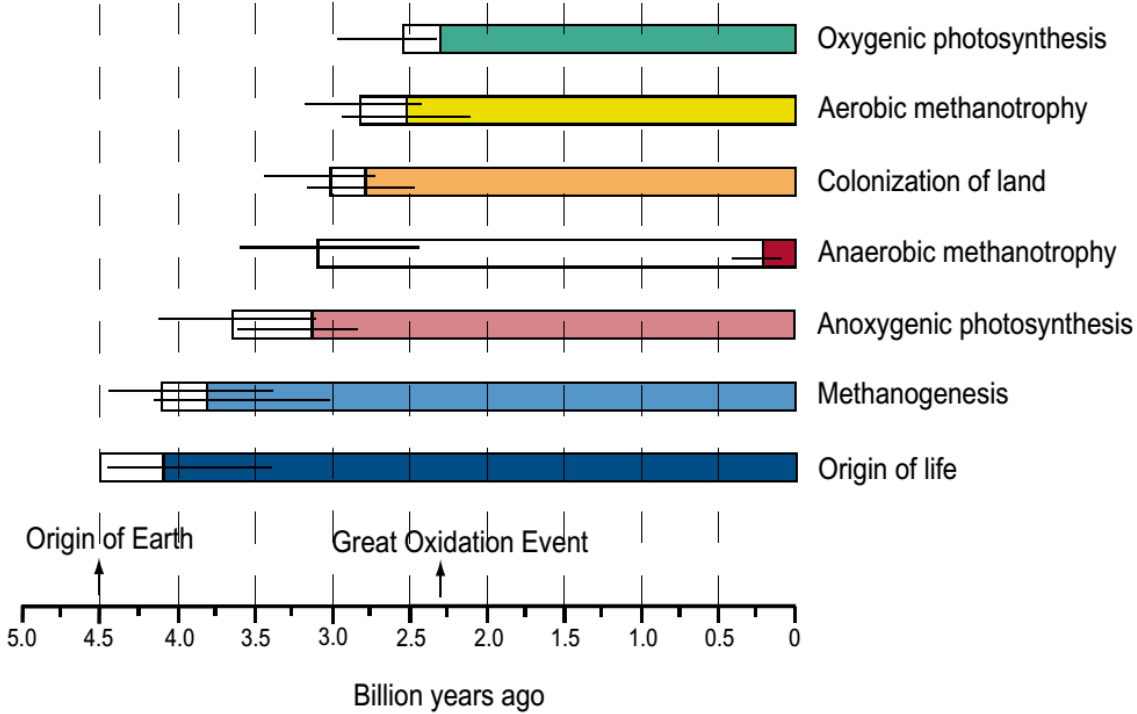
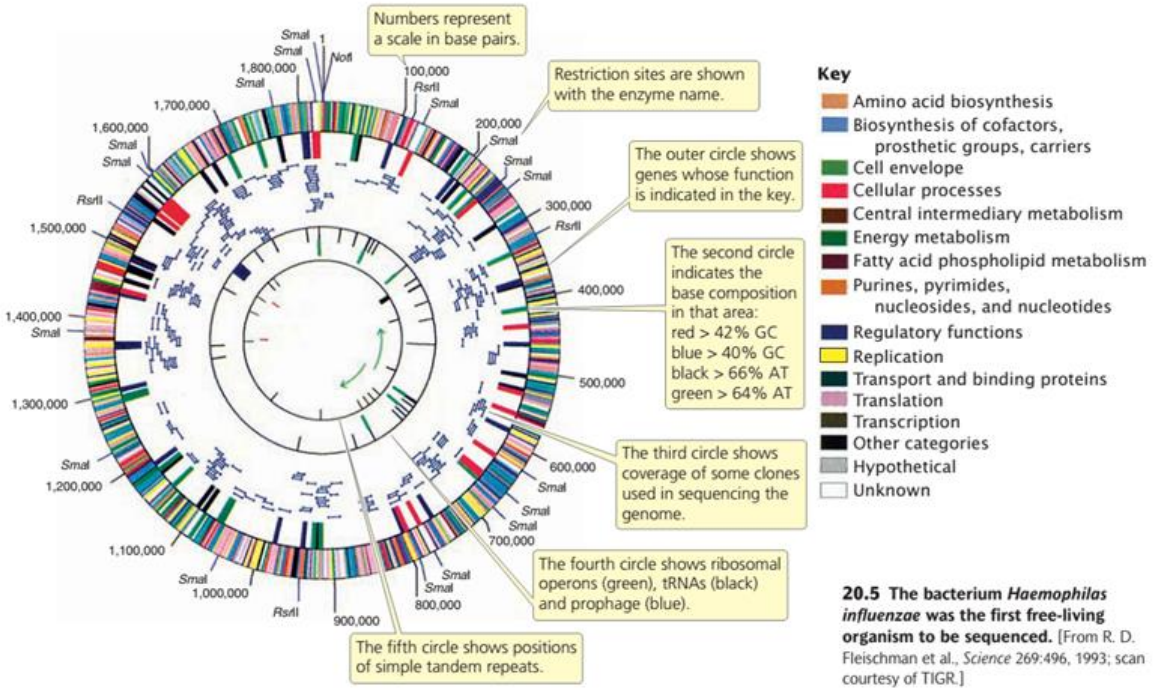


Figure 4
A time line of metabolic innovations and events on Earth. The minimum time for oxygenic photosynthesis is constrained by the Great Oxidation Event (2.3 Ga) whereas the maximum time for the origin of life is constrained by the origin of Earth (4.5 Ga). Horizontal lines indicate credibility intervals, white boxes indicate minimum and maximum time constraints on the origin of a metabolism or event, and colored boxes indicate the presence of the metabolism or event.



20.5 The bacterium *Haemophilas influenzae* was the first free-living organism to be sequenced. [From R. D. Fleischman et al., *Science* 269:496, 1993; scan courtesy of TIGR.]

Whole-Genome Random Sequencing and Assembly of *Haemophilus influenzae* Rd

Robert D. Fleischmann, Mark D. Adams, Owen White, Rebecca A. Clayton, Ewen F. Kirkness, Anthony R. Kerlavage, Carol J. Bult, Jean-Francois Tomb, Brian A. Dougherty, Joseph M. Merrick, Keith McKenney, Granger Sutton, Will FitzHugh, Chris Fields,* Jeannine D. Gocayne, John Scott, Robert Shirley, Li-Ing Liu, Anna Glodek, Jenny M. Kelley, Janice F. Weidman, Cheryl A. Phillips, Tracy Spriggs, Eva Hedblom, Matthew D. Cotton, Teresa R. Utterback, Michael C. Hanna, David T. Nguyen, Deborah M. Saudek, Rhonda C. Brandon, Leah D. Fine, Janice L. Fritchman, Joyce L. Fuhrmann, N. S. M. Geoghagen, Cheryl L. Gnehm, Lisa A. McDonald, Keith V. Small, Claire M. Fraser, Hamilton O. Smith, J. Craig Venter†

An approach for genome analysis based on sequencing and assembly of unselected pieces of DNA from the whole chromosome has been applied to obtain the complete nucleotide sequence (1,830,137 base pairs) of the genome from the bacterium *Haemophilus influenzae* Rd. This approach eliminates the need for initial mapping efforts and is therefore applicable to the vast array of microbial species for which genome maps are unavailable. The *H. influenzae* Rd genome sequence (Genome Sequence DataBase accession number L42023) represents the only complete genome sequence from a free-living organism.

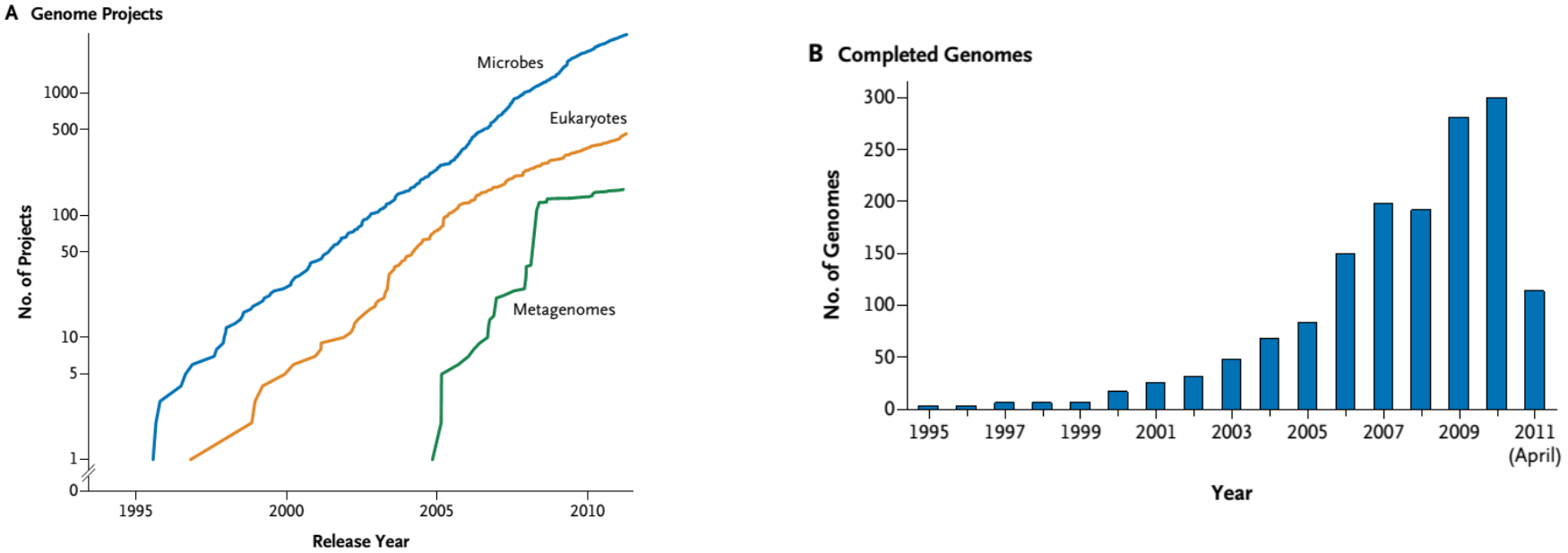
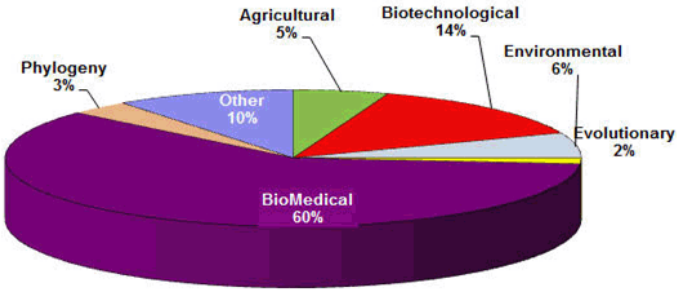
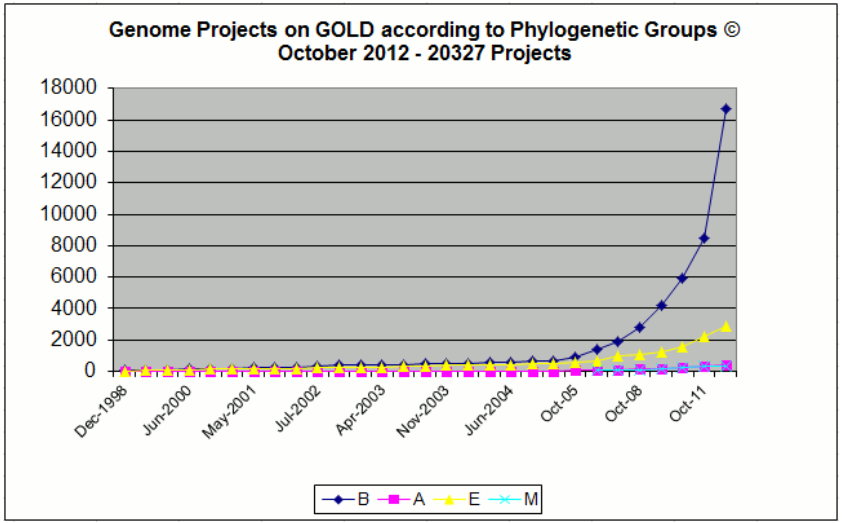


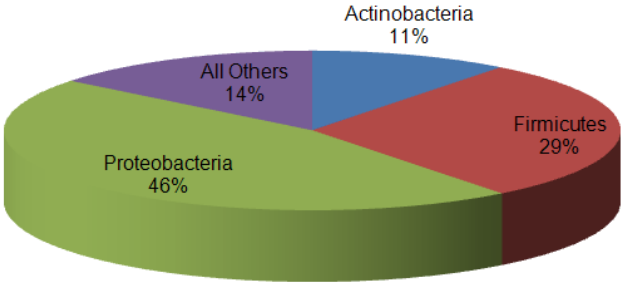
Figure 1. Genome Projects and Completed Genomes since 1995.

Panel A shows a cumulative plot of the number of genome projects — involving microbial (bacterial and archaeal), eukaryotic, and viral genomes — and metagenome projects, according to the release year at the National Center for Biotechnology Information since 1995. Panel B shows the number of completed microbial genome sequences according to year. (The most recent data were collected on April 21, 2011.)

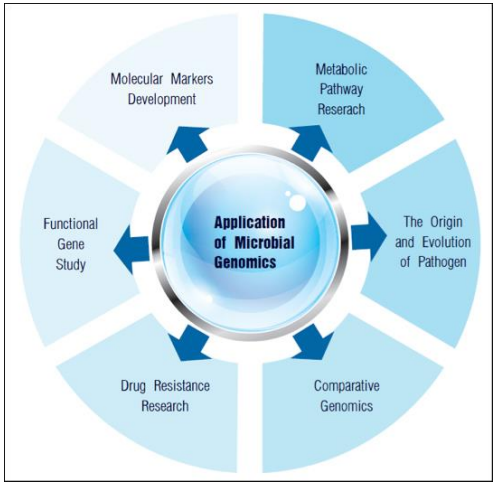


Funding Relevance of Bacterial Genome Projects





Phylogenetic distribution of Bacterial Genome Projects



Bacterial Genome Projects: **35089 projects** (12.5.2014),
Archaeal Genome Projects: **870 projects (**12.5.2014**)**

Insights on biology and evolution from microbial genome sequencing

Claire M. Fraser-Liggett

The Institute for Genomic Research, Rockville, Maryland 20850, USA

No field of research has embraced and applied genomic technology more than the field of microbiology. Comparative analysis of nearly 300 microbial species has demonstrated that the microbial genome is a dynamic entity shaped by multiple forces. Microbial genomics has provided a foundation for a broad range of applications, from understanding basic biological processes, host–pathogen interactions, and protein–protein interactions, to discovering DNA variations that can be used in genotyping or forensic analyses, the design of novel antimicrobial compounds and vaccines, and the engineering of microbes for industrial applications. Most recently, metagenomics approaches are allowing us to begin to probe complex microbial communities for the first time, and they hold great promise in helping to unravel the relationships between microbial species.

Microbiology after the genomics revolution: Genomes 2014

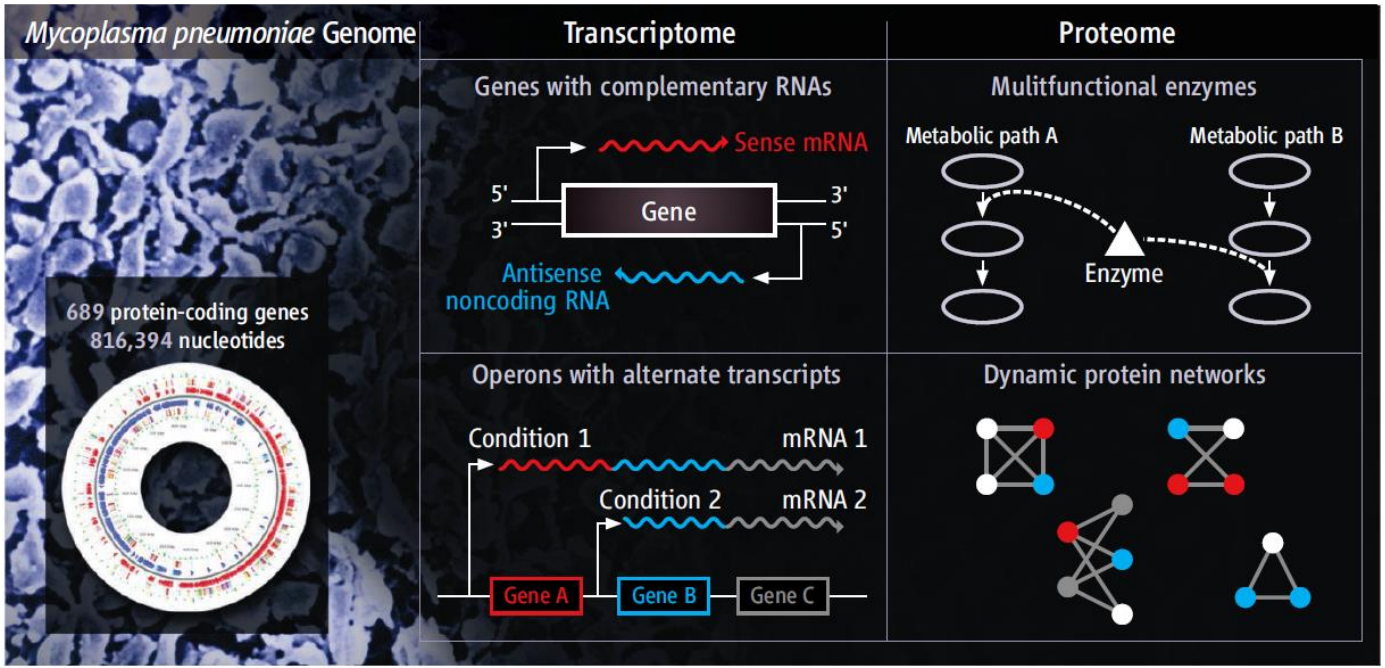
Genomics has profoundly changed our way of conducting research in microbiology. The power of high–throughput **DNA sequencing** technologies, in particular the recent development of next generation sequencing allows researchers now to **address an increasingly diverse range of biological problems**. The scale and efficiency of sequence-based analyses that can now be achieved is providing unprecedented progress in diverse areas that range from the **analyses of genomes** to related disciplines such as **transcriptional profiling** - or **protein - nucleic acid interaction studies**. **Population and metagenomics studies** can now be conducted in an unprecedented large scale, **regulatory processes** can be studied genome-wide under hundreds of different conditions. The genome wide study of the interaction of DNA or RNA with proteins brings completely new insight into regulatory processes and even **single cell analyses** become now possible. The **many diverse applications of next–generation sequencing** and the importance of the insights that are being gained through these methods are very exciting and challenging. It is the perfect time to come together and exchange new knowledge and technologies in this area.

SYSTEMS BIOLOGY

Excavating the Functional Landscape of Bacterial Cells

Functional analyses of a reduced bacterial genome suggest levels of complexity and control previously assumed to be restricted to eukaryotes.

Howard Ochman and Rahul Raghavan



Compact but complex. Analyses of the RNAs and proteins of the bacterium *M. pneumoniae* raise questions about the complexity of regulatory mechanisms in organisms with small genomes.

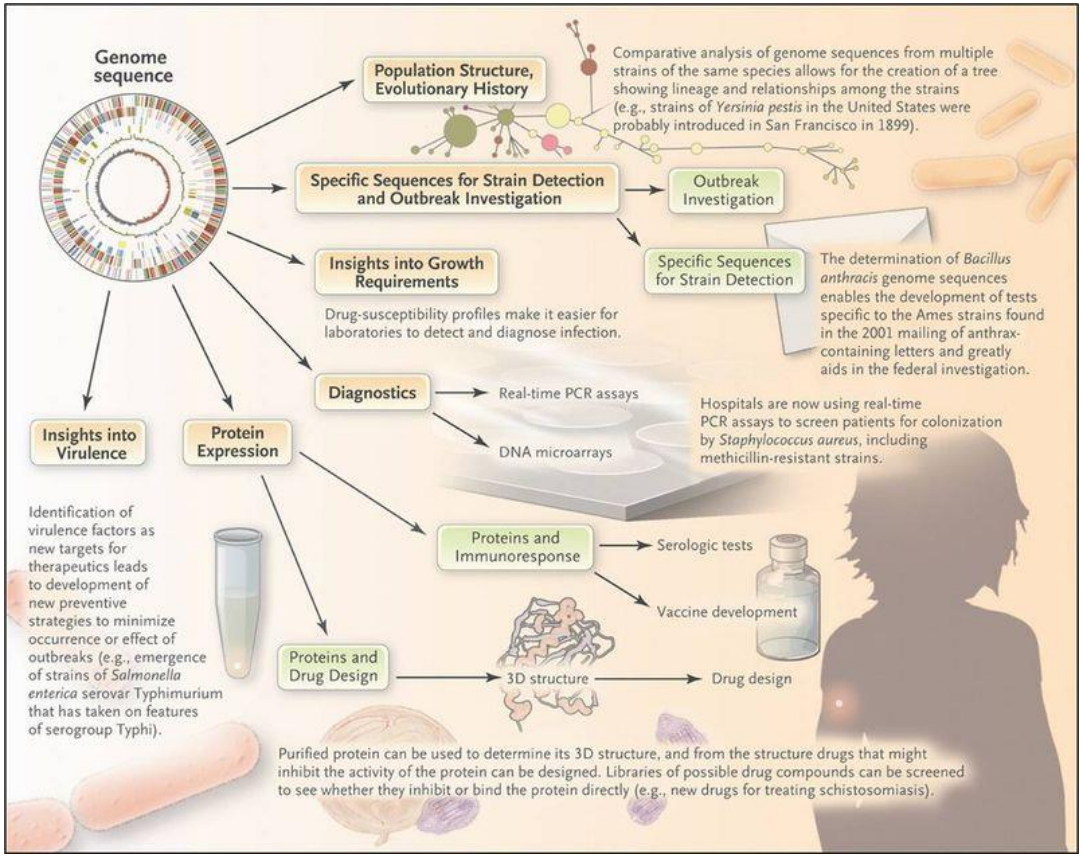
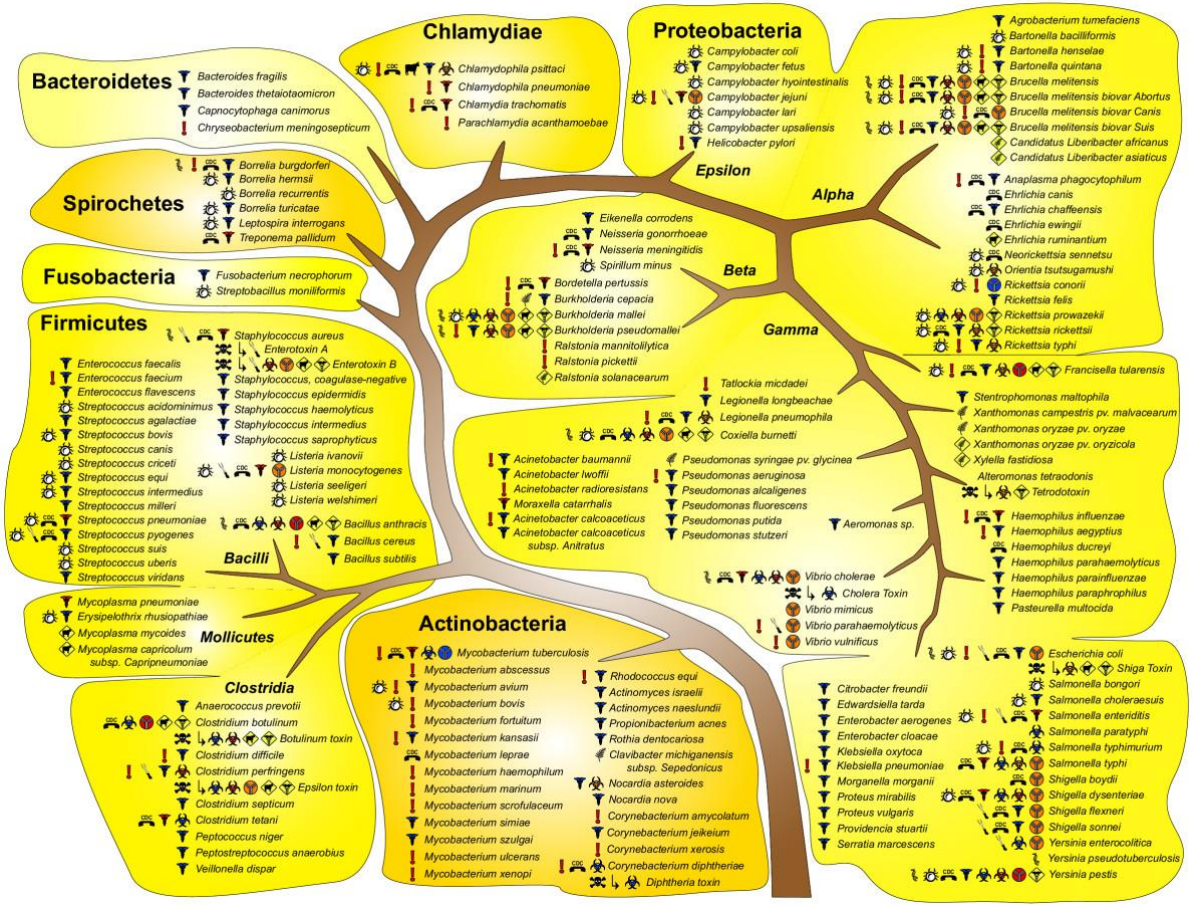


Figure 3. Microbial Genomics and Tool Development.

A genome sequence facilitates the development of a variety of tools and approaches for understanding, manipulating, and mitigating the overall effect of a microbe. The sequence provides insight into the population structure and evolutionary history of a microbe for epidemiologic investigation, information with which to develop new diagnostic tests and cultivation methods, new targets of drug development, and antigens for vaccine development.

Microbial Genomics and Infectious Diseases

Bacterial pathogens



- NIAID Category A Priority Pathogen
- NIAID Category B Priority Pathogen
- NIAID Category C Priority Pathogen
- HHS Select Agent
- USDA High Consequence Animal Pathogen
- USDA High Consequence Plant Pathogen
- Validated Biological Weapon
- Potential Biological Weapon
- Validated Biocrime Agent
- Globally Important Human Pathogen
- Medically Important Human Pathogen
- Important Animal Pathogen
- Important Plant Pathogen
- High Potential For Bioengineering
- Zoonotic Agent
- Toxin
- CDC Notifiable Agent
- Principal Foodborne Pathogen
- Emerging Infectious Agent

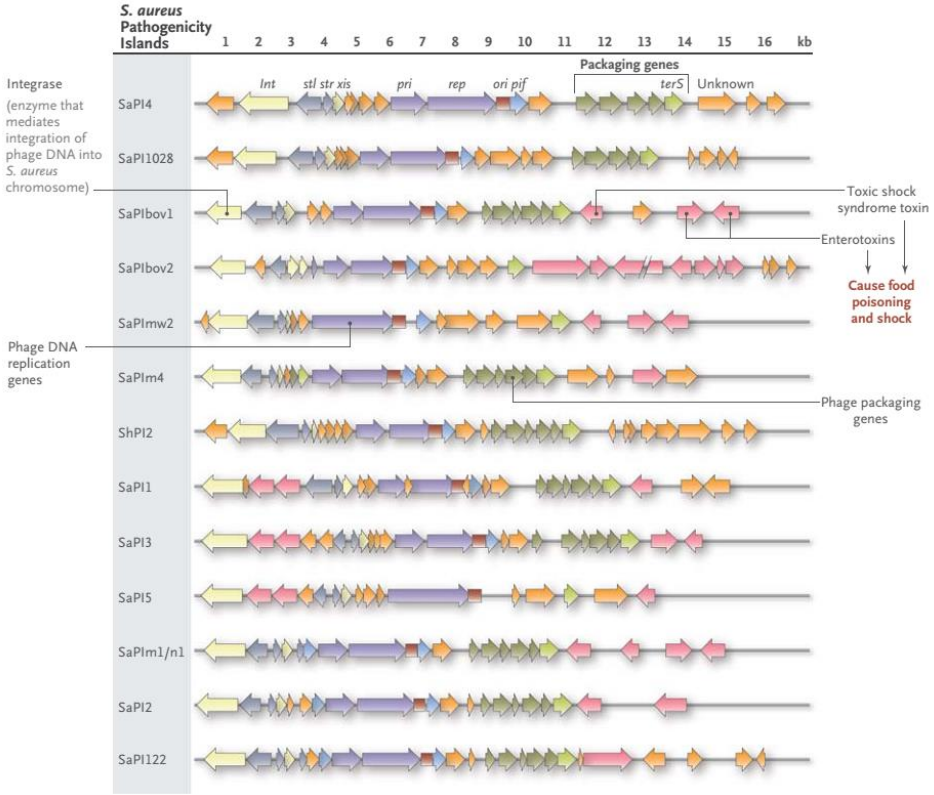


Figure 2. Unexpected Diversity of Virulence Factors for *Staphylococcus aureus*, as Shown by Comparative Genome Analysis.

