2015 ADVANCED GENOMICS, METAGENOMICS, AND BIOINFORMATICS WORKSHOP

The University of the West Indies, St. Augustine Campus, Trinidad and Tobago
February 19 & 20, 2015

Powerpoint Slide Package

The University of the West Indies
National Institute of Allergy and Infectious Diseases (NIAID)
J. Craig Venter Institute (JCVI)

UWI
ST. AUGUSTINE CAMPUS

NIAID

J. Craig Venter®
INSTITUTE

Dr. William Nierman

Presentation on the History of the Microbiome

A Brief History of Human Microbiome Research

William C. Nierman
Professor
Infectious Diseases Program Director
J. Craig Venter Institute

J. Craig Venter™
INSTITUTE

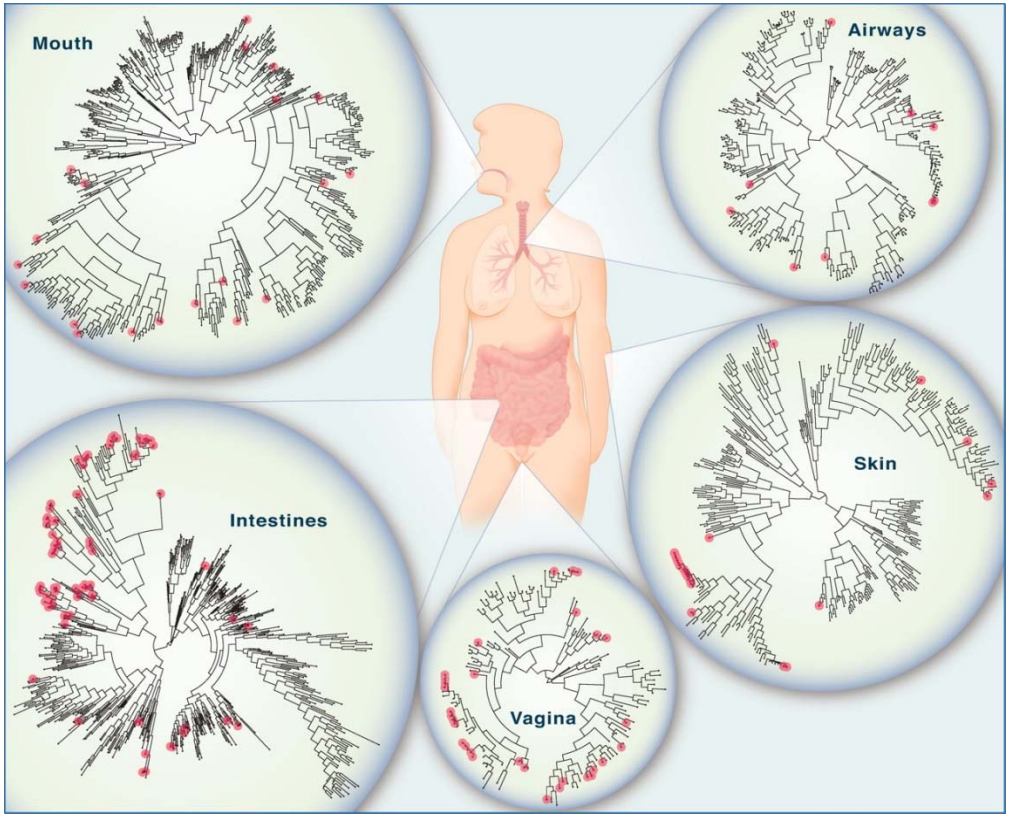
Presentation Overview

- Sargasso Sea and Global Ocean Sampling
- Early Human Microbiome Studies (Karen Nelson et al.)
- Human Microbiome Project
- JCVI Microbiome Projects
- Human Respiratory Tract Microbiome
- Modern Microbiome Analysis

J. Craig Venter™
INSTITUTE

Human Microbiome: Our Other Genome

- **Ten trillion bacterial cells**
- **Ten times more bacterial than human cells**
- **100 more bacterial than human genes**



J. Craig Venter™
INSTITUTE

The Human Microbiome, Relevance, Significance, Potential Impact

- ✓ Significant microbial diversity across the human body and across individuals - Each niche has its own ecosystem
- ✓ Estimated that ~100 trillion bacteria reside in and on the human body ~3% of body weight
- ✓ We have limited understanding on their roles but have the technologies to interrogate this diversity in health and disease
- ✓ Vast majority remain uncultured

  - ✓ Normal flora of healthy individuals can potentially be mined to identify new modulators, natural products etc. to restore normal/health conditions
  - ✓ Population changes/shifts can be used as indicators of health status
  - ✓ Microbiome can be used for disease surveillance

J. Craig Venter™ INSTITUTE

Human Microbiome Complexity

- ✓ Microbiome plays important role in immune development
- ✓ Correlations with microbiota/antibiotics and development of immune-mediated diseases demonstrated
- ✓ Improved sanitization, antibiotic usage, and immunizations are factors that can shift the microbiota
- ✓ Microbial populations can differ based on geographic location
- ✓ Microbiome may impact brain health and human behavior
- ✓ Differences observed based on weight/BMI
- ✓ Differences observed based on diet
- ✓ Differences observed based on health status

