



Next Generation Microbial Genomics: The Human Microbiome Project

George Weinstock

San Rocco: Protector from Infectious Diseases



Large genome centers

All have metagenomics programs

- Baylor College of Medicine (NIH-NHGRI)
- Beijing Genome Institute
- Broad Institute (NIH-NHGRI)
- J. Craig Venter Institute
- Joint Genome Institute (US-DOE)
- Washington University (NIH-NHGRI)
- Sanger Institute (Wellcome Trust)

What is the DNA Revolution to Genome Centers?

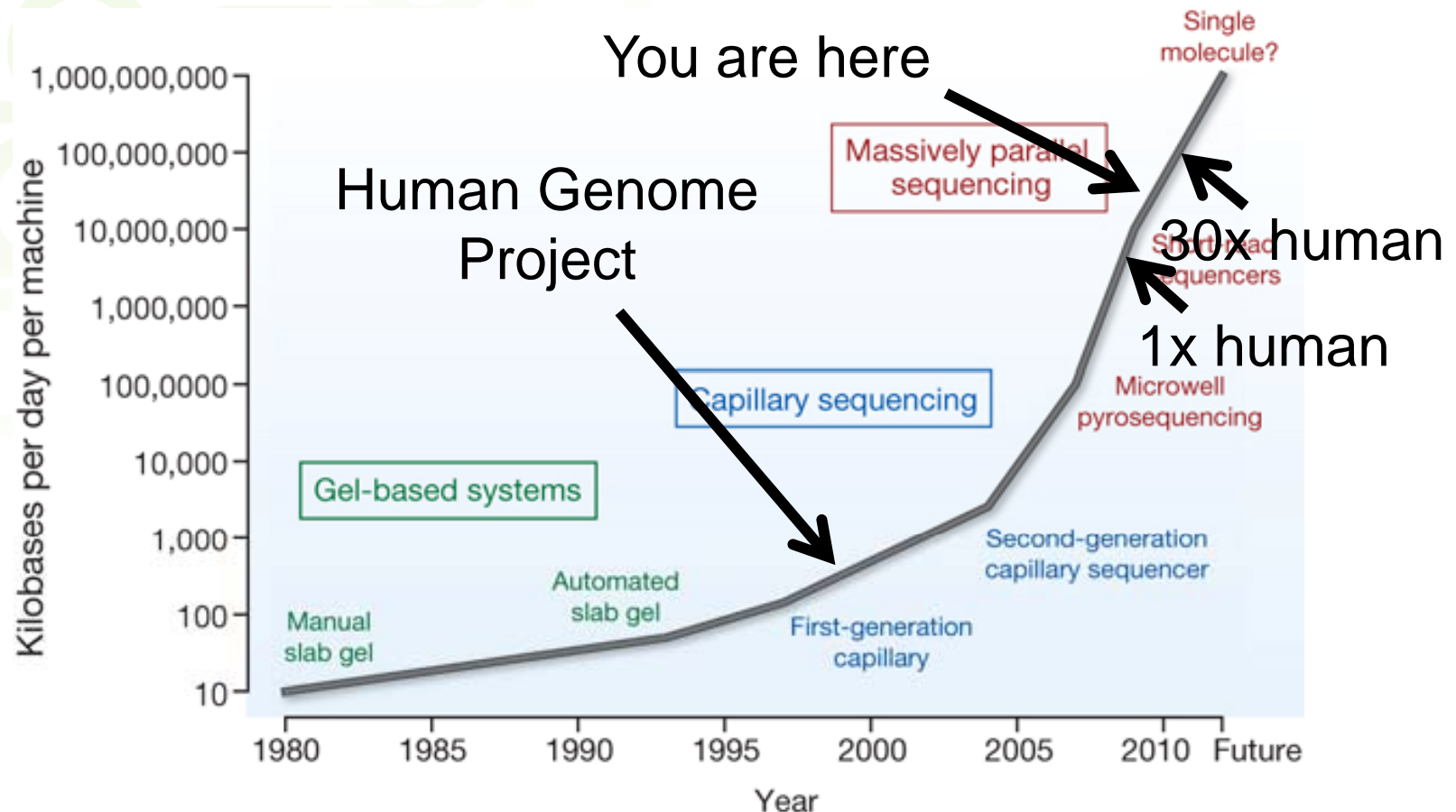
The Genome Center at Washington University



- Cancer genomics
- Human genomics
- Microbial genomics
- De novo genome projects
- Bioinformatics