Shown are phage-related, virulence-associated genomic islands from different strains of *S. aureus* (listed to the left of the genomes), with their genes colored according to functional categories. The islands are located at specific phage integration sites in the *S. aureus* genome and are responsible for the production of toxins by these strains. The numbers at the top are distance markers along the genomic islands. Data are adapted from Novick et al.[37]

From complete genome sequence to 'complete' understanding?

Michael Y. Galperin and Eugene V. Koonin

The rapidly accumulating genome sequence data allow researchers to address fundamental biological questions that were not even asked just a few years ago. A major problem in genomics is the widening gap between the rapid progress in genome sequencing and the comparatively slow progress in the functional characterization of sequenced genomes. Here we discuss two key questions of genome biology: whether we need more genomes, and how deep is our understanding of biology based on genomic analysis. We argue that overly specific annotations of gene functions are often less useful than the more generic, but also more robust, functional assignments based on protein family classification. We also discuss problems in understanding the functions of the remaining 'conserved hypothetical' genes.

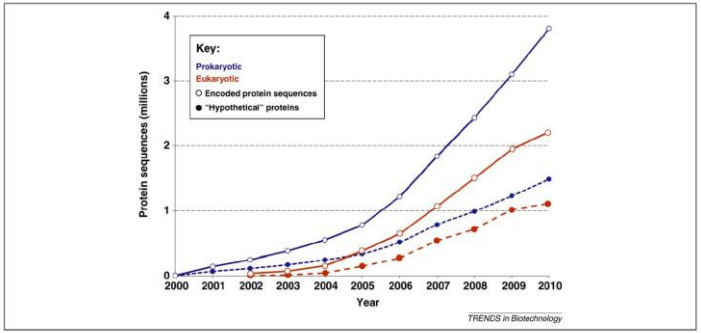
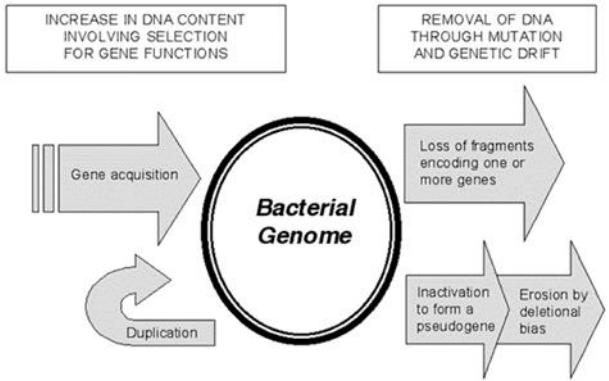


Figure 1. Accumulation of protein sequences of unknown function in the genome databases. Open symbols indicate the total number of protein sequences encoded in prokaryotic (blue) and eukaryotic (red) genomes; filled symbols indicate the number of 'hypothetical' or 'uncharacterized' proteins. The data are taken from the NCBI RefSeq database [68]; the numbers for 2010 are extrapolated from the first 4 months.

Concluding remarks

In conclusion, it might be worthwhile to make several basic generalizations regarding genomes and the understanding of gene functions:

- Functions of many widespread genes are known; all universal genes are involved in translation [9].
- Widespread genes with unknown functions remain uncharacterized for a reason: they often affect multiple processes and their mutations typically are pleiotropic (Box 1).
- The functions of a substantial fraction of genes in each sequenced genome remain unknown.
- Not every experiment on an unknown gene yields useful clues regarding function.
- Structural characterization of a protein rarely gives direct clues to its function [41,42,58].
- Analysis of gene expression rarely gives direct clues to gene functions.
- Delineation of a protein interaction network involving the gene of interest rarely gives direct clues to its function [26,59,60].
- Functional assignments for previously uncharacterized, widely conserved genes are just like any biological discoveries: they require a lot of hard work and a bit of luck.



Dinamični genomi prokariontov

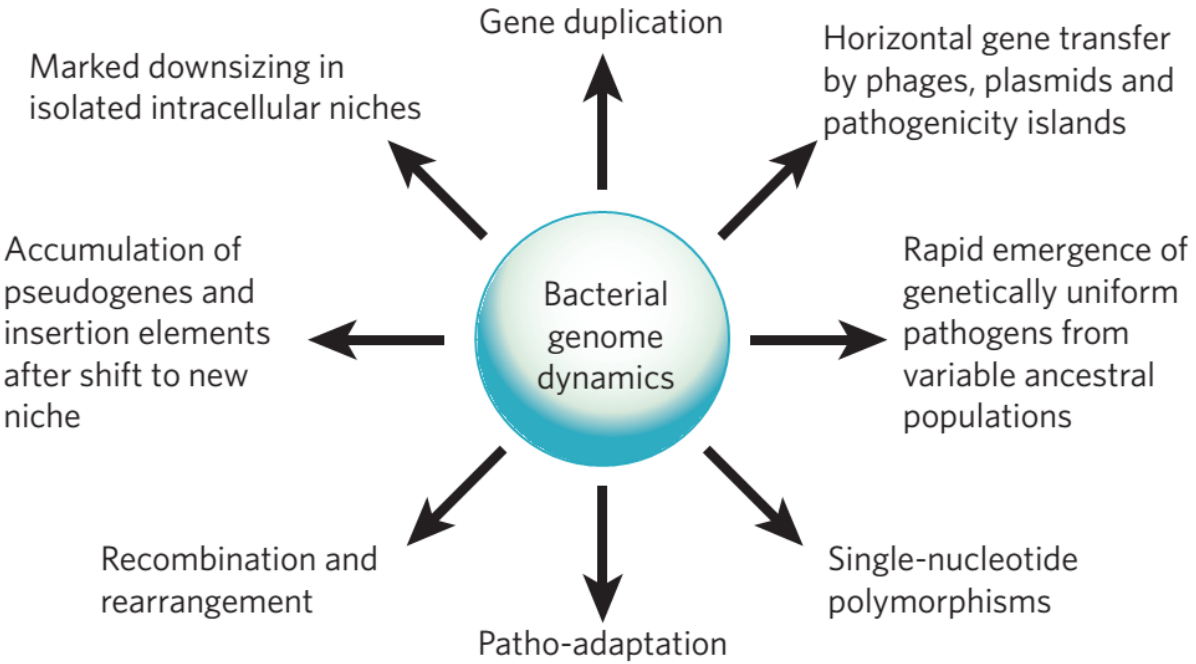


Figure 1 | Bacterial genome dynamics. There are three main forces that shape bacterial genomes: gene gain, gene loss and gene change. All three of these can take place in a single bacterium. Some of the changes that result from the interplay of these forces are shown.

Genome dynamics and its impact on evolution of *Escherichia coli*

Ulrich Dobrindt · M. Geddam Chowdary · G. Krumbholz · J. Hacker

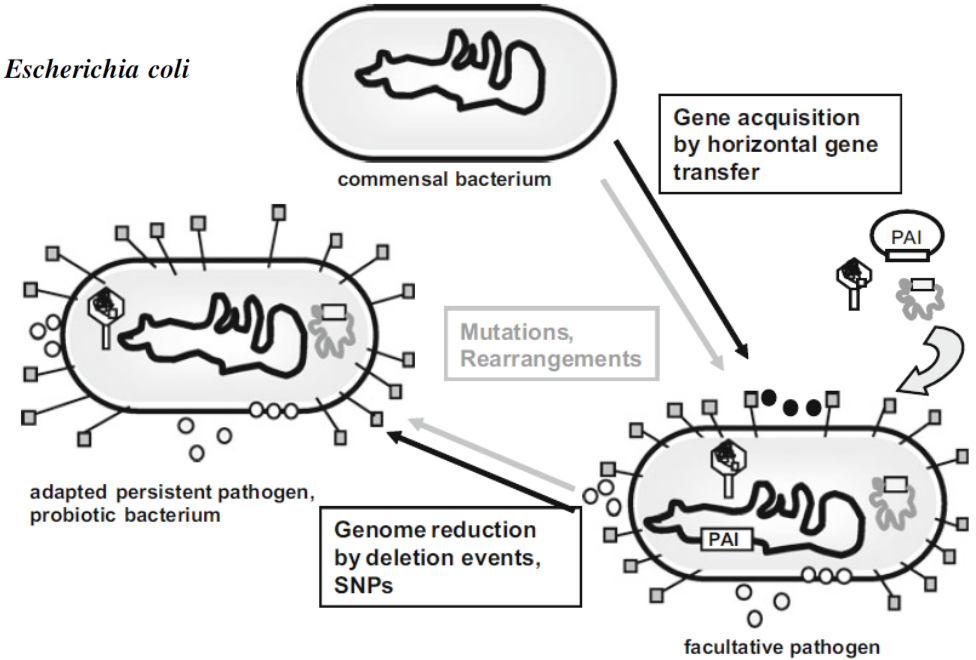


Fig. 2 Processes involved in bacterial genome dynamics and evolution. Genome evolution is based on loss and acquisition of genetic information. Mobile genetic elements such as plasmids, bacteriophages and islands encoding virulence traits (fimbriae, secreted toxins, etc.) that are horizontally transferred by transformation, transduction and conjugation play an important role in gene acquisition. Acquisition of mobile genetic elements may increase the recipient's fitness due to the availability of new traits that may contribute to bacterial adaptation under certain conditions, e.g., bacterial pathogenicity. Due to genetic alterations including deletion and rearrangement events as well as point mutations, such pathogens may lose virulence-associated traits and thus evolve into attenuated variants which may have the capacity to persist in the host. An increased fitness of such strains, shared with pathogenic variants, may contribute to the probiotic character of some *E. coli* strains. PAI, pathogenicity island; SNP, single nucleotide polymorphism

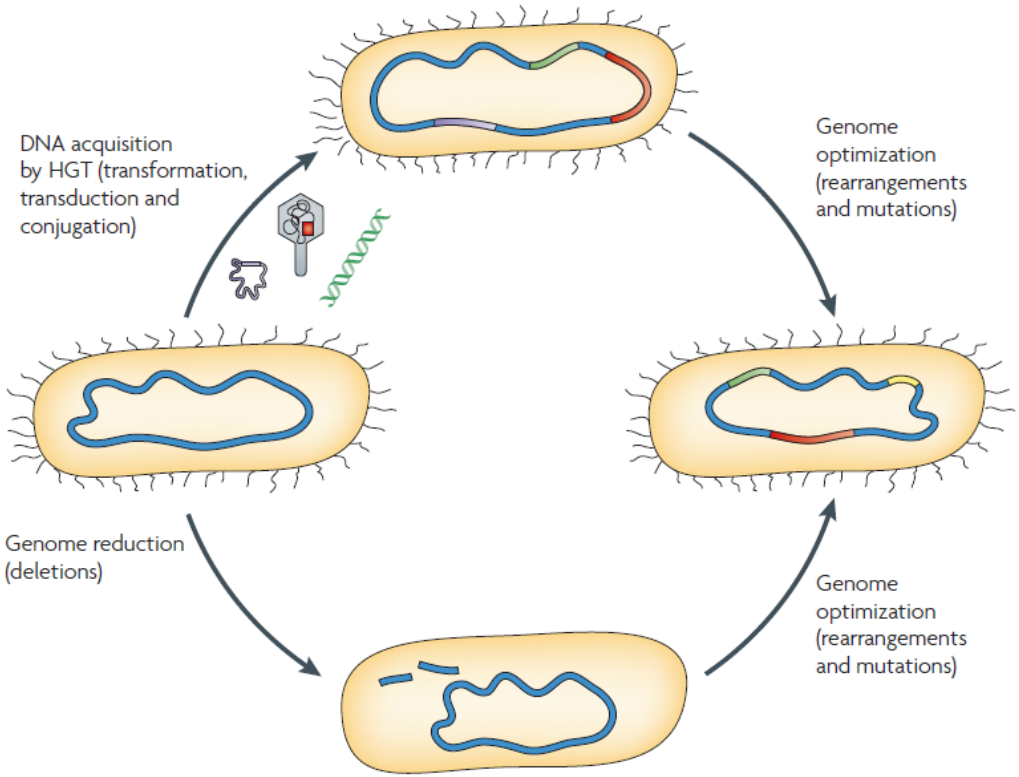
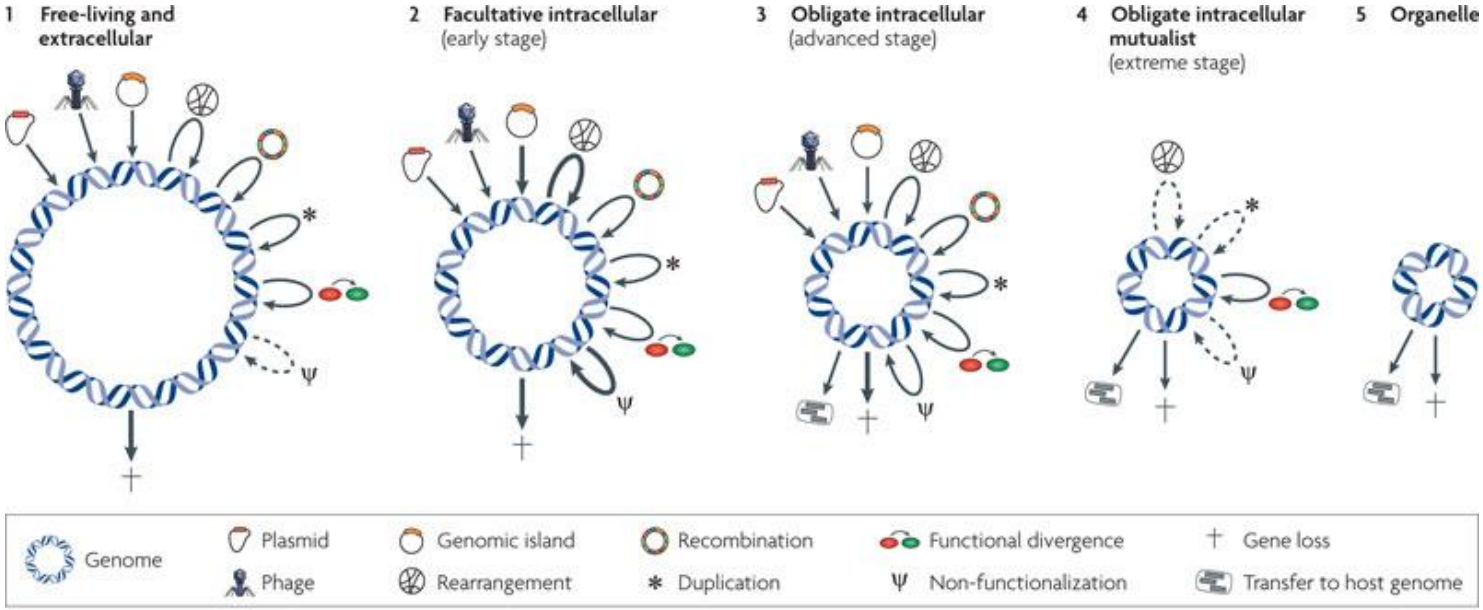
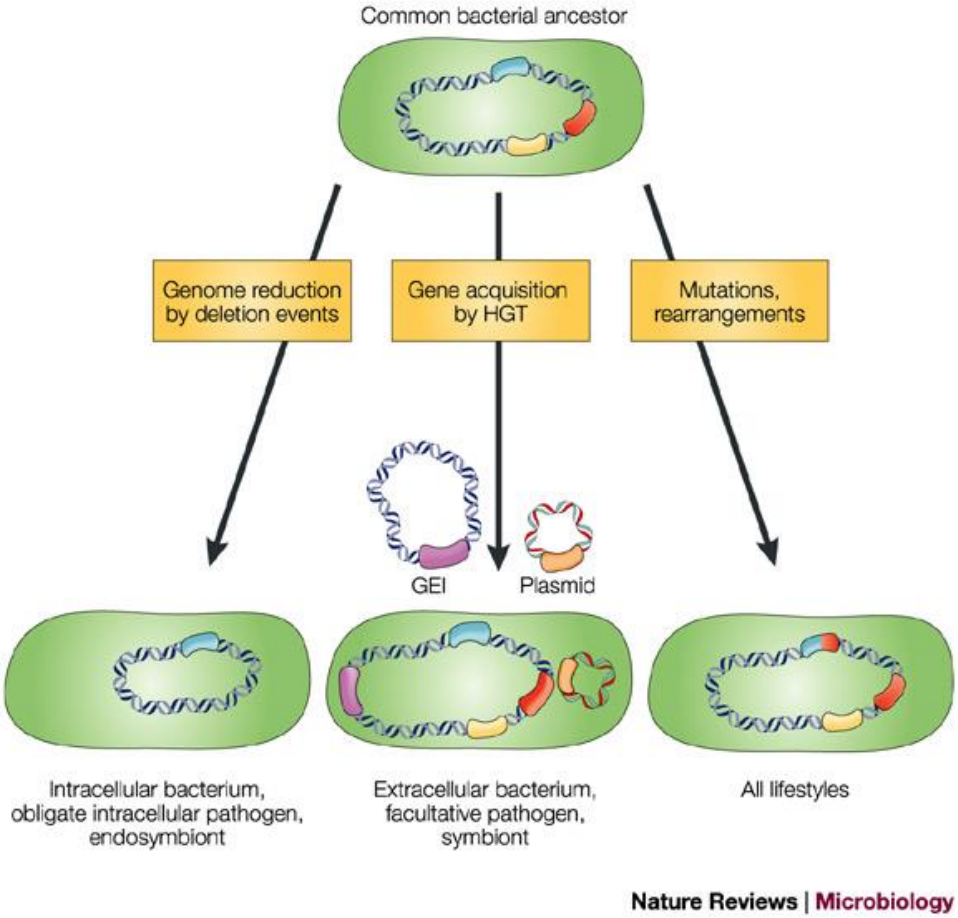


Figure 1 | **Mechanisms that contribute to bacterial genome evolution.** Genome plasticity results from DNA acquisition by horizontal gene transfer (HGT; for example, through the uptake of plasmids, phages and naked DNA) and genome reduction by DNA deletions, rearrangements and point mutations. The concerted action of DNA acquisition and gene loss results in a genome-optimization process that frequently occurs in response to certain growth conditions, including host infection or colonization.



Stages of host adaptation. The genome dynamics for different host-adaptation stages: free-living (1), facultative intracellular (2), obligate intracellular (3), obligate intracellular mutualist (4) and organelle (5). Arrows that point directly to the genomes indicate the acquisition of genes by horizontal gene transfer (through plasmids, genomic islands and/or bacteriophages). Arrows that loop back to the genome indicate changes within the genome (rearrangements, gene duplication, recombination, functional divergence (shifts) and non-functionalization). Arrows that point away from the genome indicate gene loss or gene transfer to the host genome. The relative influence of each of these types of events at the different intracellular stages is shown by the weight of the arrow.



Genomic islands in pathogenic and environmental microorganisms

Horizontal gene transfer is an important mechanism for the evolution of microbial genomes. **Pathogenicity islands** — mobile genetic elements that contribute to rapid changes in virulence potential — are known to have contributed to genome evolution by horizontal gene transfer in many bacterial pathogens.

Increasing evidence indicates that equivalent elements in non-pathogenic species — **genomic islands** — are important in the evolution of these bacteria, influencing traits such as antibiotic resistance, symbiosis and fitness, and adaptation in general.

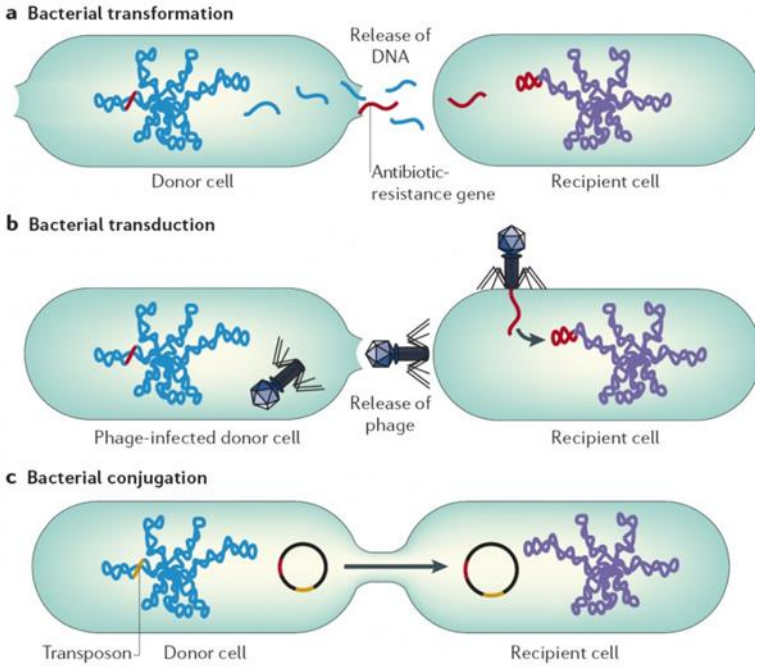
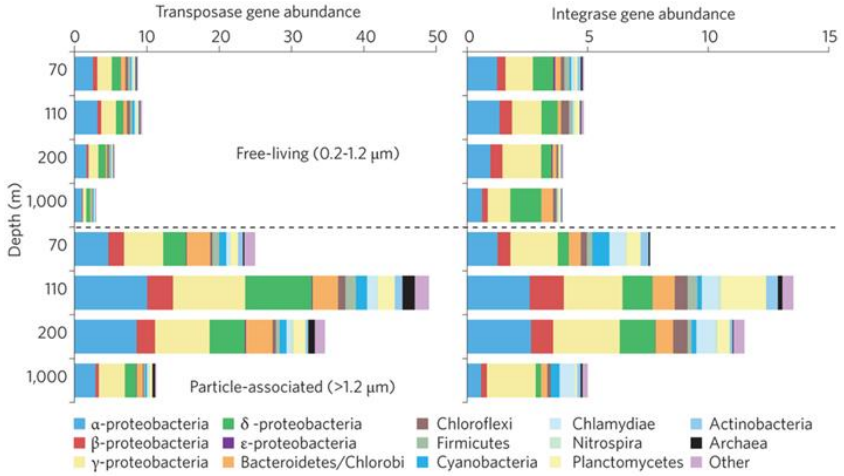


Figure 2 | **Horizontal gene transfer between bacteria. a** | Transformation occurs when naked DNA is released on lysis of an organism and is taken up by another organism. The antibiotic-resistance gene can be integrated into the chromosome or plasmid of the recipient cell. **b** | In transduction, antibiotic-resistance genes are transferred from one bacterium to another by means of bacteriophages and can be integrated into the chromosome of the recipient cell (lysogeny). **c** | Conjugation occurs by direct contact between two bacteria: plasmids form a mating bridge across the bacteria and DNA is exchanged, which can result in acquisition of antibiotic-resistance genes by the recipient cell. Transposons are sequences of DNA that carry their own recombination enzymes that allow for transposition from one location to another; transposons can also carry antibiotic-resistance genes.



Genes encoding enzymes involved in DNA transfer in marine microbes. Transposases and integrases, among the most abundant enzymes in life, facilitate the transfer of DNA within a genome or between genomes of different organisms. Depicted is the abundance of these genes — expressed as a ratio of counts of transposase or integrase genes relative to counts of a reference gene — in free-living and particle-attached marine microbial communities sampled from the oxygen minimum zone off the coast of Chile (colours indicate microbial groups from which these genes originate; see legend). The ***abundance of DNA-transfer genes is greater in particle-attached microbes than in free-living communities.*** Data are based on metagenomes from a marine oxygen minimum zone.

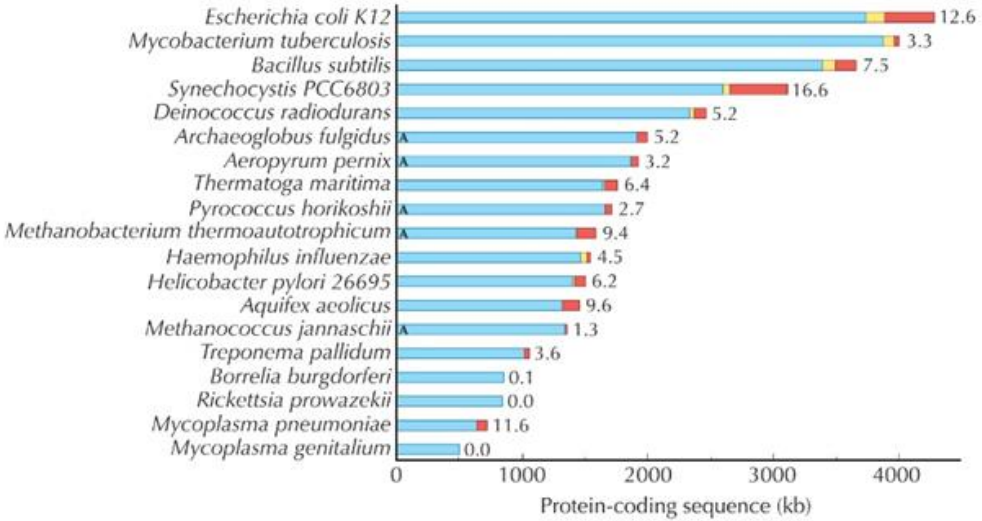
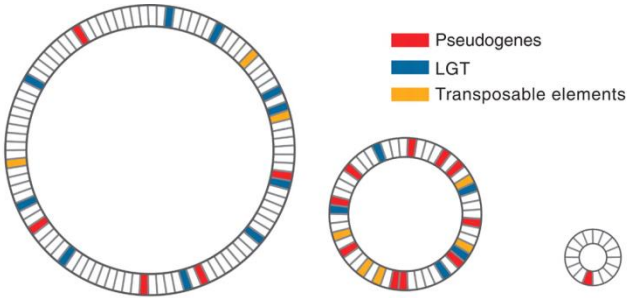


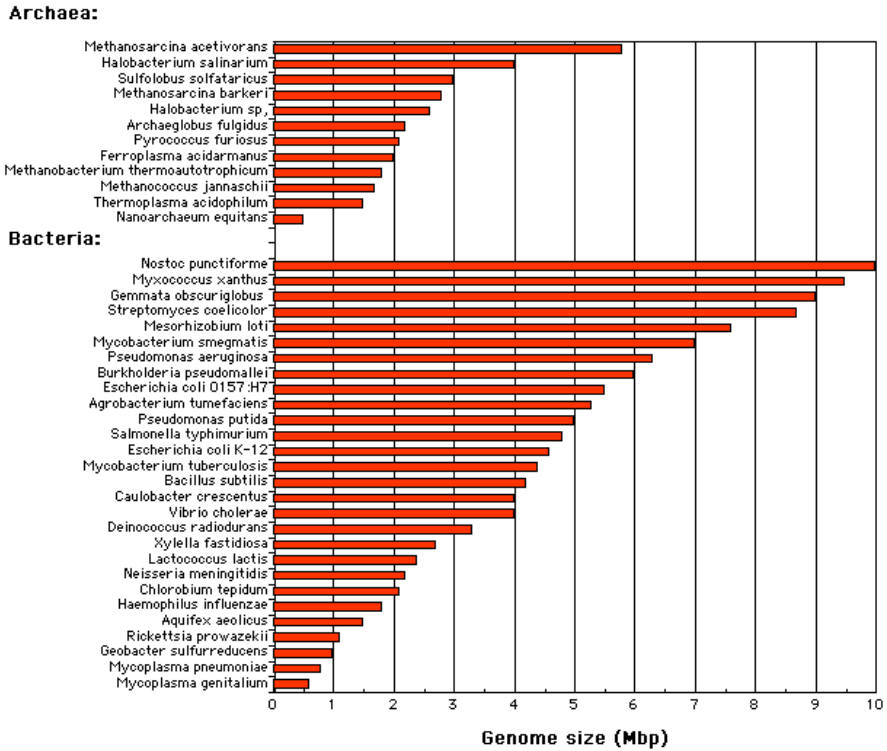
FIGURE 7.23. Estimation of the percentage of the genome that has been recently acquired by lateral transfer for different species of bacteria and archaea. *Blue,* "native" DNA (not acquired by transfer); *yellow,* known mobile DNA elements; *red,* other foreign DNA; and A, archaeal species.

