"Quantifying Your Superorganism Body Using Big Data Supercomputing"

Ken Kennedy Institute Distinguished Lecture

Rice University

Houston, TX

November 12, 2013

Dr. Larry Smarr

Director, California Institute for Telecommunications and Information Technology

Harry E. Gruber Professor,

Dept. of Computer Science and Engineering

Jacobs School of Engineering, UCSD

http://lsmarr.calit2.net

J. Craig Venter™ INSTITUTE

it²

Abstract

The human body is host to 100 trillion microorganisms, ten times the number of cells in the human body and these microbes contain 100 times the number of DNA genes that our human DNA does. The microbial component of this "superorganism" is comprised of hundreds of species spread over many taxonomic phyla. The human immune system is tightly coupled with this microbial ecology and in cases of autoimmune disease, both the immune system and the microbial ecology can have excursions far from normal. There are even some tantalizing clues that certain types of dysbiosis in the gut microbiome can be precursors of some forms of cancer. Using massive amounts of data that I collected on my own body over the last five years, I will show detailed examples of the episodic evolution of this coupled immune-microbial system. To decode the details of the microbial ecology requires high resolution genome sequencing feeding Big Data parallel supercomputers. We have also developed innovative scalable visualization systems to examine the complexities of my time-varying microbial ecology and its relations to the NIH Human Microbiome Program data on people in states of health and disease.

My View on My Own Body Was Shaped by My Lifetime of Scientific Experience

- **No Formal Training in Biology or Medicine**
- **Instead, Decades of:**
  - **Observational & Computational Astrophysics**
  - **Observing & Building Coral Reef Ecologies**

I Spent Decades Studying the Ecological Dynamics of Multi-Phyla Coral Reefs



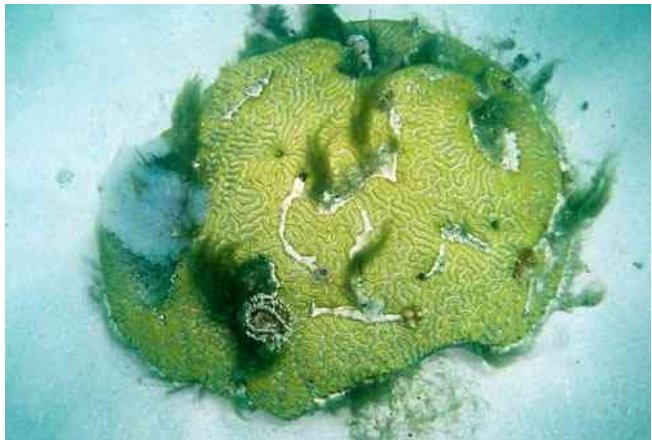


My 120 Gallon Home Salt Water Coral Reef Aquarium in Illinois



Pristine

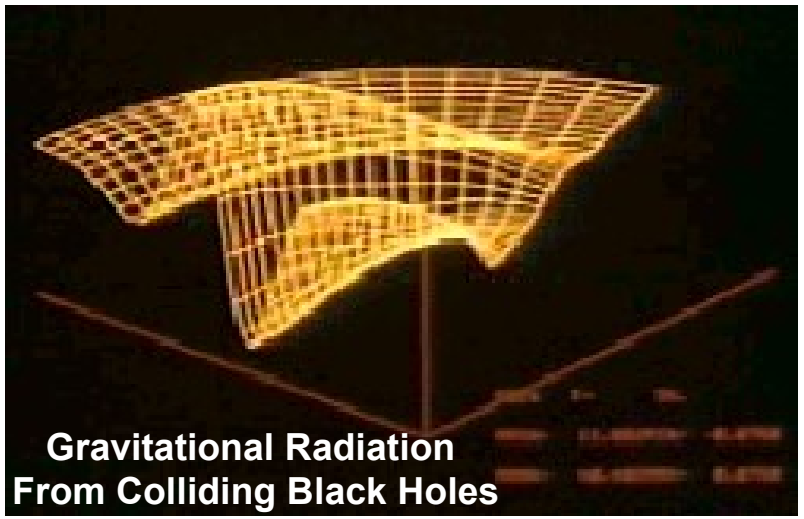
Degraded



My Snorkeling Photos From Coral Reefs

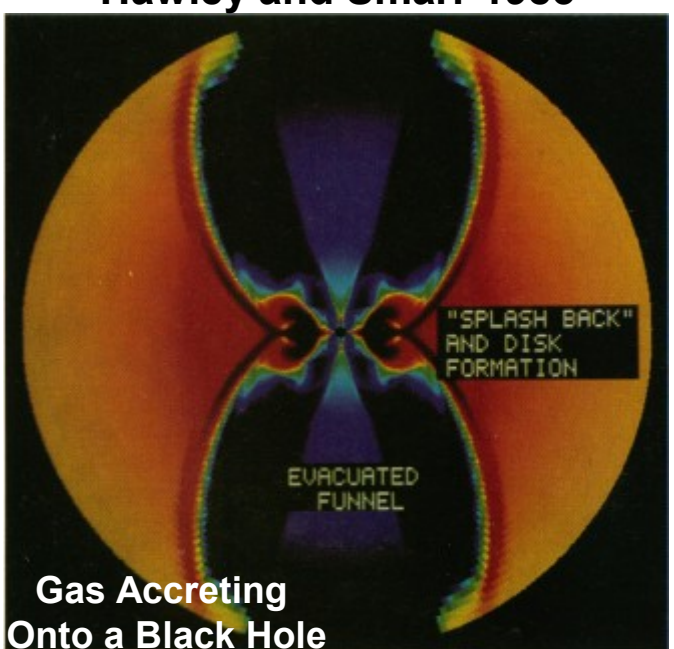
My Early Research was on Computational Astrophysics – I Learned To Think About Nonlinear Dynamic Systems

Eppley and Smarr 1977

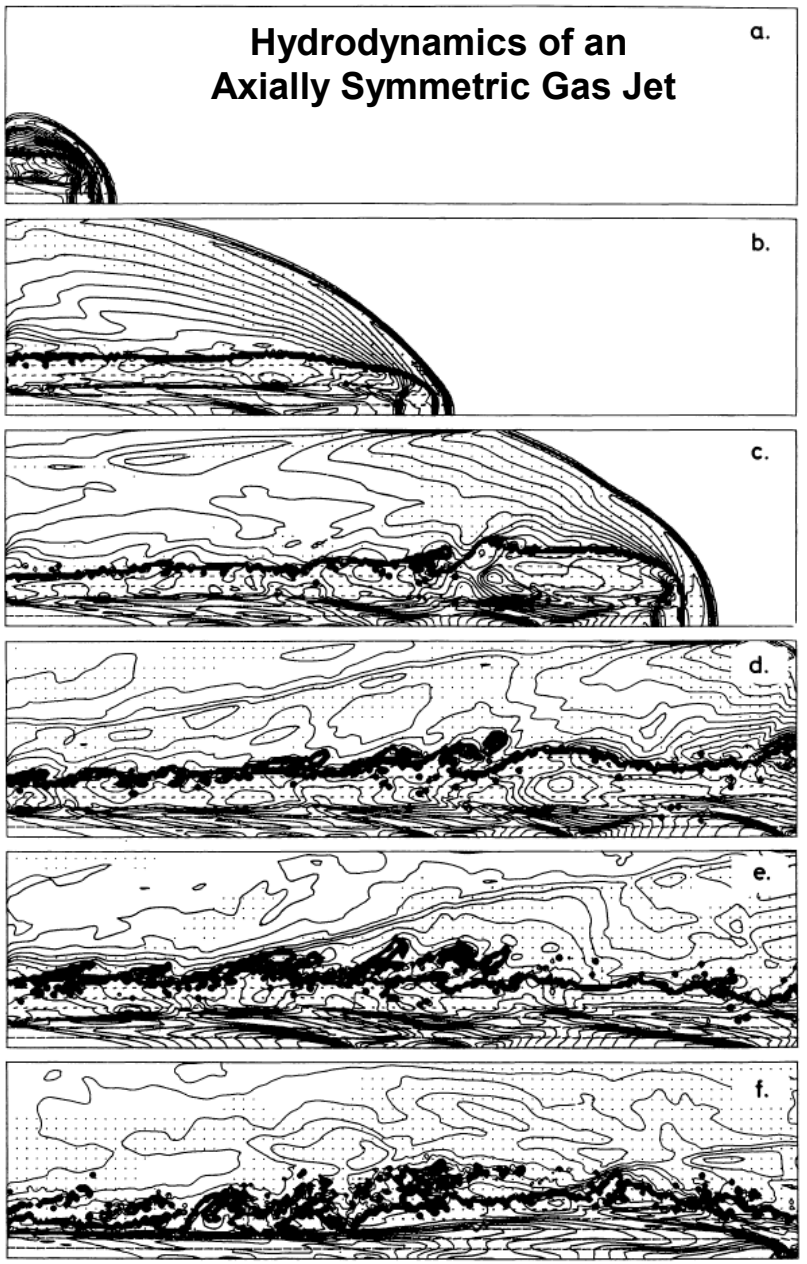


Gravitational Radiation
From Colliding Black Holes

Hawley and Smarr 1985



Gas Accreting
Onto a Black Hole



Norman, Winkler, Smarr, Smith 1982

I Arrived in La Jolla in 2000 After 20 Years in the Midwest and Decided to Move Against the Obesity Trend

Age 41



Age 51



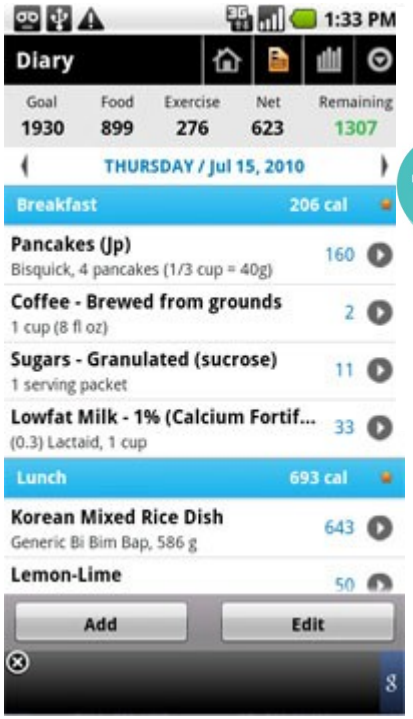
Age 61



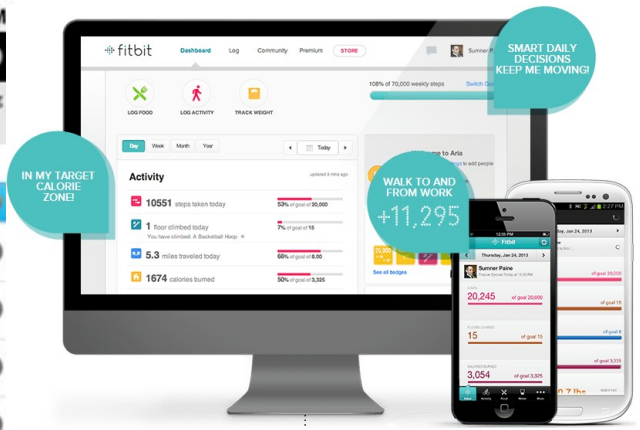
I Reversed My Body's Decline By Quantifying and Altering Nutrition and Exercise

http://lsmarr.calit2.net/repository/LS_reading_recommendations_FiRe_2011.pdf

Challenge-Develop Standards to Enable MashUps of Personal Sensor Data Across Private Clouds



MyFitnessPal-Calories Ingested



FitBit - Daily Steps & Calories Burned



Withing/iPhone-Blood Pressure





EM Wave PC-Stress



Azumio-Heart Rate



Zeo-Sleep

From Measuring Macro-Variables to Measuring Your Internal Variables

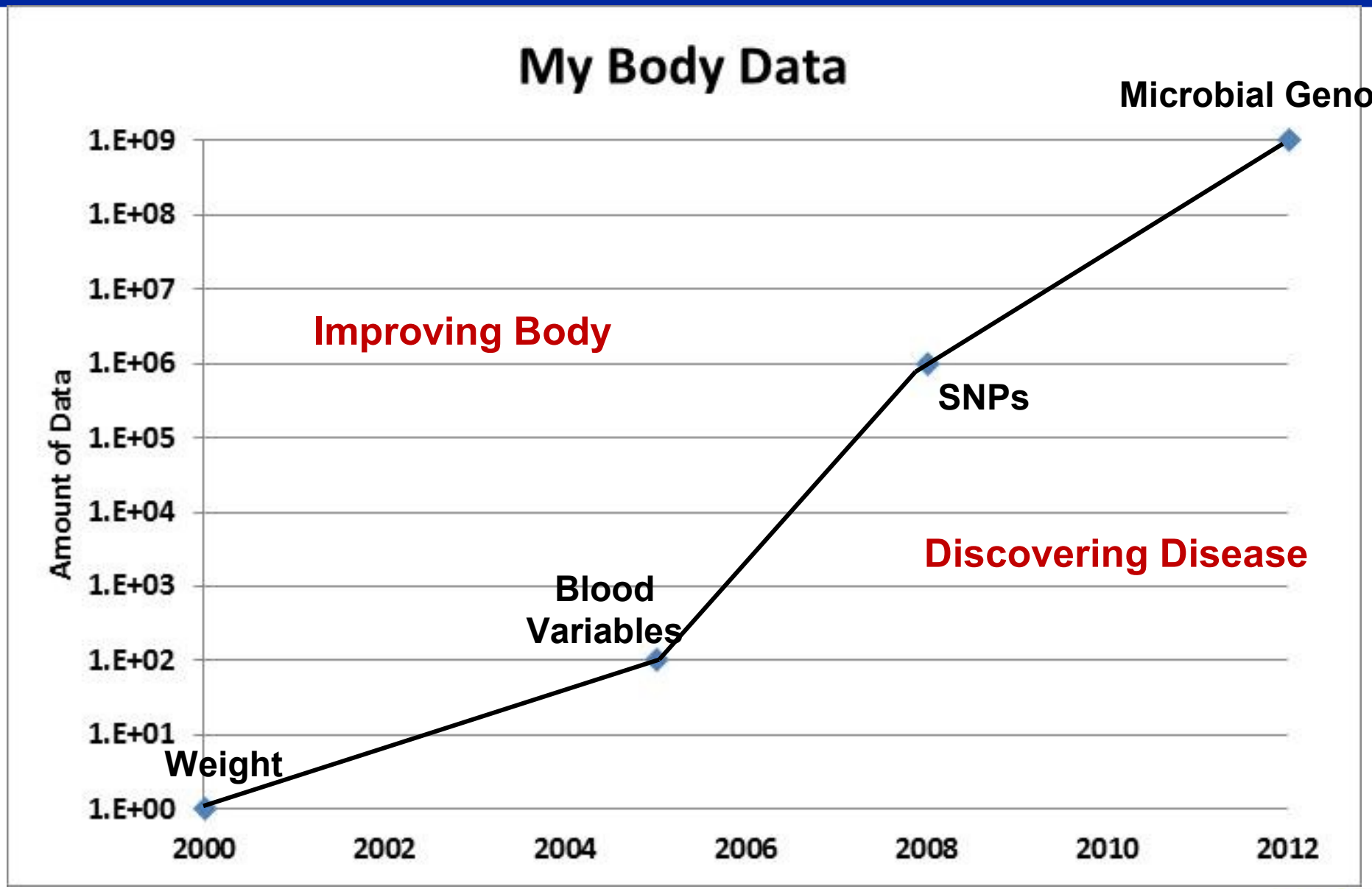


The Patient of the Future

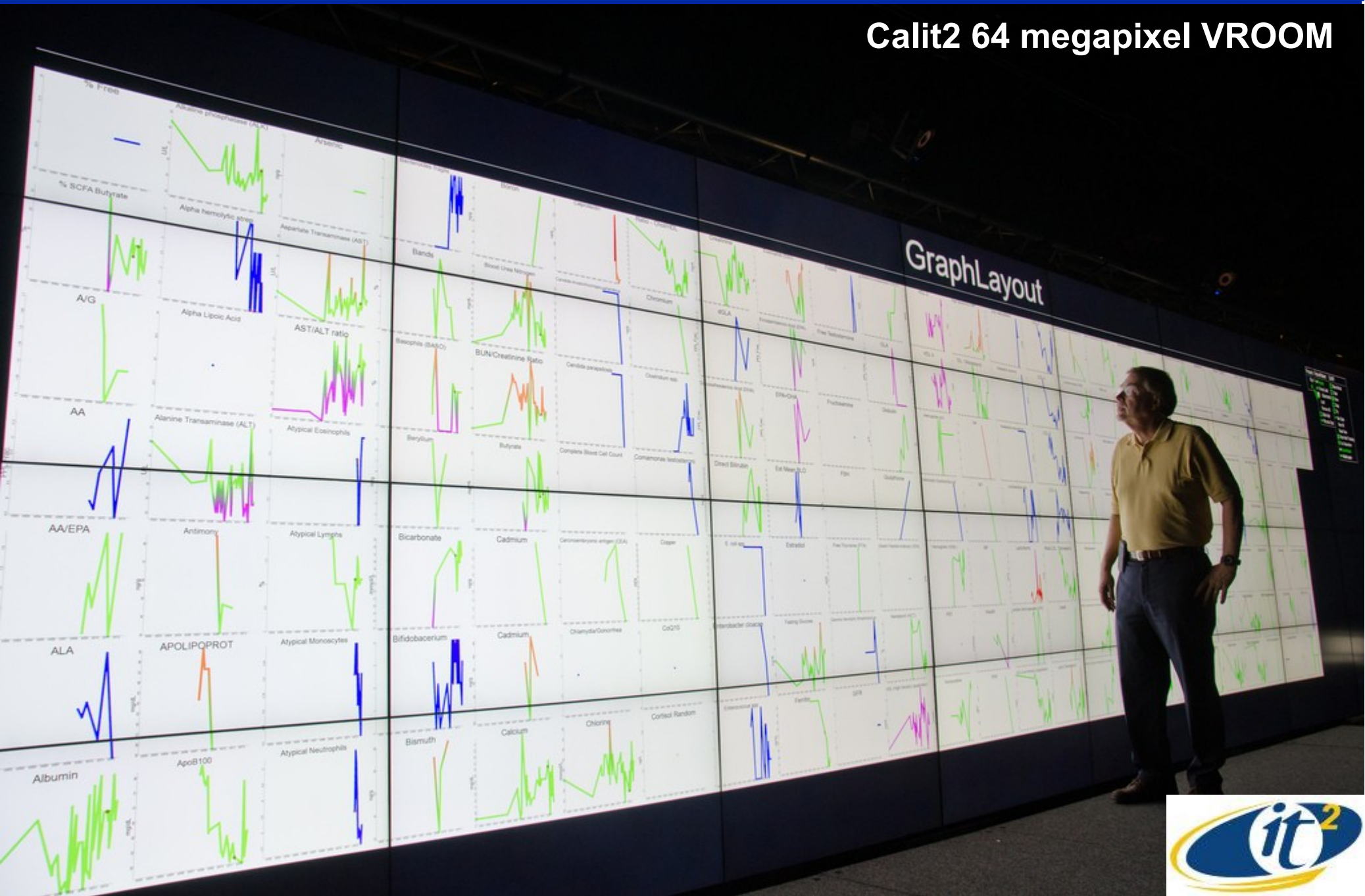
Internet pioneer Larry Smarr's quest to quantify everything about his health led him to a startling discovery, an unusual partnership with his doctor, and more control over his life.