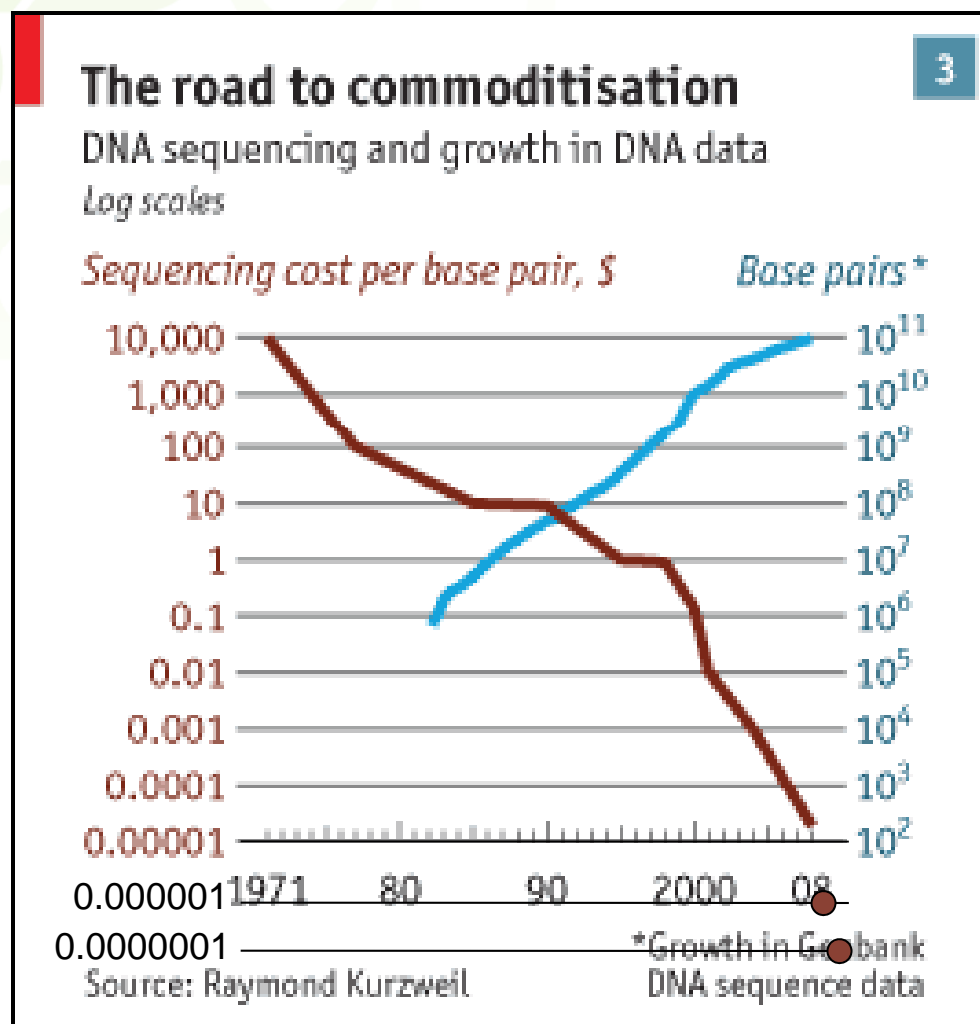
- Established in 1993.
- One of three NHGRI-funded genome centers.
- ~290 employees
- Sequenced 25% of the Human Genome; many other genome projects.
- >\$400M in grants received.
- 20 3730, 38 Illumina GAIIx, 8 454 Titanium, 1 SOLiD
- ~95 GB sequence/day
- >500 million reads/day

Improvements in the rate of DNA sequencing over the past 30 years and into the future



MR Stratton *et al. Nature* **458**, 719-724 (2009)