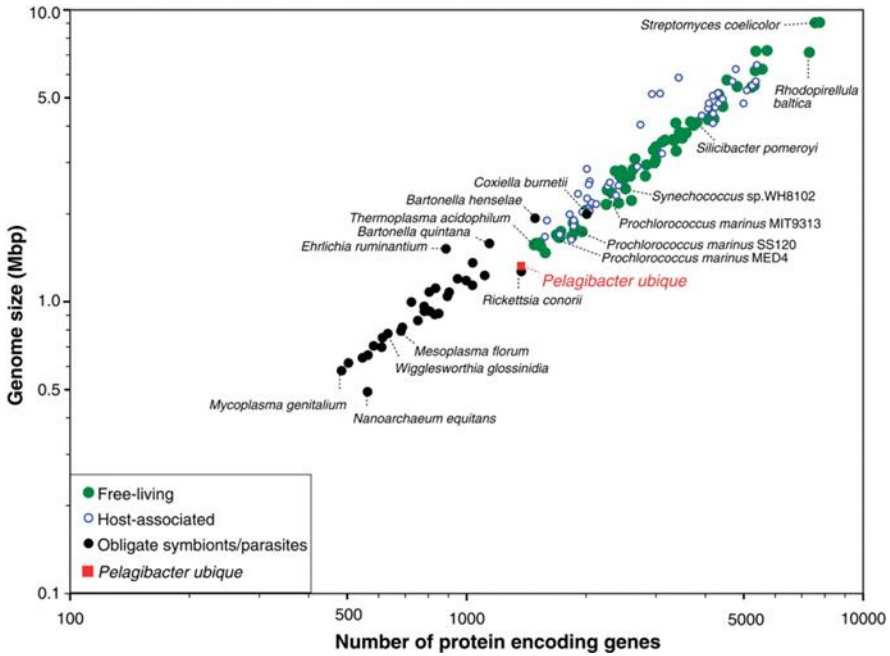
7.23, redrawn from Ochman H. et al., *Nature* **405:** 299–304, © 2000 Macmillan, www.nature.com

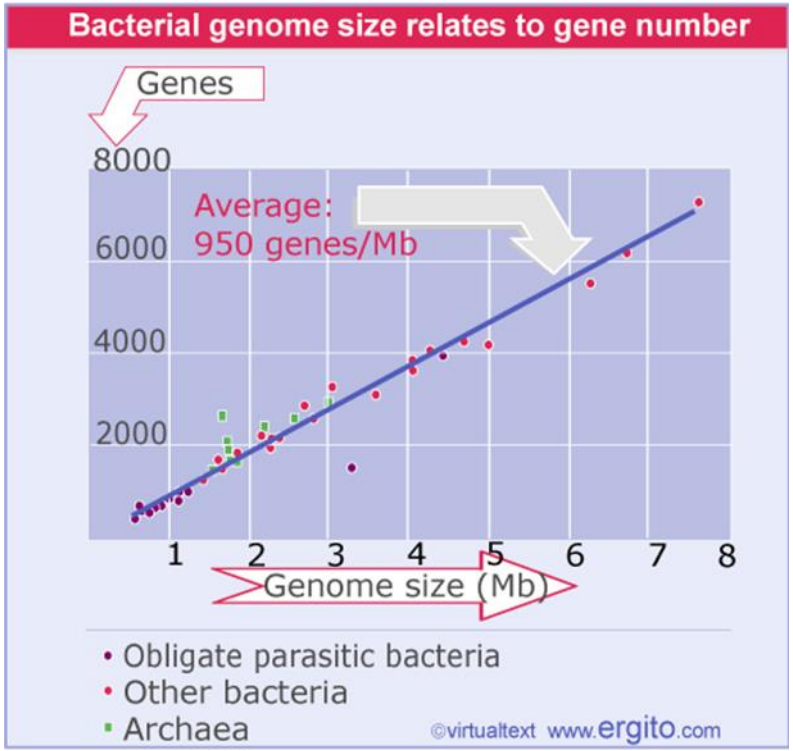
Trends in the size and contents of bacterial genomes



	Free-living	Recent or facultative pathogen	Obligate symbiont or pathogen
Genome size	Large (5-10 MB)	Intermediate (2-5 MB)	Small (0.5-1.5 MB)
Number of pseudogenes	Few	Many	Rare
Incidence of LGT	Frequent	Frequent to rare	Rare to none
Selfish genetic elements	Few	Common	Rare
Genome organization	Stable or unstable	Unstable	Stable
Effective population size	Large	Small	Small




Genome size (Mbp)
10.0
5.0
1.0
0.5
0.1
100
500
1000
5000
10000
Number of protein encoding genes
Streptomyces coelicolor
Rhodopirellula baltica
Silicibacter pomeroyi
Coxiella burnetii
Bartonella henselae
Thermoplasma acidophilum
Bartonella quintana
Ehrlichia ruminantium
Synechococcus sp.WH8102
Prochlorococcus marinus MIT9313
Prochlorococcus marinus SS120
Prochlorococcus marinus MED4
Pelagibacter ubique
Rickettsia conorii
Mesoplasma florum
Wigglesworthia glossinidia
Mycoplasma genitalium
Nanoarchaeum equitans
Free-living
Host-associated
Obligate symbionts/parasites
Pelagibacter ubique


Bacterial genome size relates to gene number
Genes
8000
6000
4000
2000
Average:
950 genes/Mb
1
2
3
4
5
6
7
8
Genome size (Mb)
Obligate parasitic bacteria
Other bacteria
Archaea
©virtualtext www.ergito.com

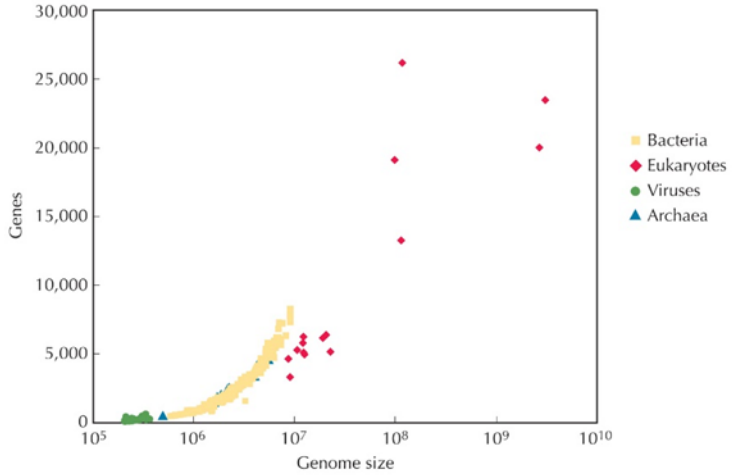


FIGURE 7.3. Genome size vs. number of protein-coding genes. The number of genes is highly correlated to genome size for bacteria, archaea, and viruses, but less so for eukaryotes. Many archaeal points (*blue triangles*) are hidden under bacterial ones (*yellow squares*).

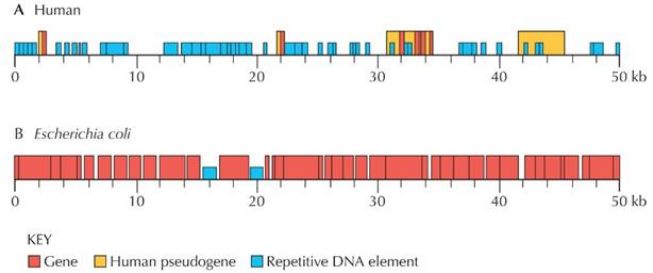


FIGURE 7.2. Genome density. Comparison of the genome density and content of humans and *Escherichia coli*. Each segment is 50 kb in length and represents (*A*) a portion of the human β T-cell receptor locus and (*B*) a region of the *E. coli* K12 genome. Note the much greater proportion of genes (*red boxes*) in *E. coli* compared to humans.

7.2, adapted from Brown T.A., *Genomes, 2e*, Fig 2.2, © 1999 Bios Scientific Publishers Ltd, by permission of Taylor & Francis Books U.K.

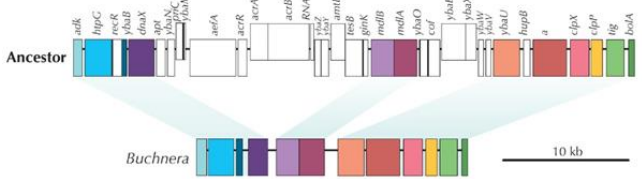
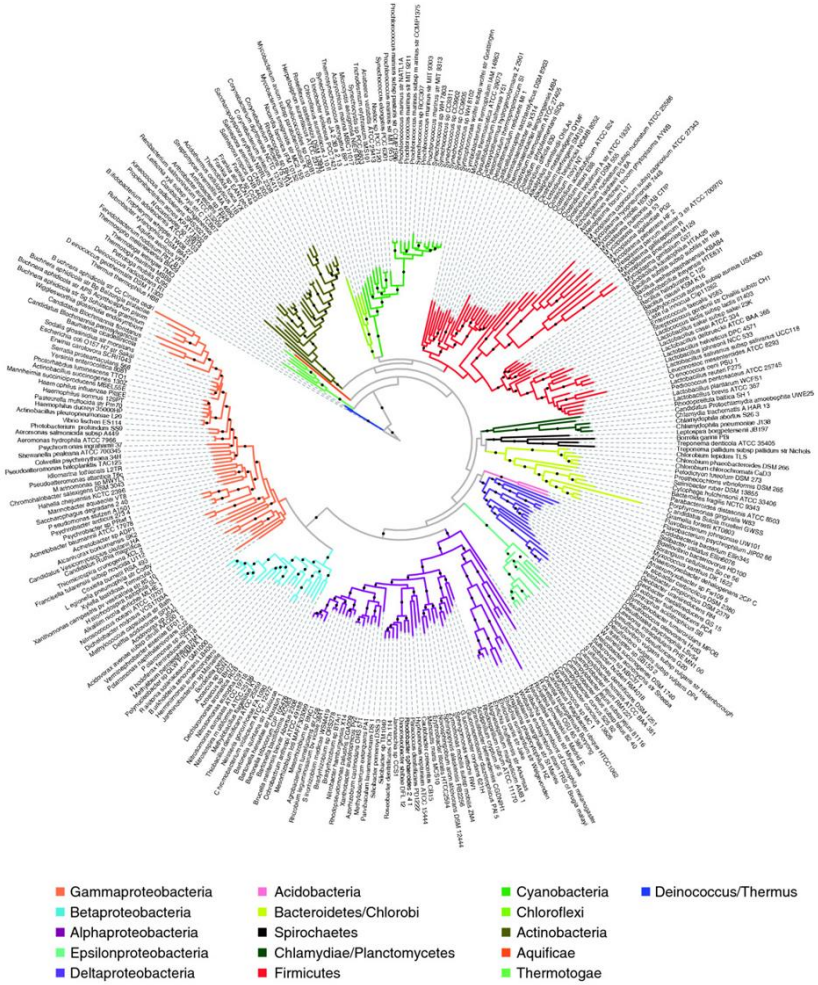


FIGURE 7.5. Genome reduction in *Buchnera* endosymbionts of aphids. A fragment of two genomes is shown. (*Top row*) The putative ancestor of all aphid endosymbionts in the *Buchnera* genus. (*Bottom row*) The genome of the symbionts today. The massive amounts of gene loss are indicated by the genes colored *white* in the ancestral genome that are missing from the modern genome below. Orthologous genes between the two genomes are shown in the same color. Note the conservation of gene order between the two genomes despite the gene loss. The direction of gene transcription is indicated by the gene box being shifted above or below the *black line*.

7.5, redrawn from Moran N.A. et al., *Genome Biol.* **2:** research0054.1–0054.12, © 2001 Nancy A. Moran

A genomic tree of bacterial life



A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea

Dongying Wu[1,2], Philip Hugenholtz[1], Konstantinos Mavromatis[1], Rüdiger Pukall[3], Eileen Dalin[1], Natalia N. Ivanova[1], Victor Kunin[1], Lynne Goodwin[4], Martin Wu[5], Brian J. Tindall[3], Sean D. Hooper[1], Amrita Pati[1], Athanasios Lykidis[1], Stefan Spring[3], Iain J. Anderson[1], Patrik D'haeseleer[1,6], Adam Zemla[6], Mitchell Singer[5], Alla Lapidus[1], Matt Nolan[1], Alex Copeland[1], Cliff Han[4], Feng Chen[1], Jan-Fang Cheng[1], Susan Lucas[1], Cheryl Kerfeld[1], Elke Lang[3], Sabine Gronow[3], Patrick Chain[1,4], David Bruce[4], Edward M. Rubin[1], Nikos C. Kyrpides[1], Hans-Peter Klenk[3] & Jonathan A. Eisen[1,2]

Sequencing of bacterial and archaeal genomes has revolutionized our understanding of the many roles played by microorganisms[1]. There are now nearly 1,000 completed bacterial and archaeal genomes available[2], most of which were chosen for sequencing on the basis of their physiology. As a result, the perspective provided by the currently available genomes is limited by a highly biased phylogenetic distribution[3–5]. To explore the value added by choosing microbial genomes for sequencing on the basis of their evolutionary relationships, we have sequenced and analysed the genomes of 56 culturable species of Bacteria and Archaea selected to maximize phylogenetic coverage. Analysis of these genomes demonstrated pronounced benefits (compared to an equivalent set of genomes randomly selected from the existing database) in diverse areas including the reconstruction of phylogenetic history, the discovery of new protein families and biological properties, and the prediction of functions for known genes from other organisms. Our results strongly support the need for systematic 'phylogenomic' efforts to compile a phylogeny-driven 'Genomic Encyclopedia of Bacteria and Archaea' in order to derive maximum knowledge from existing microbial genome data as well as from genome sequences to come.

Maximum-likelihood phylogenetic tree of the bacterial domain based on a concatenated alignment of 31 broadly conserved protein-coding genes. Phyla are distinguished by colour of the branch and GEBA genomes are indicated in red in the outer circle of species names.

Prokaryotic genomes: the emerging paradigm of genome-based microbiology

Eugene V Koonin* and Michael Y Galperin

Current Opinion in Genetics & Development 1997, **7**:757–763

Comparative analysis of the complete sequences of seven bacterial and three archaeal genomes leads to the first generalizations of emerging genome-based microbiology. Protein sequences are, generally, highly conserved, with ~70% of the gene products in bacteria and archaea containing ancient conserved regions. In contrast, there is little conservation of genome organization, except for a few essential operons. The most striking conclusions derived by comparison of multiple genomes from phylogenetically distant species are that the number of universally conserved gene families is very small and that multiple events of horizontal gene transfer and genome fusion are major forces in evolution.

Conclusions: the striking complexity of the evolutionary process

Genome comparisons suggest that, in the evolution of prokaryotes, horizontal gene transfer has been common and intense. Somewhat surprisingly, detailed analysis of the *M. jannaschii* protein sequences showed that the number of proteins with a greater similarity to bacterial homologs significantly exceeds the number of those that are closer to eukaryotic homologs ([21•]; Figure 3). In agreement with earlier observations, protein components of replication, transcription, and translation machinery resemble their eukaryotic counterparts [46,47] but the great majority of metabolic enzymes, proteins involved in cell wall biogenesis, and uncharacterized proteins appear to be of bacterial origin. Qualitatively similar relationships were observed within the available set of proteins from the Crenarchaeote *Sulfolobus solfataricus* [21•]. The prevalence of genes of apparent bacterial origin in archaeal genomes suggests a genetic basis for the prokaryotic phenotype [48].

The simplest explanation for the observed phylogenetic breakdown of sequence conservation in archaeal proteins seems to imply multiple instances of horizontal gene transfer and perhaps even a genome fusion event(s), although alternative scenarios on the basis of major variations in gene mutation rates [49] remain to be considered. Clearly, the eukaryotic genomes also are

Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world

Eugene V. Koonin* and Yuri I. Wolf

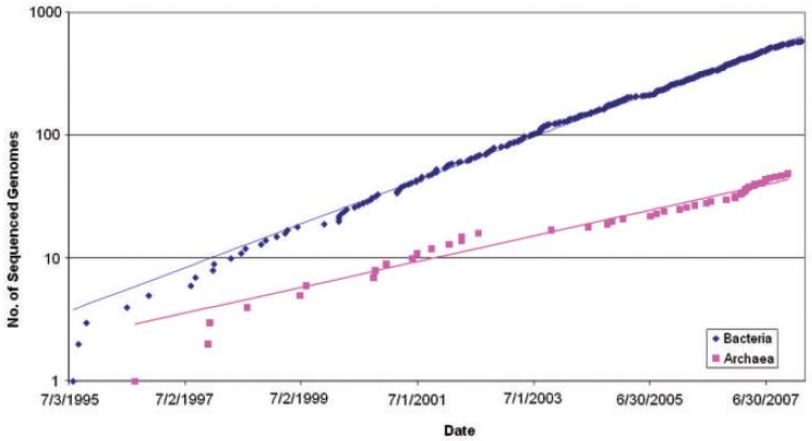


Figure 1. The temporal dynamics of genome sequencing for bacteria and archaea. Bacteria: doubling time ~20 months. Archaea: doubling time ~34 months.

The first bacterial genome was sequenced in 1995, and the first archaeal genome in 1996. Soon after these breakthroughs, an exponential rate of genome sequencing was established, with a doubling time of approximately 20 months for bacteria and approximately 34 months for archaea. Comparative analysis of the hundreds of sequenced bacterial and dozens of archaeal genomes leads to several generalizations on the principles of genome organization and evolution. A crucial finding that enables functional characterization of the sequenced genomes and evolutionary reconstruction is that the majority of archaeal and bacterial genes have conserved orthologs in other, often, distant organisms. However, comparative genomics also shows that horizontal gene transfer (HGT) is a dominant force of prokaryotic evolution, along with the loss of genetic material resulting in genome contraction. A crucial component of the prokaryotic world is the mobilome, the enormous collection of viruses, plasmids and other selfish elements, which are in constant exchange with more stable chromosomes and serve as HGT vehicles. Thus, the prokaryotic genome space is a tightly connected, although compartmentalized, network, a novel notion that undermines the 'Tree of Life' model of evolution and requires a new conceptual framework and tools for the study of prokaryotic evolution.

Table 1 | **Exploitation of bacterial genome fluidity for diagnostic and health-care applications**

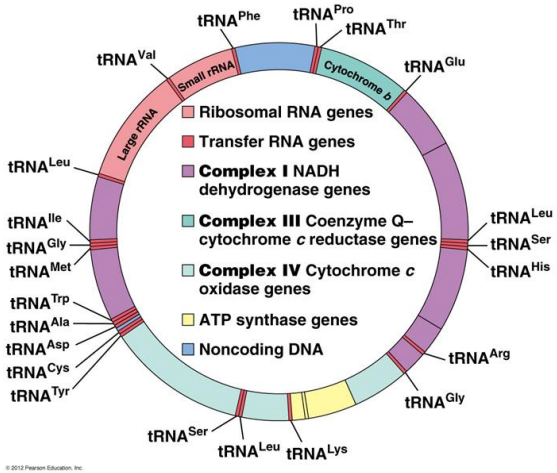
Molecular marker or assay	Genetic or molecular change	Application to health care
Escherichia coli		
Virulence genes (for example, *stx* and *kpsK1*)	Presence or absence of the locus; expression of functional virulence factors	Diagnostic markers and antigens; markers of infection
Genomic islands, plasmids and phages	Presence or absence of the locus; expression of functional virulence factors	Diagnostic markers and antigens; markers of infection
Black holes	Presence or absence of the locus; expression of functional virulence factors	Diagnostic markers and antigens; markers of infection
Core-genome MLST	Evolution of housekeeping genes	Lineage identification
SNPs	Base substitution in candidate genes	Screening for antimicrobial resistance, molecular epidemiology and evolution
Helicobacter pylori		
Indel of *cag* right junction	Small-scale insertions and deletions at the right junction of the *cag* PAI	Markers of human migration and *H. pylori* lineage identification
cagA	Presence or absence of the locus; expression of functional toxin	Diagnostic markers and antigens; markers of invasive gastric inflammation
CagA EPIYA motif	Strain-specific tyrosine phosphorylation of CagA	Geographical markers of gastric-cancer predisposition
Plasticity-region cluster	Presence or absence of loci	Putative virulence genes; possible interventional targets
Core-genome MLST	Evolution of housekeeping genes	Human-migration studies; lineage identification
Mycobacteria		
RDs	Reductive evolution	Diagnostic and genotyping applications for epidemiology; vaccine candidates (RD1)
DRs	Indel in DR region	Spoligotyping for epidemiology; strain identification
IS*6110* RFLPs	Mobile-element instability	Outbreak investigation; strain identification
DUs	Tandem duplication	Quality control of BCG vaccines
SNPs	Base substitution in candidate genes	Screening for antimicrobial resistance, molecular epidemiology and evolution
MIRUs (minisatellites)	Tandem-repeat tract expansion or shrinkage	Large-scale, high-throughput analysis that can monitor spread and transmission dynamics
Microsatellites	Replication errors	Molecular epidemiology; strain identification
PE–PPE protein family	Variable-repeat motifs and antigenic variation	Diagnostic markers and antigens; putative virulence markers
Erp protein family	Plasticity of nucleotide-repeat motifs	Diagnostic antigens and virulence markers; vaccine candidates (RD1)

BCG, *Mycobacterium bovis* bacille Calmette–Guérin; DR, direct repeat; DU, genome-duplication marker; MIRU, mycobacterial repetitive interspersed unit; MLST, multilocus sequence typing; PAI, pathogenicity island; RD, region of difference; RFLP, restriction-fragment length polymorphism; SNP, single-nucleotide polymorphism.

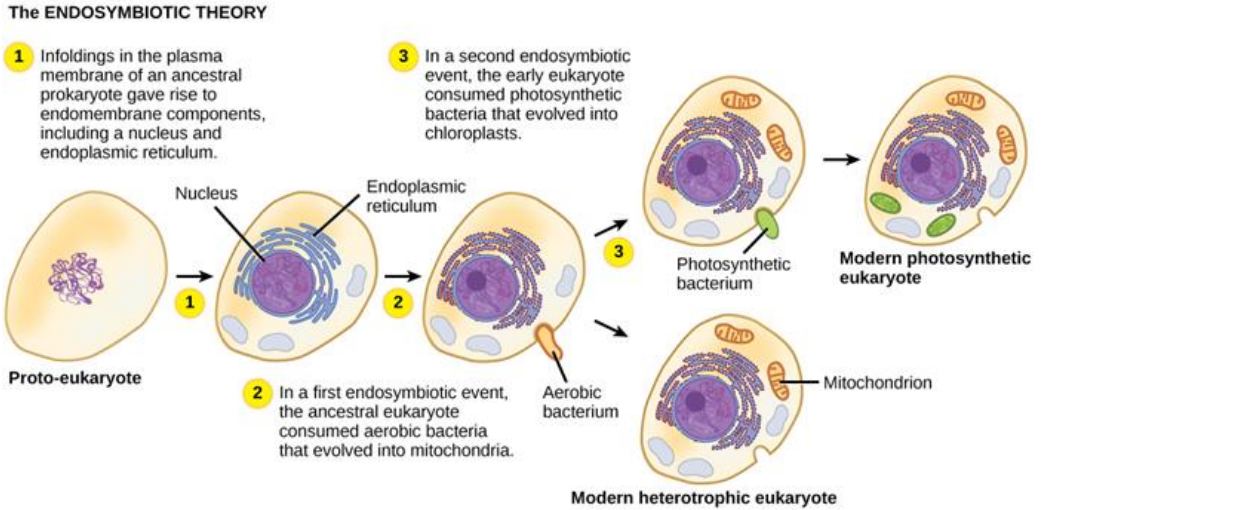
Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention

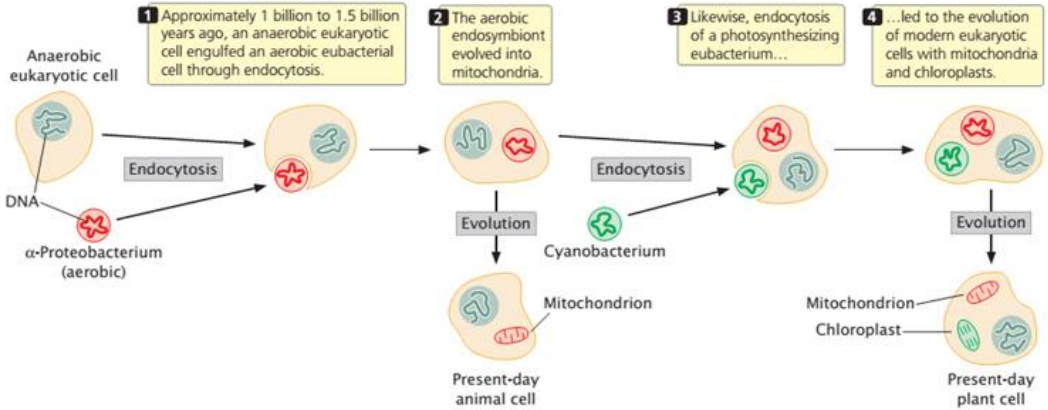
Niyaz Ahmed, Ulrich Dobrindt, Jörg Hacker and Seyed E. Hasnain

Abstract | The increasing availability of DNA-sequence information for multiple pathogenic and non-pathogenic variants of individual bacterial species has indicated that both DNA acquisition and genome reduction have important roles in genome evolution. Such genomic fluidity, which is found in human pathogens such as *Escherichia coli*, *Helicobacter pylori* and *Mycobacterium tuberculosis*, has important consequences for the clinical management of the diseases that are caused by these pathogens and for the development of diagnostics and new molecular epidemiological methods.

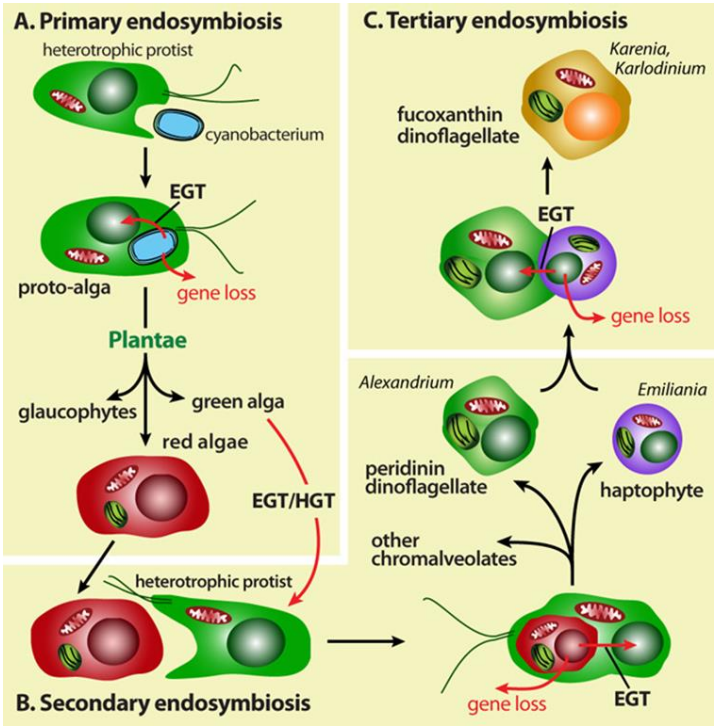


Plastids/Plastidi





21.6 The endosymbiotic theory proposes that mitochondria and chloroplasts in eukaryotic cells arose from eubacteria.



The Origin of Plastids. A three-panel schematic shows the concepts of primary, secondary, and tertiary endosymbiosis. Arrows and simplified illustrations of cells that include nuclei, mitochondria, and plastids are used to show the steps involved in endosymbiosis.