J. Craig Venter INSTITUTE

Metagenomics-Tools

- **We are capable of sequencing and analyzing the genomes of culturable species**
- **These species are estimated to represent less than 1% of total microbial diversity**

- **Culture dependent analysis**:
  - Culture and obtain pure colonies
  - Complete genome sequencing of DNA
  - Organism has to be cultured in the laboratory

- **Culture-independent analysis**
  - 16S ribosomal RNA (rRNA) sequencing
  - Whole genome sequencing, assembly, annotation

- **Metagenomics:** sequence based analysis of complete microbial communities without need for culturing

  - Made possible by number of **parallel technical developments**:
- **Assembly and data analysis capabilities developed to being able to tease apart these large datasets**
- **Sequencing capabilities capable of achieving great depths of coverage at reduced cost**
- **Demonstrated proof of concept via Sargasso Sea study**
- **Global Ocean Sampling (GOS) largest protein dataset in existence**
- ***Other "omics" technologies.*** Proteomics, metatranscriptomics, metabolomics

J. Craig Venter™ INSTITUTE

Changes in Sequencing Technologies

ABI 3730*xl* ***1-2 Mb/day***

Illumina GA IIx
50 Gb/12day run

ABI SOLiD
100Gb/12 day run

454 GS FLX +
0.6Gb/23hr run

Illumina HiSeq 2000 (2500†)
600 Gb/11day run

Ion Torrent
1Gb/2hr run

Ion Proton
100Gb/4 hr Run

†***HiSeq 2500 upgrade****: up to 120Gb/27 hour run (available now for $50K)*

J. Craig Venter™ INSTITUTE

JCVI Primary Sequencing Platforms

Higher capacity for
lower cost in
less time
than HiSeq 2000

Illumina NextSeq 500

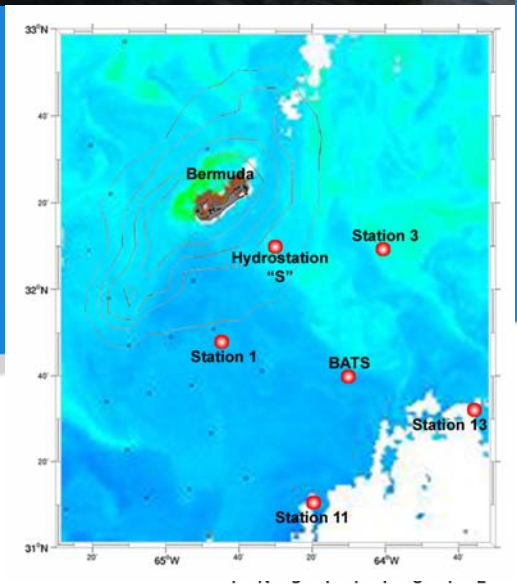
2x300 base reads
Main platform for 16S rRNA