Costs and Growth of Total Data



222:1



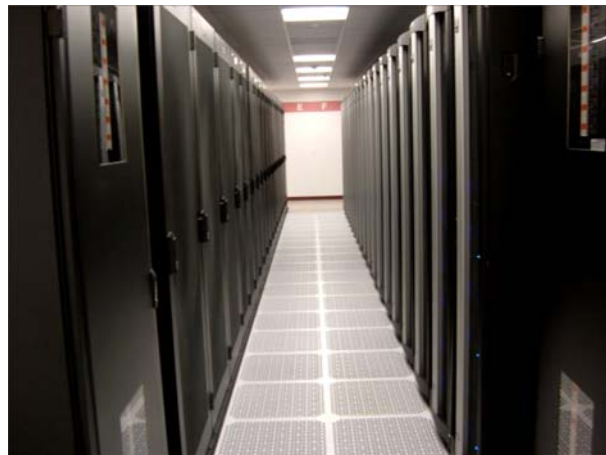
Informatics

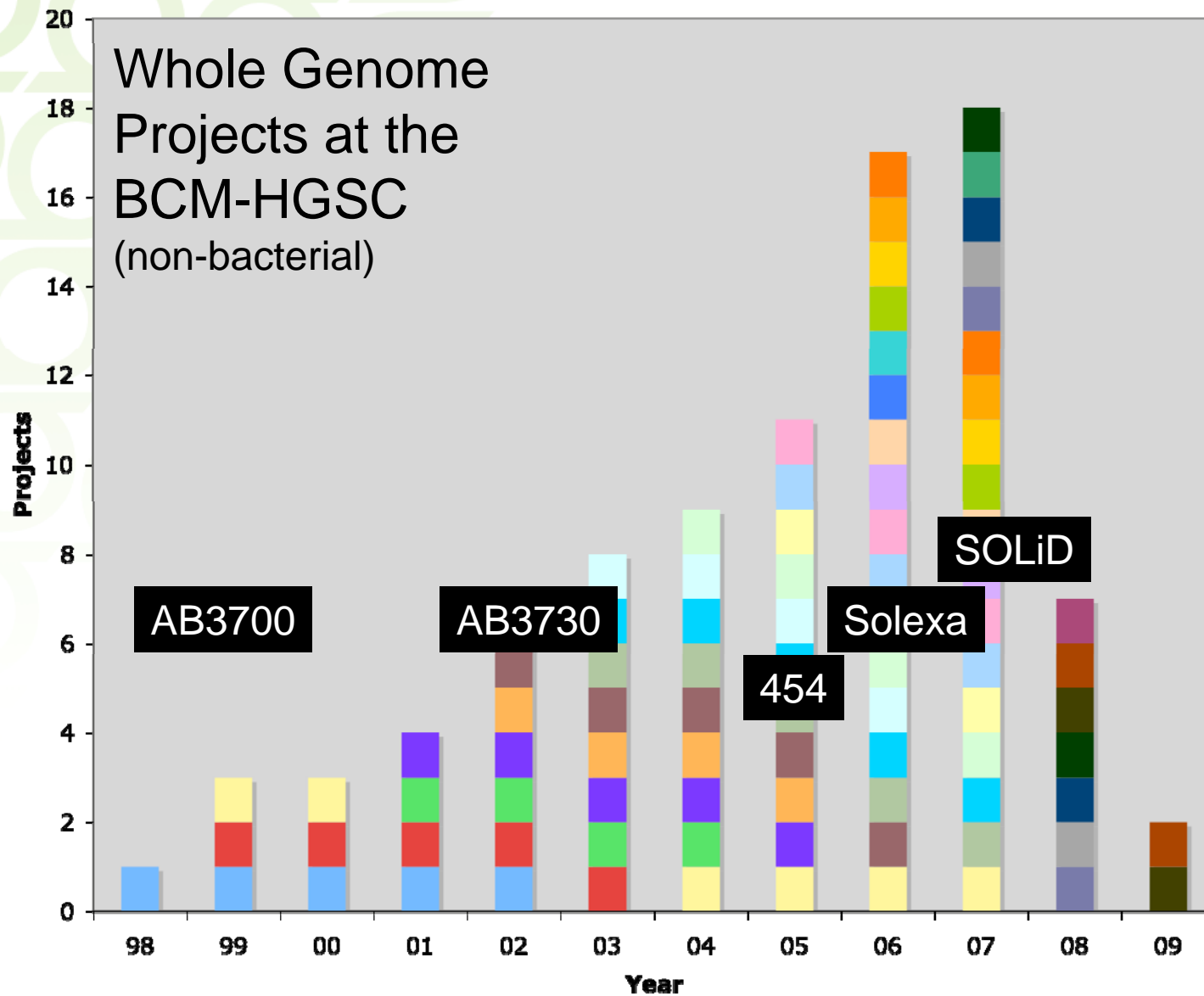


Washington Univ. Genome Data Center



- 4 PB + 3500 cores
- Additional data center space underway
- New data management models being discussed
 - Old: transfer data to central site eg NCBI
 - New: cloud computing eg to avoid data transfer





2007 was the year of the personal genome

- Watson (Roche-454, BCM-HGSC) - 454
 - Venter (JCVI) - Sanger
 - AML (Wash U) - Illumina
 - Yoruban HapMap genome – Illumina
-
- 100s of human genomes sequenced by now

Sequence a Human Genome a day

- 30x coverage
- Increase Illumina instruments from 14 to 38
- Focus on tumor-normal pairs
- Manage samples and sequencing of 100s of genomes
- Match pace of analysis with data production
- Expand computational infrastructure

Have already achieved this!