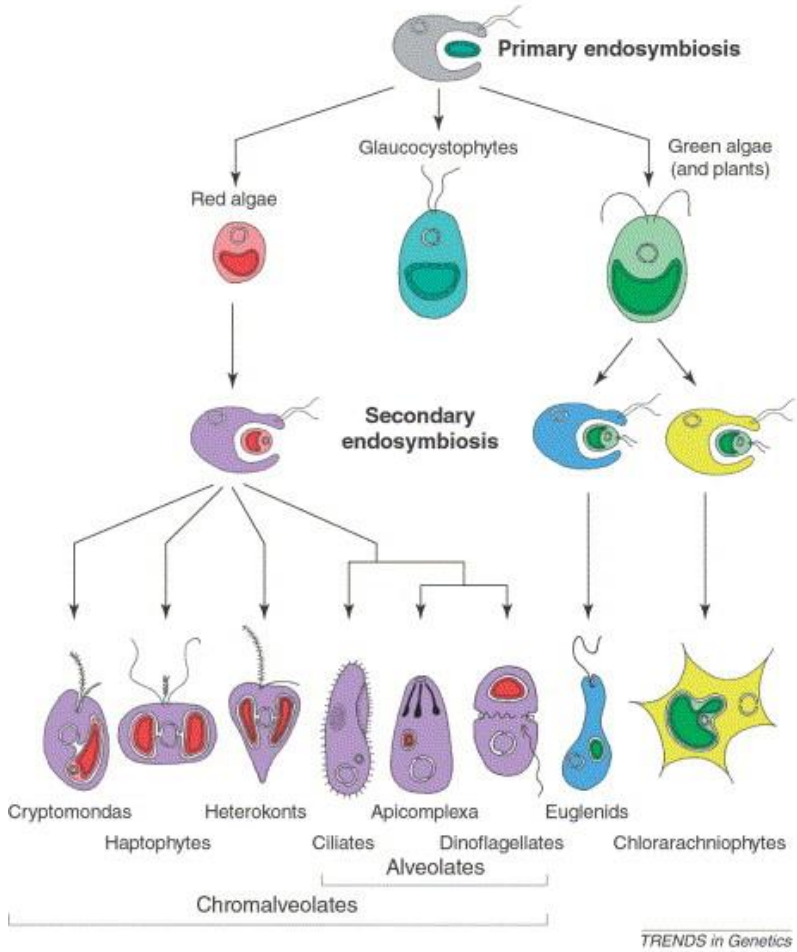


Table 1 | **Sizes and coding content of some organelle and prokaryote genomes**

Genome	Length [kbp]	Number of protein-coding genes	GenBank accession number
Algae			
cp *Porphyra purpurea*	191	200	PPU38804
cp *Cyanidium caldarium*	165	197	AF022186
cp *Guillardia theta*	122	148	AF041468
cp *Cyanophora paradoxa*	136	136	CPU30821
cp *Odontella sinensis*	120	124	OSCHLPXX
cp *Euglena gracilis*	143	58	CLEGCGA
Land plants			
cp *Marchantia polymorpha*	121	84	CHMPXX
cp *Chlorella vulgaris*	151	78	AB001684
cp *Nicotiana tabacum*	156	76	CHNTXX
cp *Oryza sativa*	134	76	X15001
cp *Zea mays*	140	76	ZMA86563
cp *Pinus thunbergii*	120	69	PINCPTRPG
Non-phosynthetic plastids			
cp *Toxoplasma gondii*	35	26	U87145
cp *Eimeria tenella*	35	28	AY217738
cp *Epifagus virginiana*	70	21	EPFCPCG
Cyanobacteria			
Synechocystis sp.	3573	3168	AB001339
Prochlorococcus marinus	1660	1884	NC_005071
Nostoc PCC 7120	6413	5368	AP003602
Nostoc punctiforme	~9000	~7400	http://www.jgi/doe/gov
Plants and algae			
mt *Pylaiella littoralis*	59	52	NC_003055
mt *Marchantia polymorpha*	187	41	MPOMTCG
mt *Laminaria digitata*	38	39	AJ344328
mt *Cyanidioschyzon merolae*	32	34	NC_000887
mt *Arabidopsis thaliana*	367	31	MIATGENA
mt *Chondrus crispus*	26	25	MTCGGNME
mt *Scenedesmus obliquus*	43	20	NC_002254
Various protists and fungi			
mt *Reclinomonas americana*	69	67	NC_001823
mt *Malawimonas jakobiformis*	47	49	AF295545
mt *Naegleria gruberi*	50	46	NC_002573
mt *Rhodomonas salina*	48	44	NC_002572
mt *Dictyostelium discoidaum*	56	40	NC_000895
mt *Phytophthora infestans*	38	40	NC_002387
mt *Acanthamoeba castellanii*	42	36	U12386
mt *Cafeteria roenbergensis*	43	34	NC_000946
mt *Monosiga brevicollis*	77	32	AF538053
mt *Physarum polyocephalum*	63	20	AB027295
mt *Harpochytrium* sp	24	14	AY182006
mt *Candida albicans*	40	13	NC_002653
mt *Cryptococcus neoformans*	25	12	NC_004336
mt *Plasmodium falciparum*	6	3	NC_001677
Anaerobic mitochondria			
mt Hydrogenosomes*	0	0	
α-proteobacteria			
Caulobacter crescentus	4017	3767	AE006573
Mesorhizobium loti	7596	7281	BA000012
Bradyrhizobium japonicum	~9100	~8300	BA000040
Yeast			
(nuclear)	13,469	6,327	http://www.ebi.ac.uk

An excellent, up-to-date list of sequenced organelle genomes is available at http://megasun.bch.umontreal.ca/ogmp/projects/other/all_list.html. Prokaryote data was gratefully received from http://dna-res.kazusa.or.jp and http://www.jgi.doe.gov/JGI_microbial/html.
*Hydrogenosomes are anaerobic forms of mitochondria that usually lack a genome. cp, chloroplast genome; mt, mitochondrial genome.

Table 21.1 Sizes of mitochondrial genomes in selected organisms

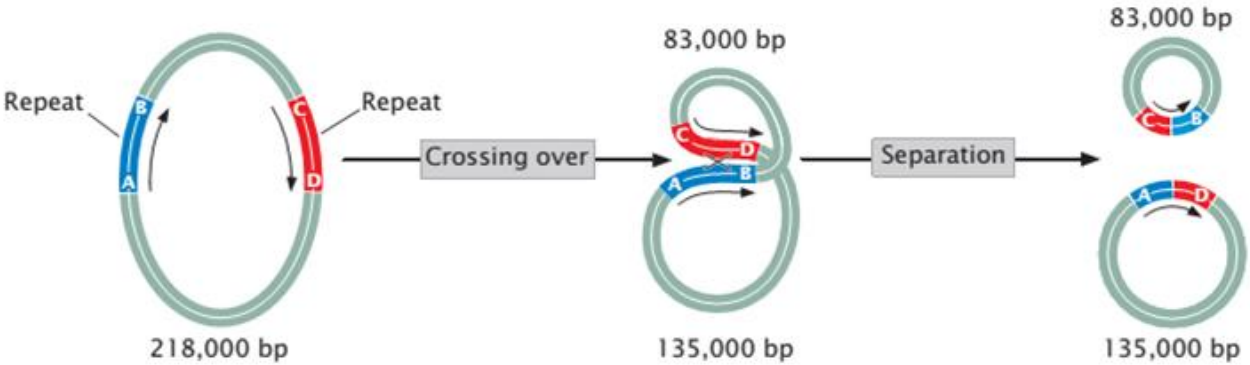
Organism	Size of mtDNA (bp)
Pichia canadensis (fungus)	27,694
Podospora anserina (fungus)	100,314
Saccharomyces cerevisiae (fungus)	85,779*
Drosophila melanogaster (fruit fly)	19,517
Lumbricus terrestris (earthworm)	14,998
Xenopus laevis (frog)	17,553
Mus musculus (house mouse)	16,295
Homo sapiens (human)	16,569
Chlamydomonas reinhardtii (green alga)	15,758
Plasmodium falciparum (protist)	5,966
Paramecium aurelia (protist)	40,469
Arabidopsis thaliana (plant)	166,924
Cucumis melo (plant)	2,400,000

*Size varies among strains.

Mitochondrial DNA Varies Widely in Size and Organization

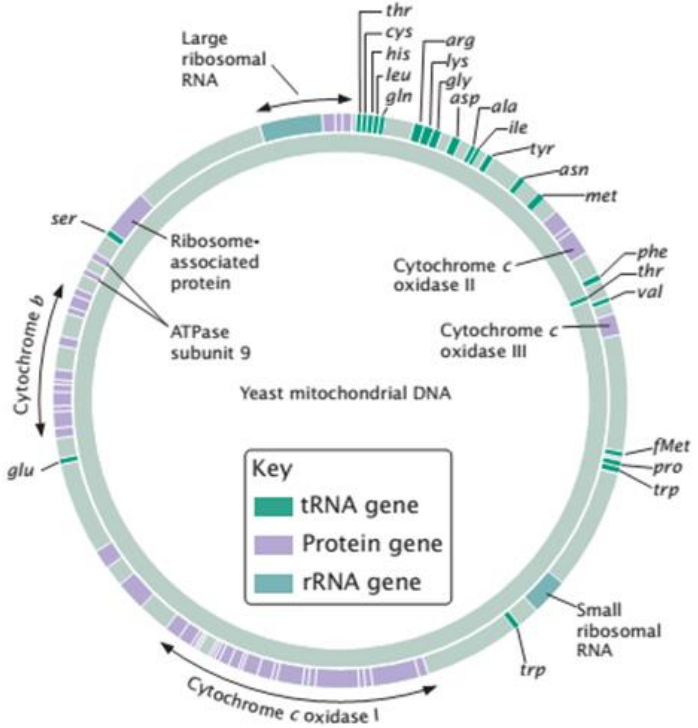
Table 21.3 Sizes of chloroplast genomes in selected organisms

Organism	Size of cpDNA (bp)
Euglena gracilis (protist)	143,172
Porphyra purpurea (red alga)	191,028
Chlorella vulgaris (green alga)	150,613
Marchantia polymorpha (liverwort)	121,024
Nicotiana tabacum (tobacco)	155,939
Zea mays (corn)	140,387
Pinus thunbergii (black pine)	119,707



21.9 Size variation in plant mtDNA can be generated through recombination between direct repeats. In turnips, the mitochondrial genome consists of a "master circle" of 218,000 bp; crossing over between the direct repeats produces two smaller circles of 135,000 bp and 83,000 nucleotide pairs.

Flowering-plant mtDNA Flowering plants (angiosperms) **have the largest and most-complex mitochondrial genomes known**; their mitochondrial genomes range in ***size from 186,000 bp in white mustard to 2,400,000 bp in muskmelon***. ***Even closely related plant species may differ greatly in the sizes of their mtDNA***. Part of the extensive size variation in the mtDNA of flowering plants can be explained by the presence of large direct repeats, which constitute large parts of the mitochondrial genome. Crossing over between these repeats can generate multiple circular chromosomes of different sizes. The mitochondrial genome in turnips, for example, consists of a "master circle" consisting of 218,000 bp that has direct repeats. Homologous recombination between the repeats can generate two smaller circles of 135,000 bp and 83,000 bp. Other species contain several **direct repeats**, providing possibilities for **complex crossing-over events that may increase or decrease the number and sizes of the circles**.



21.8 The yeast mitochondrial genome, consisting of 78,000 bp, contains much noncoding DNA.

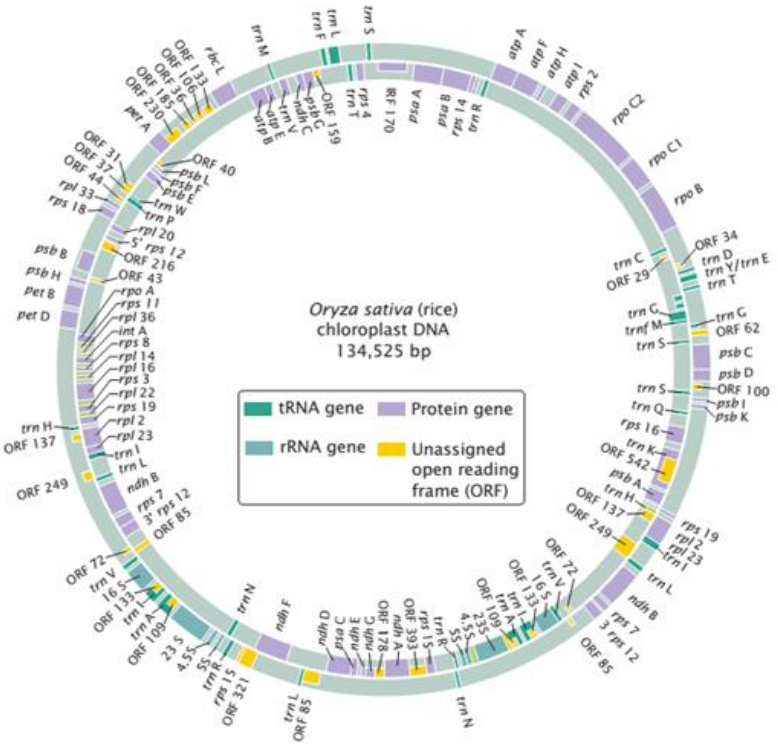
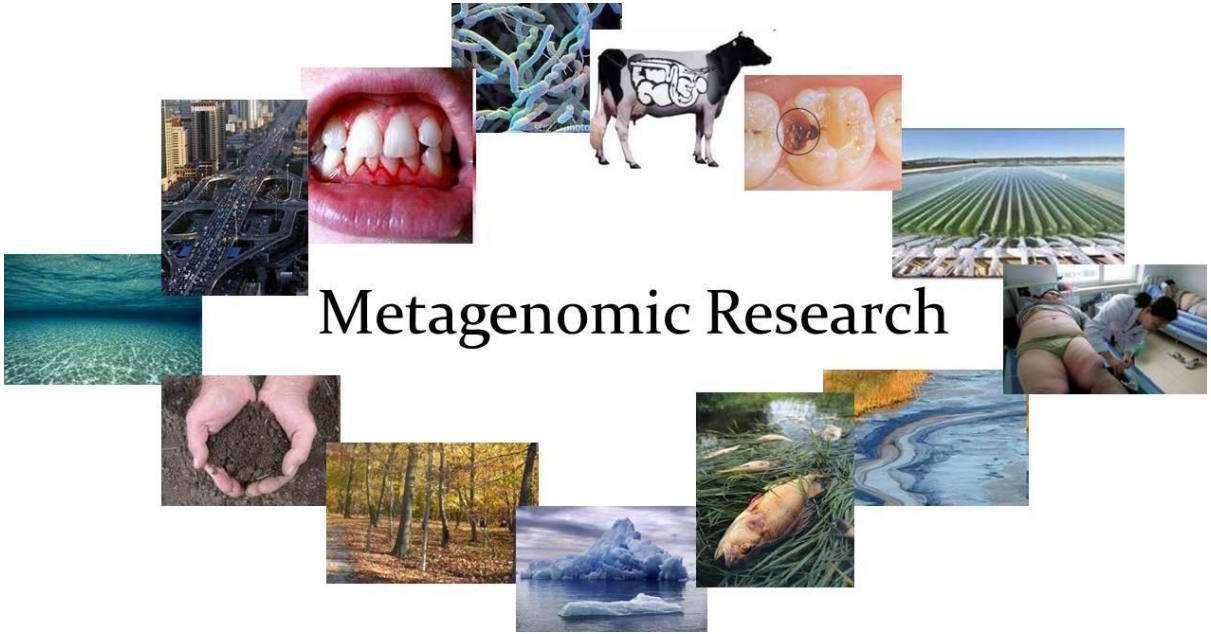


Table 21.4 Comparison of nuclear eukaryotic, eubacterial, mitochondrial, and chloroplast genomes

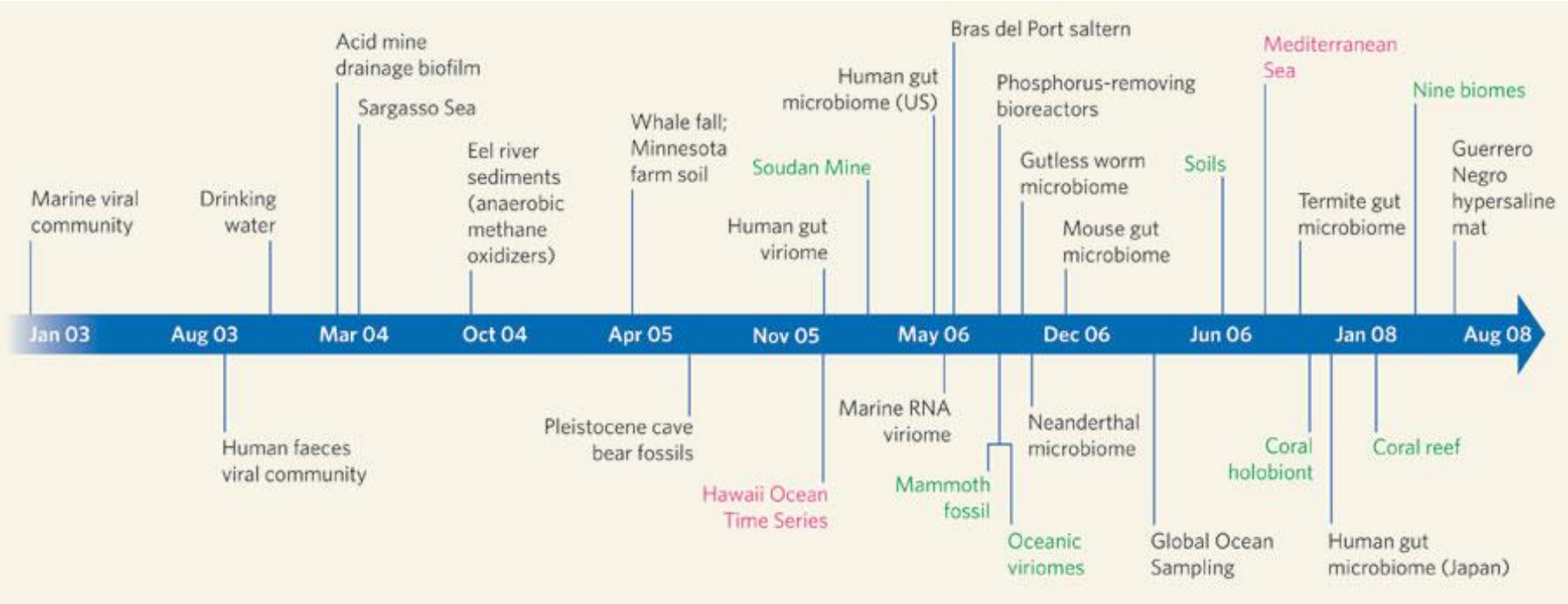
Characteristic	Eukaryotic Genome	Eubacterial Genome	Mitochondrial Genome	Chloroplast Genome
Genome consists of double-stranded DNA	Yes	Yes	Yes	Yes
Circular	No	Yes	Most	Yes
Histone proteins	Yes	No	No	No
Size	Large	Small	Small	Small
Number of molecules per genome	Several	One	One in animals; several in some plants	One
Pre-mRNA introns	Common	Absent	Absent	Absent
Group I introns	Present	Present	Present	Present
Group II introns	Absent	Present	Present	Present
Polycistronic mRNA	Uncommon	Common	Present	Common
5′ cap added to mRNA	Yes	No	No	No
3′ poly(A) tail added to mRNA	Yes	No	Some in animals	No
Shine–Dalgarno sequence in 5′ untranslated region of mRNA	No	Yes	Rare	Some
Nonuniversal codons	Rare	Rare	Yes	No
Extended wobble	No	No	Yes	No
Translation inhibited by tetracycline	No	Yes	Yes	Yes



Metagenomics/Metagenomika

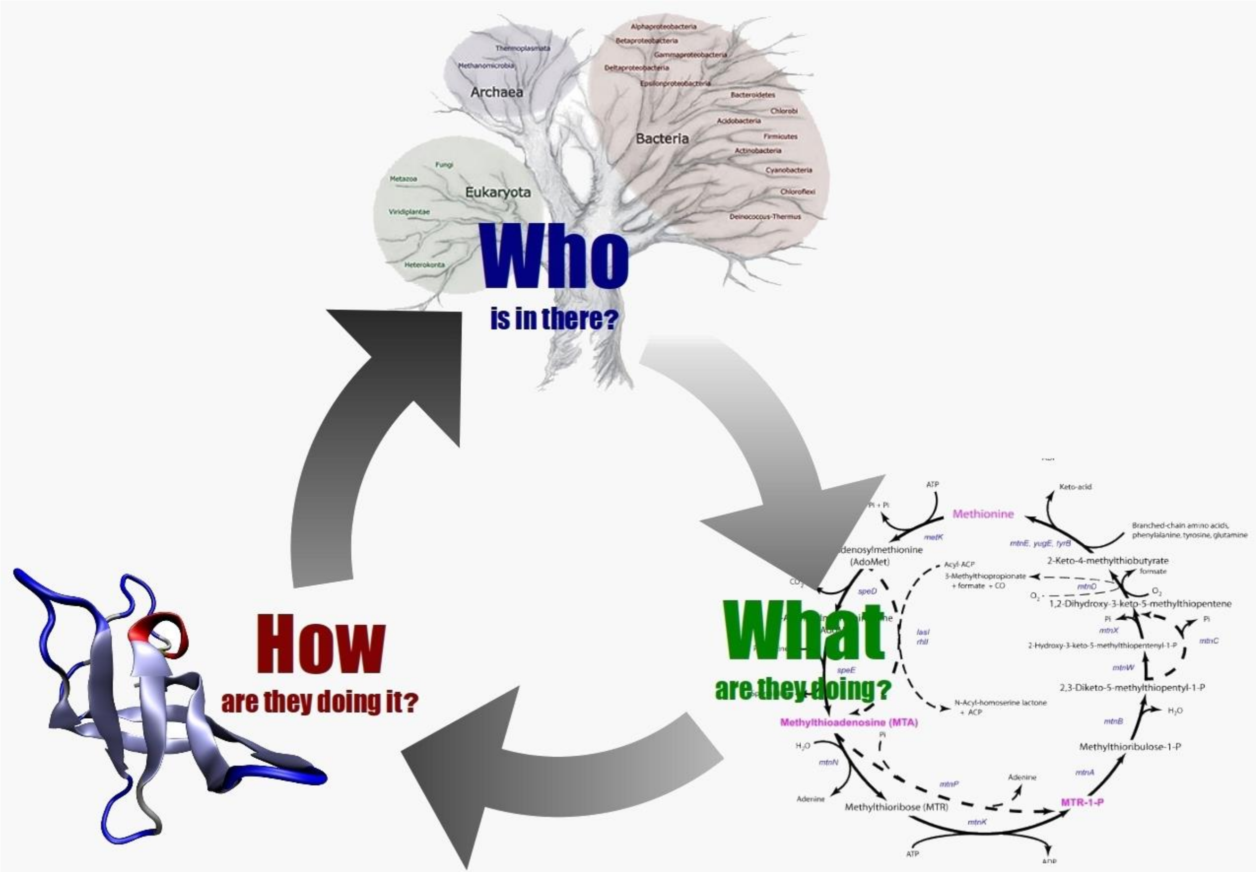
Metagenomics is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as **environmental genomics, ecogenomics or community genomics**. While traditional microbiology and microbial genome sequencing and genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes (often the 16S rRNA gene) to produce a profile of diversity in a natural sample. Such work revealed that **the vast majority of microbial biodiversity had been missed by cultivation-based methods**. Recent studies use "shotgun" Sanger sequencing or massively parallel pyrosequencing to get largely unbiased samples of all genes from all the members of the ***sampled communities***. Because of its ability to reveal the previously hidden diversity of microscopic life, **metagenomics offers a powerful lens for viewing the microbial world that has the potential to revolutionize understanding of the entire living world.**

The **term "metagenomics" was first used** by Jo Handelsman, Jon Clardy, Robert M. Goodman, and others, and first appeared in publication in **1998**. The term metagenome referenced the idea that a collection of genes sequenced from the environment could be analyzed in a way analogous to the study of a single genome. Recently, Kevin Chen and Lior Pachter (researchers at the University of California, Berkeley) defined metagenomics as "**the application of modern genomics techniques to the study of communities of microbial organisms directly in their natural environments, bypassing the need for isolation and lab cultivation of individual species**".



Timeline of sequence-based metagenomic projects showing the variety of environments sampled since 2002

The **oceanic viriomes** (all viruses in a habitat) (August 2006) were from the Sargasso Sea, Gulf of Mexico, coastal British Columbia and the Arctic Ocean. The **nine biomes** (March 2008) were ***stromatolites, fish gut, fish ponds, mosquito viriome, human-lung viriome, chicken gut, bovine gut and marine viriome***. The different technologies used are dye-terminator shotgun sequencing (black), fosmid library sequencing (pink) and pyrosequencing (green).


Who
is in there?
Archaea
Bacteria
Eukaryota
Thermoplasmata
Methanomicrobia
Alphaproteobacteria
Betaproteobacteria
Gammaproteobacteria
Deltaproteobacteria
Epsilonproteobacteria
Bacteroidetes
Chlorobi
Acidobacteria
Firmicutes
Actinobacteria
Cyanobacteria
Chloroflexi
Deinococcus-Thermus
Fungi
Metazoa
Viridiplantae
Heterokonta
What
are they doing?
Methionine
2-Keto-4-methylthiobutyrate
1,2-Dihydroxy-3-keto-5-methylthiopentene
2,3-Diketo-5-methylthiopentyl-1-P
Methylthioribulose-1-P
MTR-1-P
Methylthioribose (MTR)
Methylthioadenosine (MTA)
N-Acyl-homoserine lactone + ACP
How
are they doing it?

Applications of Metagenomics

Metagenomics has the potential to advance knowledge in a wide variety of fields. It can also be applied to **solve practical challenges in medicine, engineering, agriculture, sustainability and ecology**.

Medicine: Microbial communities play a key role in preserving human health, but their composition and the mechanism by which they do so remains mysterious. ***Metagenomic sequencing is being used to characterize the microbial communities from 15-18 body sites from at least 250 individuals***. This is part of the ***Human Microbiome initiative*** with primary goals to ***determine if there is a core human microbiome, to understand the changes in the human microbiome that can be correlated with human health, and to develop new technological and bioinformatics tools to support these goals***.

Biofuel: Bioreactors allow the observation of microbial communities as they convert biomass into cellulosic ethanol. Biofuels are fuels derived from biomass conversion, as in the conversion of cellulose contained in corn stalks, switchgrass, and other biomass into cellulosic ethanol. This process is dependent upon *microbial consortia that transform the cellulose into sugars, followed by the fermentation of the sugars into ethanol*. Microbes also produce a variety of sources of bioenergy including **methane and hydrogen**. The efficient industrial-scale deconstruction of biomass requires ***novel enzymes with higher productivity and lower cost***. Metagenomic approaches to the analysis of complex microbial communities allow the **targeted screening of enzymes with industrial applications in biofuel production, such as glycoside hydrolases**. Furthermore, knowledge of how these microbial communities function is required to control them, and metagenomics is a key tool in their understanding. Metagenomic approaches allow comparative analyses between convergent microbial systems like **biogas fermenters or insect herbivores such as the fungus garden of the leafcutter ants.**

Environmental remediation: Metagenomics can **improve strategies for monitoring the impact of pollutants on ecosystems and for cleaning up contaminated environments**. Increased understanding of how microbial communities cope with pollutants improves assessments of the **potential of contaminated sites to recover from pollution and increases the chances of bioaugmentation or biostimulation trials to succeed**.

Biotechnology: Microbial communities produce a **vast array of biologically active chemicals that are used in competition and communication**. Many of the drugs in use today were originally uncovered in microbes; recent progress in mining the rich genetic resource of non-culturable microbes has led to the discovery of new genes, enzymes, and natural products. The application of metagenomics has allowed the development of commodity and **fine chemicals, agrochemicals and pharmaceuticals** where the benefit of enzyme-catalyzed chiral synthesis is increasingly recognized. Two types of analysis are used in the bioprospecting of metagenomic data: function-driven screening for an expressed trait, and sequence-driven screening for DNA sequences of interest. In practice, experiments make use of a combination of both functional and sequence-based approaches based upon the function of interest, the complexity of the sample to be screened, and other factors.

Agriculture: The **soils** in which plants grow are inhabited by microbial communities, with **one gram of soil containing around 10^9-10^{10} microbial cells which comprise about one gigabase of sequence information**. The ***microbial communities which inhabit soils are some of the most complex known to science, and remain poorly understood despite their economic importance. Microbial consortia perform a wide variety of ecosystem services necessary for plant growth, including fixing atmospheric nitrogen, nutrient cycling, disease suppression, and sequester iron and other metals***. Functional metagenomics strategies are being used to explore the interactions between plants and microbes through cultivation-independent study of these microbial communities. By allowing insights into the role of previously uncultivated or rare community members in nutrient cycling and the promotion of plant growth, metagenomic approaches can contribute to ***improved disease detection in crops and livestock and the adaptation of enhanced farming practices which improve crop health by harnessing the relationship between microbes and plants***.