www.technologyreview.com/biomedicine/39636

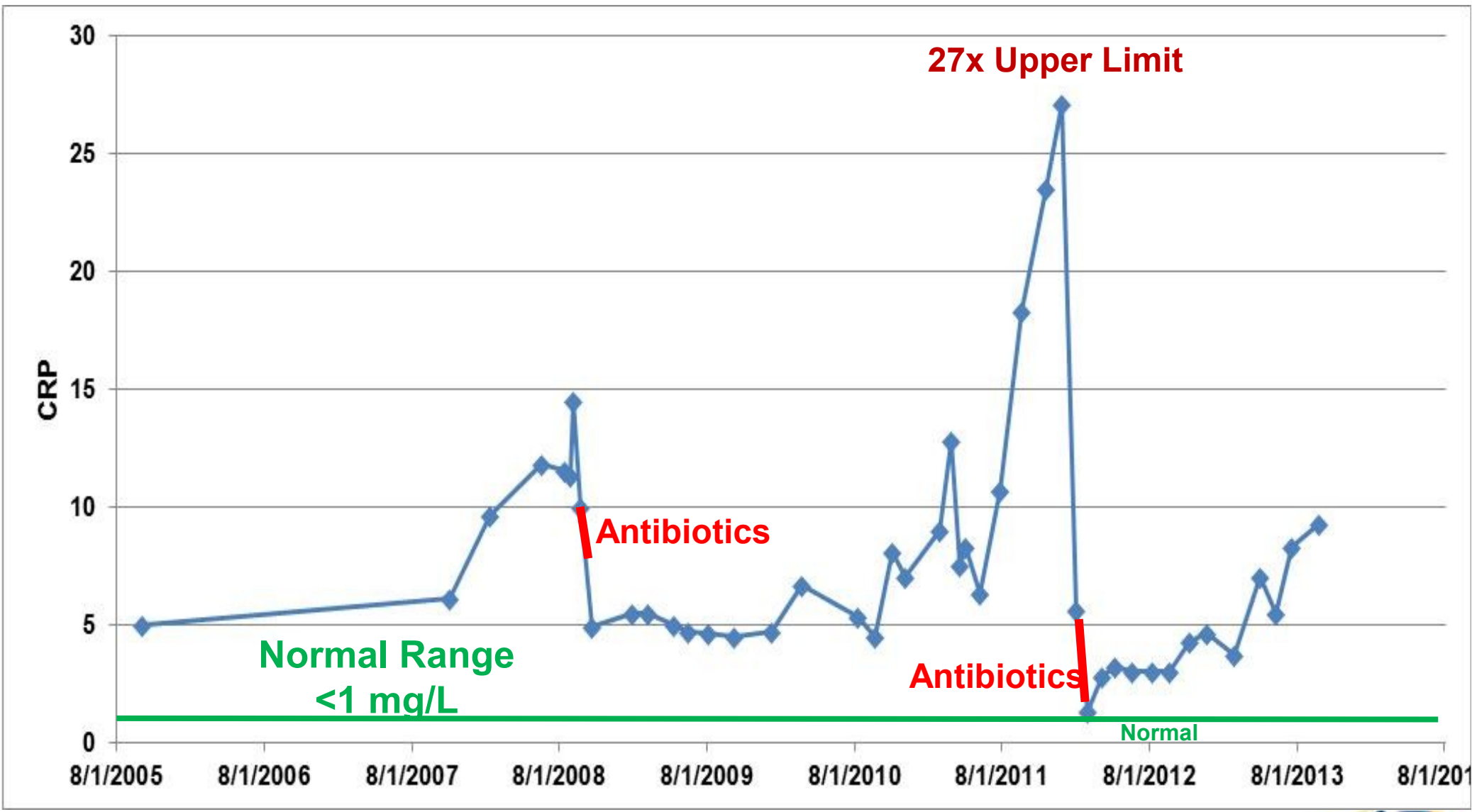
From One to a Billion Data Points Defining Me: The Exponential Rise in Body Data in Just One Decade!



Visualizing Time Series of 150 LS Blood and Stool Variables, Each Over 5-10 Years

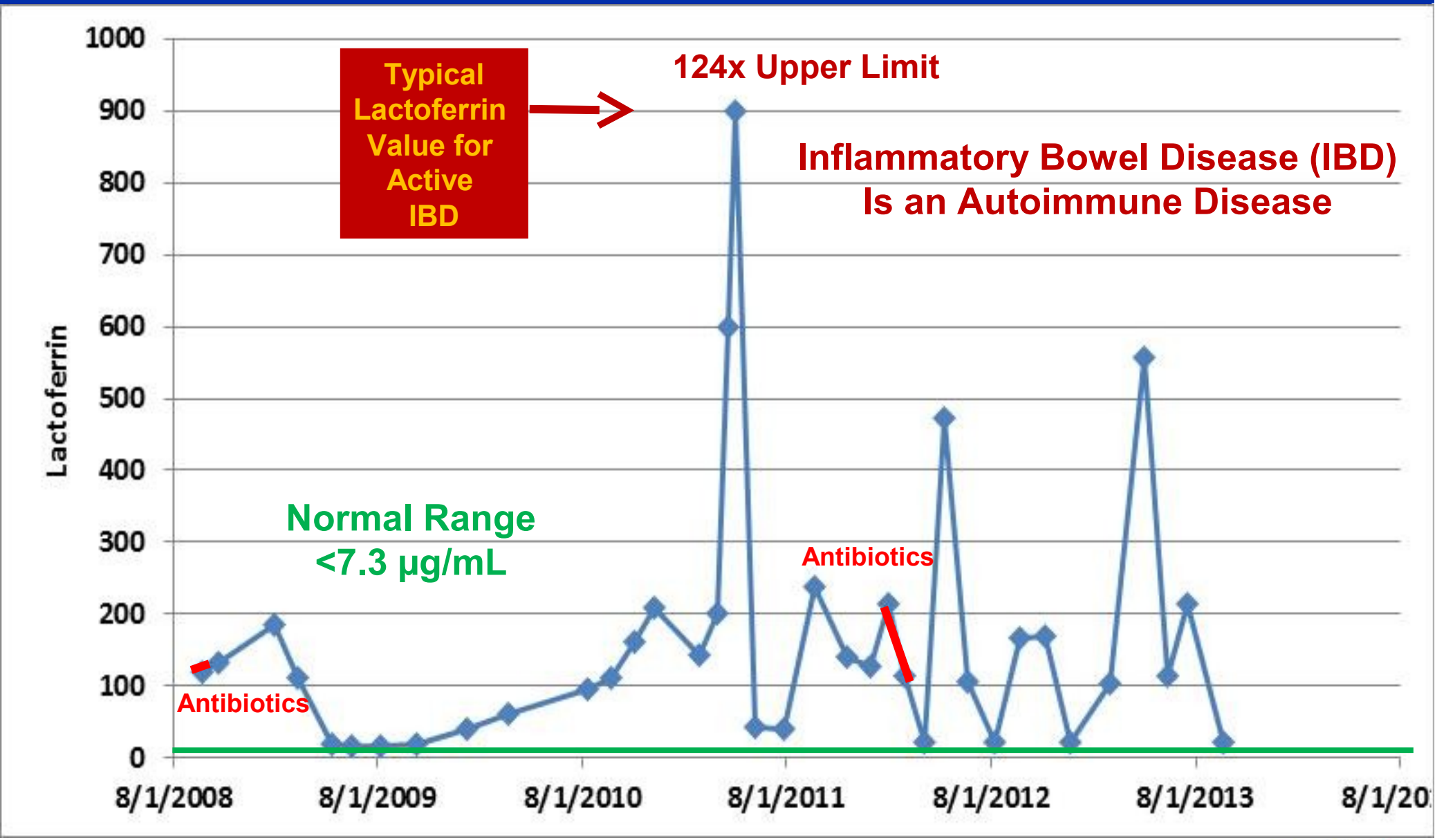


I Discovered I Had Episodic Chronic Inflammation by Tracking Complex Reactive Protein In My Blood Samples



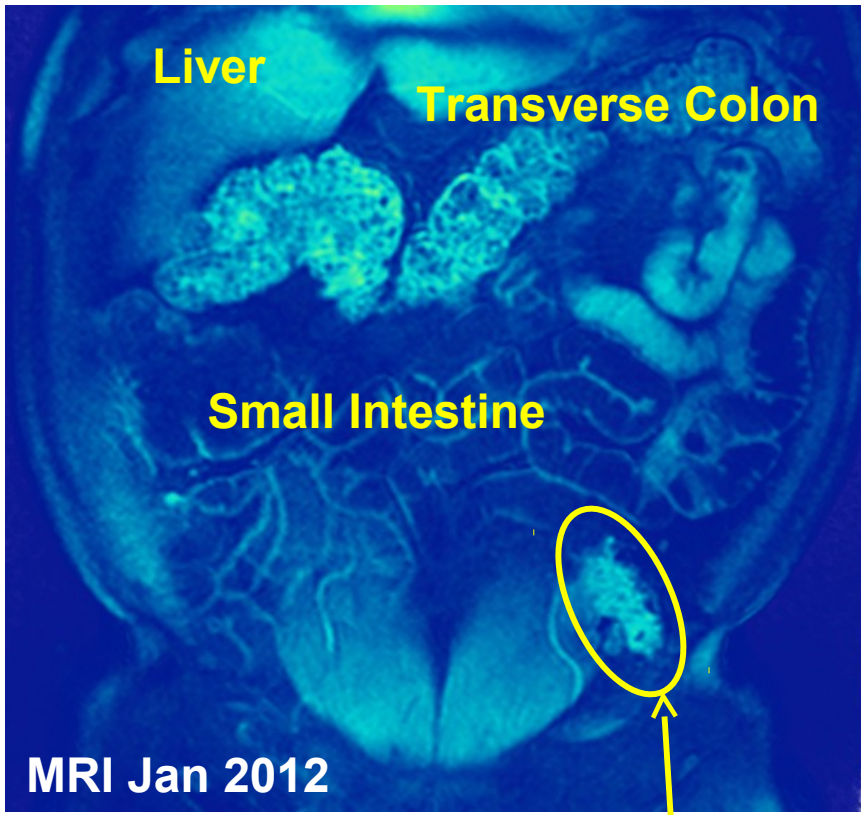
CRP is a Generic Measure of Inflammation in the Blood

My Colon's White Blood Cells Were Shedding Lactoferrin, an Antibacteria Protein Into Stool Samples



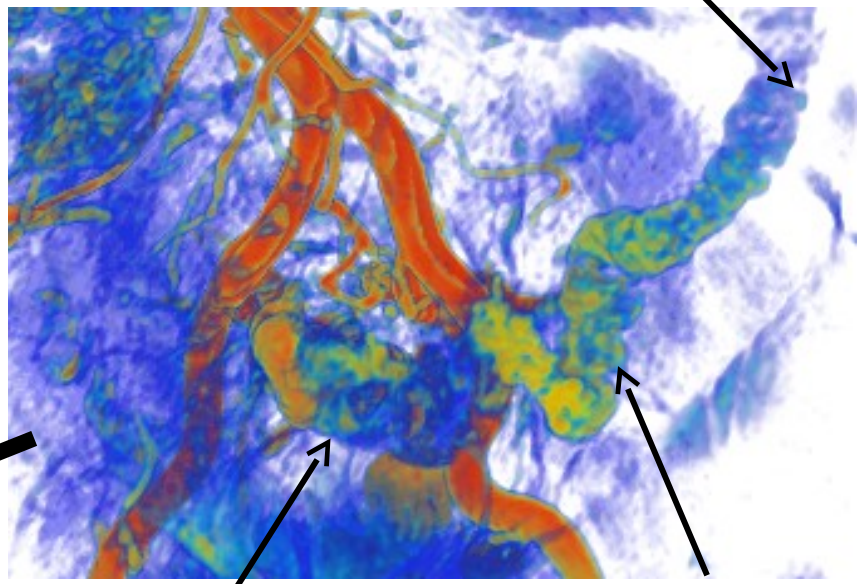
Lactoferrin is a Protein Shed from Neutrophils - An Antibacterial that Sequesters Iron

Confirming the IBD Hypothesis: Finding the "Smoking Gun" with MRI Imaging



I Obtained the MRI Slices From UCSD Medical Services and Converted to Interactive 3D Working With Calit2 Staff & DeskVOX Software

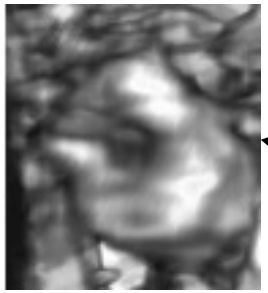
Descending Colon



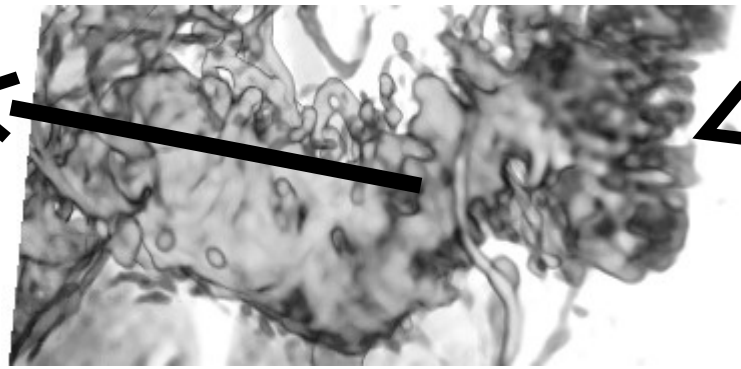
Major Kink

Sigmoid Colon Threading Iliac Arteries

Cross Section



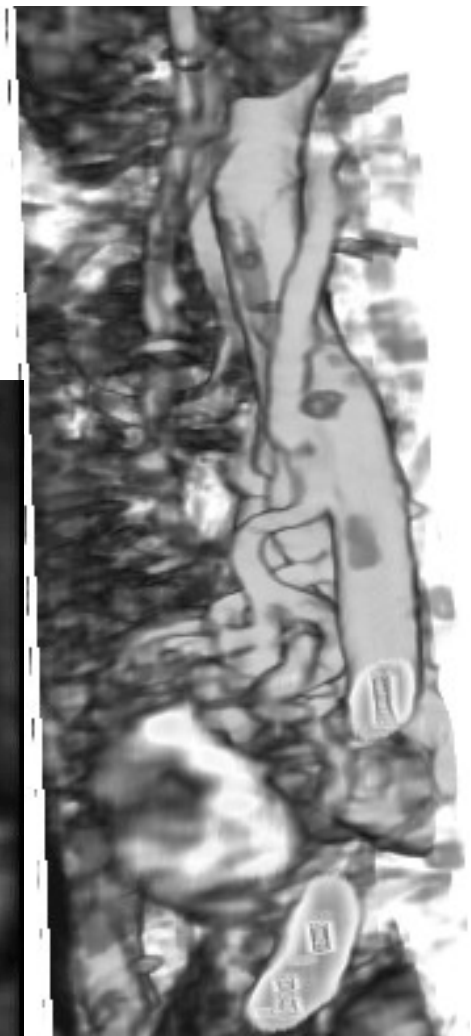
Diseased Sigmoid Colon



MRE Reveals Inflammation in 6 Inches of Sigmoid Colon Thickness 15cm – 5x Normal Thickness

**"Long segment wall thickening
in the proximal and mid portions of the sigmoid colon,
extending over a segment of approximately 16 cm,
with suggestion of intramural sinus tracts.
Edema in the sigmoid mesentery
and engorgement of the regional vasa recta."
– MD Radiologist MRI report**

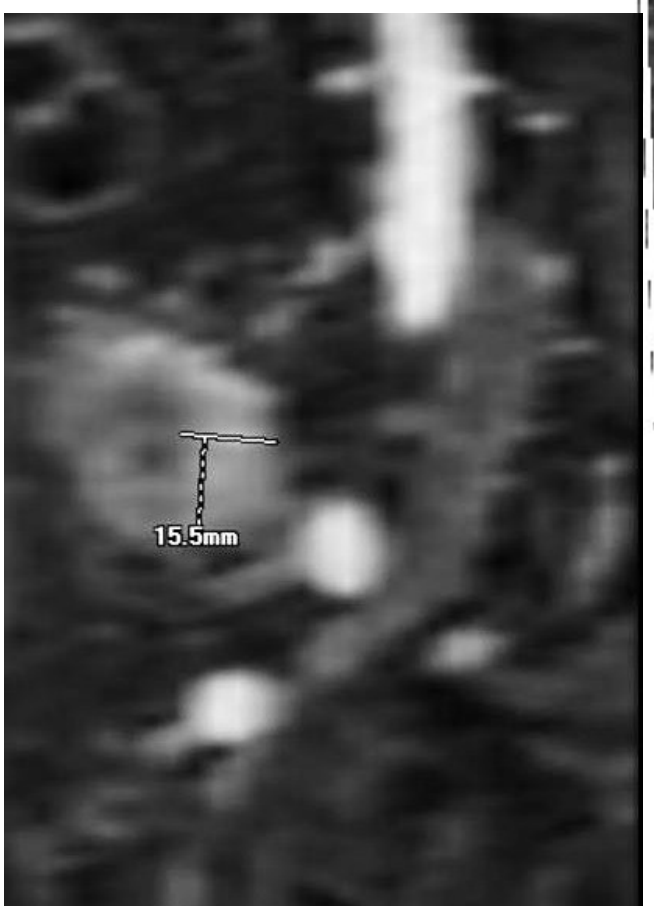
DeskVOX 3D Image



MAYO CLINIC

Crohn's disease affects the thickness of the intestinal wall.