Now aiming for 5 genomes/day

The Future

- Understand how to do “Personal Genomics”
- Whole genome sequencing for disease gene discovery
- Expect new sequencing technology in 2010
- Challenge of efficient computational solutions
- All is wasted without good clinical information

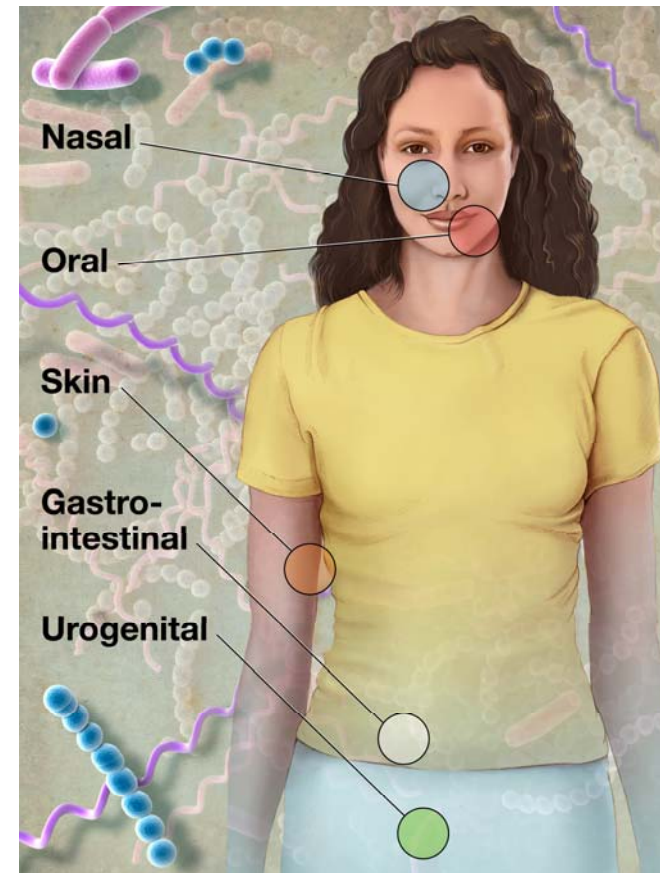
The Human Body: An extended community



- We are a composite:
 - Bacteria
 - Archaea
 - Eukaryotes
 - Viruses
 - ...*and Human* cells
- This is our microbiome
- 100 trillion cells total
 - 10% are human
 - Most are bacterial

Where did our microbiome come from?

- Microbes have been on earth for 3 billion years
 - We evolved in the midst of microbes
- We evolved organs to house our microbes
- Each body site has its own microbial community