Ecology: Metagenomics can provide valuable insights into the **functional ecology of environmental communities**. Metagenomic analysis of the bacterial consortia found in the defecations of Australian sea lions suggests that nutrient-rich sea lion faeces may be an important nutrient source for coastal ecosystems. This is because the bacteria that are expelled simultaneously with the defecations are adept at breaking down the nutrients in the faeces into a bioavailable form that can be taken up into the food chain. DNA sequencing can also be used more broadly to **identify species present in a body of water, debris filtered from the air, or sample of dirt. This can establish the range of invasive species and endangered species, and track seasonal populations.**

THE METAGENOMICS PROCESS

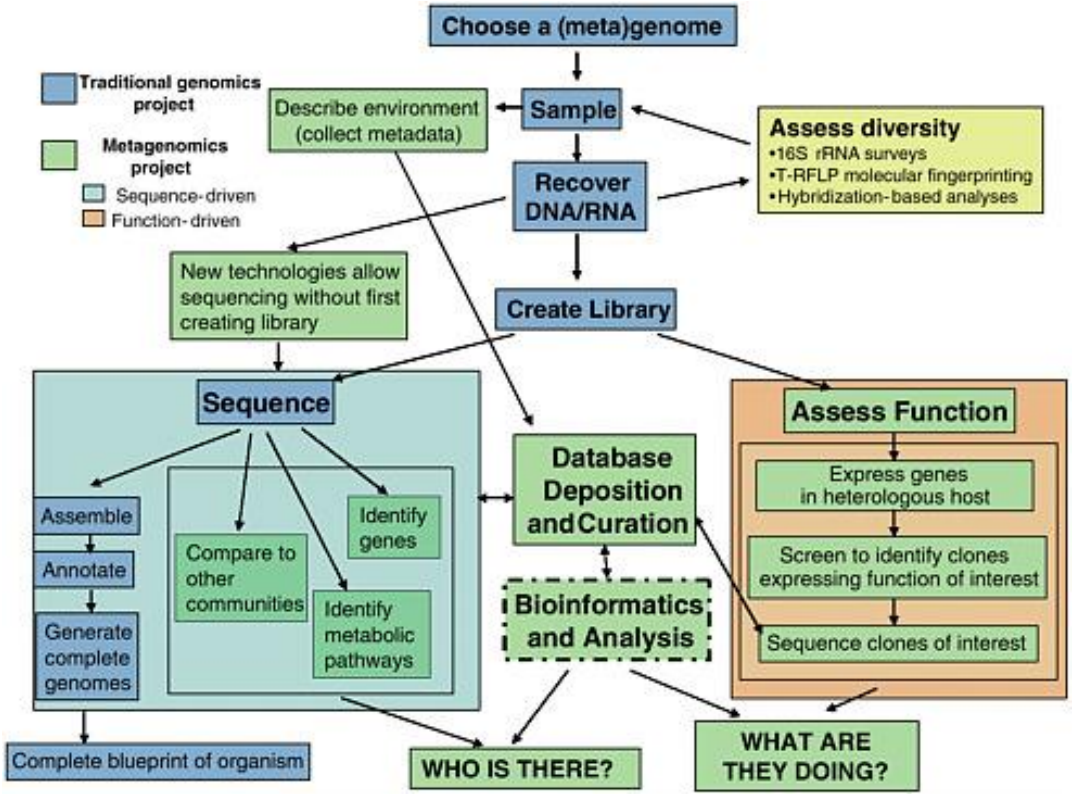
Extract all DNA from microbial community in sampled environment

DETERMINE WHAT THE GENES ARE
(Sequence-based metagenomics)

- Identify genes and metabolic pathways
- Compare to other communities
- and more…

DETERMINE WHAT THE GENES DO
(Function-based metagenomics)

- Screen to identify functions of interest, such as vitamin or antibiotic production
- Find the genes that code for functions of interest
- and more…



Metagenomics differs from traditional genomic sequencing in many ways. The dark blue boxes show the typical steps in the sequencing of a single organism's genome. **Metagenomics requires greater attention to sampling, and assessing the diversity of the sample by various means (yellow box) is necessary to ensure that the sample is representative.** Extracting the appropriate nucleic acids from the sample is another step that can be challenging in a metagenomics project. Preparation of a library is often the next step, but new sequencing technology can bypass this step. The DNA from metagenomics samples can then either be sequenced (blue box) or assessed for the functions it encodes (orange box). The sequence can sometimes be assembled into complete genomes of community members, but can also be analysed in other ways (light blue box). Data storage and computational analyses are critical steps in metagenomics projects and must be integrated throughout the project. **Overall, a metagenomics project can answer the questions "Who is there?" and "What are they doing?" in addition to assembling genomes.**

Metagenomics and industrial applications

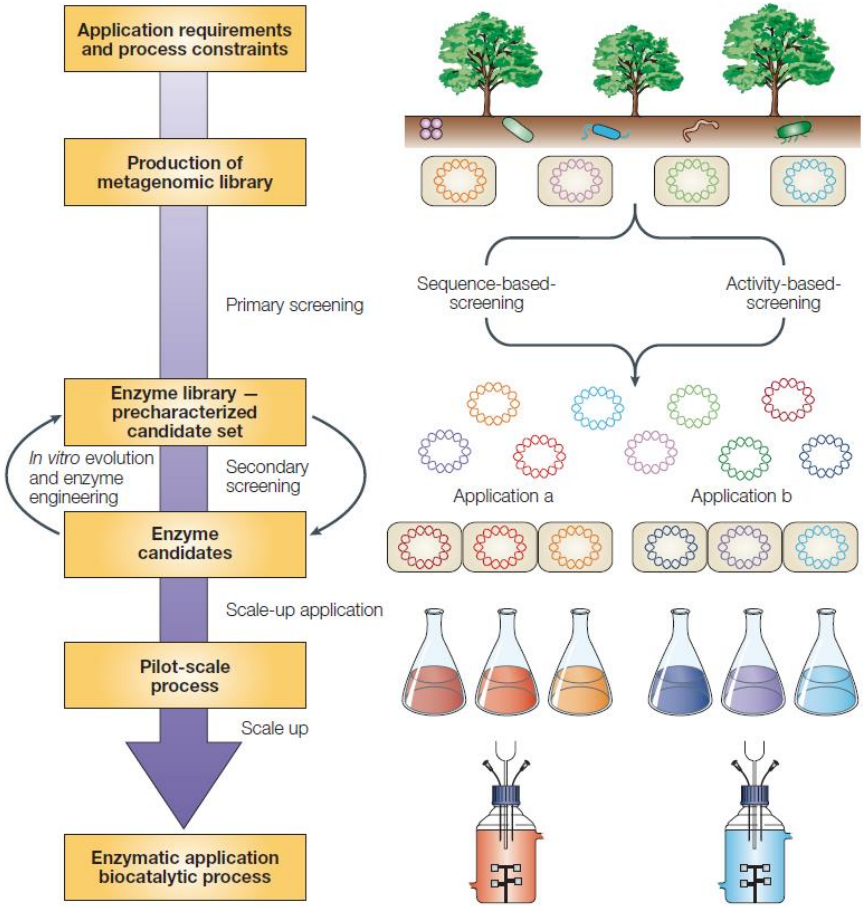


Figure 2 | **Industrial enzymes — from the metagenome to applications and processes.** This figure illustrates industrial rational bioprospecting of the metagenome. A library of cloned DNA is produced and primary screening, based on application requirements such as the conversion of an indicator substrate, produces enzyme libraries (different clones, which encode different enzymes, are indicated by different colours) that serve as platforms for subsequent development. Secondary screening of the enzyme library can identify process-specific properties such as substrate specificity, activity or stability. Primary and secondary screening both involve several stages based on different criteria, therefore the screening stages are multi-layered. Enzyme engineering and *in vitro* gene evolution can form a part of the development process. Subsets of cloned enzymes (red clones represent one subset, blue clones represent another subset) are then used in scale-up application or process testing to identify suitable enzyme candidates. Finally, process improvement by enzyme optimization and process engineering are carried out. Economic feasibility must be proven in a pilot-scale process environment as a prerequisite for scale-up to production or final application scale. Examples of the use of enzyme libraries in an industrial context include nitrilases[36] (Diversa), alcohol dehydrogenases[9] (Schering Plough) and glycosylhydrolases[39] (BRAIN AG).

Table 1 | **Activity-based screening for industrially relevant enzymes and biocatalysts from metagenomic libraries**

Function	Habitat	Library type	Average insert size (kb)	Number of clones screened	Library size (Mb)	Substrate	Number of hits	Hit rate (hit per Mb)	Ref.
Esterase/lipase	Forest soil	Plasmid	8	67,000	536	Tributyrin	98	1/5.5	*
Esterase/lipase	Forest soil	Fosmid	40	19,968	799	Tributyrin	47	1/17	*
Esterase/lipase	Sandy ecosystem	Fosmid	30	29,884	903	Tributyrin	49	1/18	*
Esterase/lipase	Sandy ecosystem	Fosmid	40	25,344	1,014	Tributyrin	29	1/35	*
Esterase/lipase	Soil	Plasmid	6	286,000	1,716	Tributyrin	3	1/572	63
Esterase/lipase	Soil	Plasmid	6	730,000	4,380	Triolein	1	1/4,380	63
Esterase/lipase	Soil	BAC	27	3,648	100	Bacto Lipid	2	1/50	64
Oxidation of polyols	Soil	Plasmid	3	900,000	2,700	1,2-ethanediol; 1,2-propanediol; 2,3-butanediol	15	1/180	65
Alcohol oxidoreductase	Soil/enrichment	Plasmid	4	400,000	1,600	Glycerol/1,2-propanediol	10	1/160	66
Amidase	Soil/enrichment	Plasmid	5	193,000	965	D-phenylglycine-L-leucine	7	1/138	67
Amylase	Soil	Plasmid	5	80,000	400	Starch	1	1/400	68
Amylase	Soil	BAC	27	3,648	100	Starch	8	1/12	64
Biotin production	Soil/excrement enrichment	Cosmid	35	50,000	1,750	Biotin-deficient medium	7	1/250	69
Protease	Soil	Plasmid	10	100,000	1,000	Skimmed milk	1	1/1,000	70
Cellulase	Sediment enrichment	λ phage	6	310,000	1,860	Carboxymethyl-cellulose	3	1/620	71
Chitinase	Seawater	λ phage	5	825,000	4,125	Methylumbelliferyl-diacetylchitobioside	11	1/375	72
Dehydratase	Soil/sediment enrichment	Plasmid	4	560,000	2,240	Glycerol	2	1/1,120	38
4-Hydroxybutyrate conversion	Soil	Plasmid	6	930,000	5,580	4-Hydroxybutyrate	5	1/1,116	73
β-Lactamase	Soil	Plasmid	5	80,000	400	Ampicillin	4	1/100	68

The screening host for all studies was *Escherichia coli*. *Unpublished results, K. Liebeton *et al.*, BRAIN AG. BAC, bacterial artificial chromosome.

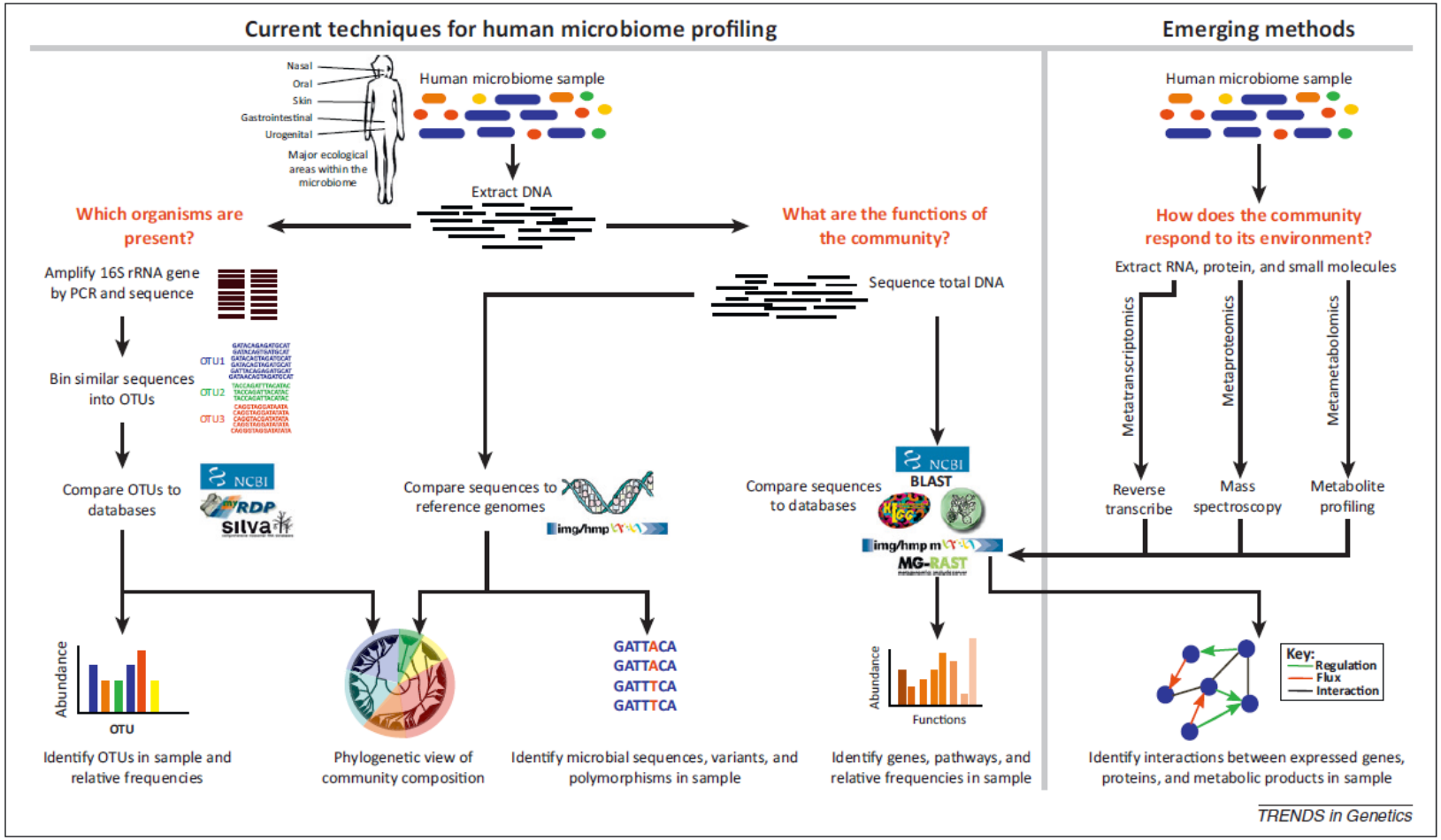
Metagenomics — an industrial future

As the excitement about genetic access to the boundless realms of microbial diversity slowly gives way to the reality of tapping into this diversity, the usual challenge of heterologous gene expression needs to be addressed to turn metagenomic technologies into commercial successes, particularly in applications for which bulk enzyme or product quantities have to be produced at competitive prices. Metagenomics, together with *in vitro* evolution and high-throughput screening technologies, provides industry with an unprecedented chance to bring biomolecules into industrial application. That this is possible has been shown by Diversa. A full cycle from the discovery of novel molecular scaffolds from multiple resources, including metagenomes, to *in vitro* evolution technology to generate an improved biocatalyst for a specific application was realized to produce α-amylases for applications in the hot and acidic process of starch liquefaction[62], showing the feasibility of the 'ideal biocatalyst' concept.

TABLE 6.4. Tolerances of life-forms

Type of Environment	Examples of Environments	Mechanism(s) of Survival	Practical Uses
High temperature (thermophiles)	Hydrothermal vents, hot springs, volcanoes	Amino acid changes, increase H-bonds, metal cofactors	Thermostable enzymes
Low temperature (psychrophiles)	Antarctic Ocean, glacier surfaces	Antifreeze proteins, solutes	Enhancing cold tolerance of crops
High hydrostatic pressure (barophiles)	Deep sea	Solute changes	
High salinity (halophiles)	Evaporating ponds and seas, salt evaporators	Solute changes, ion transport, protein amino acid adaptation	Industrial enzymes; soy sauce production
High pH (alkaliphiles)	Soda lakes	Transporters	Detergents
Low pH (acidophiles)	Mine tailings	Transporters	Bioremediation
Desiccation (xerophiles)	Evaporating ponds, deserts	Spore formation, solute changes starvation tolerance, DNA repair, scavenge free radicals	Freeze-drying additives
High radiation (radiophiles)	Nuclear reactors or waste sites, high-altitude surfaces	Absorb radiation, enhance DNA repair, scavenge free radicals	Bioremediation

Overview of bioinformatic methods for functional metagenomics



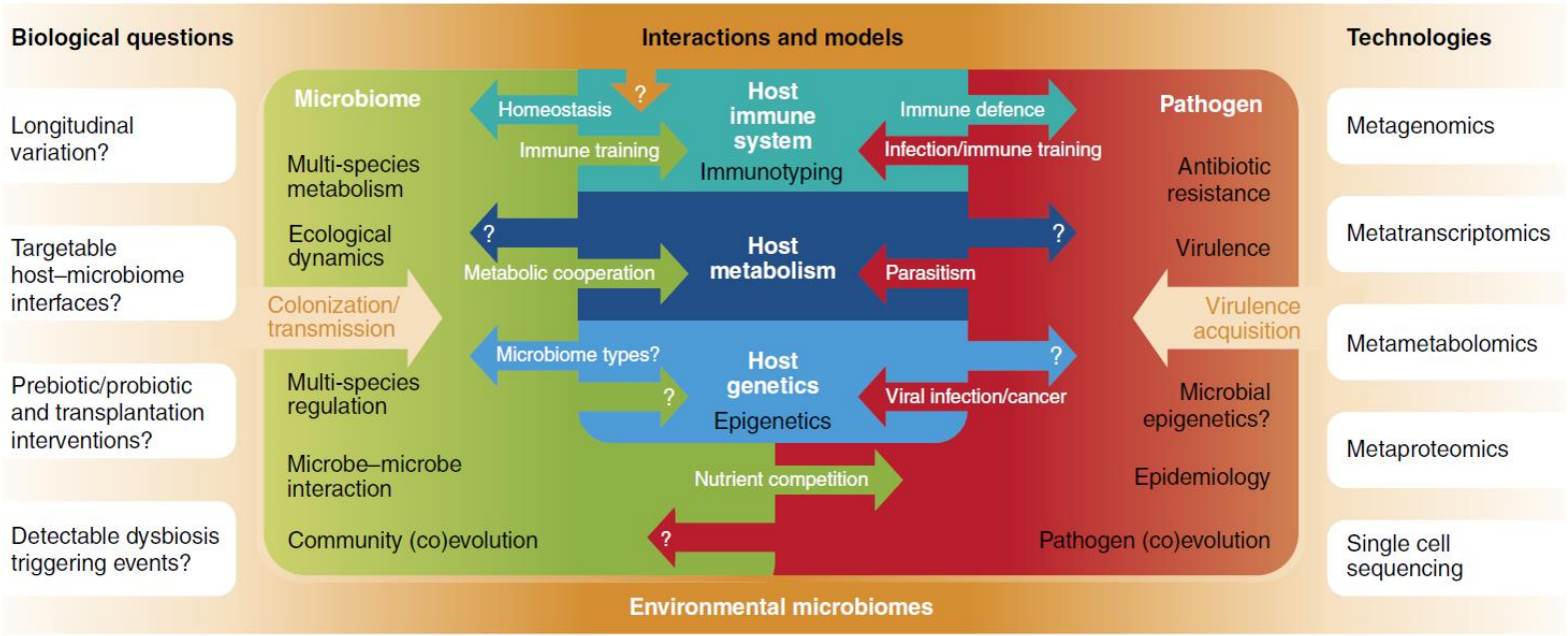
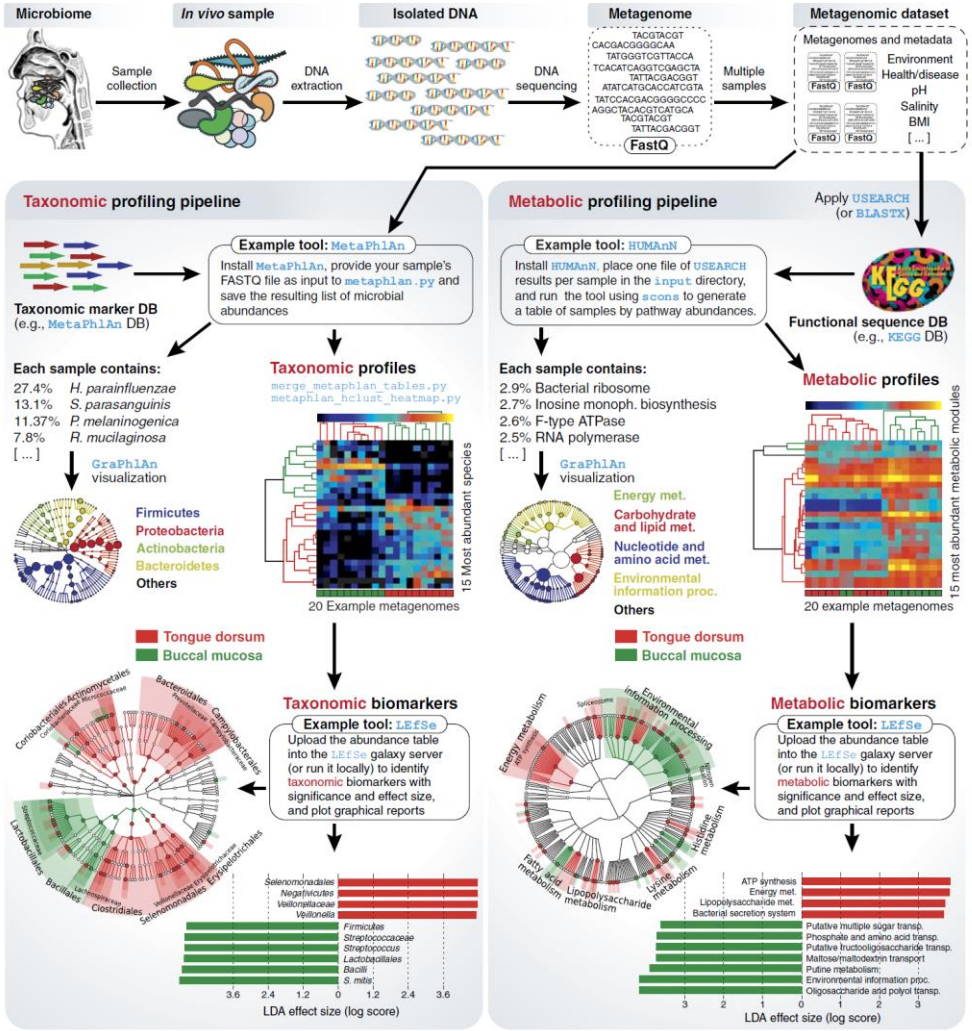
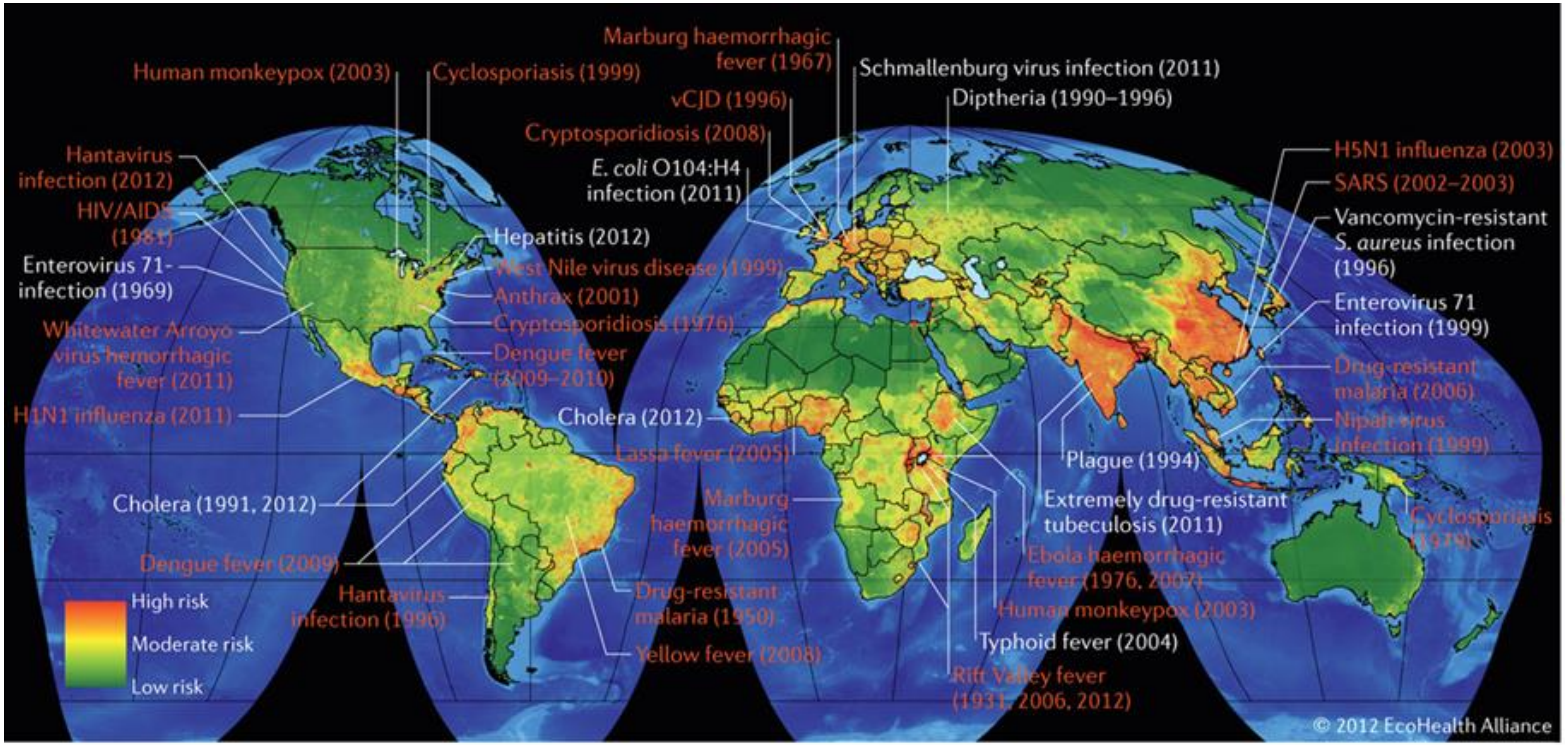


Figure 1 Open biological questions in microbial community biology, and emerging technologies and models for their exploration. Microbial communities are complex biological entities interacting with the environment, host organisms, and transient microbes. Predictive models for most of the interactions within these ecosystems are currently rare, but several studies have begun to provide key insights.



A typical current computational meta'omic pipeline to analyze and contrast microbial communities.

After collecting microbiome samples, community DNA or RNA is extracted and sequenced, generating WMS samples (i.e., metagenomes) generally consisting of several million short reads each. This example uses 20 WMS samples from the oral cavity (10 from the buccal mucosa, and 10 from the tongue dorsum). Complementary methods reconstruct the taxonomic characteristics (left) and metabolic potential (right) of the microbial communities. MetaPhlAn is one of many alternatives to detect and quantify microbial clades with species-level resolution, whereas HUMAnN quantitatively characterizes genes, pathways, and metabolic modules from each community. Differentially abundant clades or pathways can then be identified and assessed by tools such as LEfSe and represented graphically (e.g., here by GraPhlAn).