Having Crohn's disease that affects your colon increases your risk of colon cancer.



Clinical MRI Slice Program

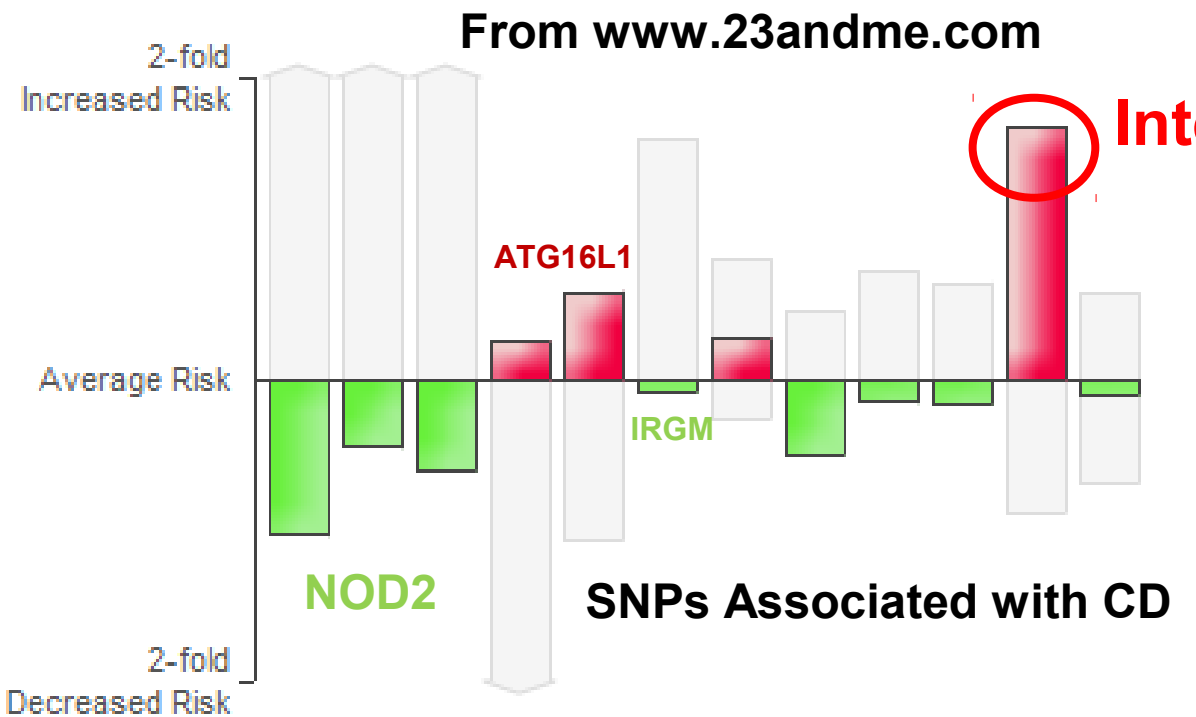
Why Did I Have an Autoimmune Disease like IBD?

Despite decades of research,
the etiology of Crohn's disease
remains unknown.
Its pathogenesis may involve
a complex interplay between
host genetics,
immune dysfunction,
and microbial or environmental factors.
--*The Role of Microbes in Crohn's Disease*

So I Set Out to Quantify All Three!

Paul B. Eckburg & David A. Relman
Clin Infect Dis. 44:256-262 (2007)

I Wondered if Crohn's is an Autoimmune Disease, Did I Have a Personal Genomic Polymorphism?



Polymorphism in Interleukin-23 Receptor Gene — 80% Higher Risk of Pro-inflammatory Immune Response

Now Comparing 163 Known IBD SNPs with 23andme SNP Chip and My Full Human Genome

News and Views

Nature Medicine **13**, 26 - 28 (2007)
doi:10.1038/nm0107-26

IL-23: a master regulator in Crohn disease

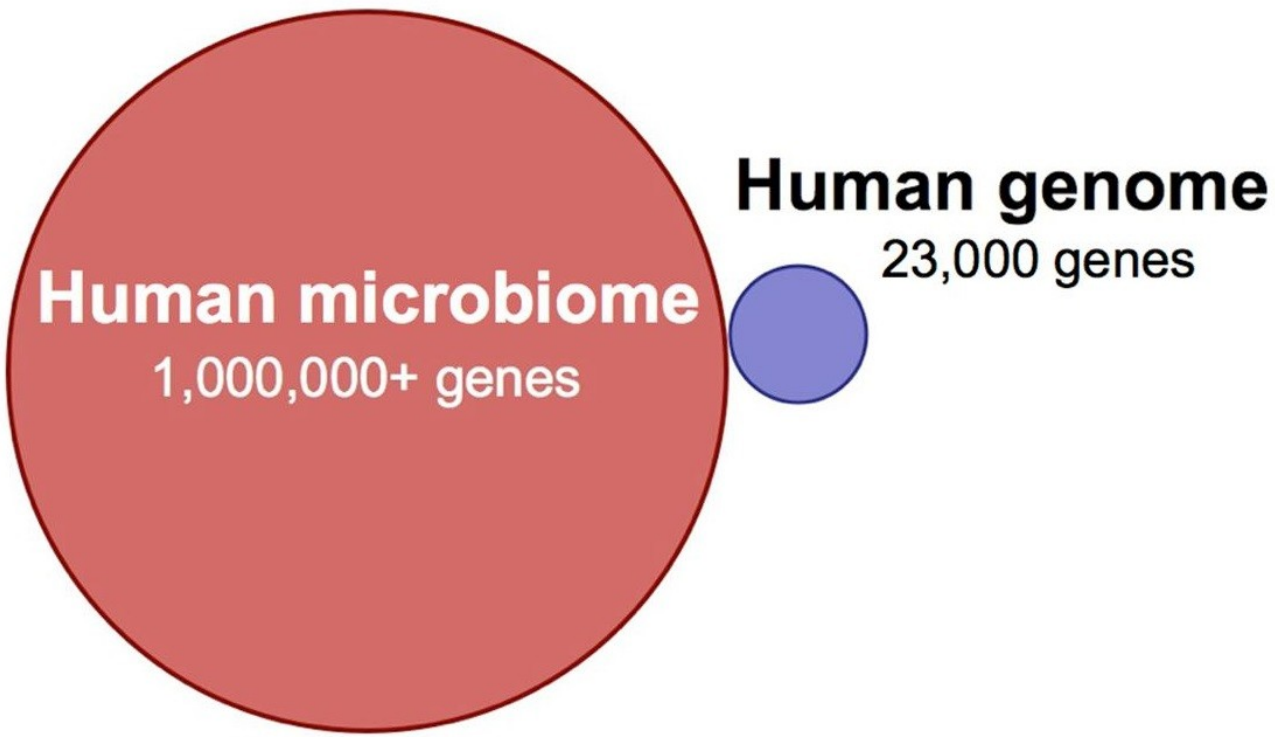
Markus F Neurath[1]

I Had Carried Out Observations in Optical, Radio, and X-Ray on the Andromeda Galaxy in the 1980s



Now I am Observing the 100 Trillion Non-Human Cells in My Body

Your Body Has 10 Times As Many Microbe Cells As Human Cells



99% of Your DNA Genes Are in Microbe Cells Not Human Cells

Inclusion of the Microbiome Will Radically Change Medicine

J. Craig Venter™ INSTITUTE

it²

When We Think About Biological Diversity We Typically Think of the Wide Range of Animals






But All These Animals Are in One SubPhylum Vertebrata of the Chordata Phylum

All images from Wikimedia Commons.
Photos are public domain or by Trisha Shears & Richard Bartz

Think of These Phyla of Animals When You Consider the Biodiversity of Microbes Inside You



Phylum Chordata



Phylum Cnidaria



Phylum Echinodermata



Phylum Annelida



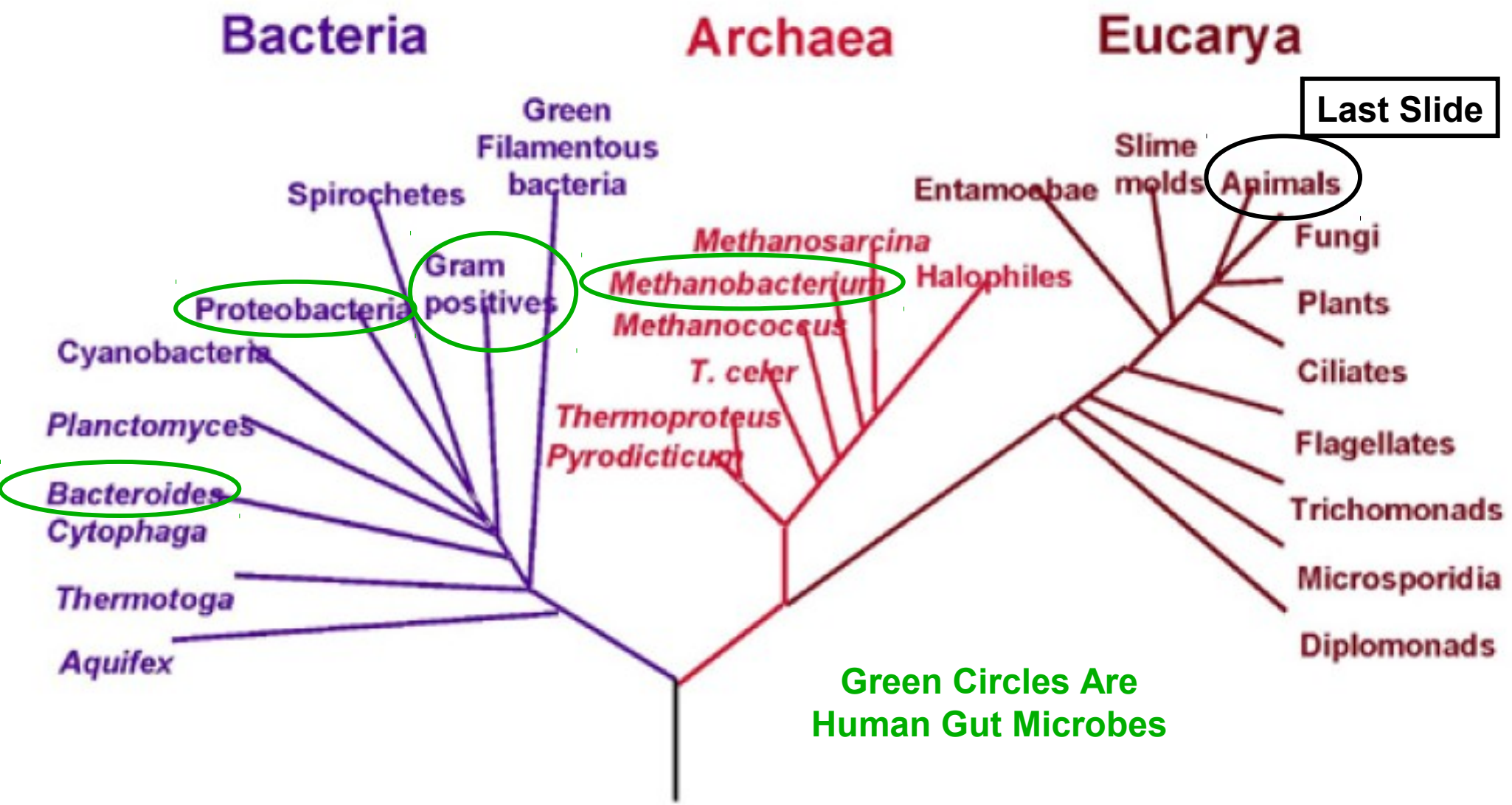
Phylum Mollusca



Phylum Arthropoda

All images from WikiMedia Commons.
Photos are public domain or by Dan Hershman, Michael Linnenbach, Manuae, B_cool

However, The Evolutionary Distance Between Your Gut Microbes Is Much Greater Than Between All Animals



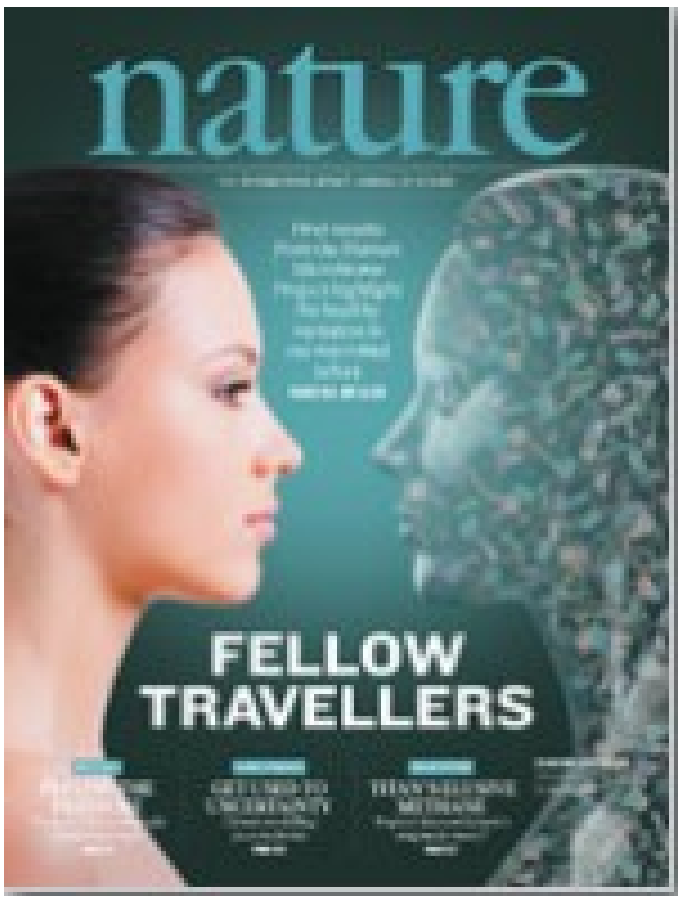
Evolutionary Distance Derived from Comparative Sequencing of 16S or 18S Ribosomal RNA