Illumina MiSeq

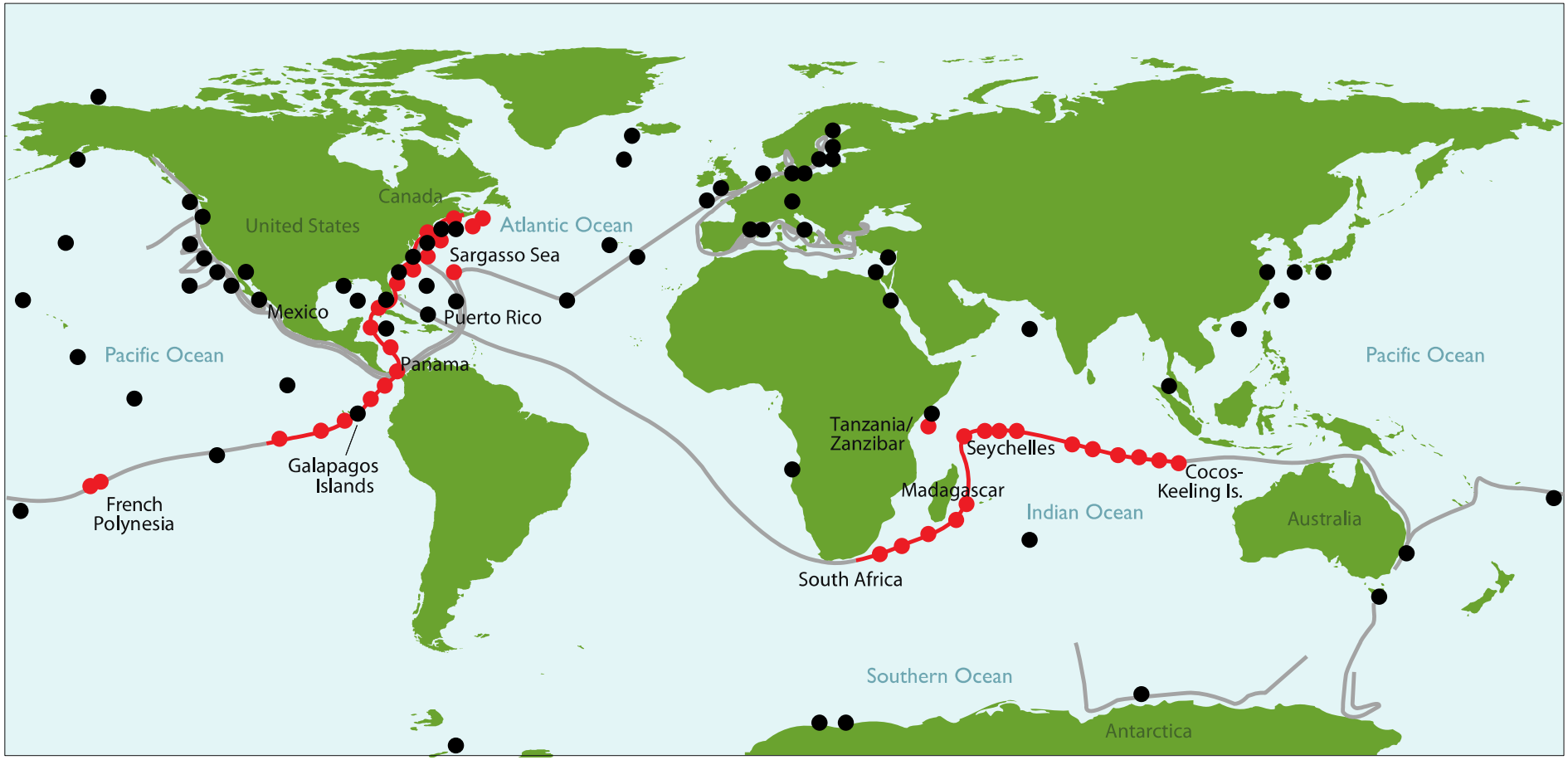
J. Craig Venter™
INSTITUTE

Sargasso Sea Study

- ✓ Venter and colleagues at the JCVI
- ✓ Generated 1,987,936 DNA reads
- ✓ Approximately 1,625 Mb of DNA
- ✓ 1.2 million new genes identified
- ✓ ~1,412 rRNA genes
- ✓ Estimated 1,800 species
- ✓ 12 complete genomes recovered
- ✓ Demonstration of the power of genomics