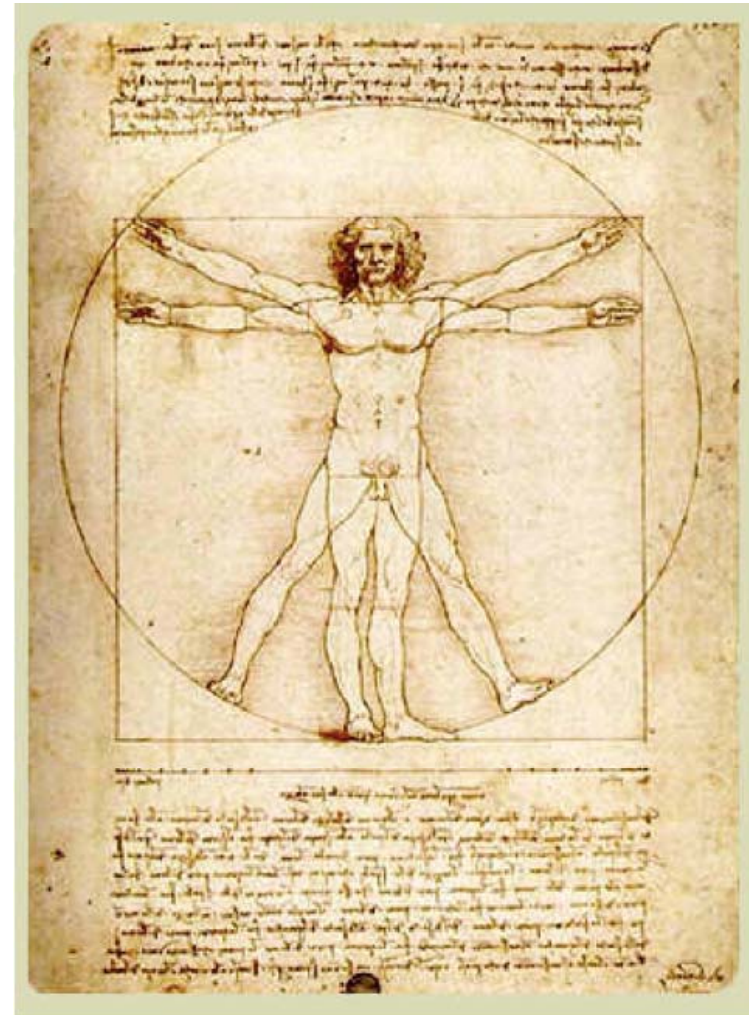


- The microbial genomes make up a metagenome – *our second genome*.
- Our metagenome has 100-fold more microbial genes than human genes.



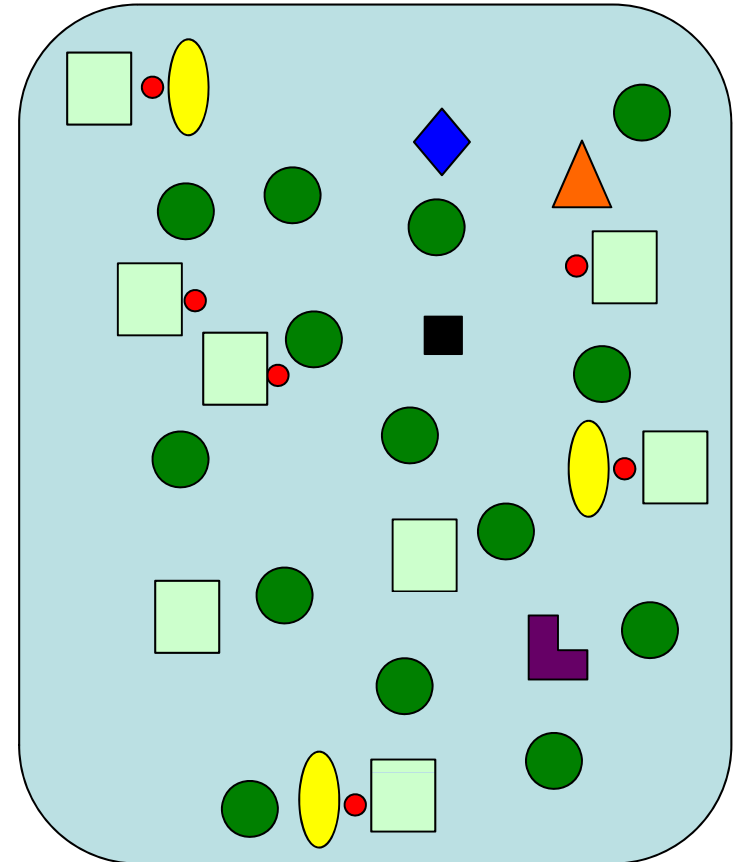
An extended view of ourselves as a life form

- The microbiome is an integral part of our genetic landscape *and of our genetic evolution* – we are a superorganism