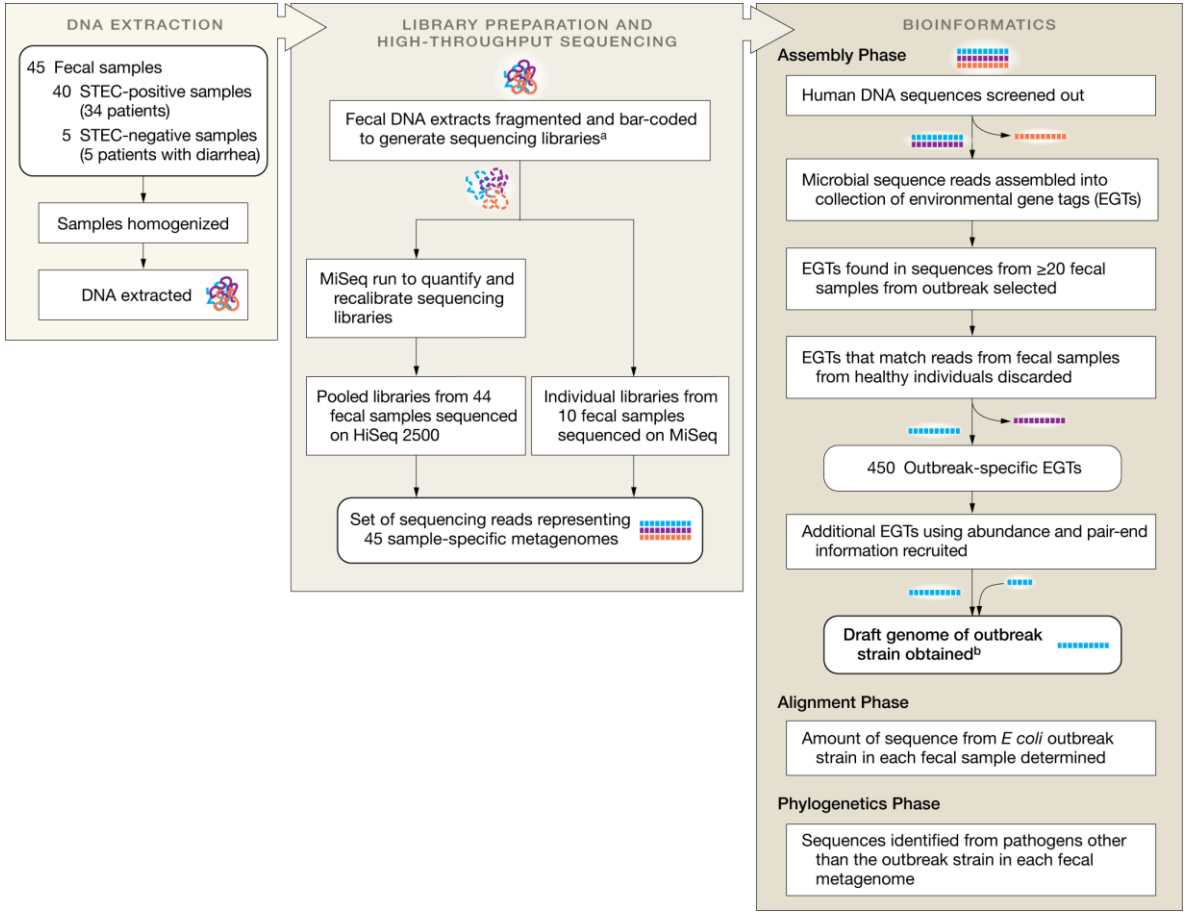


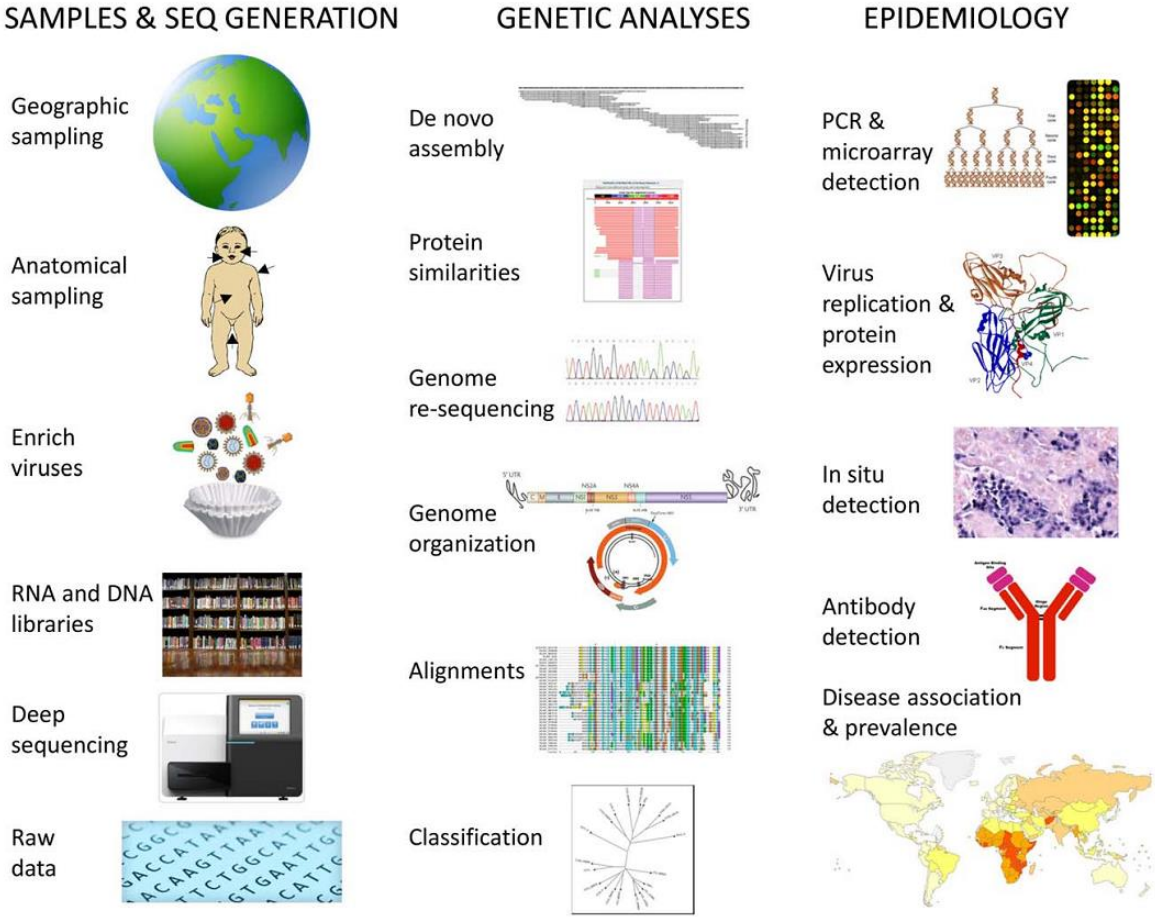
Nature Reviews | Microbiology

Hot spots of outbreaks for recently emerging and reemerging infectious diseases

Zoonotic infections are highlighted in red text. E. coli, Escherichia coli; S. aureus, Staphylococcus aureus; vCJD, variant Creutzfeldt–Jakob disease.

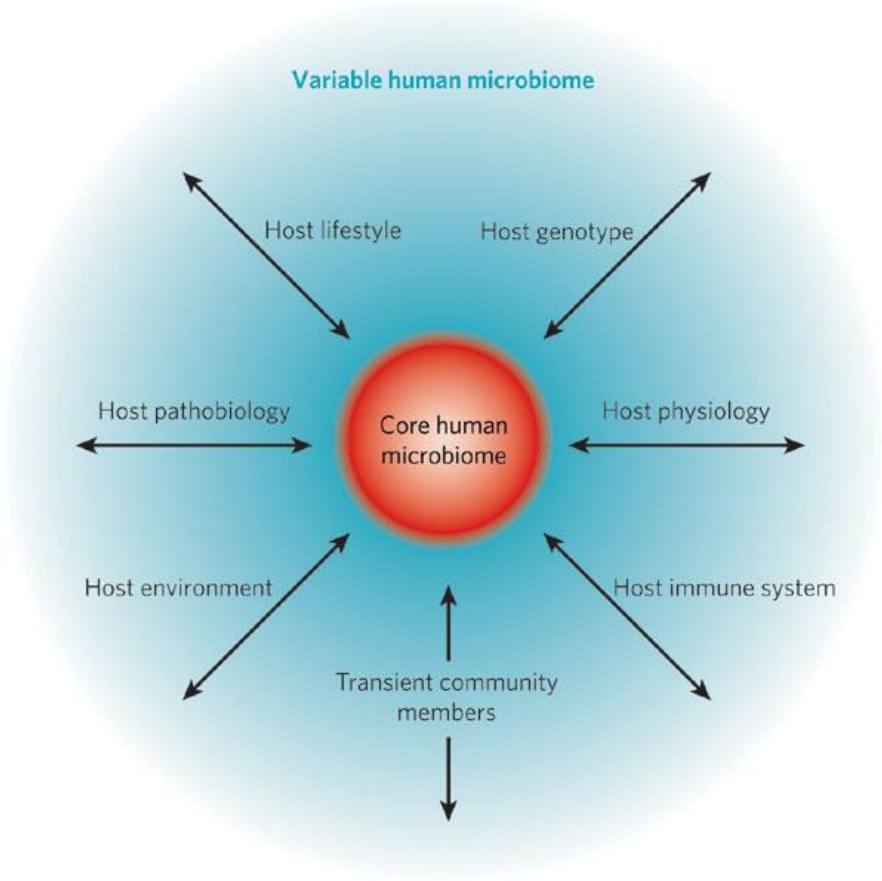
Workflow for Identification and Characterization of an Outbreak Strain Using Metagenomics





Schematic steps for determination of human virome and its impact on health, including biological samples and data acquisition, genetic analysis, and epidemiology of viral infections.

Human microbiome/Človeški mikrobiom



The concept of a core human microbiome

The **core human microbiome (red) is the set of genes present in a given habitat in all or the vast majority of humans**. **Habitat** can be defined over a range of scales, from the ***entire body to a specific surface area, such as the gut or a region within the gut***. The ***variable human microbiome (blue) is the set of genes present in a given habitat in a smaller subset of humans.*** This *variation could result from a combination of factors such as host genotype, host physiological status (including the properties of the innate and adaptive immune systems), host pathobiology (disease status), host lifestyle (including diet), host environment (at home and/or work) and the presence of transient populations of microorganisms that cannot persistently colonize a habitat.* The gradation in colour of the core indicates the possibility that, ***during human micro-evolution, new genes might be included in the core microbiome, whereas other genes might be excluded***.

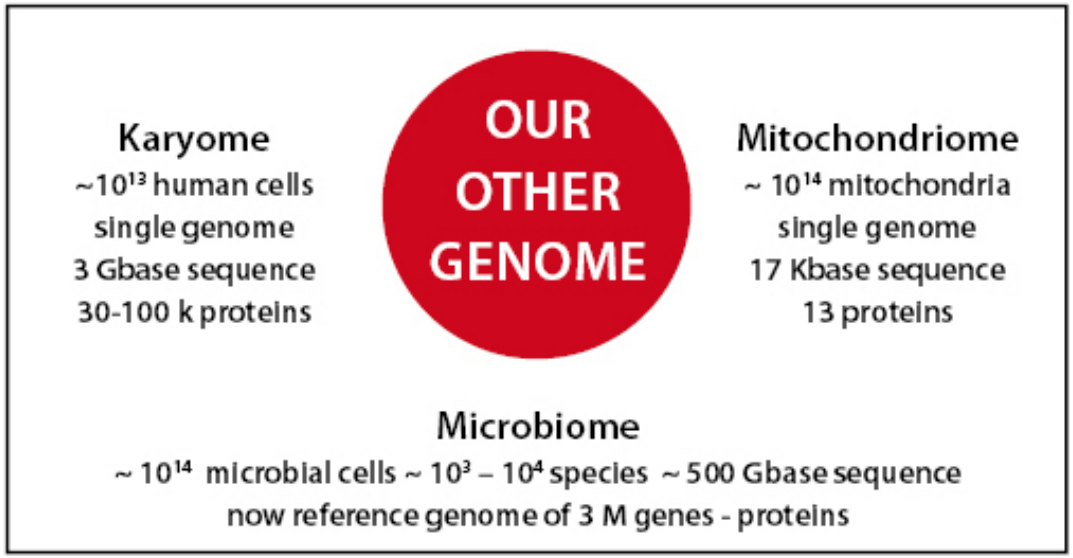
Peter J. Turnbaugh, Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight & Jeffrey I. Gordon. **The Human Microbiome Project.** Nature 449, 804-810 (2007)

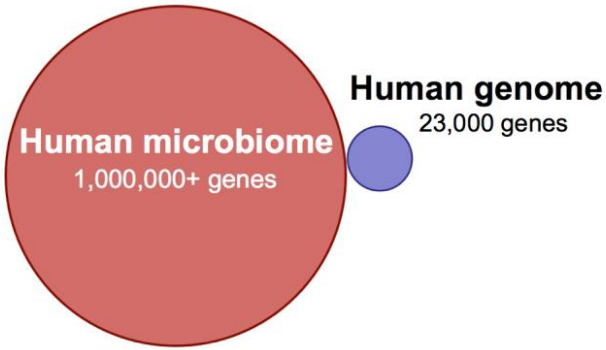
doi:10.1038/nature11234

Structure, function and diversity of the healthy human microbiome

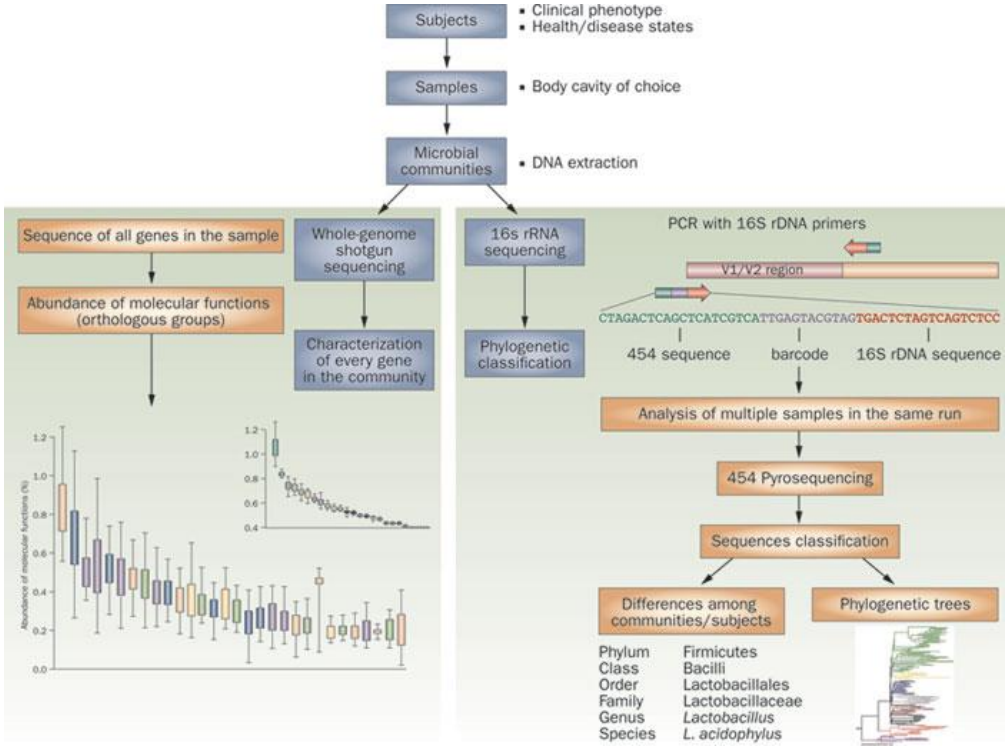
The Human Microbiome Project Consortium*

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81–99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.


Karyome
~10^13 human cells
single genome
3 Gbase sequence
30-100 k proteins
OUR OTHER GENOME
Mitochondriome
~ 10^14 mitochondria
single genome
17 Kbase sequence
13 proteins
Microbiome
~ 10^14 microbial cells ~ 10^3 – 10^4 species ~ 500 Gbase sequence
now reference genome of 3 M genes - proteins

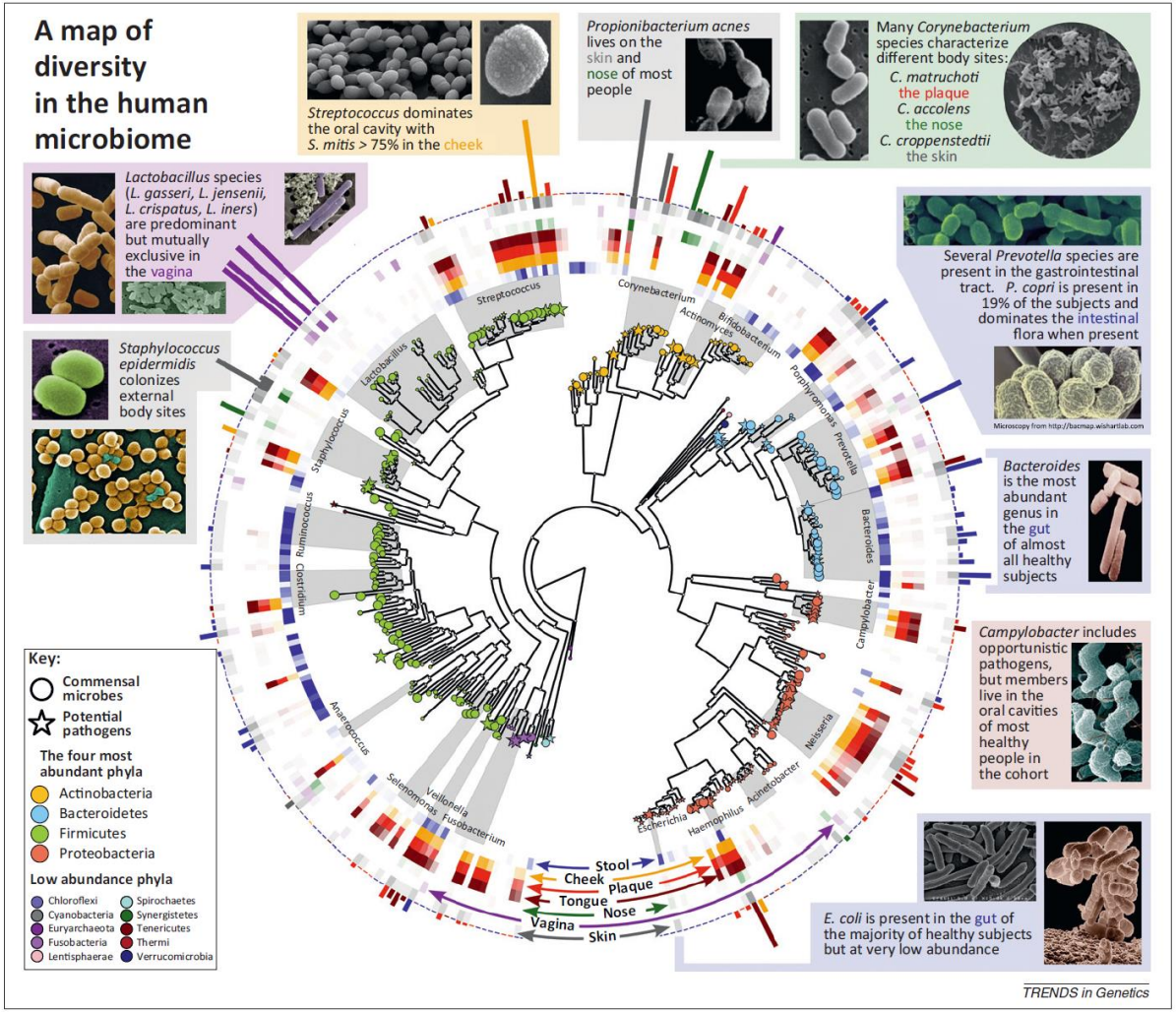

Human microbiome
1,000,000+ genes
Human genome
23,000 genes

Culture-independent genomic analysis of the human microbiome



Culture-independent techniques have advanced our capacity to survey complex microbial communities in human samples. ***Well-characterized individuals (healthy and diseased) are asked to donate samples for microbiome analyses. Two metagenomic sequencing approaches are utilized. Conserved and variable 16S rRNA genomic regions are amplified and subjected to pyrosequencing***. The resulting sequences are then aligned, filtered and compared to publicly available databases of ***16S rRNA sequences, enabling taxonomic classification of bacteria present or absent in a given sample***. ***Whole genome shotgun sequencing provides information that enables identification of genes present and allows for subsequent comparison of enzymatic pathways and functions represented among different samples***. ***Enzymatic databases are also available to assist in the identification of protein function, enabling the richness and diversity of functional capacities provided by the microbiome to be assessed***.

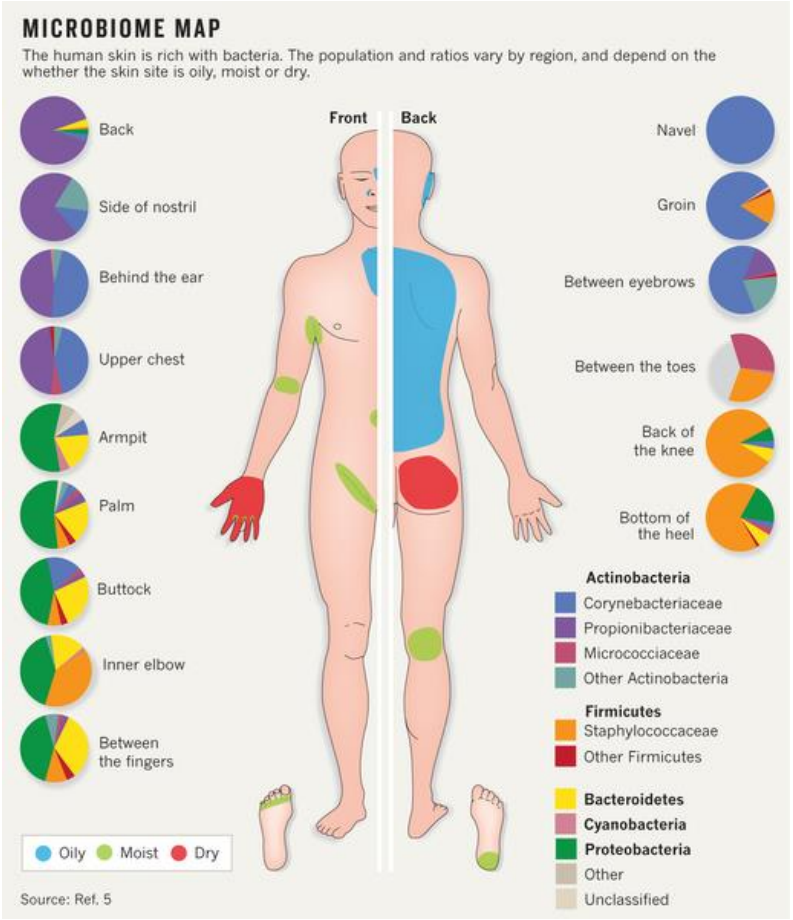
Abbreviations: PCR, polymerase chain reaction; rRNA, ribosomal RNA.



A map of diversity in the human microbiome

The **human microbiome is dominated by four phyla: Actinobacteria, Bacteroidetes, Firmicutes, and Proteobacteria**. In the center is a **phylogenetic tree of organisms abundant in the human microbiome**. Commensal microbes are indicated by circles, and potential pathogens are indicated by stars. The **middle ring corresponds to body sites at which the various taxa are abundant and is color-coded by site** [e.g., Ruminococcus (blue) is found mostly in the gut, whereas Lactobacillus (purple) is found mostly in the vagina]. The **bar heights on the outside of the circle are proportional to taxa abundance at the body site of greatest prevalence** [e.g., Streptococcus mitis (yellow) dominates the inside of the cheek, whereas the gut is abundant in a variety of Bacteroides]. The **intensity of external colors corresponds to species prevalence in each body site**.

Skin Microbiome



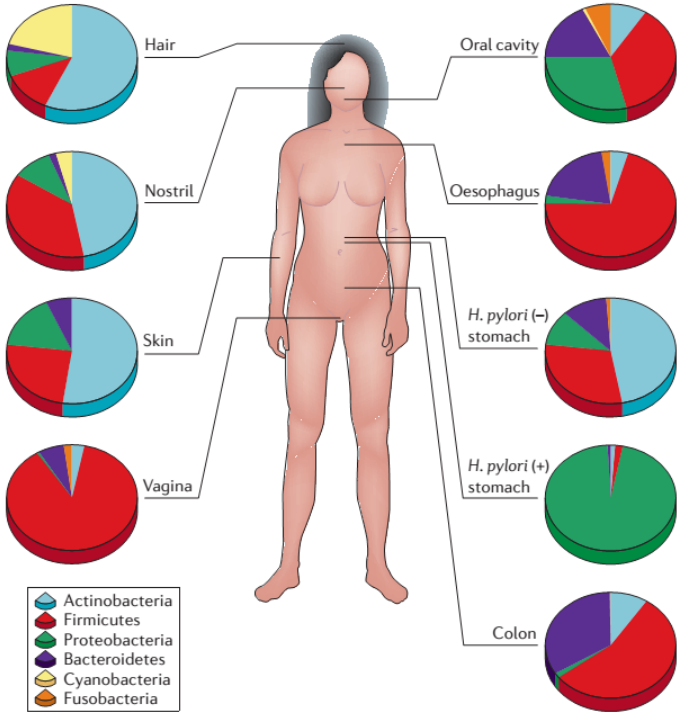


Figure 1 | **Compositional differences in the microbiome by anatomical site.** High-throughput sequencing has revealed substantial intra-individual microbiome variation at different anatomical sites, and inter-individual variation at the same anatomical sites[4,5,25,52,89,93]. However, higher-level (for example, at the level of phyla) taxonomic features display temporal (longitudinal) stability in individuals at specific anatomical sites. Such site-specific differences and the observed conservation between human hosts provide an important framework to determine the biological and pathological significance of a particular microbiome composition. The figure indicates the relative proportion of sequences determined at the taxonomic phylum level at eight anatomical sites. Certain features, such as the presence (+) or absence (–) of *Helicobacter pylori*, can lead to permanent and marked perturbations in community composition[93].

The characteristics of human microbiota change over time in response to varying environmental conditions and life stages

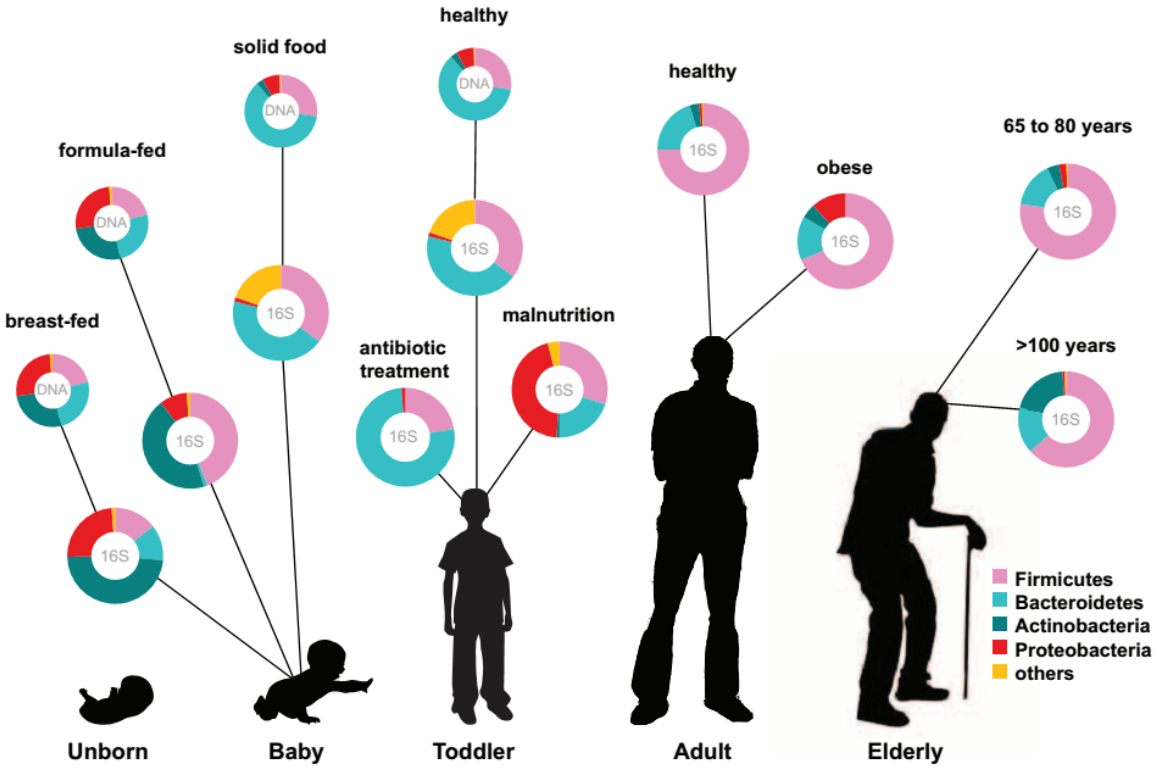


FIGURE 1 | Human microbiota: onset and shaping through life stages and perturbations. The graph provides a global overview of the relative abundance of key phyla of the human microbiota composition in different stages of life. Measured by either 16S RNA or metagenomic approaches (DNA). Data arriving from: Babies breast- and formula-fed (Schwartz et al., 2012), baby solid food (Koenig et al., 2011), toddler antibiotic treatment (Koenig et al., 2011), toddler healthy or malnourished (Monira et al., 2011), adult, elderly, and centenarian healthy (Biagi et al., 2010), and adult obese (Zhang et al., 2009).

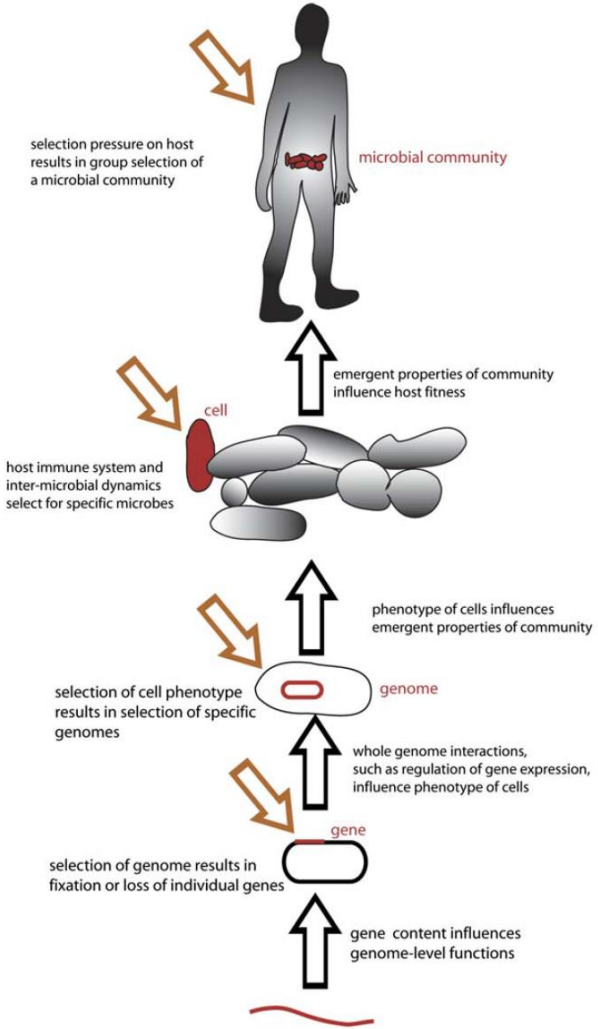


Figure 1. Schematic Diagram of the Selection Pressures Operating at Different Levels in the Human-Microbial Hierarchy

Brown arrows indicate selection pressures and point to the unit under selection (red). Black arrows indicate emergent properties of one level that affect higher levels in the hierarchy. According to hierarchy theory, higher levels place constraints on possible organizational solutions at lower levels. Ecologic principles predict that host-driven ("top-down") selection for functional redundancy would result in a community composed of widely divergent microbial lineages (divisions) whose genomes contain functionally *similar* suites of genes. Another prediction is the widespread occurrence of, and abundant mechanisms for, lateral gene transfer. In contrast, competition between members of the microbiota would exert "bottom-up" selection pressure that results in specialized genomes with functionally *distinct* suites of genes (metabolic traits). Once established, these lineage-specific traits can be maintained by barriers to homologous recombination (Majewski et al., 2000).