Source: Carl Woese, et al

Intense Scientific Research is Underway on Understanding the Human Microbiome



June 8, 2012



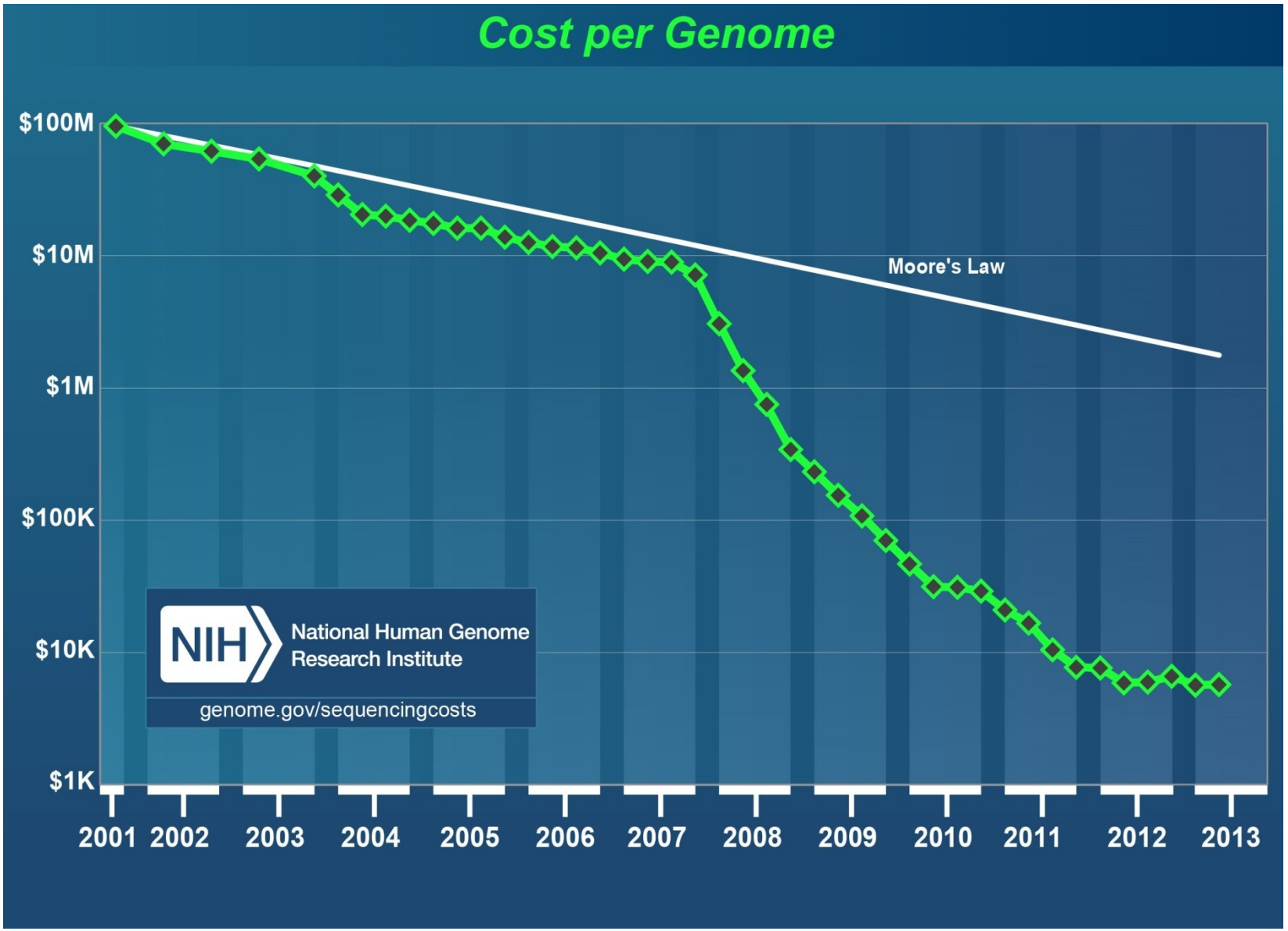
June 14, 2012

From Culturing Bacteria to Sequencing Them

J. Craig Venter™ INSTITUTE

it²

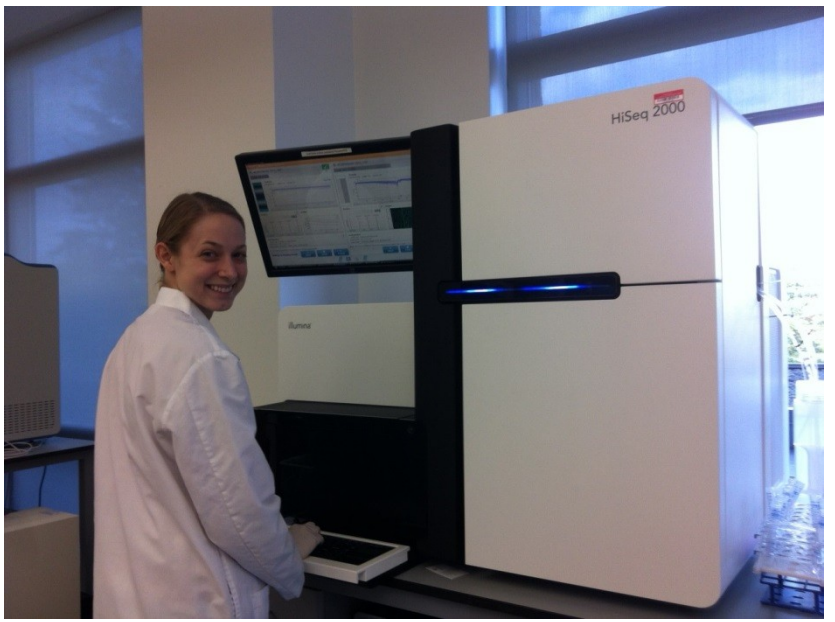
The Cost of Sequencing a Human Genome Has Fallen Over 10,000x in the Last Ten Years!



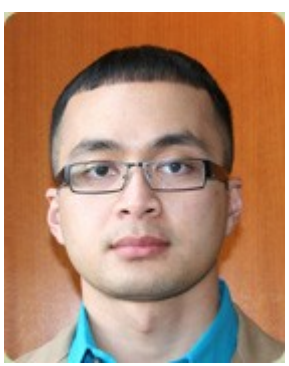
This Has Enabled Sequencing of Both Human and Microbial Genomes

To Map Out the Dynamics of My Microbiome Ecology I Partnered with the J. Craig Venter Institute

- JCVI Did Metagenomic Sequencing on Six of My Stool Samples Over 1.5 Years
- Sequencing on Illumina HiSeq 2000
  - Generates 100bp Reads
  - Run Takes ~14 Days
  - My 6 Samples Produced
    - 190.2 Gbp of Data
- JCVI Lab Manager, Genomic Medicine
  - Manolito Torralba
- IRB PI Karen Nelson
  - President JCVI



Illumina HiSeq 2000 at JCVI


Manolito Torralba, JCVI


Karen Nelson, JCVI

J. Craig Venter INSTITUTE

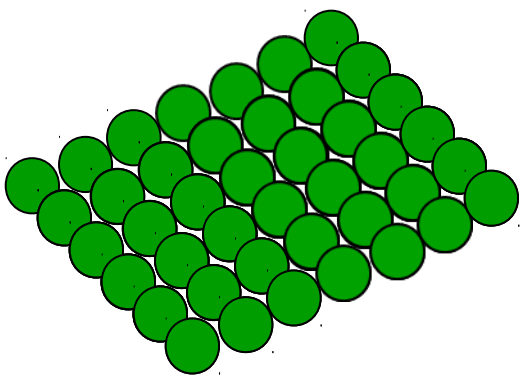
it2

We Downloaded Additional Phenotypes from NIH HMP For Comparative Analysis

Download Raw Reads
~100M Per Person

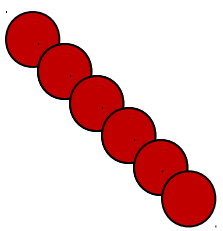
"Healthy" Individuals

35 Subjects
1 Point in Time



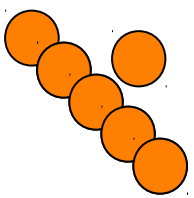
Larry Smarr

6 Points in Time

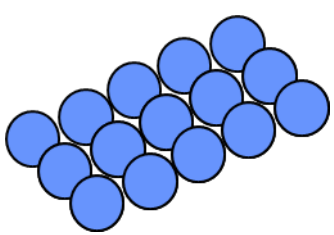


IBD Patients

2 Ulcerative Colitis Patients,
6 Points in Time



5 Ileal Crohn's Patients,
3 Points in Time



NIH

HMP

Total of 5 Billion Reads

Source: Jerry Sheehan, Calit2
Weizhong Li, Sitao Wu, CRBS, UCSD

it²

We Created a Reference Database Of Known Gut Genomes

- **NCBI April 2013**
  - **2471 Complete + 5543 Draft Bacteria & Archaea Genomes**
  - **2399 Complete Virus Genomes**
  - **26 Complete Fungi Genomes**
  - **309 HMP Eukaryote Reference Genomes**
- **Total 10,741 genomes, ~30 GB of sequences**

NCBI National Center for Biotechnology Information

HMP

Now to Align Our 5 Billion Reads Against the Reference Database

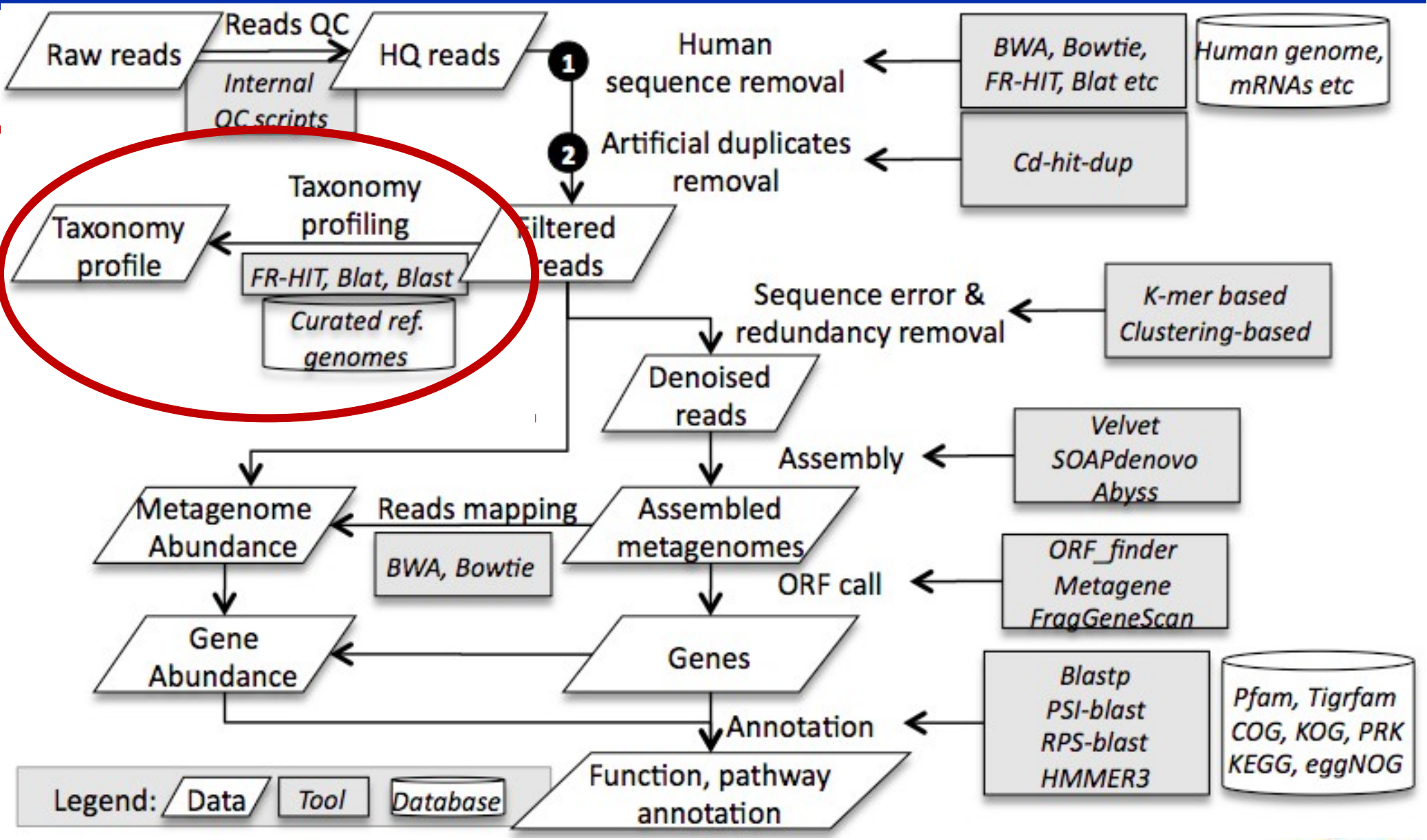
NIH

J. Craig Venter INSTITUTE

Source: Weizhong Li, Sitao Wu, CRBS, UCSD

it²

Computational NextGen Sequencing Pipeline: From "Big Equations" to "Big Data" Computing



National Human Genome Research Institute

HMP

PI: (Weizhong Li, CRBS, UCSD):
NIH R01HG005978 (2010-2013, $1.1M)

CRBS

it²

Computing and Parallelization Requirements of the Computational Tools in Our Workflow

Analysis	Tool	Computing requirement			Parallelization	
		Data	CPU	RAM	Multi-threading	Map Reduce
Quality control	QC script	√				
Human DNA removal	Bowtie	√			√	
Duplicates removal	CD-HIT-DUP	√		√	√	
Mapping	FR-HIT	√	√		√	
Read clustering	CD-HIT-EST	√		√	√	
Assembly	Velvet, SOAPdenovo	√		√	√	
ORF call	Metagene, Fraggenescan	√				
ORF clustering	CD-HIT	√		√	√	
COG, KOG annotation	RPS-BLAST	√	√		√	√
KEGG annotation	BLAST	√	√		√	√
Pfam, Tigrfam annotation	HMMER3	√	√		√	√

Source: Weizhong Li, CRBS, UCSD

We Used SDSC's Gordon Data-Intensive Supercomputer to Analyze a Wide Range of Gut Microbiomes

- ~180,000 Core-Hrs on Gordon
  - KEGG function annotation: 90,000 hrs
  - Mapping: 36,000 hrs
    - Used 16 Cores/Node and up to 50 nodes
  - Duplicates removal: 18,000 hrs
  - Assembly: 18,000 hrs
  - Other: 18,000 hrs
- Gordon RAM Required
  - 64GB RAM for Reference DB
  - 192GB RAM for Assembly
- Gordon Disk Required
  - Ultra-Fast Disk Holds Ref DB for All Nodes
  - 8TB for All Subjects



SDSC

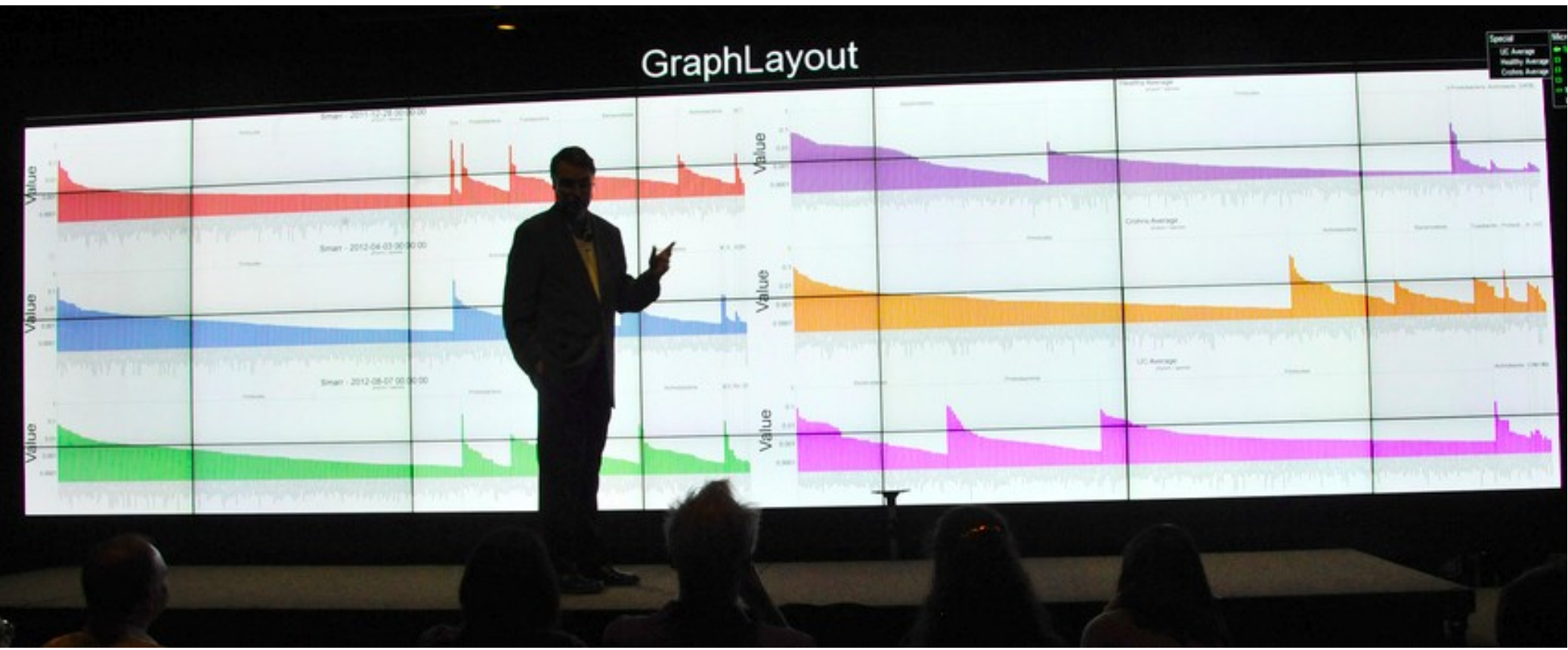
Enabled by
a Grant of Time
on Gordon from SDSC
Director Mike Norman

NSF NIH HMP

J. Craig Venter INSTITUTE

CRBS it²

Using Scalable Visualization Allows Comparison of the Relative Abundance of 200 Microbe Species