Global Ocean Sampling Expeditions

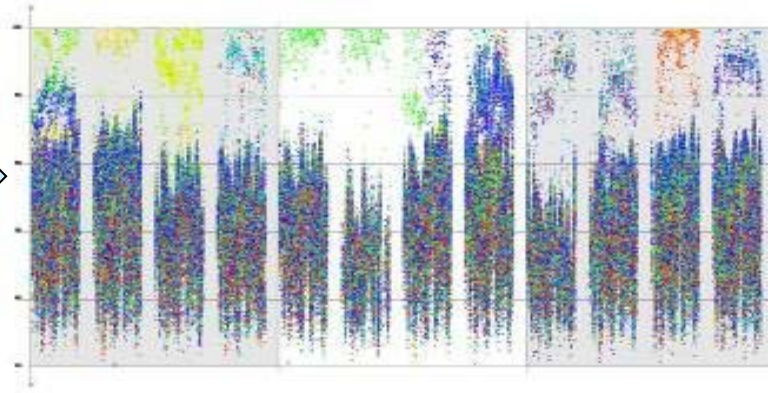


J. Craig Venter®
INSTITUTE

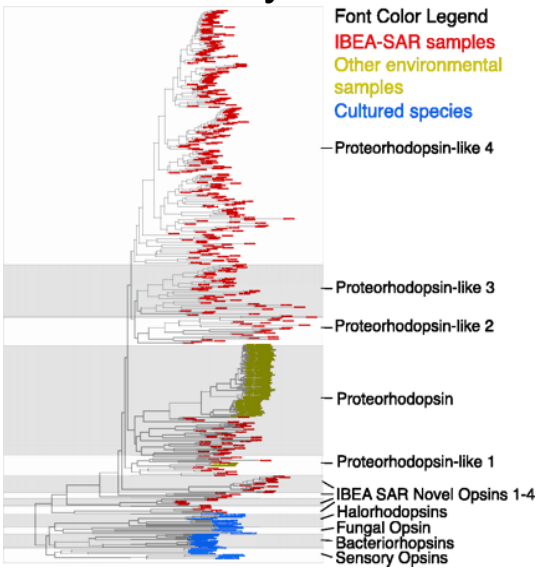
Global Ocean Sampling and Analysis

Sampling and Sequencing

Tool Development

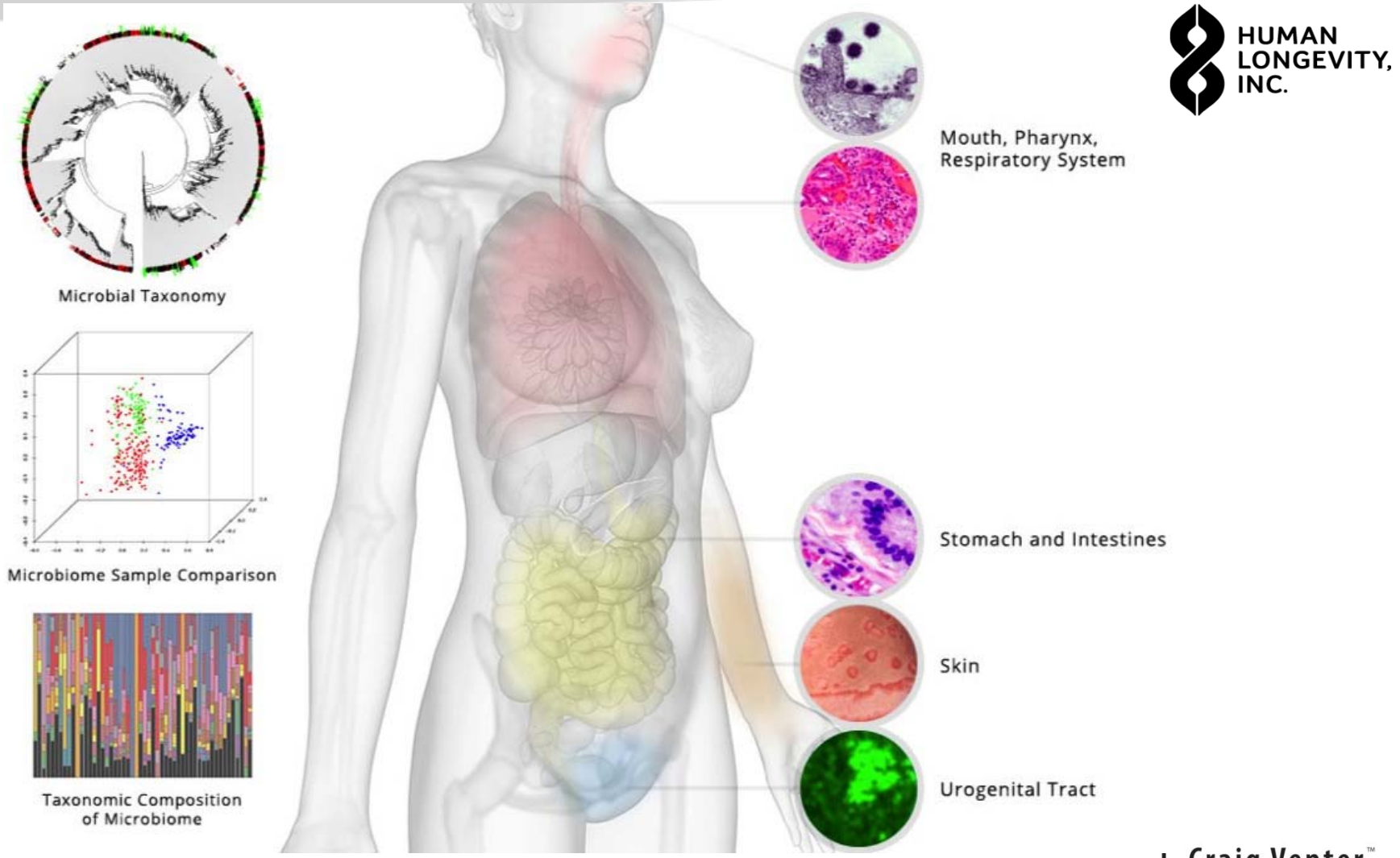


Data Analysis



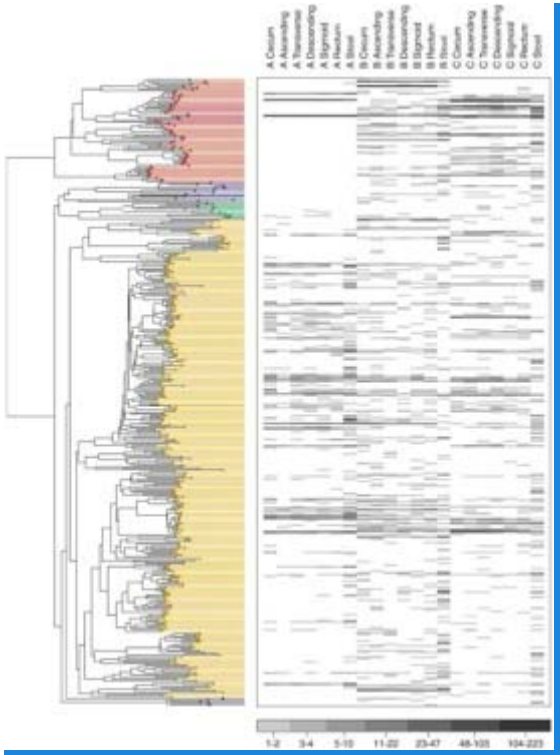
J. Craig Venter™ INSTITUTE

Human Microbiome Profiling by High Throughput Sequencing

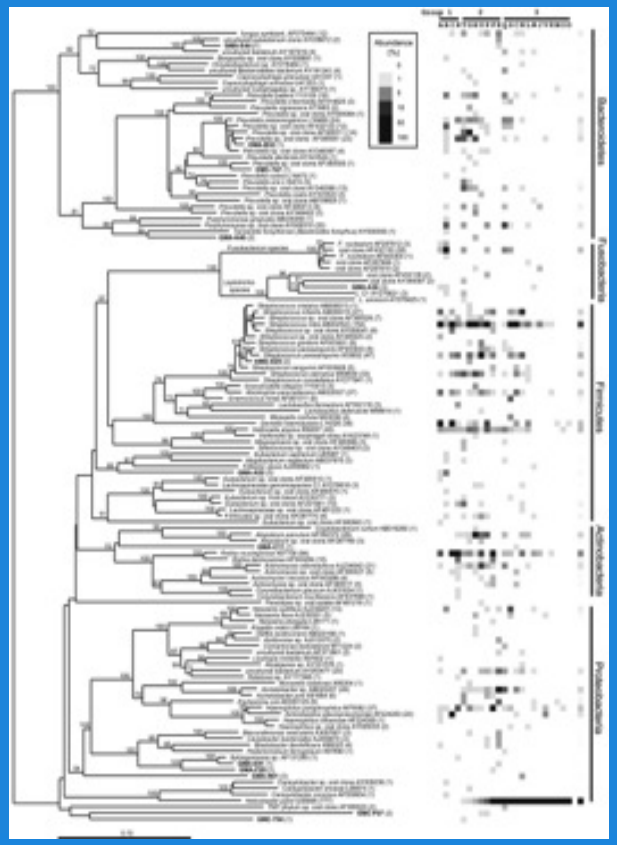


HUMAN LONGEVITY, INC.

J. Craig Venter™ INSTITUTE



First Published Study **Fecal** samples were collected from each subject 1 month following colonoscopy. From 11,831 bacterial and 1524 archaeal 16S sequences, identified 395 phylotypes. *Eckburg et at., 2005 Science 308(5728):1635-8.*



Described 128 16S rDNA phylotypes. Derived from 23 human subjects

First human metagenomic study
Investigated the gastrointestinal tract of two healthy adults

J. Craig Venter™
INSTITUTE

First human metagenomic paper

- Investigated the gastrointestinal tract (via fecal samples) of two healthy adults
- 78 Mbp
- 2062 amplified 16S rDNA

RESEARCH ARTICLE

Metagenomic Analysis of the Human Distal Gut Microbiome

Steven R. Gill,[1]*‡ Mihai Pop,[1]† Robert T. DeBoy,[1] Paul B. Eckburg,[2,3,4] Peter J. Turnbaugh,[5] Buck S. Samuel,[5] Jeffrey I. Gordon,[5] David A. Relman,[2,3,4] Claire M. Fraser-Liggett,[1,6] Karen E. Nelson[1]