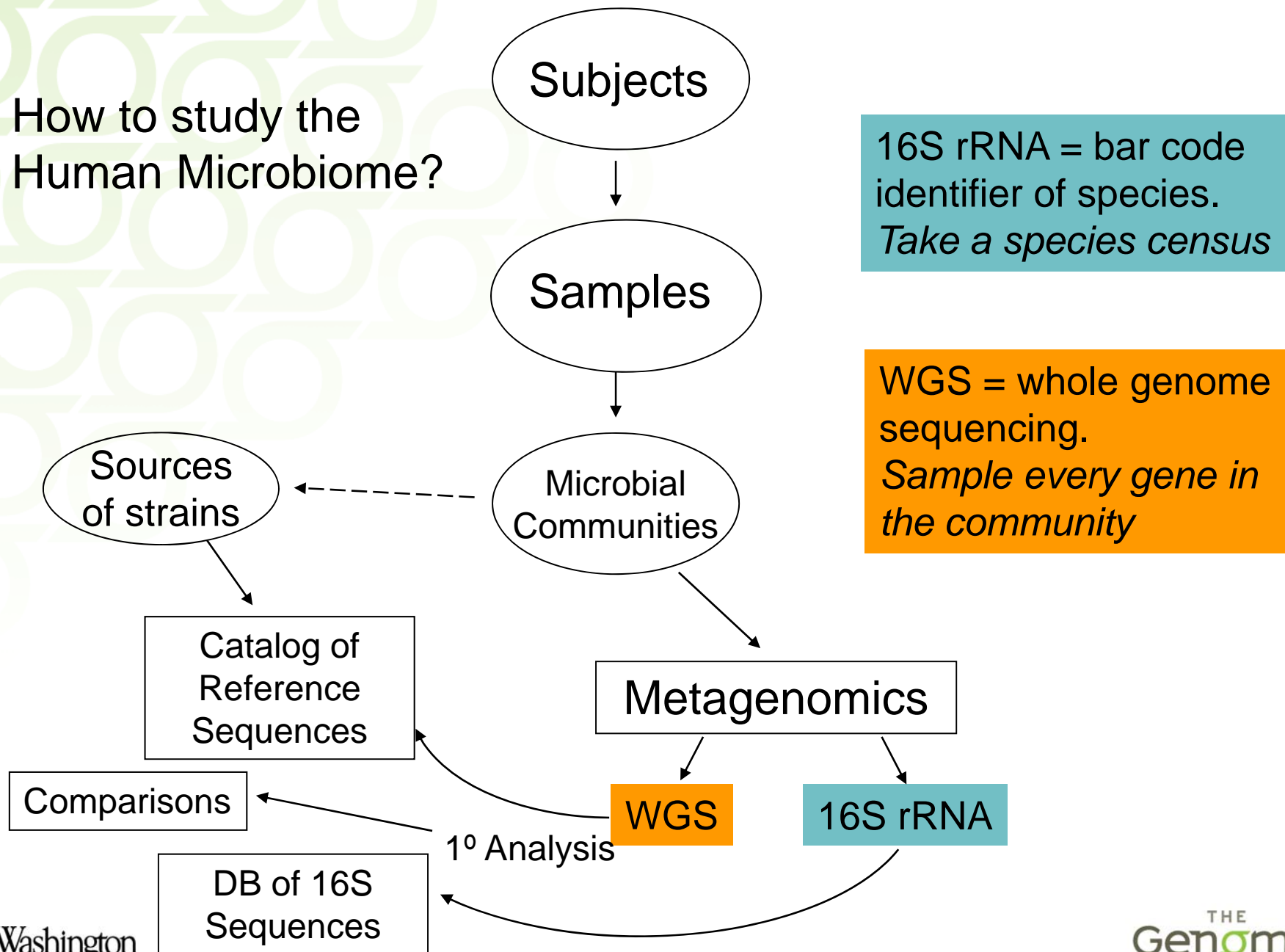


What is a microbial community?

- Hundreds of species, Millions+ of cells
- Mainly not culturable
- Varying abundances
- Sequence the metagenome
 - Sequence a mix of genomes
 - Like a Human Genome Project in scale