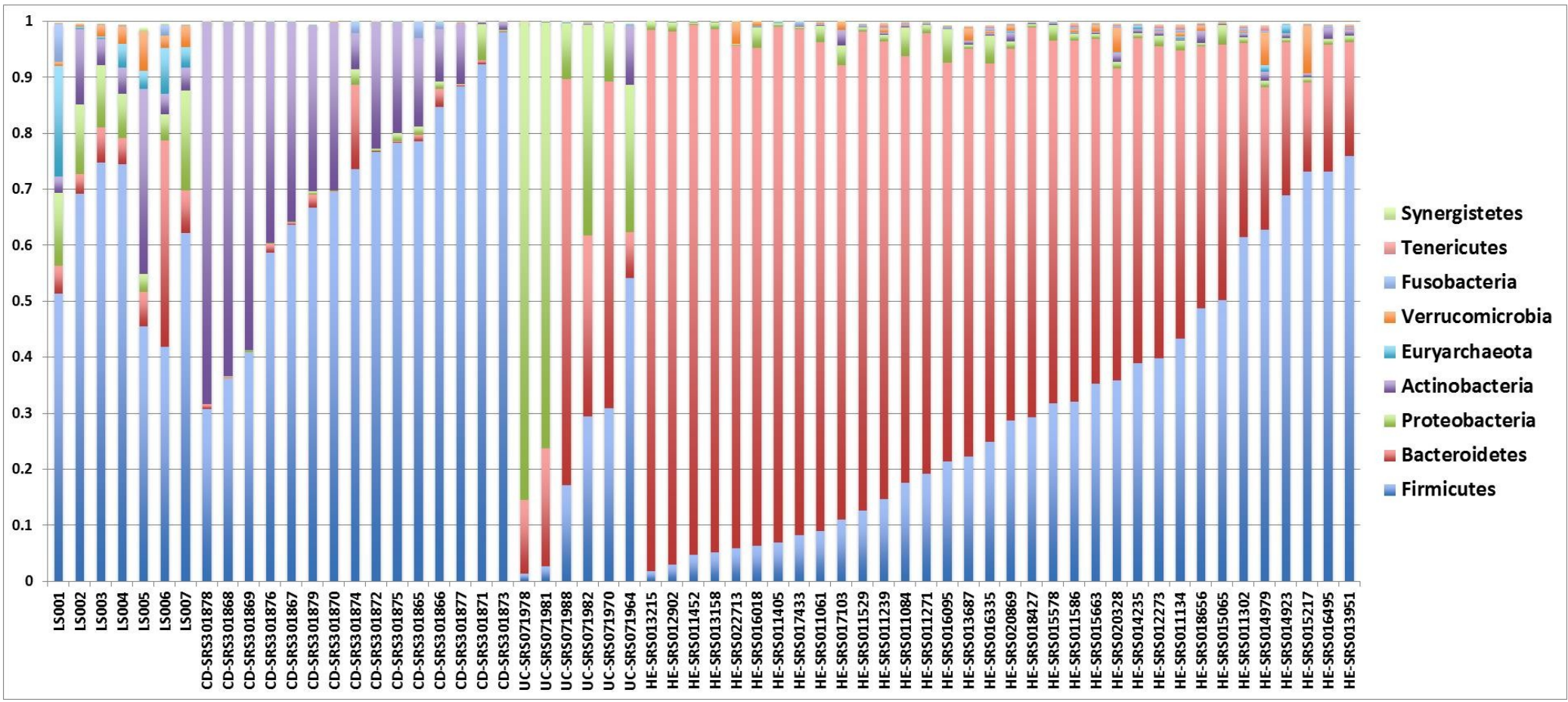


Comparing 3 LS Time Snapshots (Left)
with Healthy, Crohn's, UC (Right Top to Bottom)

Calit2 VROOM-FuturePatient Expedition

Phyla Gut Microbial Abundance Without Viruses: LS, Crohn's, UC, and Healthy Subjects

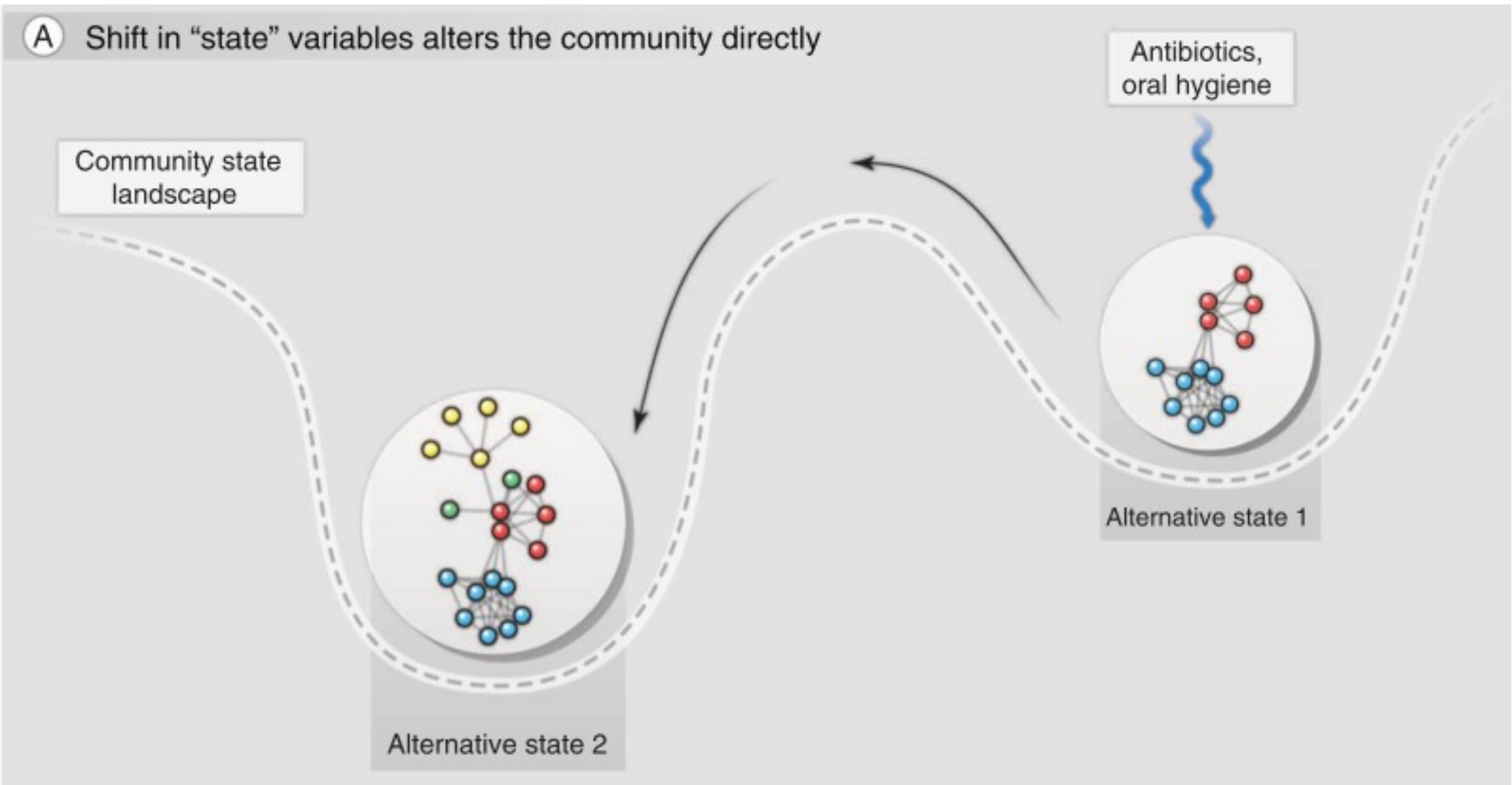
Source: Weizhong Li, Sitao Wu, CRBS, UCSD



LS **Crohn's** **Ulcerative Colitis** **Healthy**

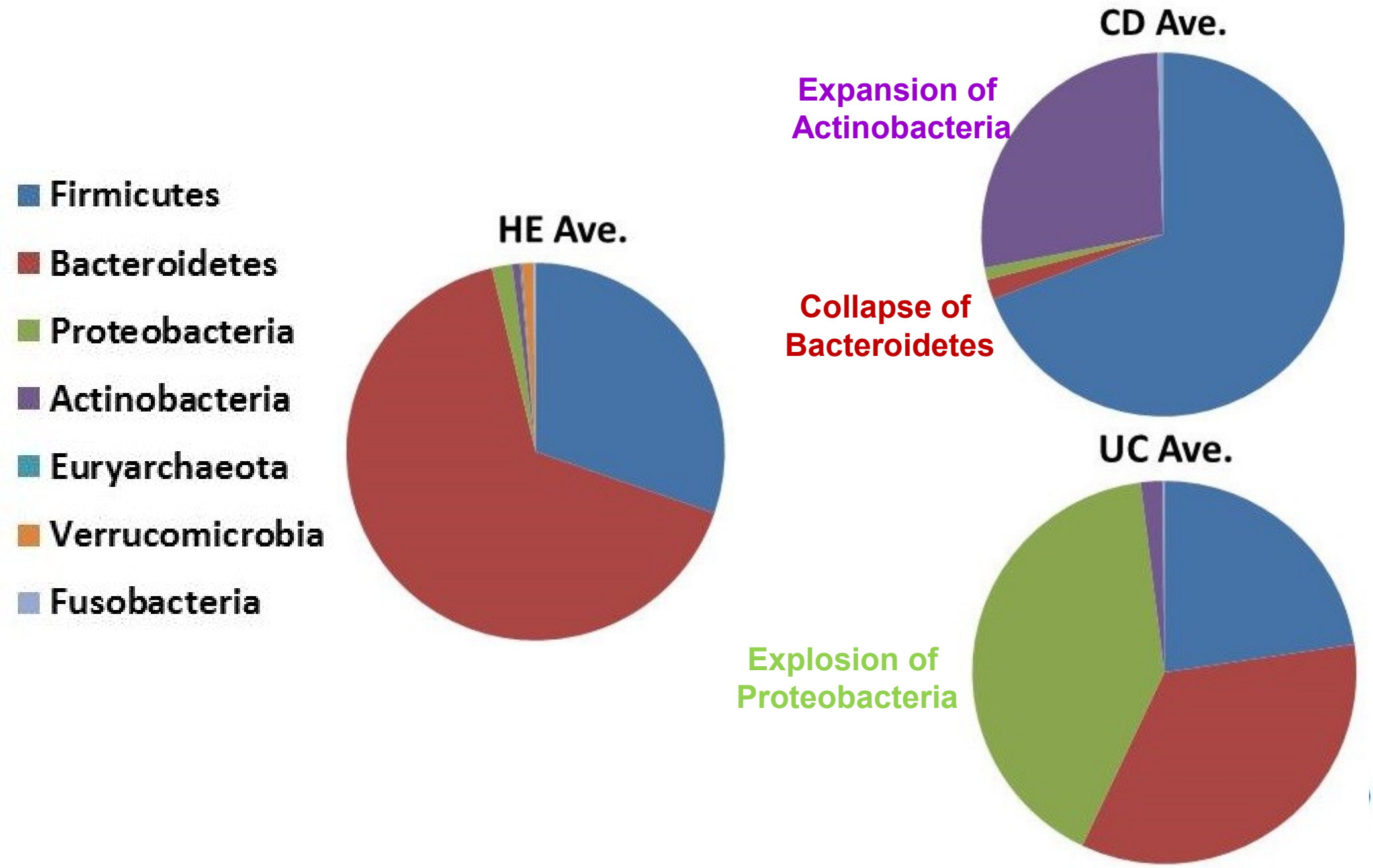
Toward Noninvasive Microbial Ecology Diagnostics

Lessons from Ecological Dynamics I: Gut Microbiome Has Multiple Relatively Stable Equilibria



"The Application of Ecological Theory Toward an Understanding of the Human Microbiome,"
Elizabeth Costello, Keaton Stagaman, Les Dethlefsen, Brendan Bohannan, David Relman
Science 336, 1255-62 (2012)

Comparison of 35 Healthy to 15 CD and 6 UC Gut Microbiomes at the Phyla Level



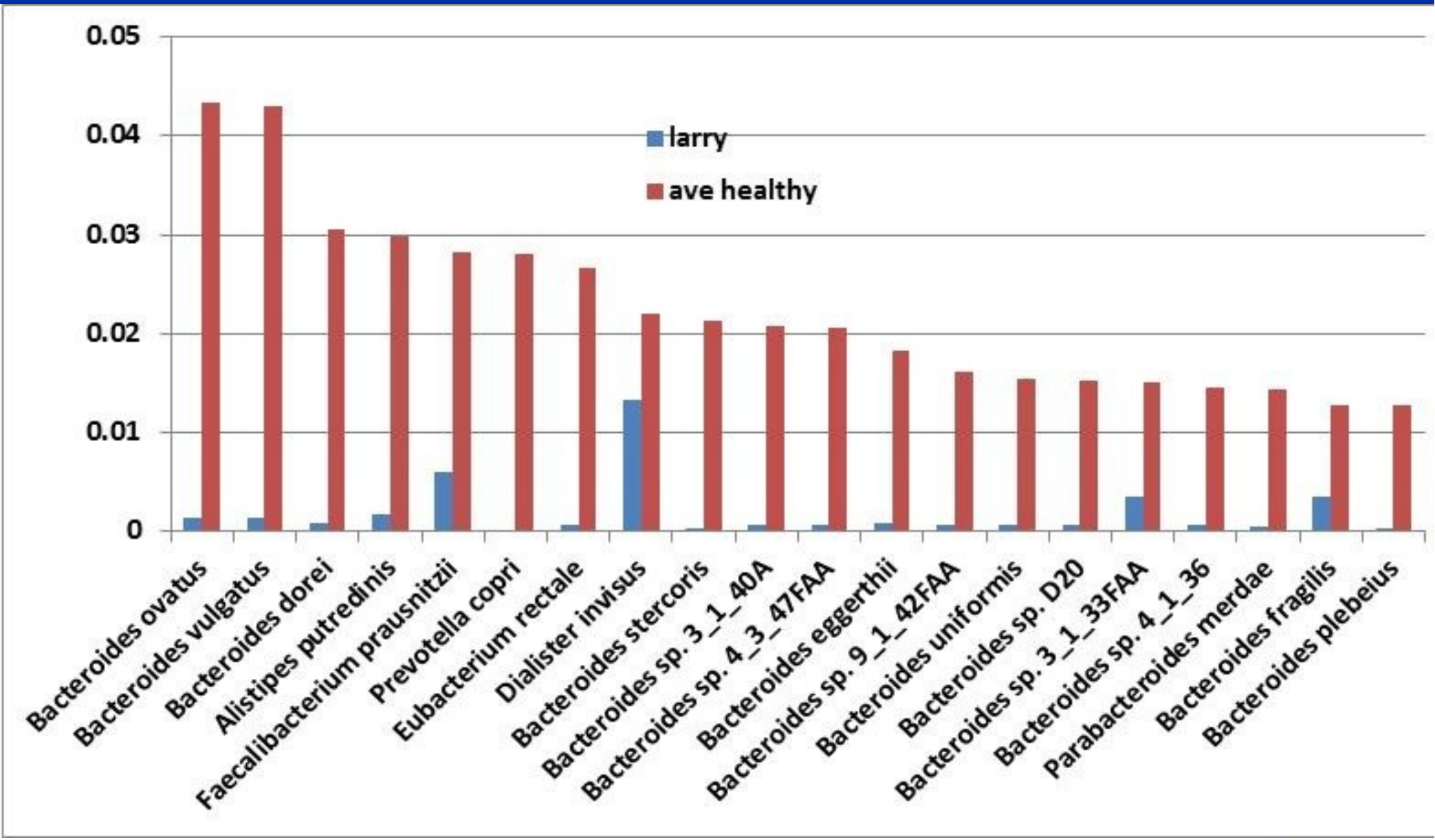
Lessons From Ecological Dynamics II:
Invasive Species Dominate After Major Species Destroyed



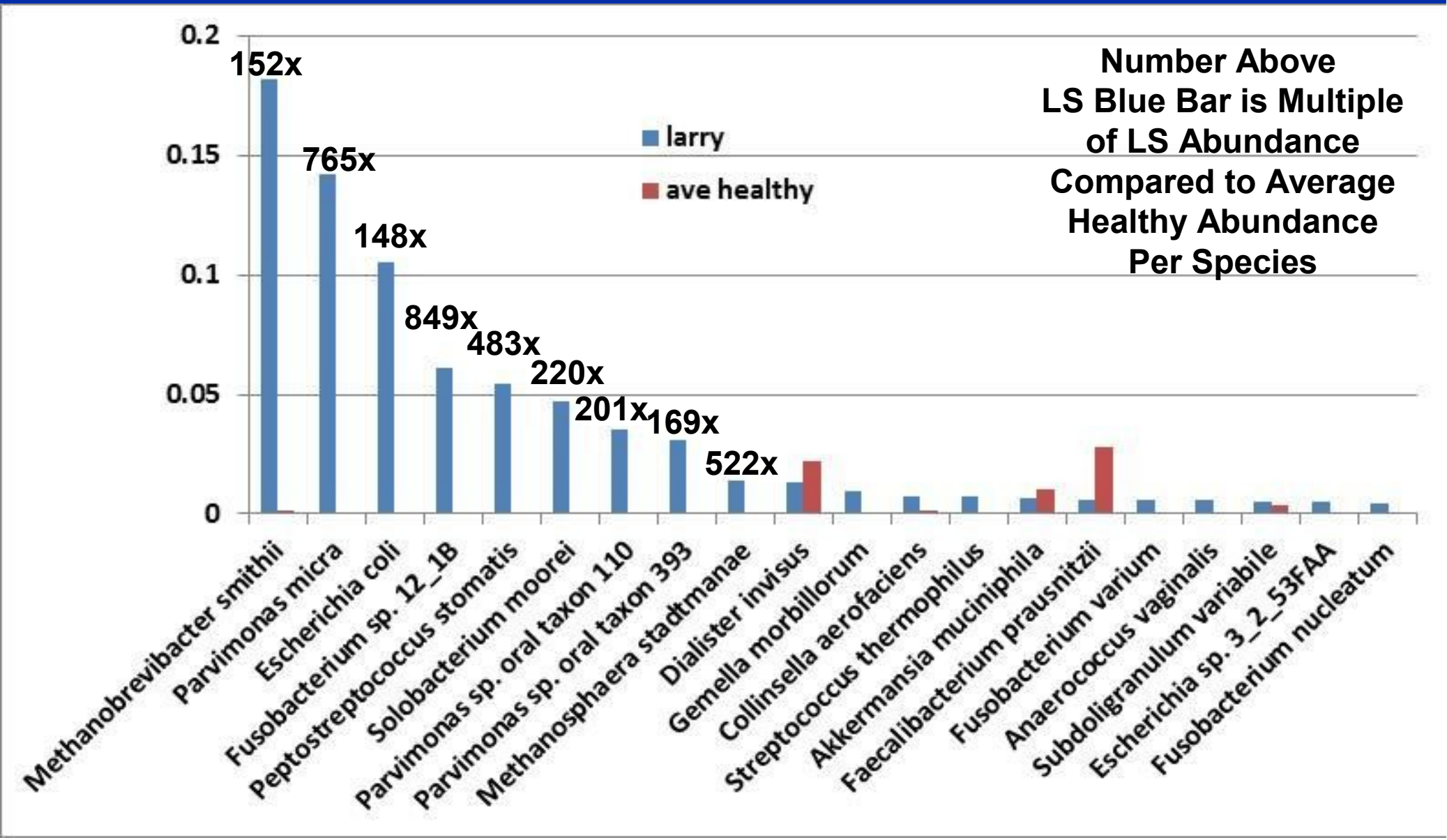
”In many areas following these burns invasive species are able to establish themselves, crowding out native species.”