The human intestinal microbiota is composed of 10^{13} to 10^{14} microorganisms whose collective genome ("microbiome") contains at least 100 times as many genes as our own genome. We analyzed ~78 million base pairs of unique DNA sequence and 2062 polymerase chain reaction–amplified 16S ribosomal DNA sequences obtained from the fecal DNAs of two healthy adults. Using metabolic function analyses of identified genes, we compared our human genome with the average content of previously sequenced microbial genomes. Our microbiome has significantly enriched metabolism of glycans, amino acids, and xenobiotics; methanogenesis; and 2-methyl-D-erythritol 4-phosphate pathway–mediated biosynthesis of vitamins and isoprenoids. Thus, humans are superorganisms whose metabolism represents an amalgamation of microbial and human attributes.

Our body surfaces are home to microbial communities whose aggregate ≥100 times as many genes as our 2.85–billion base pair (bp) human genome (*1*). Therefore, a

of single organisms, recent reports from Venter *et al.* (*9*) and Baker *et al.* (*10*) have demonstrated the utility of this approach for studying mixed microbial communities. Variations in the relative abundance of each member of the microbial community and their respective genome sizes determine the final depth of sequence coverage for any organism at a particular level of sequencing. This means that the genome sequences of abundant species will be well represented in a set of random shotgun reads, whereas lower abundance species may be represented by a small number of sequences. In fact, the size and depth of coverage (computed as the ratio between the total length of the reads placed into contigs and the total size of the contigs) of genome assemblies generated from a metagenomics project can provide information on relative species abundance.