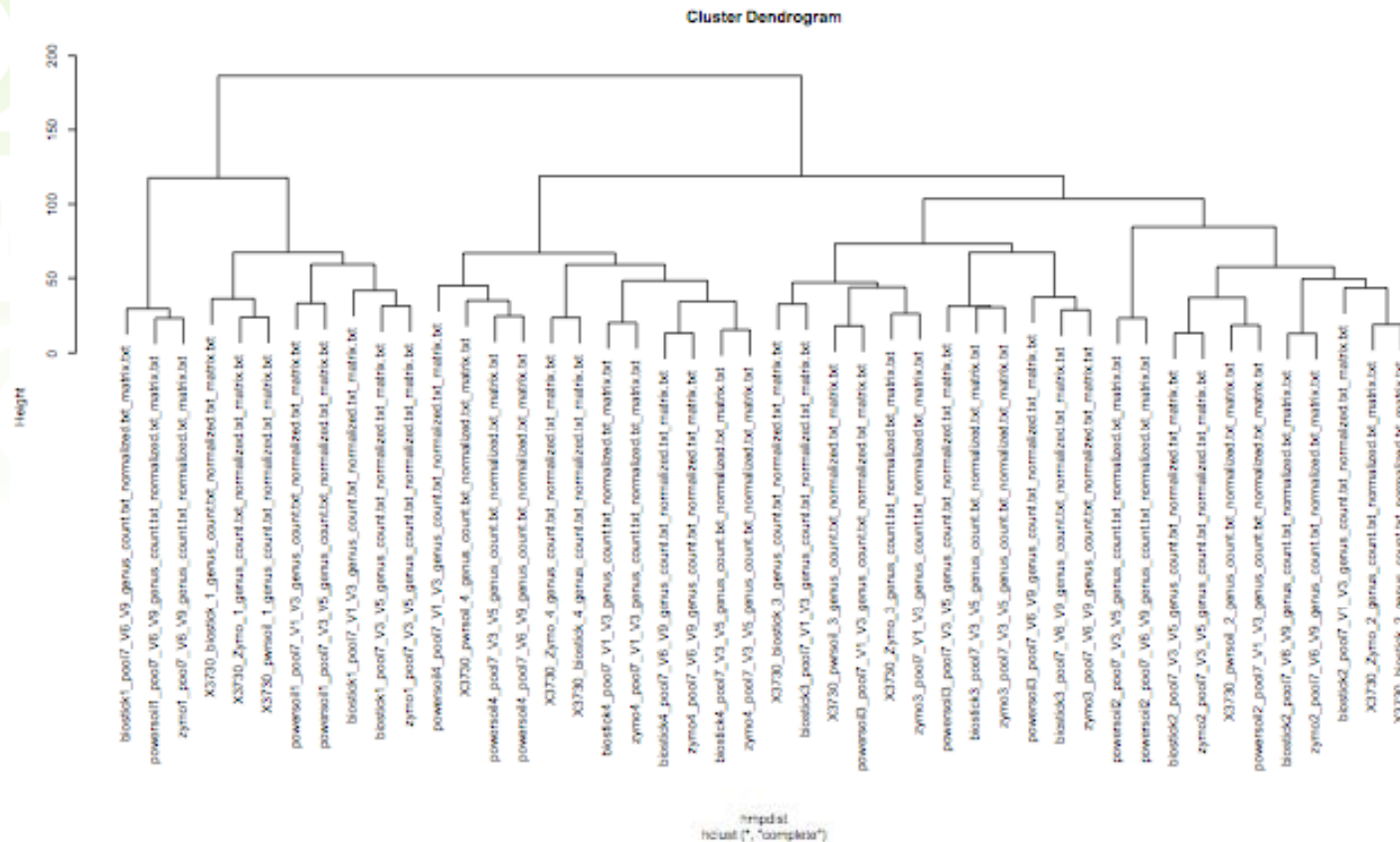


How to study the Human Microbiome?