Source: Ponderosa Pine Fire Ecology
http://cpluhna.nau.edu/Biota/ponderosafire.htm

Almost All Abundant Species (≥1%) in Healthy Subjects Are *Severely Depleted* in Larry's Gut Microbiome



Top 20 Most Abundant Microbial Species In LS vs. Average Healthy Subject



Source: Sequencing JCVI; Analysis Weizhong Li, UCSD
LS December 28, 2011 Stool Sample

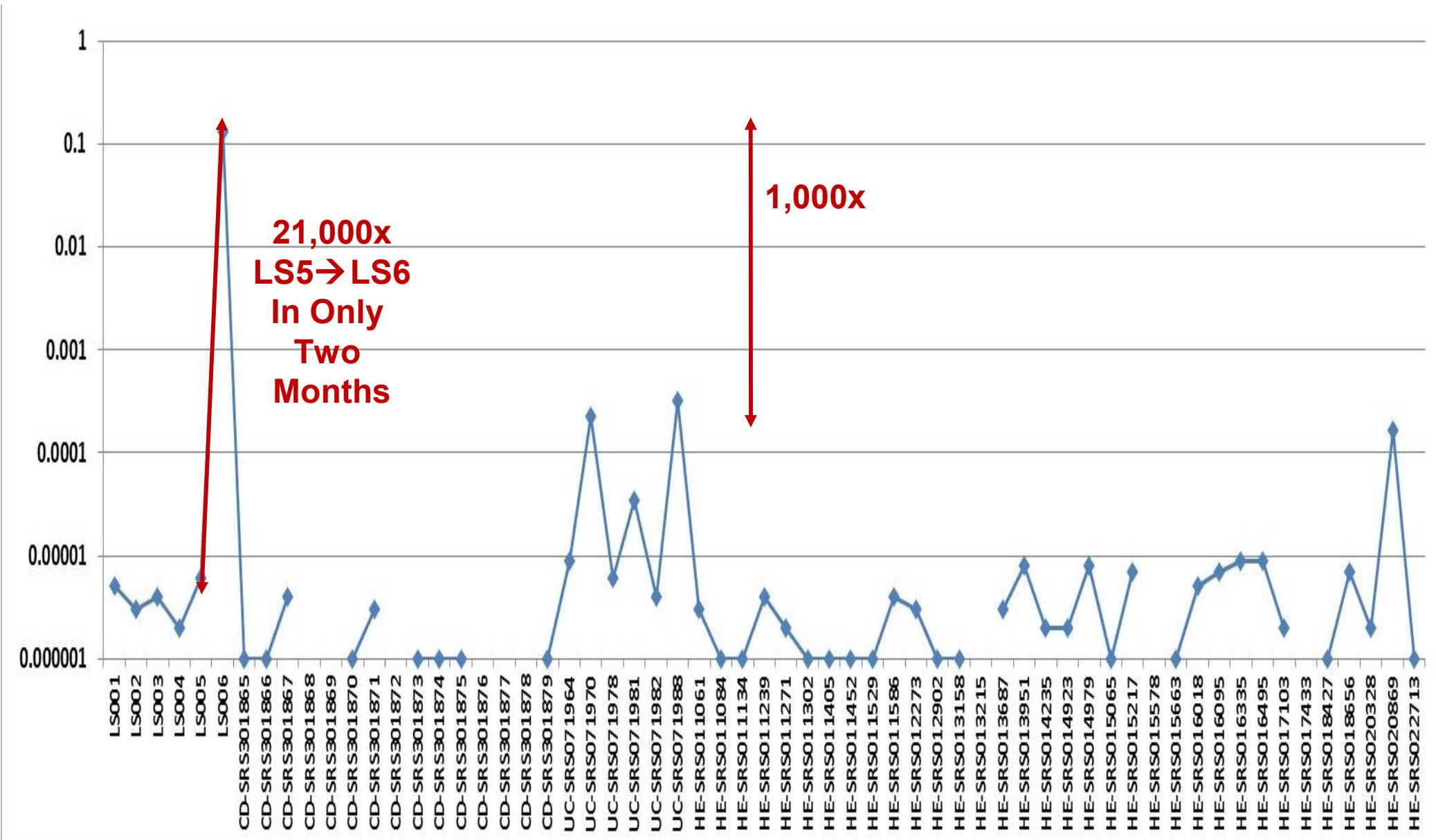
Lessons From Ecological Dynamics III: From Equilibrium to Chaos

In addition to chaos,
other forms of complex dynamics,
such as regular oscillations & quasiperiodic oscillations,
are preeminent features of many biological systems.

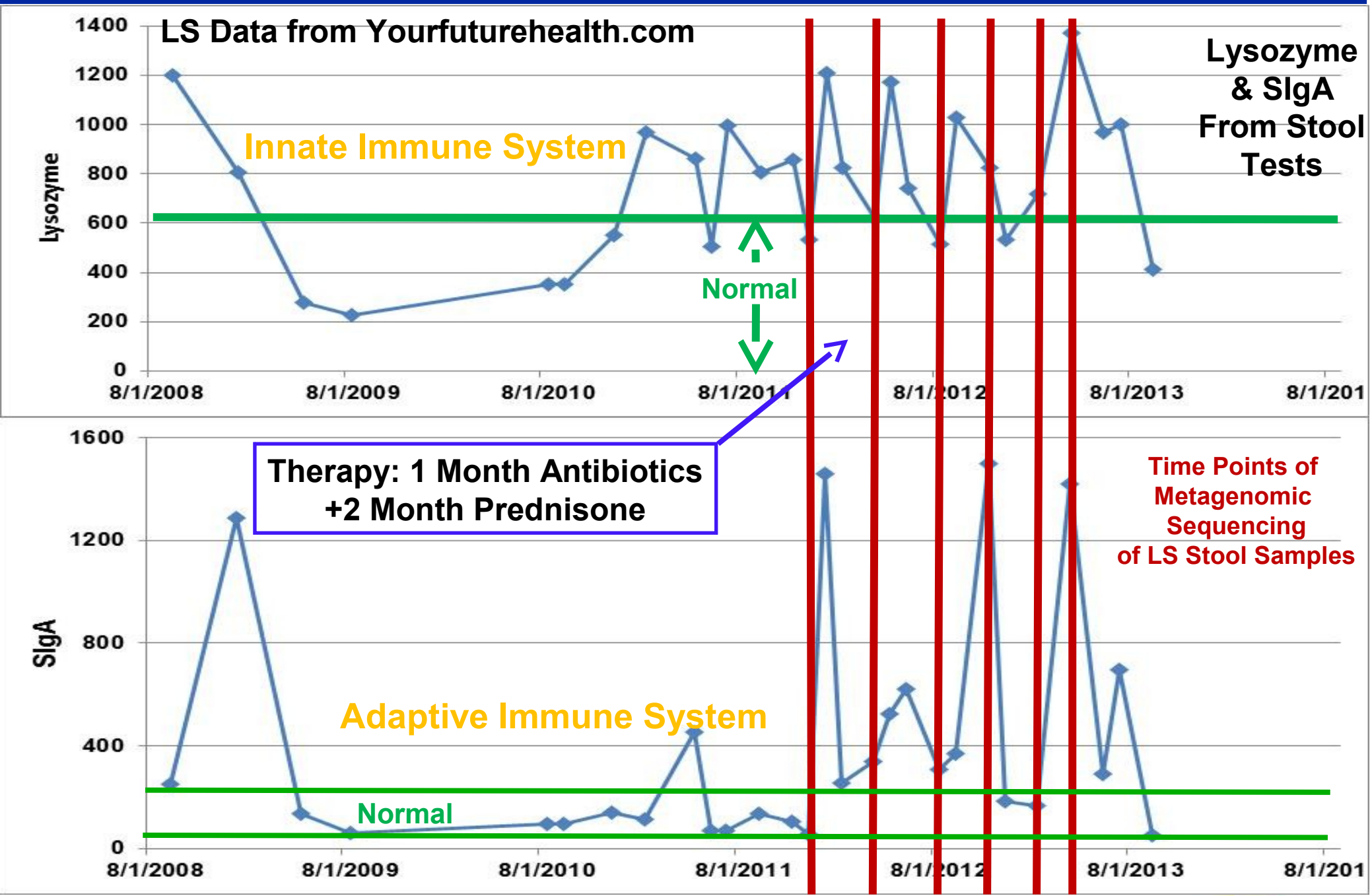
- From "Biological Chaos and Complex Dynamics"
David A. Vasseur
Oxford Bibliographies Online

Chaos: Large Fast Changes From Small Initial Conditions: Dramatic Bloom of Enterobacteriaceae bacterium 9_2_54FAA

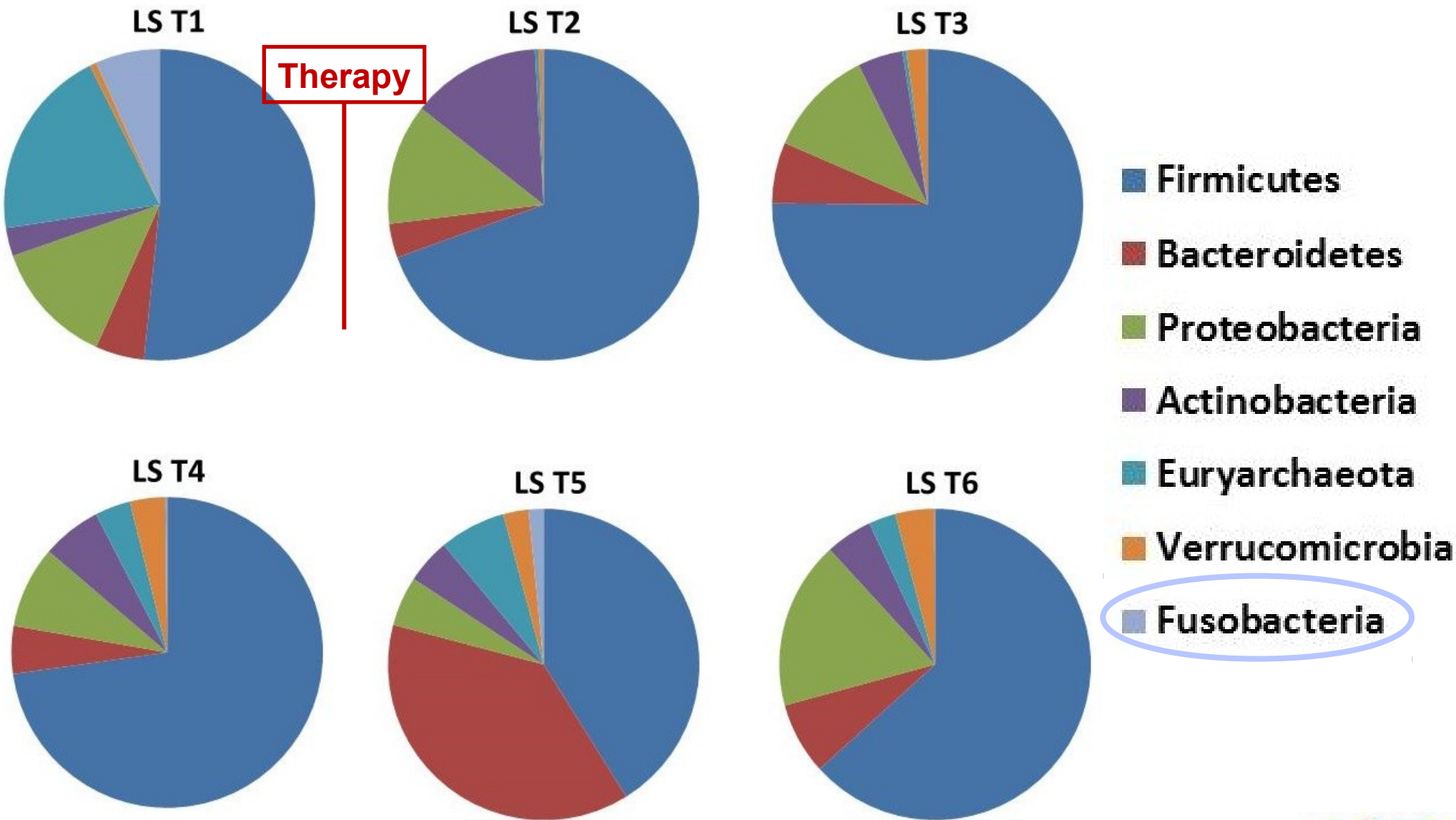
This Microbe is a Proteobacteria Targeted by the NIH HMP



Fine Time Resolution Sampling Revealed Regular Oscillations of the Innate and Adaptive Immune System



Time Series Reveals Autoimmune Dynamics of Gut Microbiome by Phyla



Six Metagenomic Time Samples Over 16 Months

Fusobacteria Are Found To Be More Abundant In Colonrectal Carcinoma (CRC) Tissue

Two Cancer Studies Find Bacterial Clue in Colon

The New York Times

By GINA KOLATA
Published: October 17, 2011

Genomic analysis identifies association of *Fusobacterium* with colorectal carcinoma Aleksandar D. Kostic, **et al.**

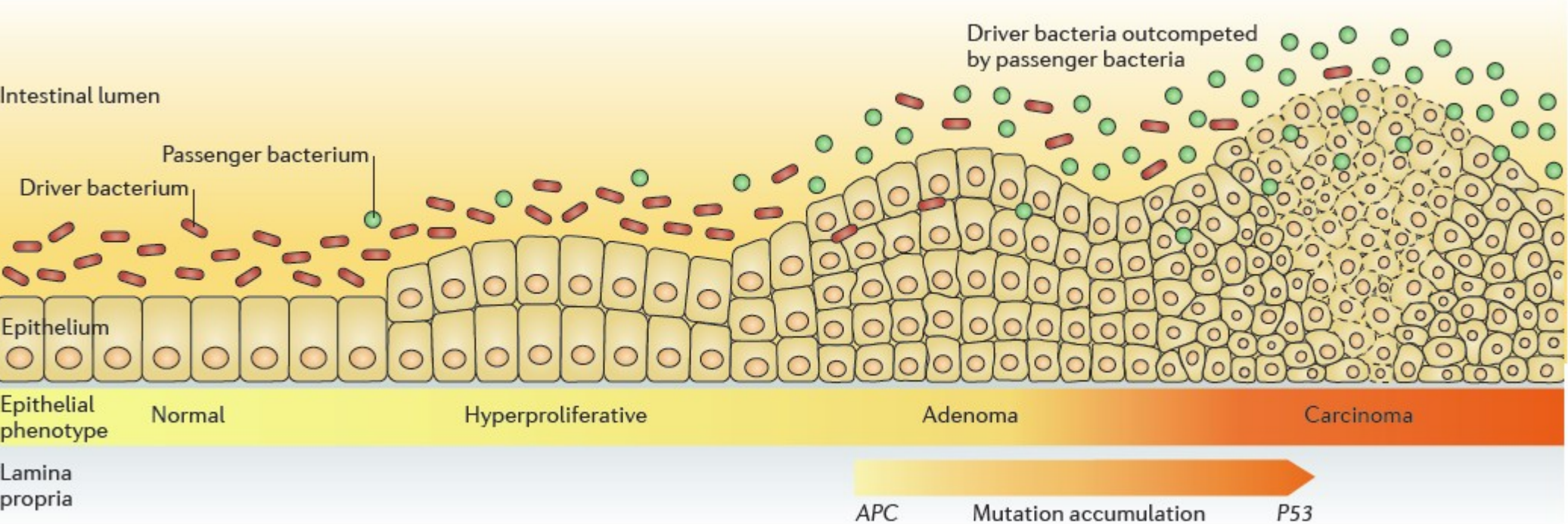
Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma Mauro Castellarin, **et al.**

CSH PRESS

GENOME RESEARCH

The Bacterial Driver-Passenger Model for Colorectal Cancer Initiation

Is Fusobacterium nucleatum a "Driver" or a "Passenger"

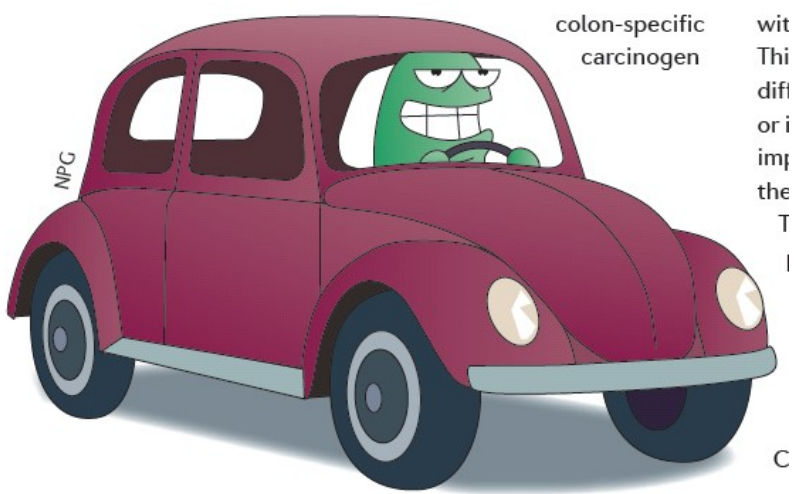


"Early detection of Colorectal Cancer (CRC) is one of the greatest challenges in the battle against this disease & the establishment of a CRC-associated microbiome risk profile could aid in the early identification of individuals who are at high risk and require strict surveillance."