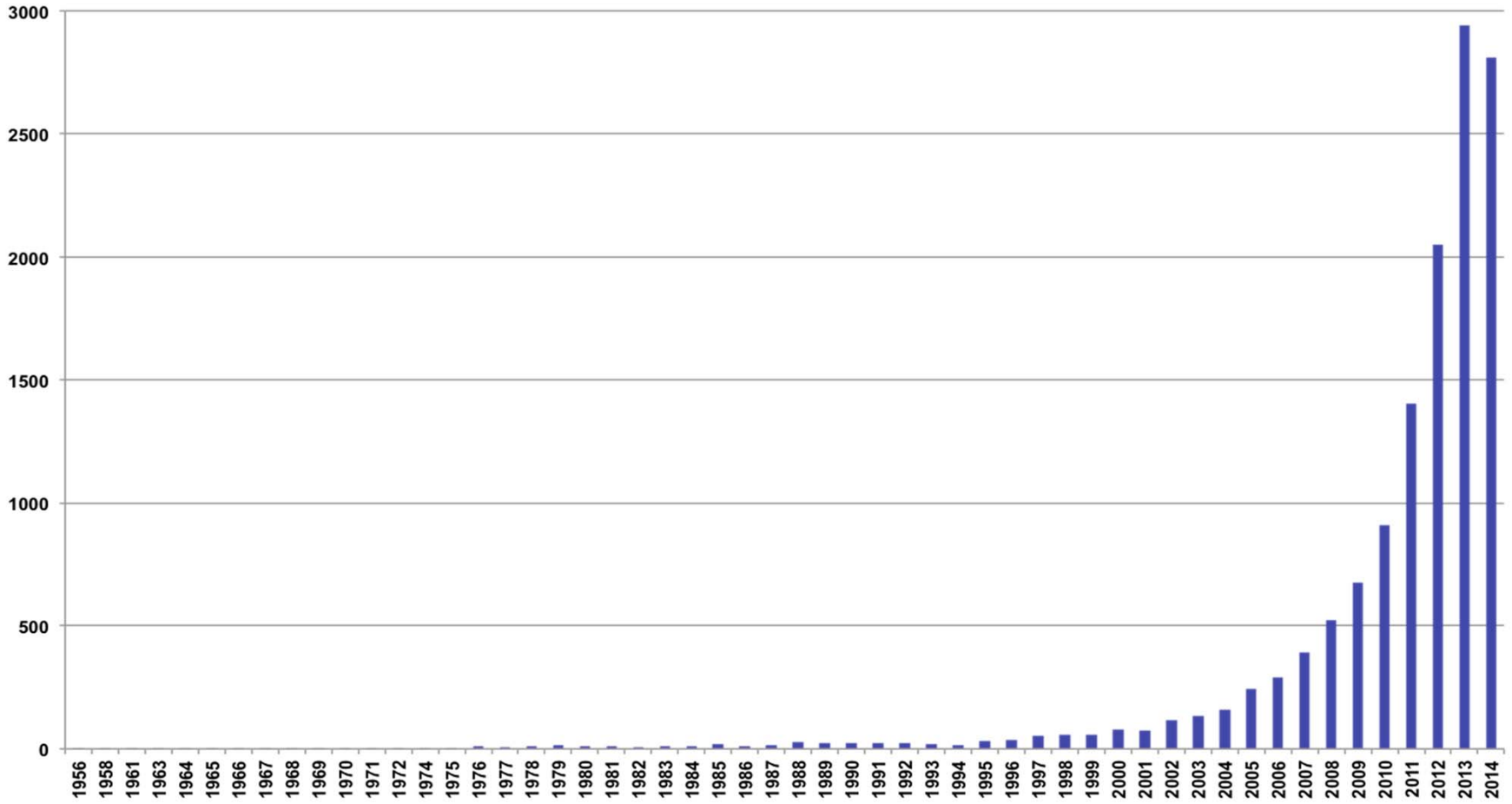
A total of 65,059 and 74,462 high-quality sequence reads were generated from random DNA libraries created with fecal specimens of two healthy humans (subjects 7 and 8). These two subjects, ages 28 and 37, female and male,

Gill et al, *Science* 2006

J. Craig Venter™ INSTITUTE

Expanding Number of Microbiome Publications

PubMed – Oct 2014



J. Craig Venter™
INSTITUTE

NIH Road Map Initiative
Human Microbiome Project (HMP)