Clustering 454 16S rRNA data

Fecal samples, E. coli infections



Patient 1

Patient 4

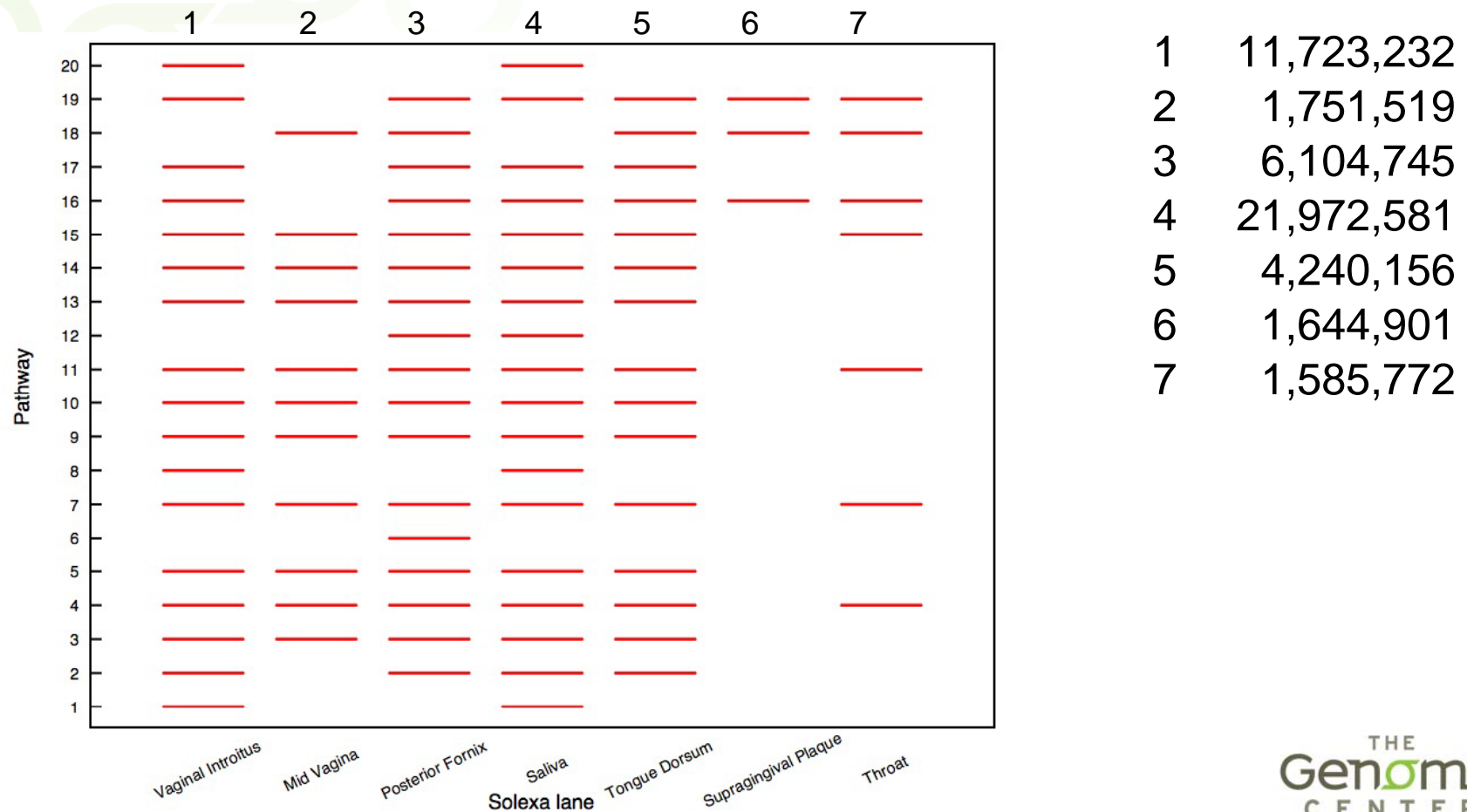
Patient 3

Patient 2

Metabolic Pathway Distribution

Shotgun sequencing of communities

- 131 pathways present in every sample;
- 19 pathways absent in every sample;
- 20 pathways were restricted to one or more body sites.



NIH Human Microbiome Project

- Sequencing 100s/1000s of genomes (catalog)
- Sequencing metagenomes (communities)
 - 375 healthy subjects, 18 body sites, 12,000 specimens
- Discovery of new organisms (uncultured)
 - bacteria, eukaryotes, viruses
- Linking to disease
 - Lurking pathogens
 - Community imbalances (over or undergrowth)
- 5 years, \$140 million+
- 4 Genome Centers, 15 other groups

Sugar cane

Maize

Rice

Dental biofilm
(plaque)

Carbos $\xrightarrow{\text{Fermentation}}$ Lactic acid

Streptococcus mutans
Lactobacillus spp.
Actinomyces spp.

Demineralization
(favored pH<5.5)

Mineralization

Demonstration Projects

Psoriasis	Skin	NYU
Crohn's Disease	GI	WUSTL
Crohn's Disease	GI	U Md
Crohn's Disease	GI	Penn
Ulcerative colitis	GI	U Mich
Vaginitis	Vagina	VCU
Urethritis	Urethra	Indiana U/WUSTL
Obesity	GI	U Md
Fever unknown origin	Systemic	WUSTL
Acne	Skin	UCLA/WUSTL
Necrotizing enterocolitis	GI	WUSTL
Eczema	Skin	NHGRI
Abdominal pain children	GI	Baylor
Vaginitis	GI	U Md
Esophageal cancer	Respiratory	NYU

Other Wash. Univ.: HIV, Periodontitis, Primate gut, NALFD

The Future

- The Virome
- Diagnostics
 - Measuring the microbiome
- Therapeutics
 - Manipulating the microbiome
 - Probiotics
- Commercialization opportunities