Tjalsma, et al. Nature Reviews Microbiology v. 10, 575-582 (2012)

E. coli claims the driving seat for cancer

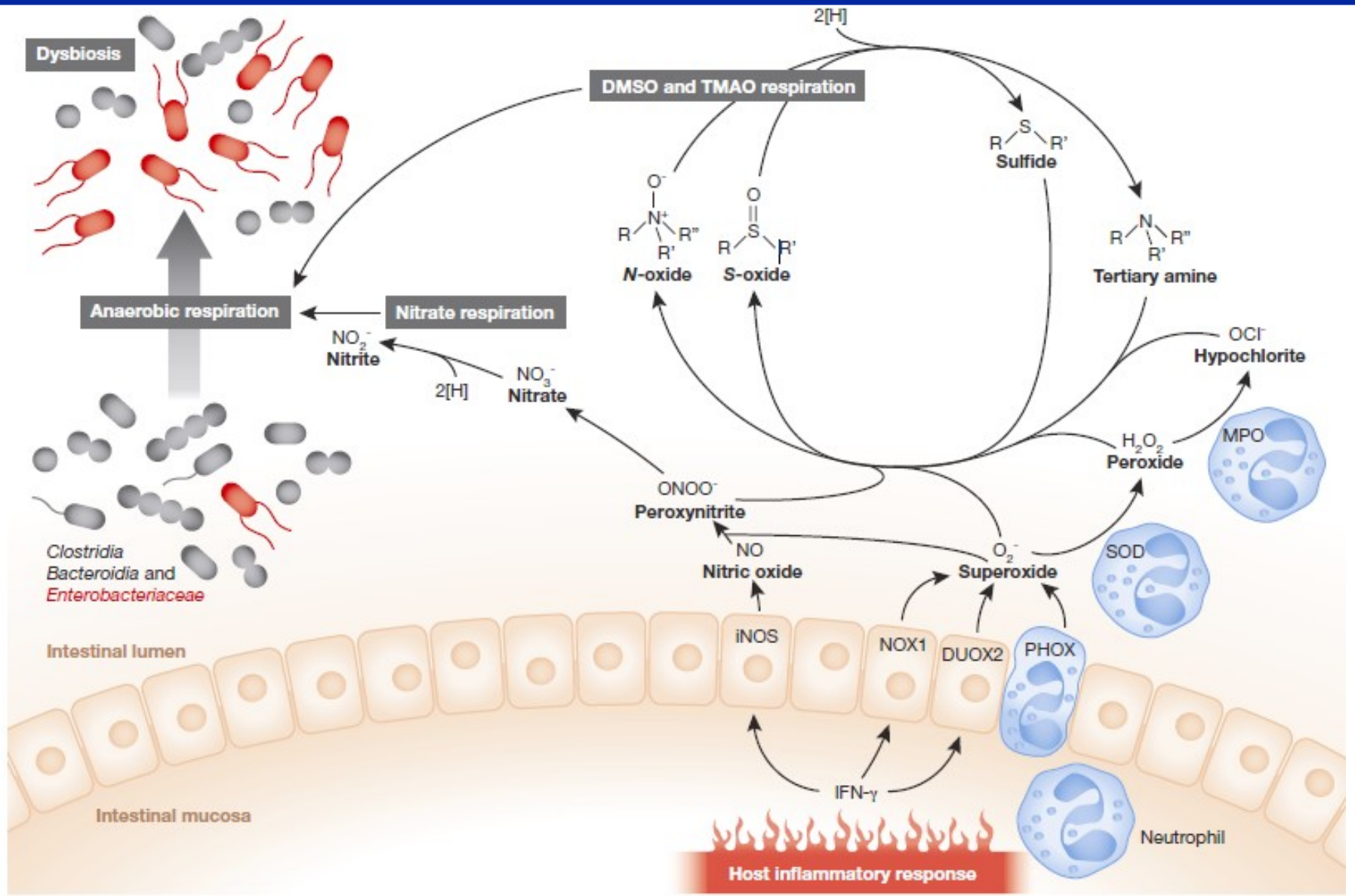
"Arthur *et al.* provide evidence that inflammation alters the intestinal microbiota by favouring the proliferation of genotoxic commensals, and that the *Escherichia coli* genotoxin colibactin promotes colorectal cancer (CRC)."



Christina Tobin Kåhrström
Associate Editor,
Nature Reviews Microbiology

VOLUME 12 | OCTOBER 2012
NATURE REVIEWS | CANCER

Inflammation Enables Anaerobic Respiration Which Leads to Phylum-Level Shifts in the Gut Microbiome



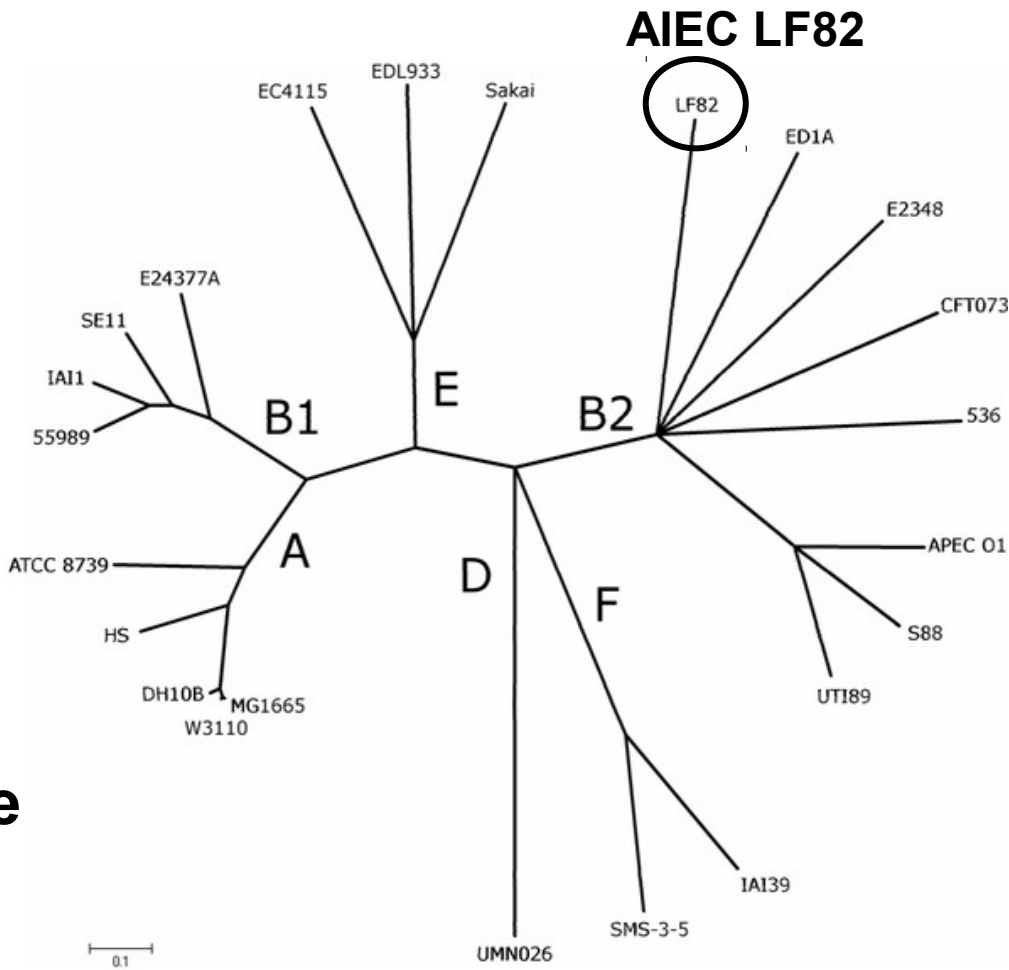
Sebastian E. Winter, Christopher A. Lopez & Andreas J. Bäumler, EMBO reports VOL 14, p. 319-327 (2013)

Does Intestinal Inflammation Select for Pathogenic Strains That Can Induce Further Damage?

"Adherent-invasive E. coli (AIEC) are isolated more commonly from the intestinal mucosa of individuals with Crohn's disease than from healthy controls."

"Thus, the mechanisms leading to dysbiosis might also select for intestinal colonization with more harmful members of the Enterobacteriaceae* —such as AIEC— thereby exacerbating inflammation and interfering with its resolution."

Sebastian E. Winter , et al.,
EMBO reports VOL 14, p. 319-327 (2013)