J. Craig Venter™
INSTITUTE

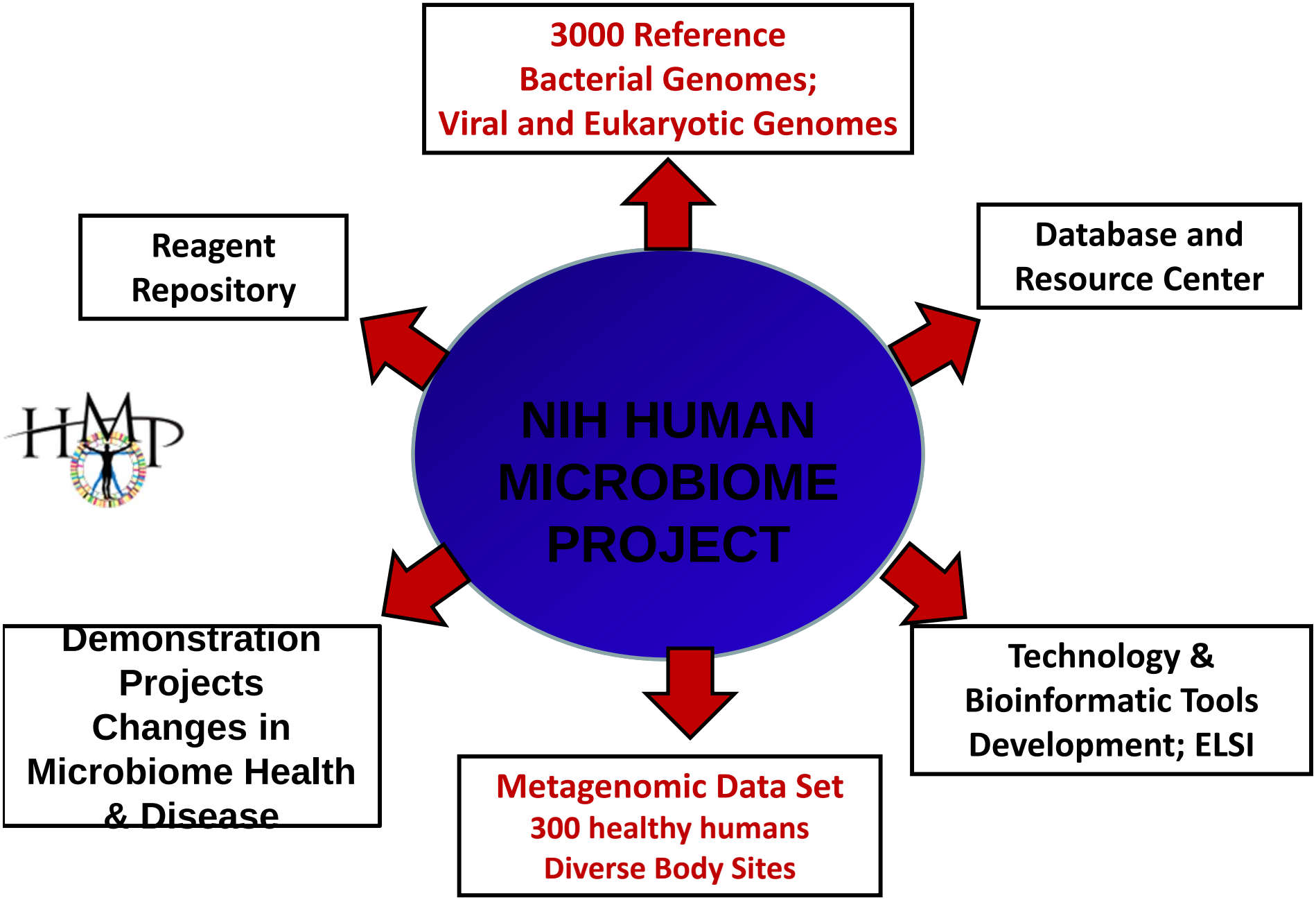
NIH Roadmap HMP

- ✓ **Budget > $175 million 2007-2014**
- ✓ **Main teams were 4 large-scale sequencing centers**
  - ✓ **Broad, Wash U, JCVI, Baylor**
- ✓ **Goal: Characterize the microbes that inhabit the human body and examine whether changes in the microbiome can be related to health and disease**
- ✓ **Project designed to determine the value of microbial metagenomics in biomedical research**
- ✓ **Develop Standard Operating Procedures, Reagents and Data sets all rapidly placed in public domain**
- ✓ **Continuous Scientific Community Input**
  - ✓ **External Scientific Advisory Group, Workshops.**

- ✓ http://nihroadmap.nih.gov/hmp
- ✓ http://www.human-microbiome.org/#

J. Craig Venter™ INSTITUTE

Slide courtesy Maria Giovanni-NIAID



Slide courtesy Maria Giovanni-NIAID

“Healthy Cohort” Body Sites

Oral
- Saliva
- Tongue dorsum
- Hard palate
- Buccal mucosa
- Keratinized (attached) gingiva
- Palatine tonsils
- Throat
- Supragingival plaque
- Subgingival plaque

Skin
- Retroauricular crease, both ears (2)
- Antecubital fossa (inner elbow), both arms (2)