Ecological and Evolutionary Forces Shaping Microbial Diversity in the Human Intestine

Ruth E. Ley,[1] Daniel A. Peterson,[1] and Jeffrey I. Gordon[1,*]
[1]Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108, USA
*Contact: jgordon@molecool.wustl.edu
DOI 10.1016/j.cell.2006.02.017

The human gut is populated with as many as 100 trillion cells, whose collective genome, the microbiome, is a reflection of evolutionary selection pressures acting at the level of the host and at the level of the microbial cell. The ecological rules that govern the shape of microbial diversity in the gut apply to mutualists and pathogens alike.

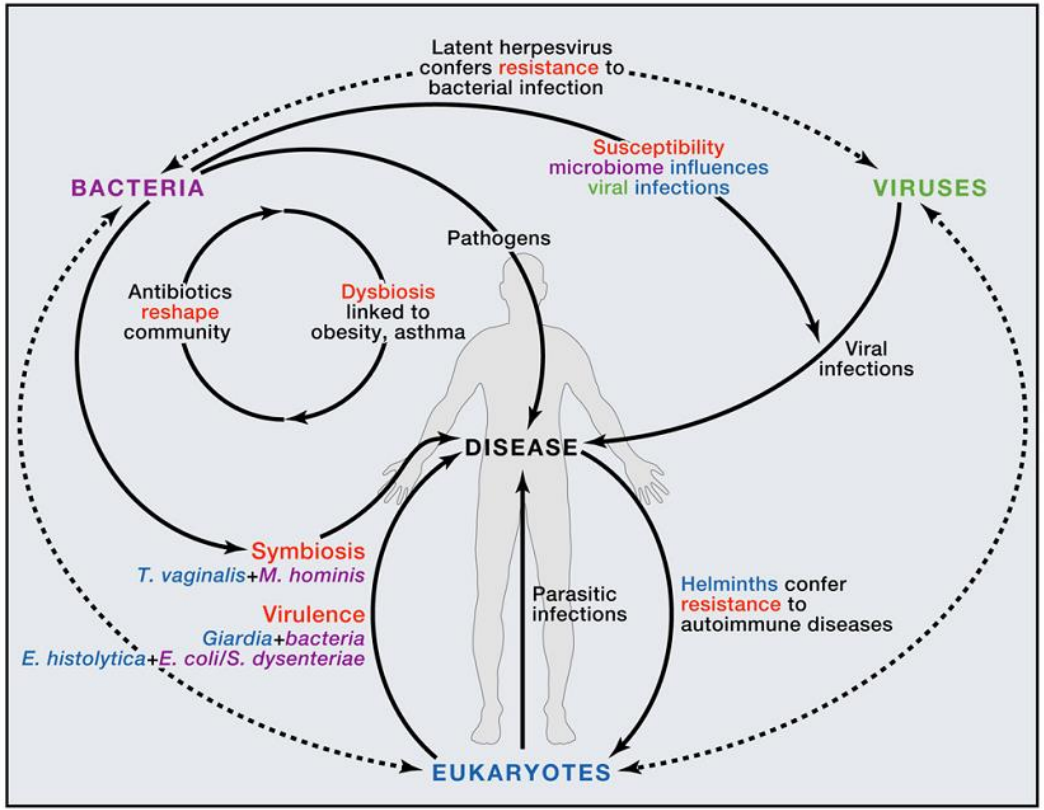
Cell

Leading Edge
Review

The Impact of the Gut Microbiota on Human Health: An Integrative View

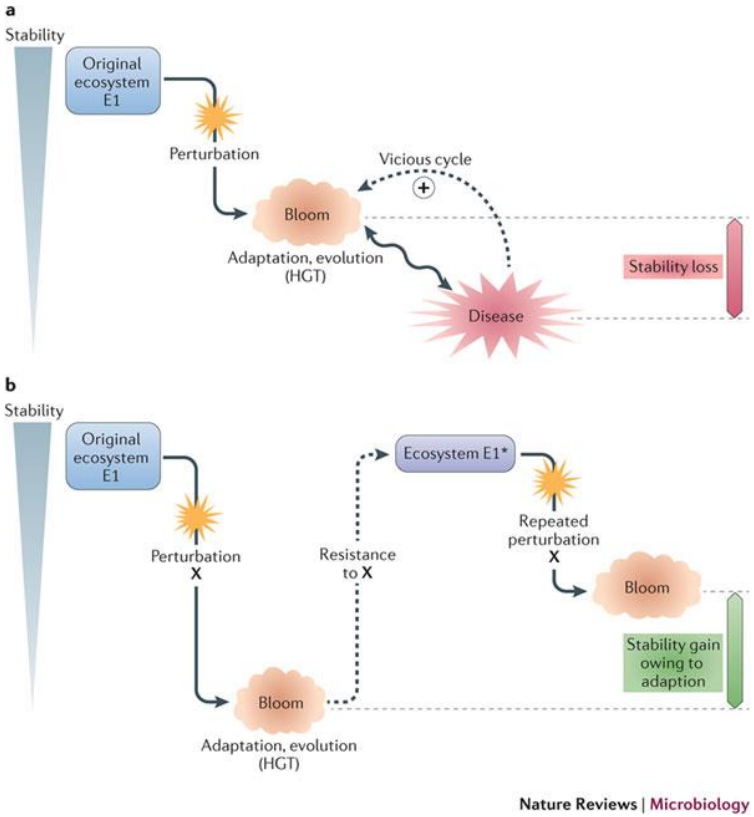
Jose C. Clemente,[1] Luke K. Ursell,[1] Laura Wegener Parfrey,[1] and Rob Knight[1,2,*]
[1]Department of Chemistry & Biochemistry, University of Colorado at Boulder, Boulder, CO 80309, USA
[2]Howard Hughes Medical Institute, Boulder, CO 80309, USA
*Correspondence: rob.knight@colorado.edu
DOI 10.1016/j.cell.2012.01.035

The human gut harbors diverse microbes that play a fundamental role in the well-being of their host. The constituents of the microbiota—bacteria, viruses, and eukaryotes—have been shown to interact with one another and with the host immune system in ways that influence the development of disease. We review these interactions and suggest that a holistic approach to studying the microbiota that goes beyond characterization of community composition and encompasses dynamic interactions between all components of the microbiota and host tissue over time will be crucial for building predictive models for diagnosis and treatment of diseases linked to imbalances in our microbiota.

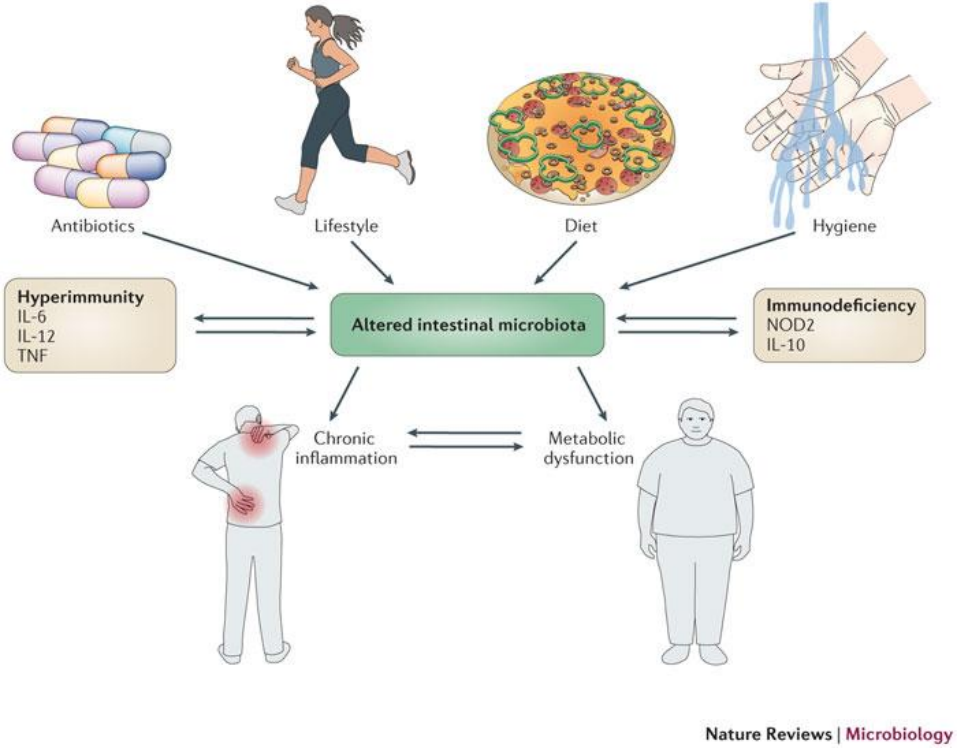


Effect of Interactions of Bacteria, Viruses, and Eukaryotes in Health and Disease

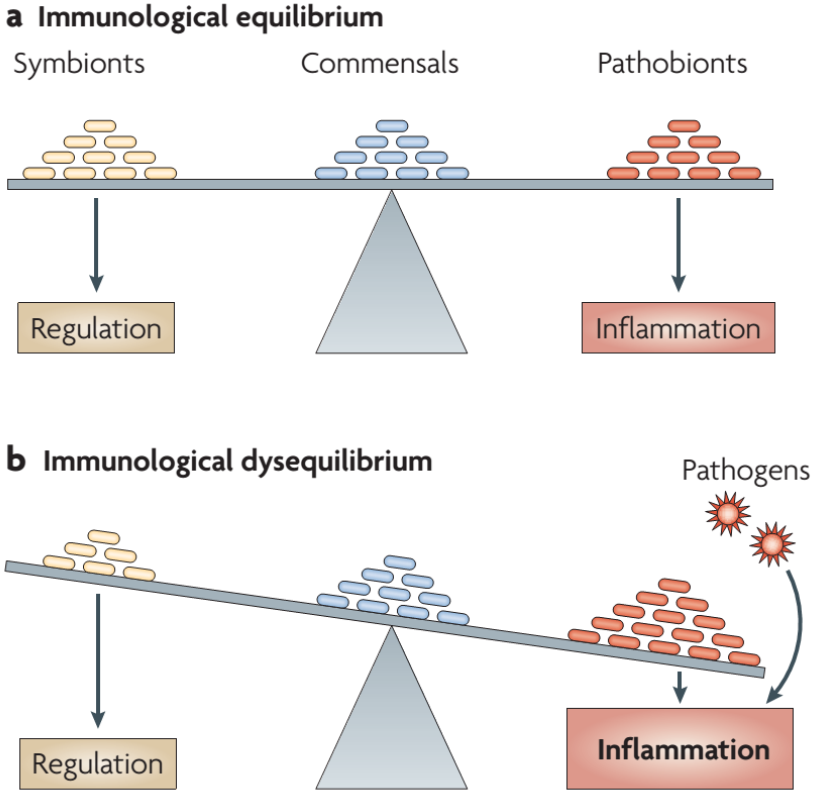
Diseases have been traditionally studied under a paradigm of ''one microbe, one disease.'' However, a **new understanding is emerging on how disease phenotypes are actually a result of complex interactions between bacteria, viruses, and eukaryotes, as well as their interactions with the host or with certain drugs**. Virulence of some eukaryotes is, for instance, linked to the presence of certain bacteria, such as in the case of E. histolytica and E. coli or S. dysenteriae. The susceptibility of the host to viral infections is conditioned by the particular configuration of the microbiota, whereas herpesvirus infection can confer resistance to certain bacterial infections. ***Antibiotics can significantly reshape the composition of the microbiota***. ***As a clear correlation has been observed between many diseases and dysbiosis, the widespread use of antibiotics may be linked to the dramatic increase observed in autoimmune diseases over the last years***. Conversely, helminthes confer resistance to autoimmune diseases.



Bacteria might adapt to growth in dysbiotic conditions and acquire even higher pathogenic potential by horizontal gene transfer (HGT) of virulence factors.



The composition of the gut microbiota is influenced by various environmental factors, including the use of antibiotics, lifestyle, diet and hygiene preferences.

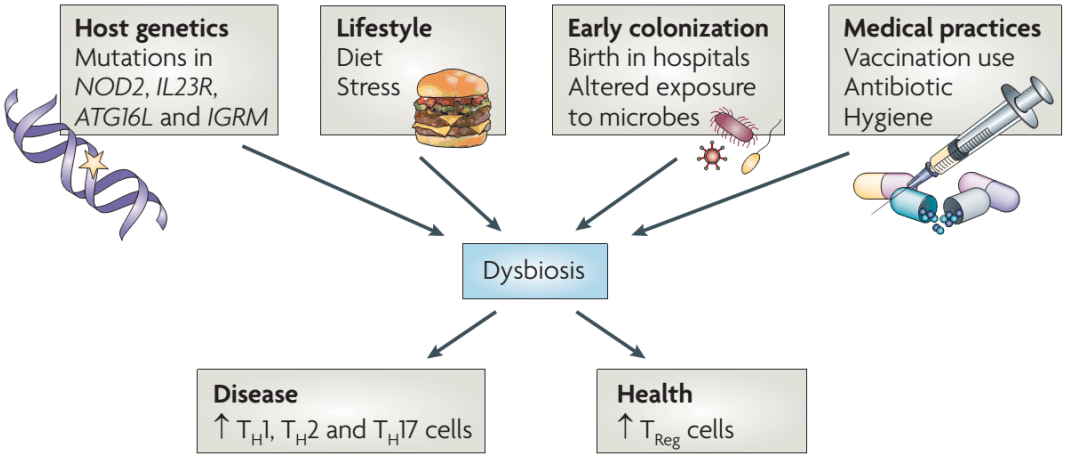


Immunological dysregulation associated with dysbiosis of the microbiota

a| A healthy microbiota contains a balanced composition of many classes of bacteria. Symbionts are organisms with known health-promoting functions. Commensals are permanent residents of this complex ecosystem and provide no benefit or detriment to the host (at least to our knowledge). Pathobionts are also permanent residents of the microbiota and have the potential to induce pathology.

b| In conditions of dysbiosis there is an unnatural shift in the composition of the microbiota, which results in either a reduction in the numbers of symbionts and/or an increase in the numbers of pathobionts. The causes for this are not entirely clear, but are likely to include recent societal advances in developed countries. The result is non-specific inflammation, which may predispose certain genetically susceptible people to inflammatory disease and may be caused by pathogens, which are opportunistic organisms that cause acute inflammation.

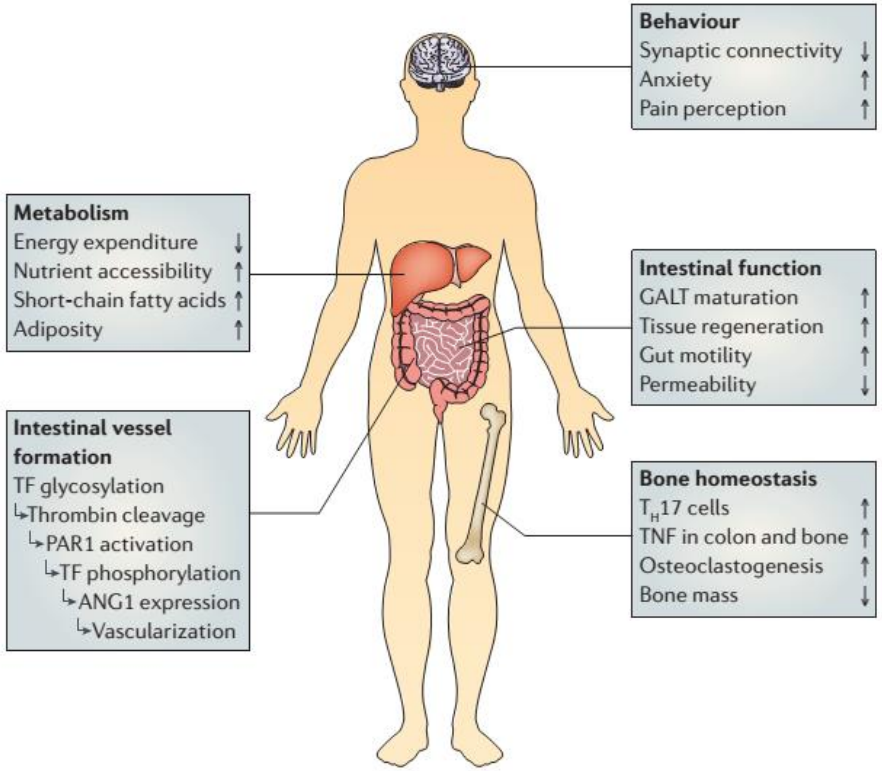
Dysbiosis: An imbalance in the structural and/or functional configuration of the microbiota, leading to a disruption of host–microorganism homeostasis.



Proposed causes of dysbiosis of the microbiota

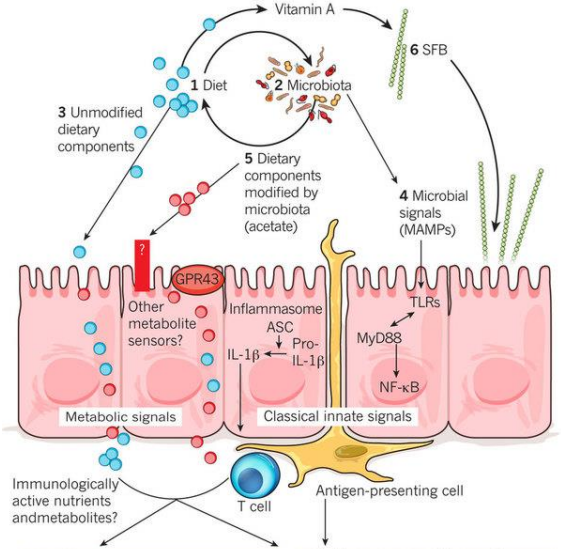
We propose that the ***composition of the microbiota can shape a healthy immune response or predispose to disease***. Many factors can contribute to dysbiosis, including host genetics, lifestyle, exposure to microorganisms and medical practices. **Host genetics** can potentially influence dysbiosis in many ways. An individual with mutations in genes involved in immune regulatory mechanisms or pro-inflammatory pathways could lead to unrestrained inflammation in the intestine. It is possible that inflammation alone influences the composition of the microbiota, skewing it in favour of pathobionts. Alternatively, a host could 'select' or exclude the colonization of particular organisms. This selection can be either active (as would be the case of an organism recognizing a particular receptor on the host) or passive (the host environment is more conducive to fostering the growth of select organisms). Selection of pathobionts by the host could tip the balance in favour of inflammation. **Diet and stress** also have the potential to influence the microbiota. **Birth in the sterile environment of hospitals can protect from exposure to dangerous pathogens, but can also prevent early exposure to health-promoting bacteria. Overuse of vaccination and antibiotics, which do not distinguish between pathogenic or symbiotic microorganisms, could adversely alter the microbiota**.

ATG16L, autophagy-related gene 16-like; IGRM, immunity-related GTPase family, M; IL23R, interleukin-23 receptor; NOD2, nucleotide-binding oligomerization domain 2; TH, T helper; TReg, regulatory T.



Microbial impact on host physiology

The **gut microbiota has been shown to affect several aspects of host physiology; arrows represent either stimulatory or inhibitory effects of the gut microbiota on host physiological processes**. The microbiota has been shown to influence intestinal function in the host, promoting gut-associated lymphoid tissue (GALT) maturation, tissue regeneration (in particular of the villi) and gut motility, and reducing the permeability of epithelial cells lining the gut, thus promoting barrier integrity. Similarly, the gut microbiota influences the morphogenesis of the vascular system surrounding the gut. This is associated with increased glycosylation of tissue factor (TF), which leads to cleavage of thrombin, in turn activating proteinase-activated receptor 1 (PAR1). This then phosphorylates TF to promote epithelial expression of angiopoietin 1 (ANG1), which promotes increased vascularization. Changes in the microbiota composition or a complete lack of a gut microbiota has been shown to affect metabolism, behaviour and tissue homeostasis, suggesting that the microbiota also regulates these processes. Specifically, the **gut microbiota can influence the host's nervous system, decreasing synaptic connectivity and promoting anxiety-like behaviour and pain perception**. In the case of host metabolism, the gut microbiota has been shown to facilitate energy harvest from the diet, to modulate host metabolism (for example, by decreasing energy expenditure) and to promote host adiposity. Finally, the gut microbiota can influence tissue homeostasis, for example decreasing bone mass by promoting the function of osteoclasts (which cause bone resorption) and increasing the numbers of pro-inflammatory T helper 17 (TH17) cells.



T cell

mTOR
Promotes T_H1, T_H2, T_H17 cell differentiation; inhibits T_{reg} cell differentiation

RAR–RXR
Promotes intestinal T-cell homing; promotes T_H2 and T_{reg} cell differentiation

VDR–RXR
Promotes T_{reg} cell differentiation; inhibits T_H1 and T_H17 cell differentiation

AHR
Promotes T_H17 and T_{reg} cell differentiation

LXR and PPAR
Control T-cell differentiation

Antigen-presenting cell

TLRs

Inflammasomes

mTOR
Modulates DC function and differentiation

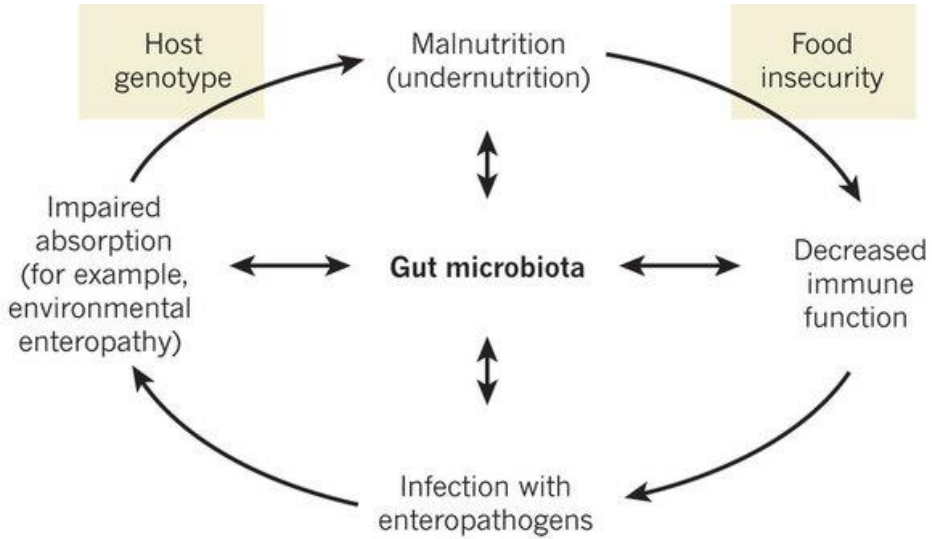
RAR–RXR and VDR–RXR

AHR
Modulates DC differentiation

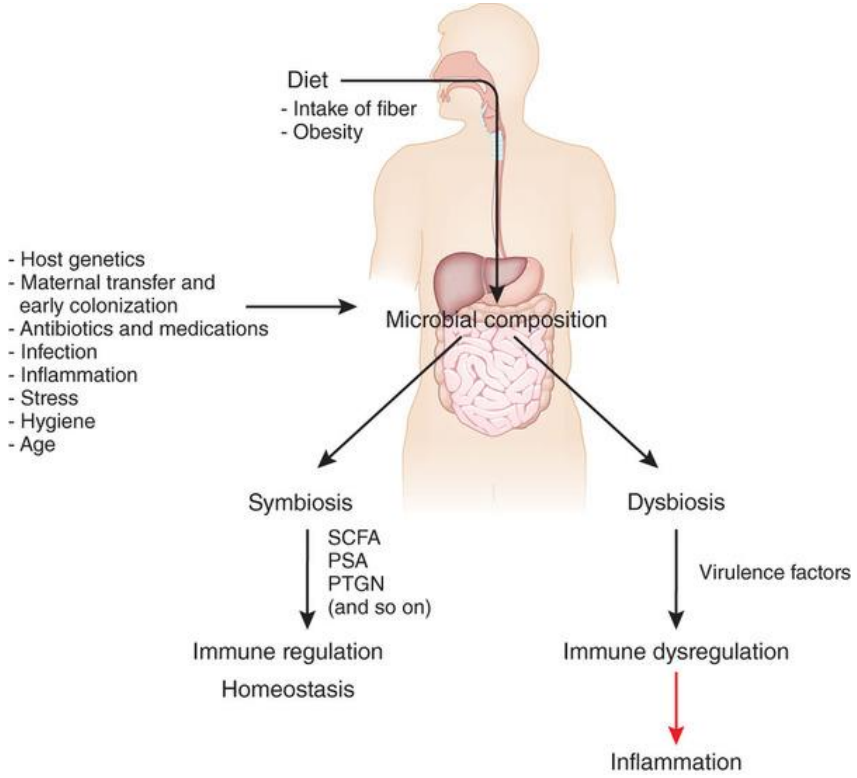
PKR
Regulates inflammatory responses

GPR120
Inhibits inflammatory responses in macrophages

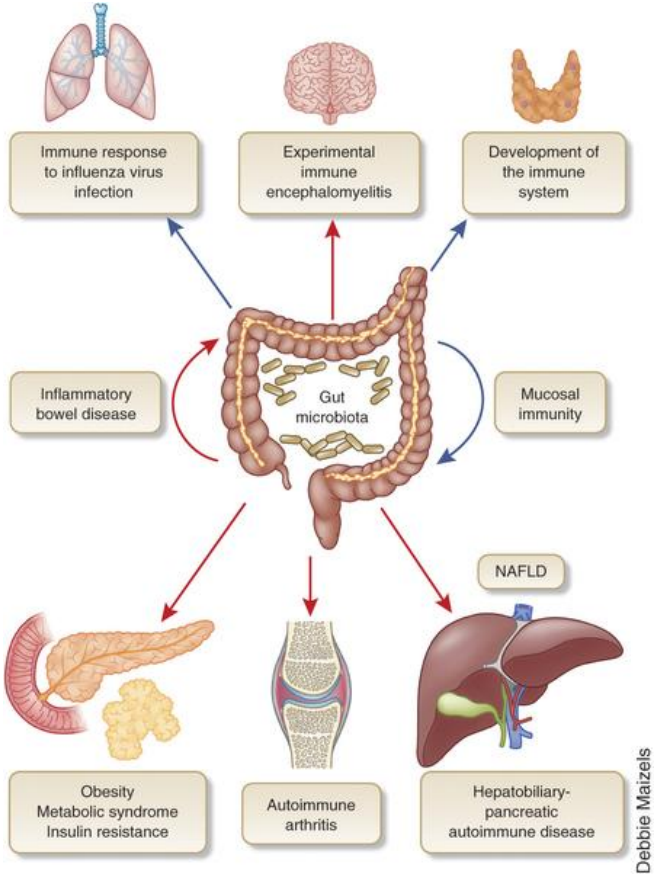
Metabolite sensors that help to coordinate immune responses



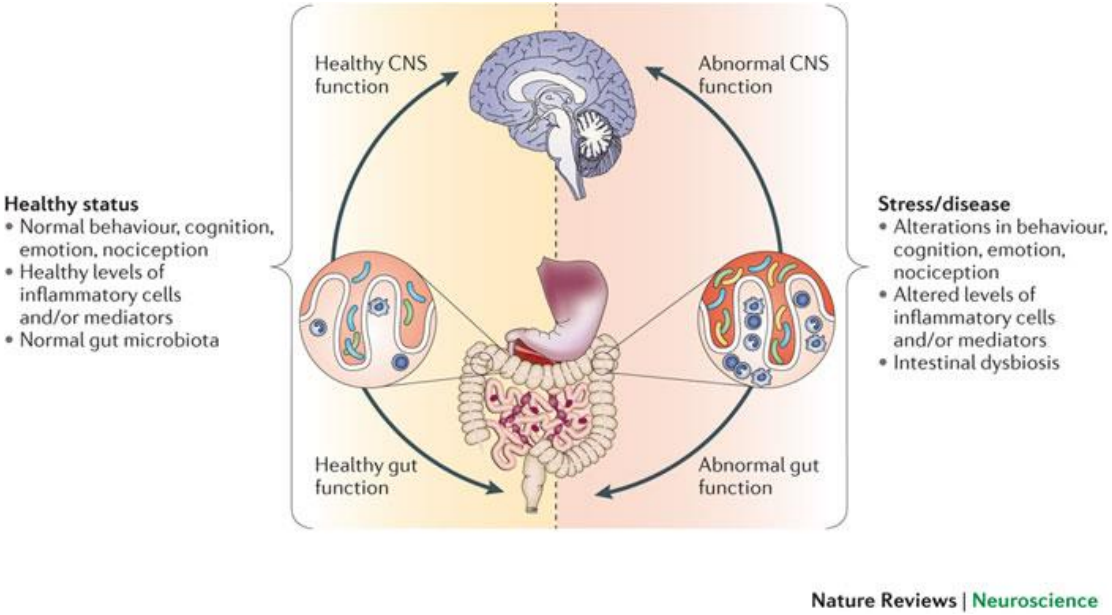
A schematic of the proposed relationships between the gut microbiota, the immune system and the diet, which underlie the development of malnutrition.