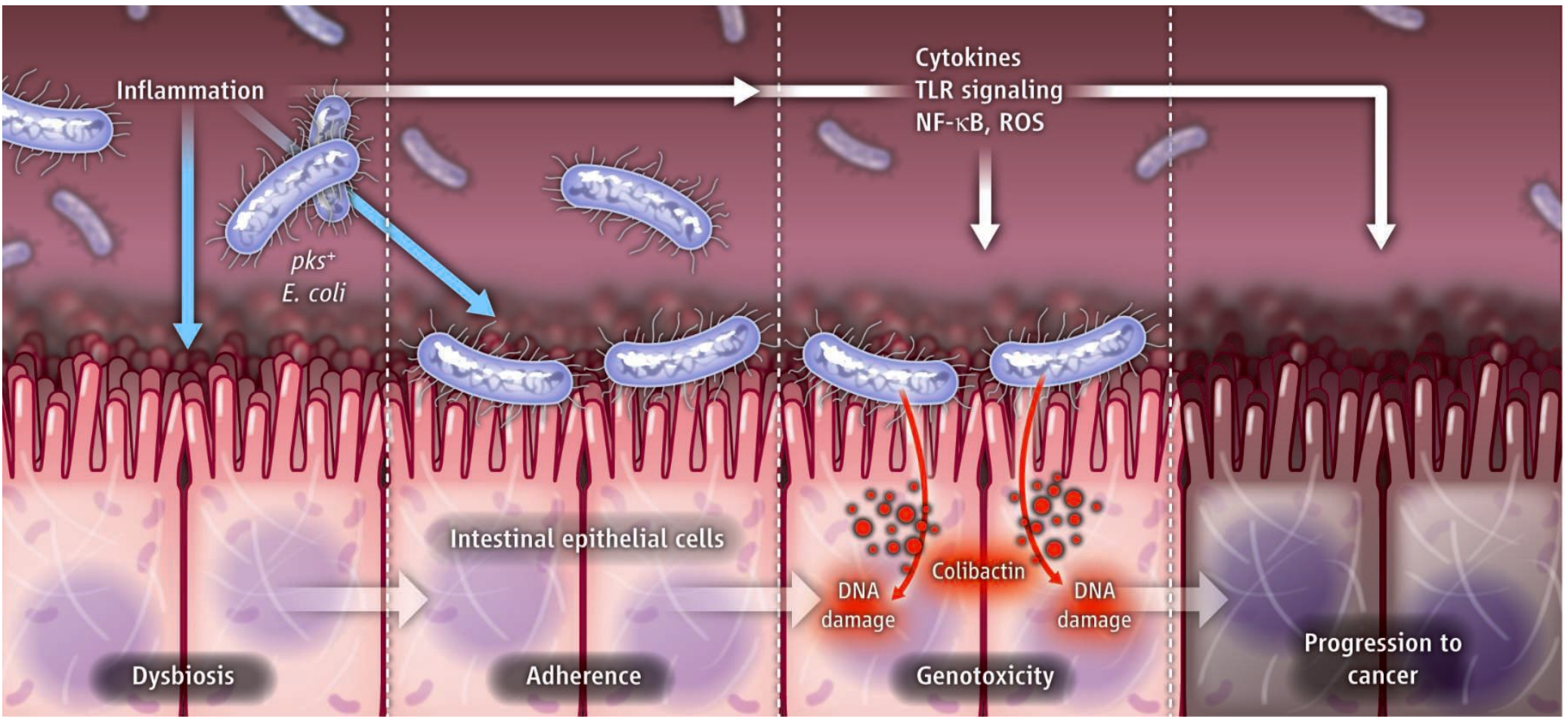


E. coli/Shigella Phylogenetic Tree
Miquel, et al.
PLOS ONE, v. 5, p. 1-16 (2010)

*Family Containing E. coli

Chronic Inflammation Can Accumulate Cancer-Causing Bacteria in the Human Gut

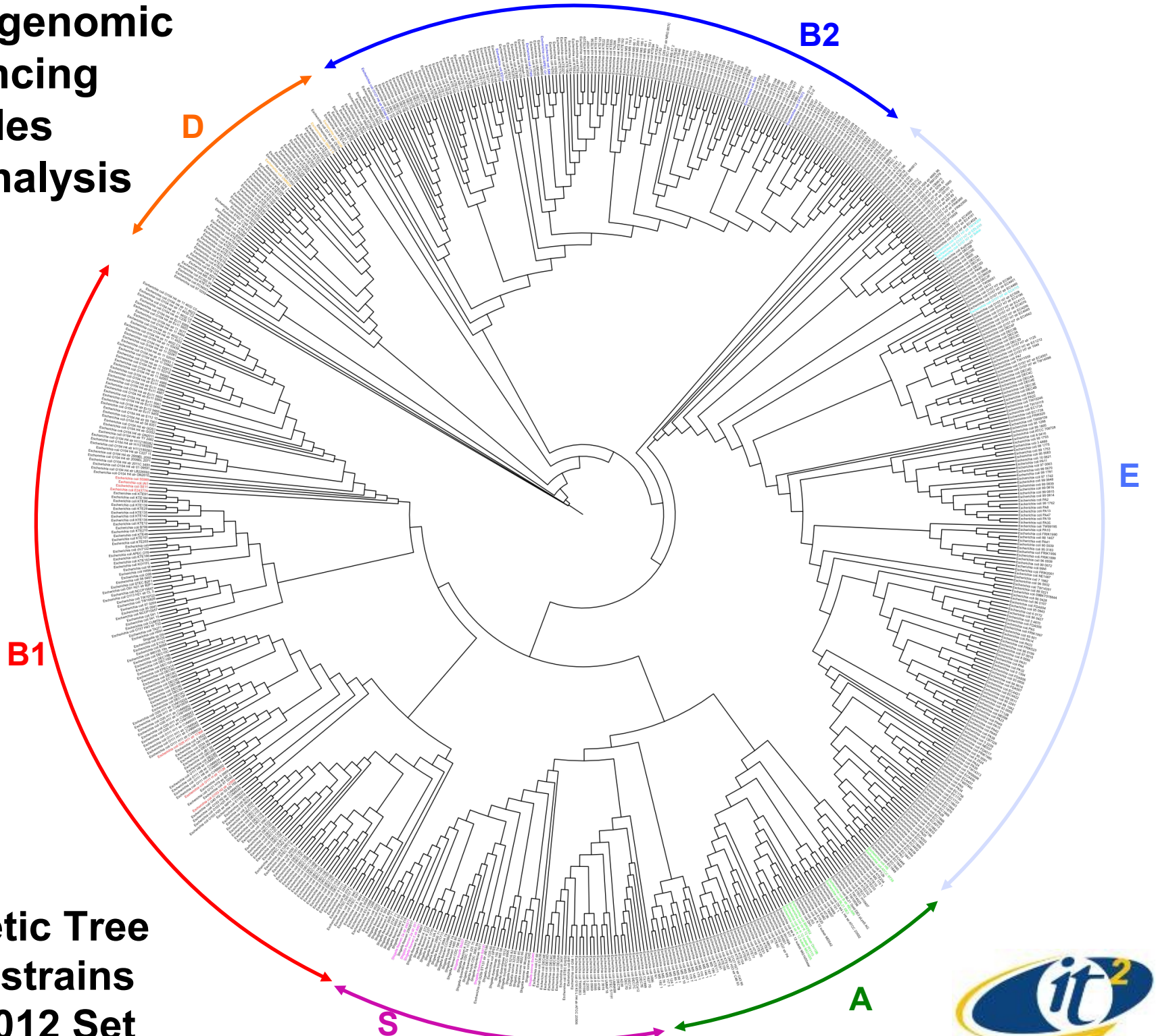
Escherichia coli Strain NC101



Science
AAAS

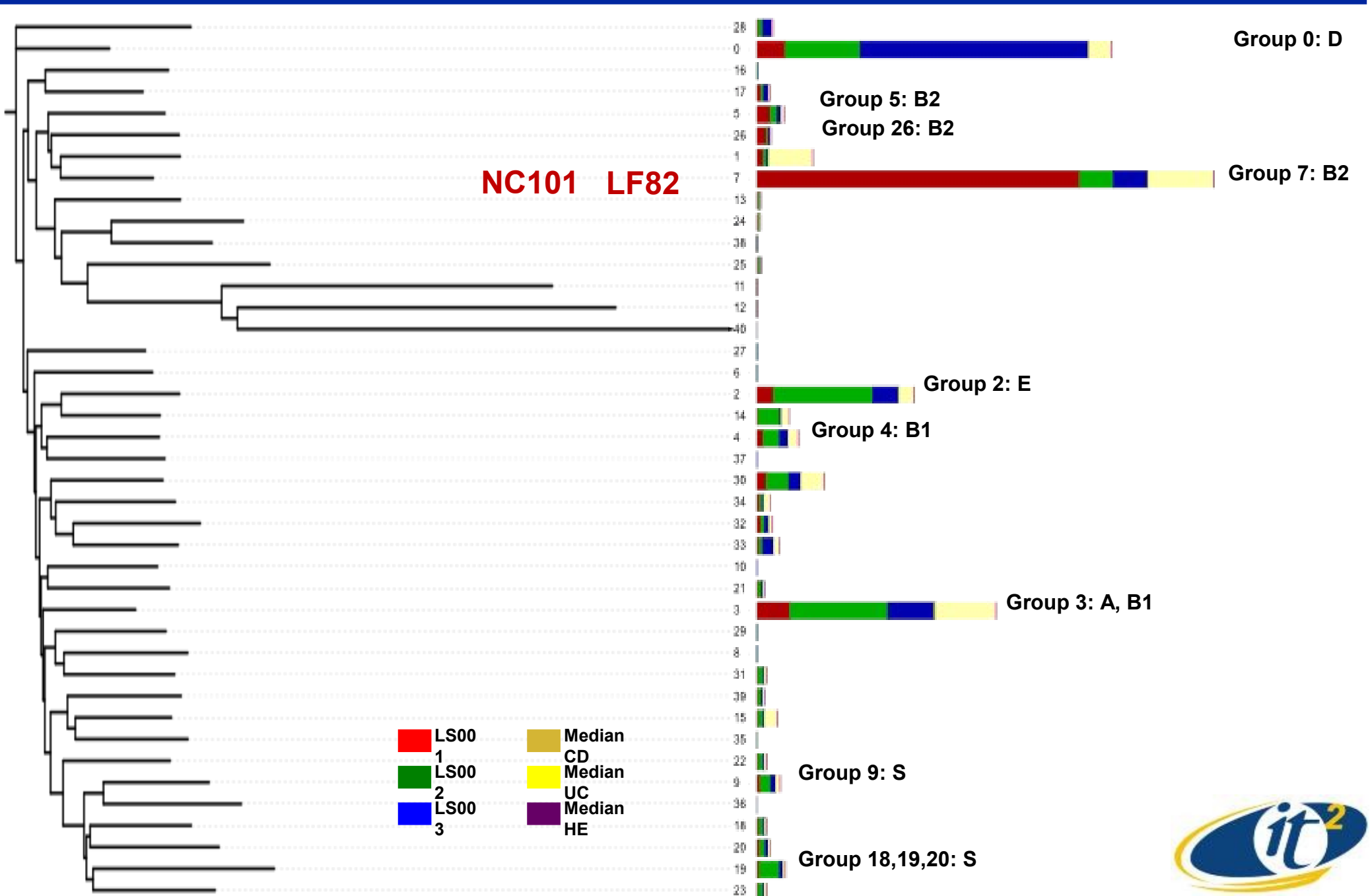
Bacteria Deliver a Genotoxic Hit
Robert F. Schwabe and Timothy C. Wang
Science **338**, 52 (2012);
DOI: 10.1126/science.1229905

Deep Metagenomic Sequencing Enables Strain Analysis

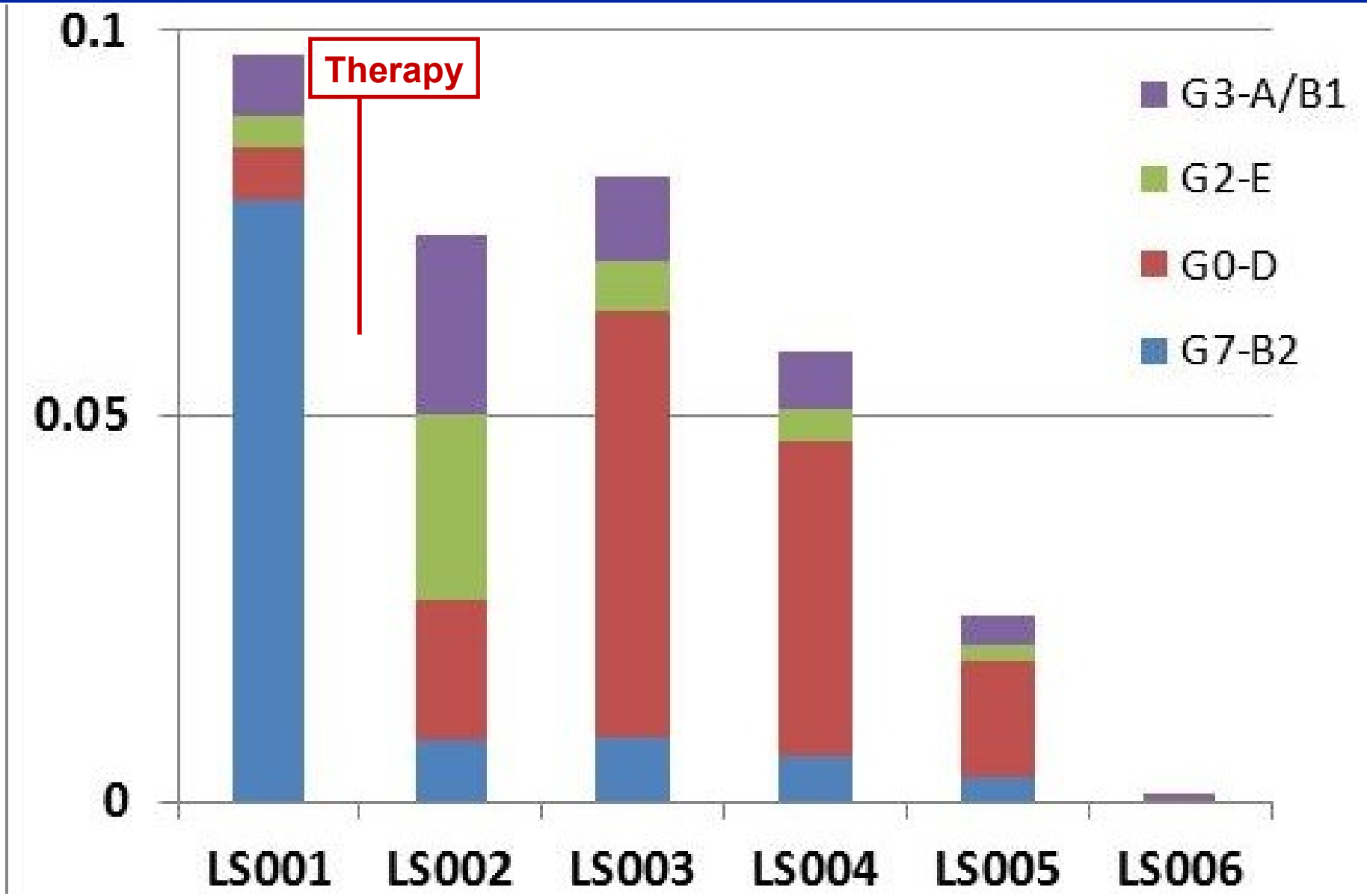


Phylogenetic Tree
778 Ecoli strains
=6x our 2012 Set

We Divided the 778 E. coli Strains into 40 Groups, Each of Which Had 80% Identical Genes



Reduction in E. coli Over Time
With Major Shifts in Strain Abundance



Strains >0.5% Included

Early Attempts at Modeling the Systems Biology of the Gut Microbiome and the Human Immune System

Trends in Biotechnology June 2011, Vol. 29, No. 6

Prospects for systems biology and modeling of the gut microbiome

Fredrik H. Karlsson, Intawat Nookaew, Dina Petranovic and Jens Nielsen

Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE-412 96 Göteborg, Sweden

2012 IEEE 26th International Parallel and Distributed Processing Symposium

High-Performance Interaction-Based Simulation of Gut Immunopathologies with ENteric Immunity SImulator (ENISI)

Keith Bisset*, Md. Maksudul Alam*, Josep Bassaganya-Riera*, Adria Carbo*, Stephen Eubank*, Raquel Hontecillas*, Stefan Hoops*, Yongguo Mei*, Katherine Wendelsdorf*, Dawen Xie*, Jae-Seung Yeom*†, and Madhav Marathe*†

**Virginia Bioinformatics Institute*

†Department of Computer Science

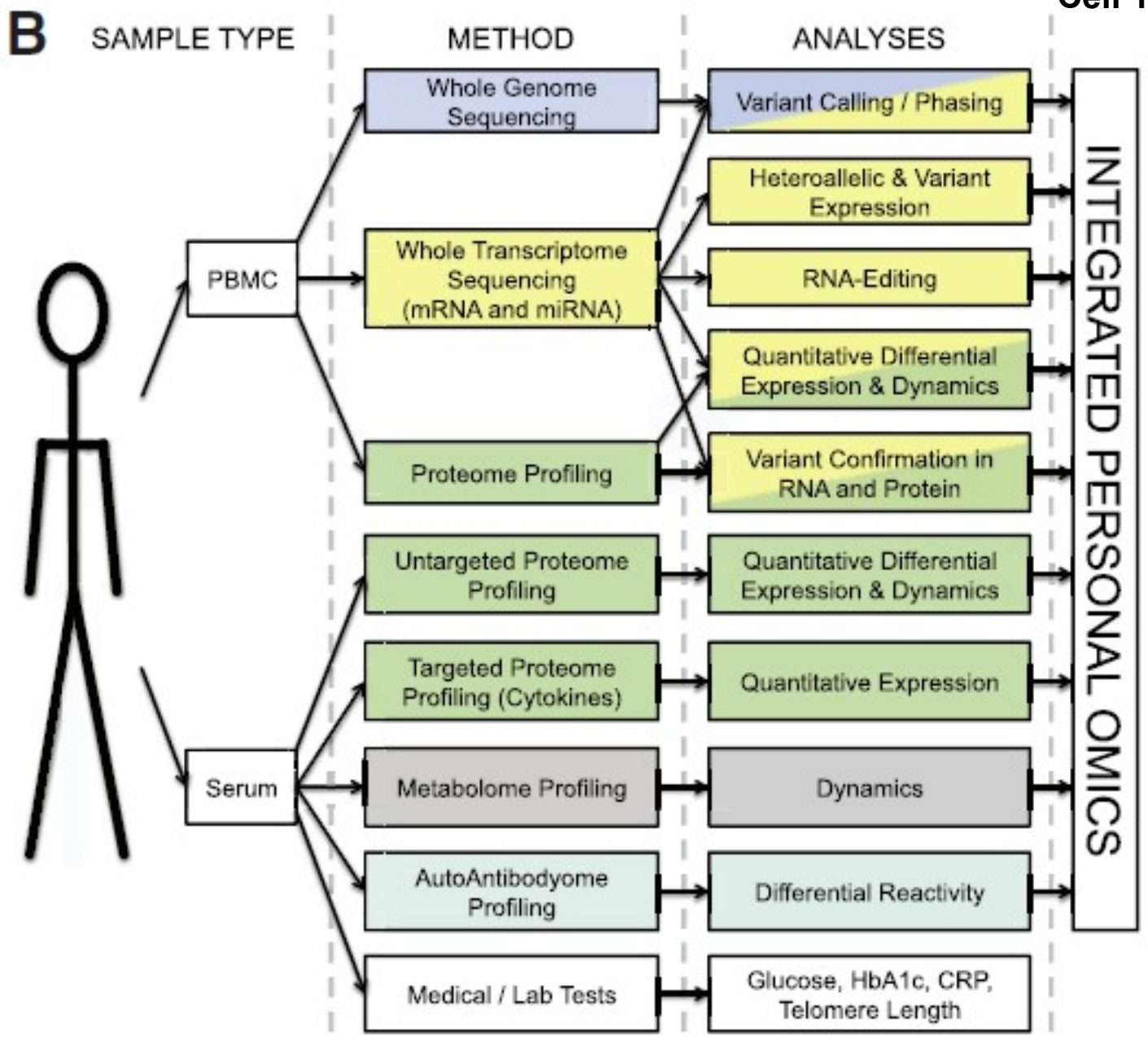
Virginia Tech, Blacksburg VA 24061

Next Step: Time Series of Metagenomic Gut Microbiomes and Immune Variables in an N=100 Clinic Trial

Goal: Understand
The Coupled Human Immune-Microbiome Dynamics
In the Presence of Human Genetic Predispositions

Next Level of Monitoring - Integrative Personal Omics Profiling Using 100x My Quantifying Biomarkers

Cell 148, 1293–1307, March 16, 2012



- **Michael Snyder**, Chair of Genomics Stanford Univ.
- Genome 140x Coverage
- Blood Tests 20 Times in 14 Months
  - tracked nearly 20,000 distinct transcripts coding for 12,000 genes
  - measured the relative levels of more than 6,000 proteins and 1,000 metabolites in Snyder's blood

From Quantified Self to National-Scale Biomedical Research Projects

DELSA

DATA-ENABLED LIFE SCIENCES ALLIANCE

Accelerate the impact of

Data-Enabled Life Sciences Research

on the pressing needs of the global society.

The Quantified Human Initiative is an effort to combine our natural curiosity about self with new research paradigms. **Rich datasets of two individuals, Drs. Smarr and Snyder**, serve as 21st century personal data prototypes.

My Anonymized Human Genome is Available for Download

Personal Genome Project



www.delsaglobal.org

www.personalgenomes.org

it²

Where I Believe We are Headed: Predictive, Personalized, Preventive, & Participatory Medicine

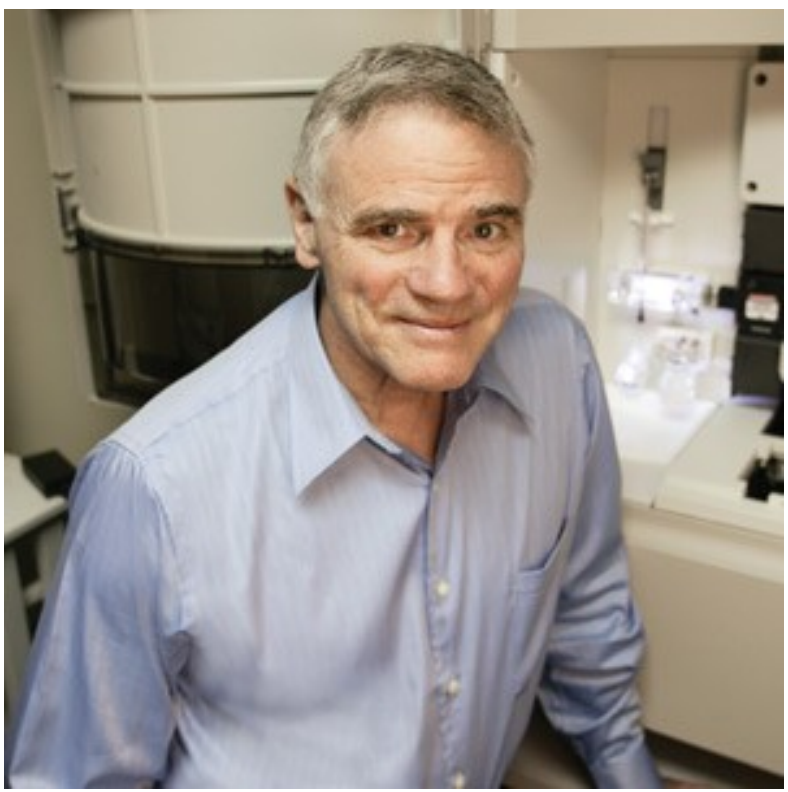
Newsweek

I am Lee Hood's Lab Rat!

A Doctor's Vision of the Future of Medicine

Leroy Hood
NEWSWEEK

From the magazine issue dated Jul 13, 2009



www.newsweek.com/2009/06/26/a-doctor-s-vision-of-the-future-of-medicine.html

Thanks to Our Great Team!

UCSD Metagenomics Team

Weizhong Li
Sitao Wu

JCVI Team

Karen Nelson
Shibu Yooseph
Manolito Torralba

Calit2@UCSD Future Patient Team

Jerry Sheehan
Tom DeFanti
Kevin Patrick
Jurgen Schulze
Andrew Prudhomme
Philip Weber
Fred Raab
Joe Keefe
Ernesto Ramirez



SDSC Team

Michael Norman
Mahidhar Tatineni
Robert Sinkovits

UCSD Health Sciences Team

William J. Sandborn
Elisabeth Evans
John Chang
Brigid Boland
David Brenner