Diet, microbial composition and regulation of the immune system. Diet and other environmental and host factors have a major effect on gut microbial composition.



Crosstalk between an organism and its gut commensal microbiota has both potentiating and detrimental effects on the immune response.



Impact of the gut microbiota on the gut–brain axis in health and disease.

It is now generally accepted that **a stable gut microbiota is essential for normal gut physiology and contributes to appropriate signalling along the gut–brain axis and, thereby, to the healthy status of the individual, as shown on the left-hand side of the figure**. As shown on the ***right-hand side of the figure, intestinal dysbiosis can adversely influence gut physiology, leading to inappropriate gut–brain axis signalling and associated consequences for CNS functions and resulting in disease states***. Conversely, stress at the level of the CNS can affect gut function and lead to perturbations of the microbiota.

Biotehnološka uporaba mikrobnih genomov

Microbial Genomics at the U.S. Department of Energy (http://genomics.energy.gov/)

DOE's Microbial Genome Program (MGP -(1994-2005)). The MGP was begun in 1994 as a spinoff from the HGP. The program sequenced the genomes of a number of **nonpathogenic microbes** useful in solving DOE's mission challenges in **environmental-waste cleanup, energy production, carbon cycling, and biotechnology**.

Why Microbes? Microbes, which make up most of the earth's biomass, have evolved for some **3.8 billion years**. They have been found in virtually every environment, ***surviving and thriving in extremes of heat, cold, radiation, pressure, salt, acidity, and darkness***. Often in these environments, no other forms of life are found and the ***only nutrients come from inorganic matter***. The ***diversity and range of their environmental adaptations indicate that microbes long ago »solved« many problems for which scientists are still actively seeking solutions***.

Potential Microbial Applications. Researchers have only scratched the surface of microbial biodiversity. Knowledge about the enormous range of microbial capacities has broad and far-reaching implications for environmental, energy, health, and industrial applications.

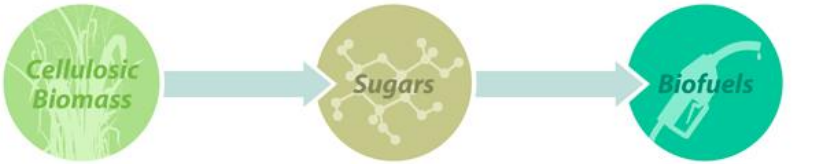
-Cleanup of toxic-waste sites worldwide.
-Production of novel therapeutic and preventive agents and pathways.
-Energy generation and development of renewable energy sources (e.g., methane and hydrogen).
-Production of chemical catalysts, reagents, and enzymes to improve efficiency of industrial processes.
-Management of environmental carbon dioxide, which is related to climate change.
-Detection of disease-causing organisms and monitoring of the safety of food and water supplies.
-Use of genetically altered bacteria as living sensors (biosensors) to detect harmful chemicals in soil, air, or water.
-Understanding of specialized systems used by microbial cells to live in natural environments with other cells.

Genomic Science Program Goal and Objectives

Genome Sequence → **System-Wide Biological Investigations** → **Predictive Understanding**

Goal: Achieve a predictive, system-level understanding of plants, microbes, and biological communities, via integration of fundamental science and technology development, to enable biological solutions to DOE mission challenges in energy, environment, and climate.

- **Objective 1:** Determine the genomic properties, molecular and regulatory mechanisms, and resulting functional potential of microbes, plants, and biological communities central to DOE missions.
- **Objective 2:** Develop the experimental capabilities and enabling technologies needed to achieve a genome-based, dynamic system-level understanding of organism and community function.
- **Objective 3:** Develop the knowledgebase, computational infrastructure, and modeling capabilities to advance the understanding, prediction, and manipulation of complex biological systems.

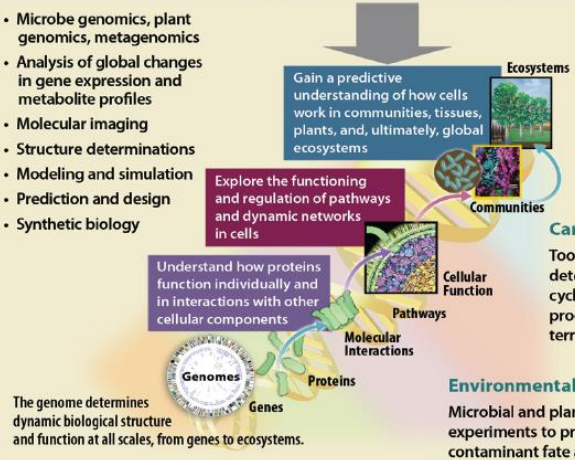


DOE Genomic Science Program
A Mission-Inspired Fundamental Research Approach

Technologies and Methods for Systems Biology

- Microbe genomics, plant genomics, metagenomics
- Analysis of global changes in gene expression and metabolite profiles
- Molecular imaging
- Structure determinations
- Modeling and simulation
- Prediction and design
- Synthetic biology

Fundamental Research Needs



Mission Grand Challenges for Biology

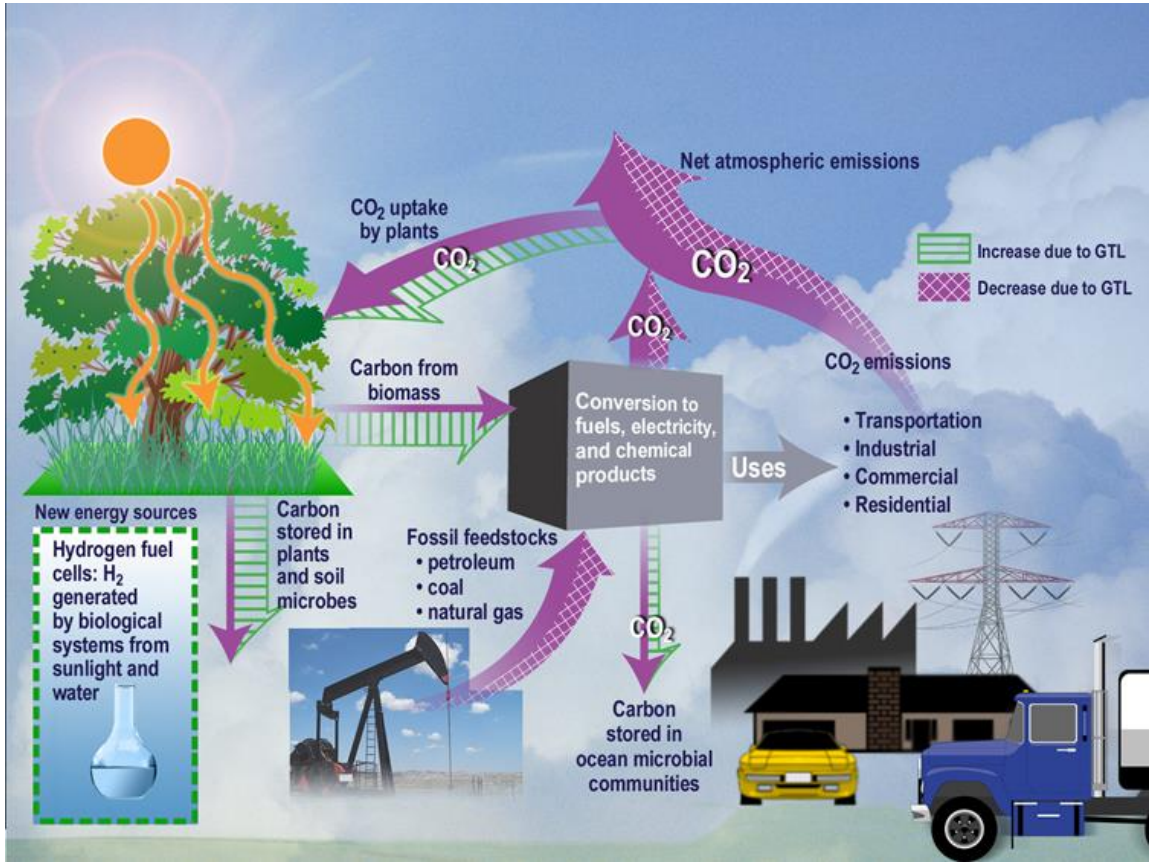
Energy
Tools and concepts for designing and engineering bioenergy plant and microbial systems, including the mechanistic bases.

Carbon Cycle
Tools and concepts to determine the carbon cycling and biosequestration processes of ocean and terrestrial ecosystems.

Environmental Remediation
Microbial and plant modeling and experiments to predict and control contaminant fate and transport.

Technology Endpoints
Payoffs for the Nation

- Sustainable and Viable Biofuel Technologies
- Earth System Modeling and Biosequestration Strategies
- Improved Strategies for Environmental Remediation and Long-Term Stewardship



GTL (Genomes to Life) funded by DOE

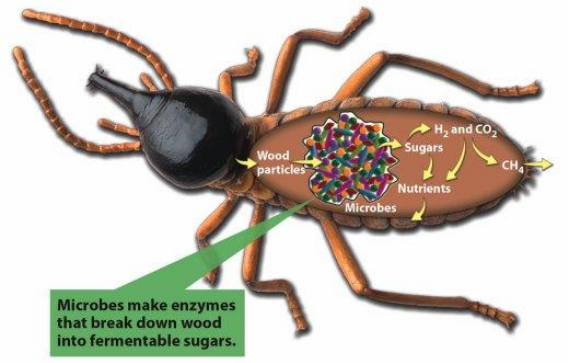
Andreas Brune. **Symbiotic digestion of lignocellulose in termite guts**. Nature Reviews Microbiology 12, 168–180 (2014)

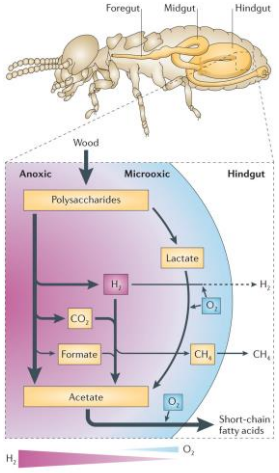
Termites and biofuels

Although the industrial fermentation of sugar-rich and starch-rich crops to ethanol is well established, the **production of so-called second-generation biofuels from agricultural wastes is still inefficient**. A better understanding of lignocellulose digestion by termites may help to overcome challenges in the conversion of lignified plant cell walls into soluble sugars.

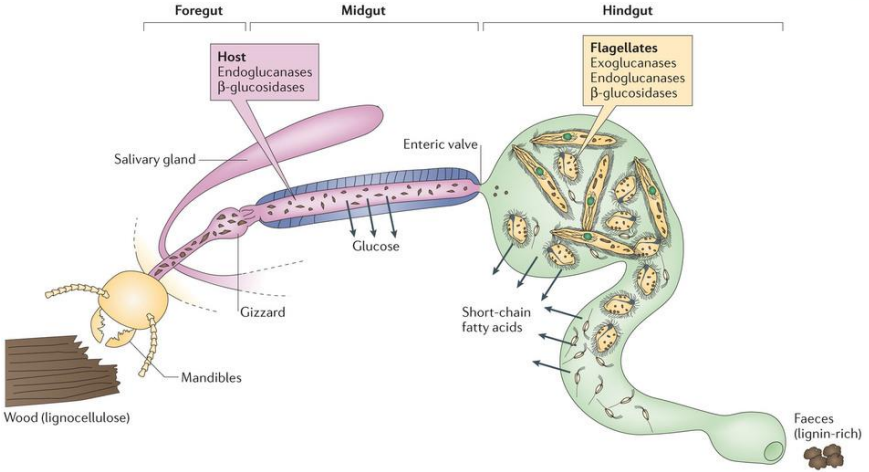
Models for technical processes The **strategies that termites use for the breakdown of lignified plant cell walls resemble technical processes much more closely than those found in other environments**. Mandibles and gizzards are powerful mechanical mills, the midgut is an enzymatic reaction chamber with a permeation filter (the peritrophic membrane) for product recovery and the hindgut paunch is an anaerobic digester that converts polymers to microbial products. The consecutive gut compartments of higher termites form sequential reactors that use the same alkaline pretreatment of lignocellulose as the paper industry. However, other properties of the digestive system are more difficult to mimic. In particular, the minute size of the hindgut bioreactor cannot be scaled up without loss of its intrinsic properties. It creates a delicate balance between the influx and removal of oxygen, which enables oxidative processes and anaerobic fermentations to occur in close juxtaposition. Interactions between the gut lumen, periphery and epithelium do not require radial mixing; diffusion alone suffices as a means of metabolite transport. **Understanding the basis for the suppression of methanogenesis in the wood-feeding species may hold the key to increasing the yields of hydrogen or other valuable products in technical fermentations of plant biomass.**

Sources of novel enzymes Although termites probably cannot be directly used in the processing of agricultural wastes, they are a **promising reservoir of microbial symbionts and enzymes that have biotechnological potential**. Most research has been done on the **endogenous endoglucanases of termites**. They have been heterologously expressed, and their thermostability and catalytic properties have been improved by genetic engineering. **Transgenic enzymes with proper glycosylation and catalytic activities that are superior to those of endoglucanases from bacteria or fungi have been produced in eukaryotic expression systems**. In addition, **some cellulases from gut flagellates have been expressed in different hosts; however, they may require codon optimization to avoid premature polyadenylation. Except for a few xylanases, enzymes of bacterial origin have only been poorly investigated**.

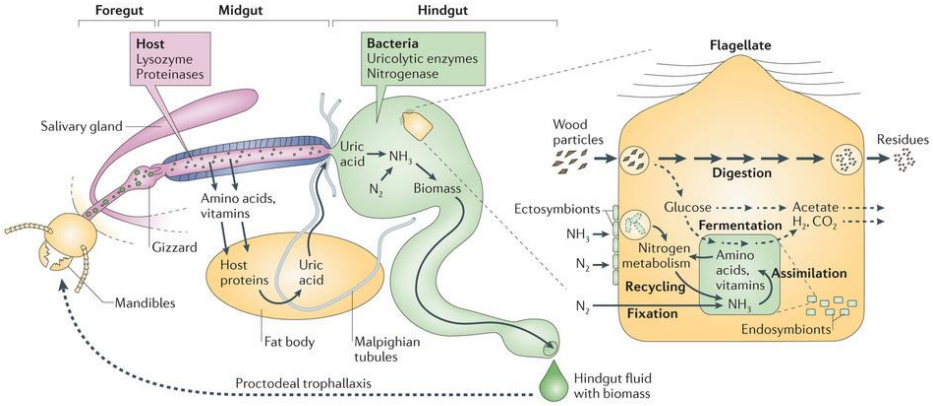




Nature Reviews | Microbiology



Nature Reviews | Microbiology



Nature Reviews | Microbiology

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 Kim,[3] Harshal Chokhawala,[3] Gary Schroth,[4] Shujun Luo,[4] Douglas S. Clark,[3,5] Feng Chen,[1,2] Tao Zhang,[1,2] Roderick I. Mackie,[6] Len A. Pennacchio,[1,2] Susannah G. Tringe,[1,2] Axel Visel,[1,2] Tanja Woyke,[1,2] Zhong Wang,[1,2] Edward M. Rubin[1,2]†

The paucity of enzymes that efficiently deconstruct plant polysaccharides represents a major bottleneck for industrial-scale conversion of cellulosic biomass into biofuels. Cow rumen microbes specialize in degradation of cellulosic plant material, but most members of this complex community resist cultivation. To characterize biomass-degrading genes and genomes, we sequenced and analyzed 268 gigabases of metagenomic DNA from microbes adherent to plant fiber incubated in cow rumen. From these data, we identified 27,755 putative carbohydrate-active genes and expressed 90 candidate proteins, of which 57% were enzymatically active against cellulosic substrates. We also assembled 15 uncultured microbial genomes, which were validated by complementary methods including single-cell genome sequencing. These data sets provide a substantially expanded catalog of genes and genomes participating in the deconstruction of cellulosic biomass.

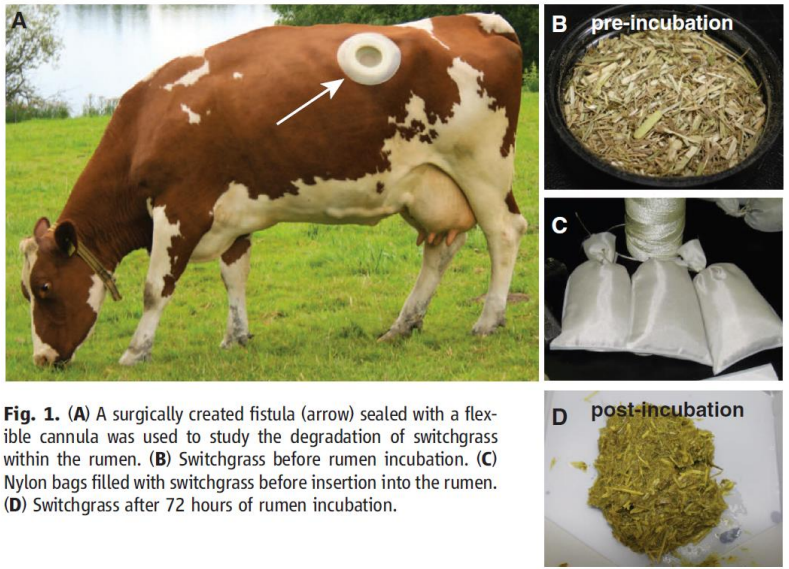
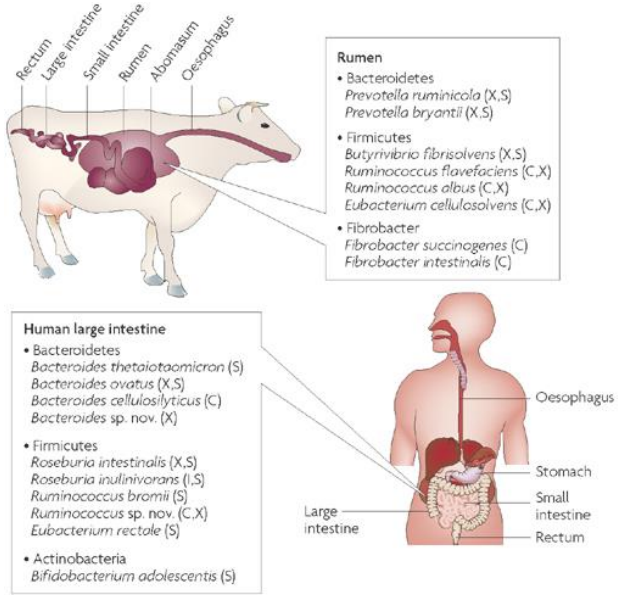


Fig. 1. (**A**) A surgically created fistula (arrow) sealed with a flexible cannula was used to study the degradation of switchgrass within the rumen. (**B**) Switchgrass before rumen incubation. (**C**) Nylon bags filled with switchgrass before insertion into the rumen. (**D**) Switchgrass after 72 hours of rumen incubation.



Nature Reviews Microbiology

Polysaccharide-degrading bacteria in the ruminant and human gastrointestinal tracts

The **major sites of microbial breakdown of dietary polysaccharides, which also support the highest densities of bacteria, are the rumen in ruminant animals and the large intestine in humans**. Examples of cultured polysaccharide-degrading species are shown for these sites, together with the phylum to which they belong (Firmicutes or Bacteroidetes) and their characteristic polysaccharide-utilizing abilities. Much of the diversity remains undefined, however, and new species of polysaccharide-utilizing bacteria have been described recently in the human colon.

C, cellulose; I, inulin; S, starch; X, xylan.

	Bacteroides thetaiotaomicron 5482	*Bifidobacterium longum* NCC2705	*Ruminococcus flavefaciens* FD1	*Fibrobacter succinogenes* S85
Location	Human colon	Human colon	Rumen, cellulolytic	Rumen, cellulolytic
Genome size	6.26 Mb (complete genome)	2.26 Mb (complete genome)	Approximately 4 Mb (partial genome; dockerin-encoding genes only)	3.8 Mb (complete genome)
Number of glycoside hydrolases*	236 (40)	47 (17)	65 (14)	104
Number of polysaccharide lyases*	15 (7)	0	12 (4)	4
Number of carbohydrate esterases*	20 (9)	1	23 (5)	14
Number of carbohydrate-binding modules*	16 (3)	10 (5)	61 (12)	Limited information available
References	77	106	58	68,69

*The number of enzyme families that are represented is shown in brackets (CAZY (carbohydrate-active enzymes) database; see Further information). Mb, megabases.

Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis

Harry J. Flint, Edward A. Bayer, Marco T. Rincon, Raphael Lamed & Bryan A. White

Nature Reviews Microbiology 6, 121-131 (February 2008)

Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*

Gregory Bokinsky[a,b], Pamela P. Peralta-Yahya[a,b], Anthe George[a,c], Bradley M. Holmes[a,c], Eric J. Steen[a,d], Jeffrey Dietrich[a,d], Taek Soon Lee[a,e], Danielle Tullman-Ercek[a,f], Christopher A. Voigt[g], Blake A. Simmons[a,c], and Jay D. Keasling[a,b,d,e,f,1]

One approach to reducing the costs of advanced biofuel production from cellulosic biomass is to engineer a single microorganism to both digest plant biomass and produce hydrocarbons that have the properties of petrochemical fuels. Such an organism would require pathways for hydrocarbon production and the capacity to secrete sufficient enzymes to efficiently hydrolyze cellulose and hemicellulose. To demonstrate how one might engineer and coordinate all of the necessary components for a biomass-degrading, hydrocarbon-producing microorganism, we engineered a microorganism naïve to both processes, *Escherichia coli*, to grow using both the cellulose and hemicellulose fractions of several types of plant biomass pretreated with ionic liquids. Our engineered strains express cellulase, xylanase, beta-glucosidase, and xylobiosidase enzymes under control of native *E. coli* promoters selected to optimize growth on model cellulosic and hemicellulosic substrates. Furthermore, our strains grow using either the cellulose or hemicellulose components of ionic liquid-pretreated biomass or on both components when combined as a coculture. Both cellulolytic and hemicellulolytic strains were further engineered with three biofuel synthesis pathways to demonstrate the production of fuel substitutes or precursors suitable for gasoline, diesel, and jet engines directly from ionic liquid-treated switchgrass without externally supplied hydrolase enzymes. This demonstration represents a major advance toward realizing a consolidated bioprocess. With improvements in both biofuel synthesis pathways and biomass digestion capabilities, our approach could provide an economical route to production of advanced biofuels.

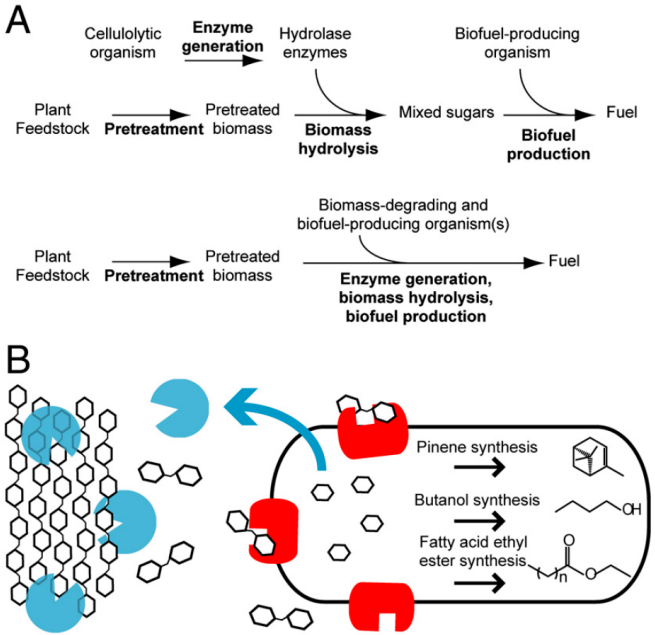


Fig. 1. Consolidated bioprocessing of plant biomass into biofuels by *E. coli*. (*A*) Two processes for biofuel production. Typically, cellulase and hemicellulase enzymes are produced in a process step separate from biomass hydrolysis and biofuel production (top). Consolidated bioprocessing (bottom) combines enzyme generation, biomass hydrolysis, and biofuel production into a single stage. (*B*) Engineering *E. coli* for use in consolidated bioprocessing. Cellulose and hemicellulose are hydrolyzed by secreted cellulase and hemicellulose enzymes (cyan) into soluble oligosaccharides. β-glucosidase enzymes (red) further hydrolyze the oligosaccharides into monosaccharides, which are metabolized into biofuels via heterologous pathways.

Genomics, metagenomics and proteomics in biomining microorganisms

Lissette Valenzuela[a], An Chi[b], Simon Beard[a], Alvaro Orell[a], Nicolas Guiliani[a], Jeff Shabanowitz[b], Donald F. Hunt[b], Carlos A. Jerez[a,*]

The use of acidophilic, chemolithotrophic microorganisms capable of oxidizing iron and sulfur in industrial processes to recover metals from minerals containing copper, gold and uranium is a well established biotechnology with distinctive advantages over traditional mining. A consortium of different microorganisms participates in the oxidative reactions resulting in the extraction of dissolved metal values from ores. Considerable effort has been spent in the last years to understand the biochemistry of iron and sulfur compounds oxidation, bacteria–mineral interactions (chemotaxis, quorum sensing, adhesion, biofilm formation) and several adaptive responses allowing the microorganisms to survive in a bioleaching environment. All of these are considered key phenomena for understanding the process of biomining. The use of genomics, metagenomics and high throughput proteomics to study the global regulatory responses that the biomining community uses to adapt to their changing environment is just beginning to emerge in the last years. These powerful approaches are reviewed here since they offer the possibility of exciting new findings that will allow analyzing the community as a microbial system, determining the extent to which each of the individual participants contributes to the process, how they evolve in time to keep the conglomerate healthy and therefore efficient during the entire process of bioleaching.

How can we use genomic data?

Microbial Genomics: vaccines, antibiotics, and diagnostics for infectious diseases, new energy fuels (biofuels); environmental monitoring to detect pollutants; protection from biological and chemical warfare; and safe, efficient toxic-waste cleanup.

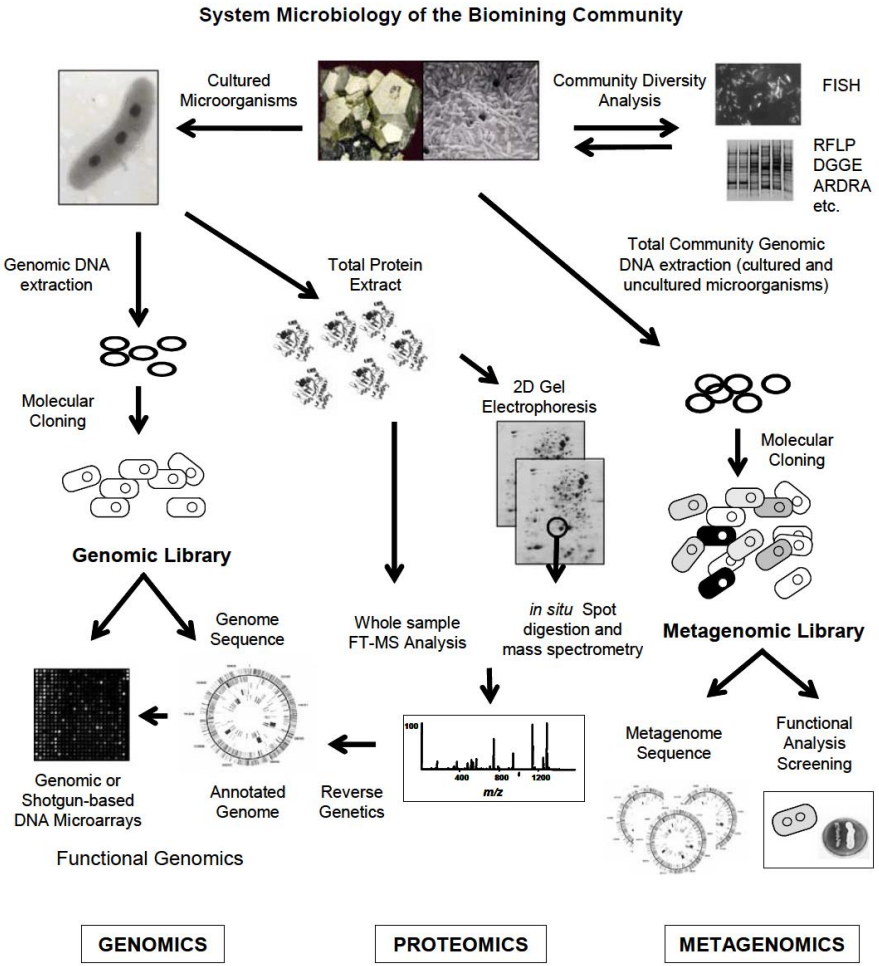


Fig. 1. Overview of the application of genomics, proteomics and metagenomics to biomining microorganisms.