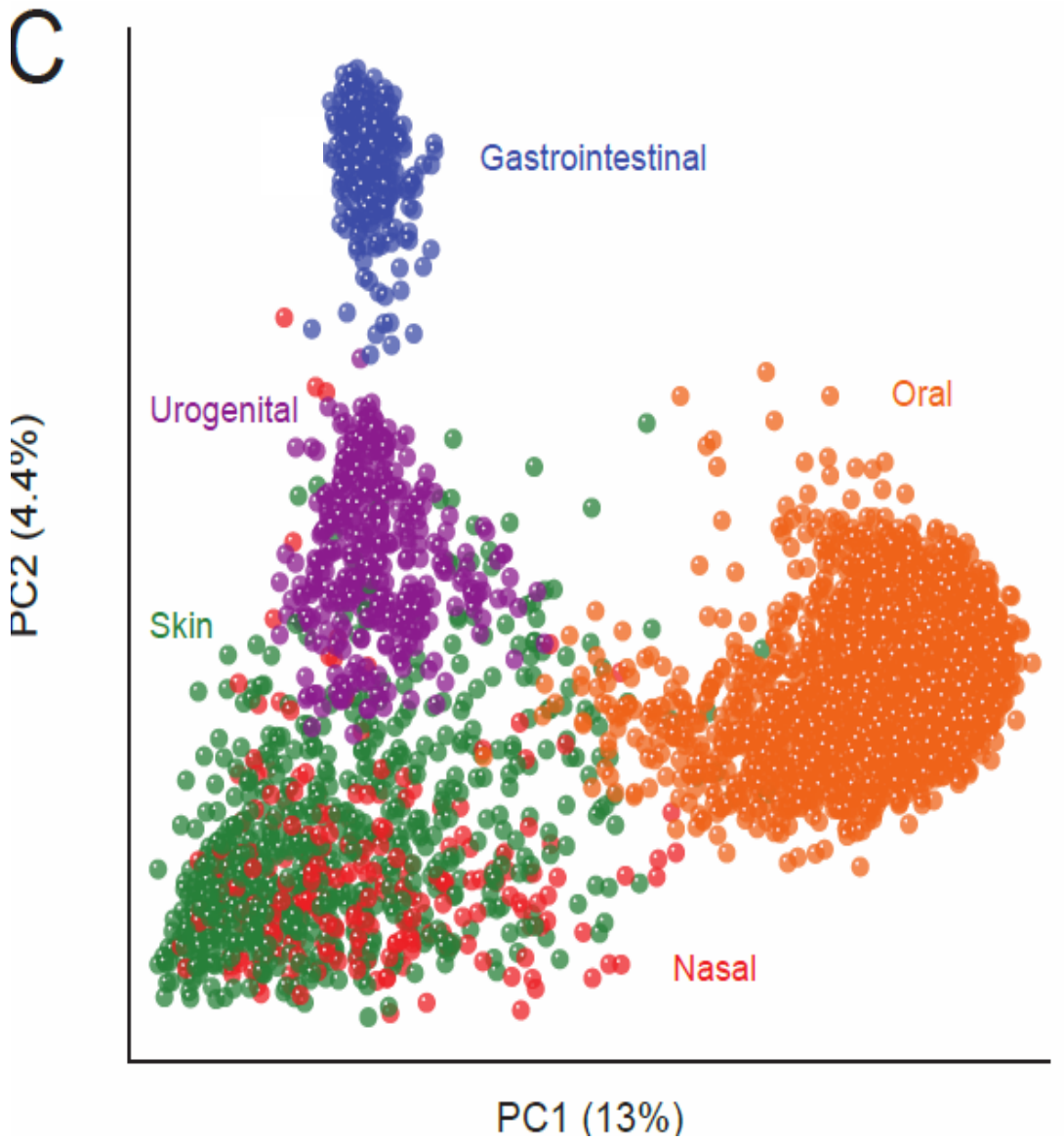
Nasal
- Anterior right and left nares (pooled)

Gut

Vaginal
- Posterior fornix, vagina
- Midpoint, vagina
- Vaginal introitus

Slide courtesy of NHGRI

J. Craig Venter INSTITUTE



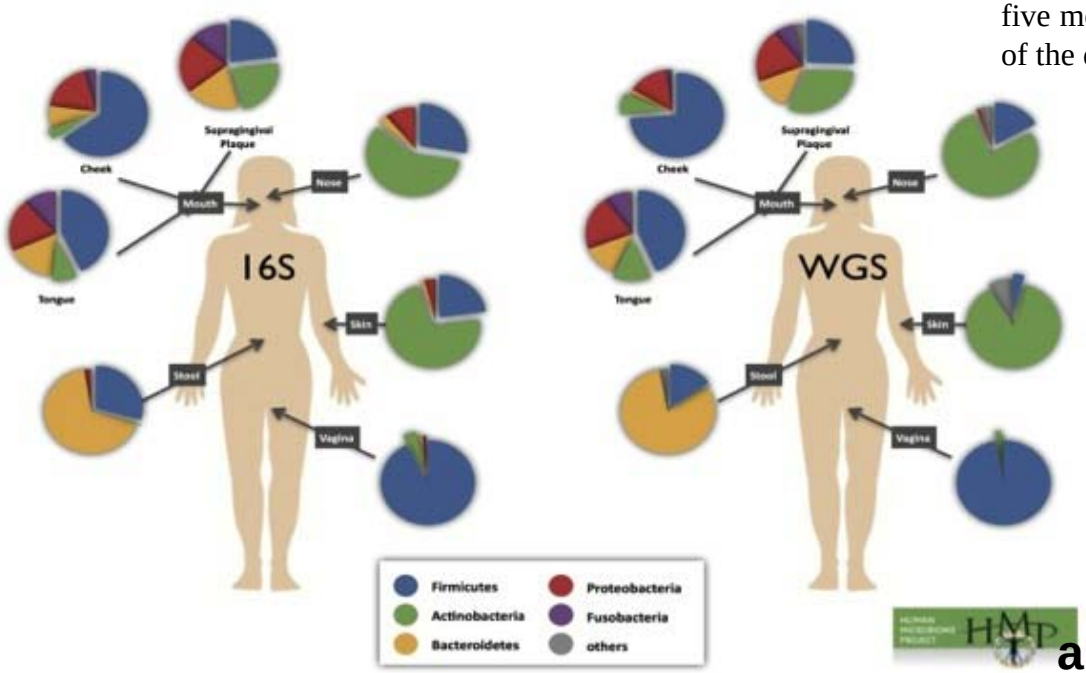
In adults, each part of the body supports a distinct microbial community.

With no apparent relationship with gender, age, weight, ethnicity or race.

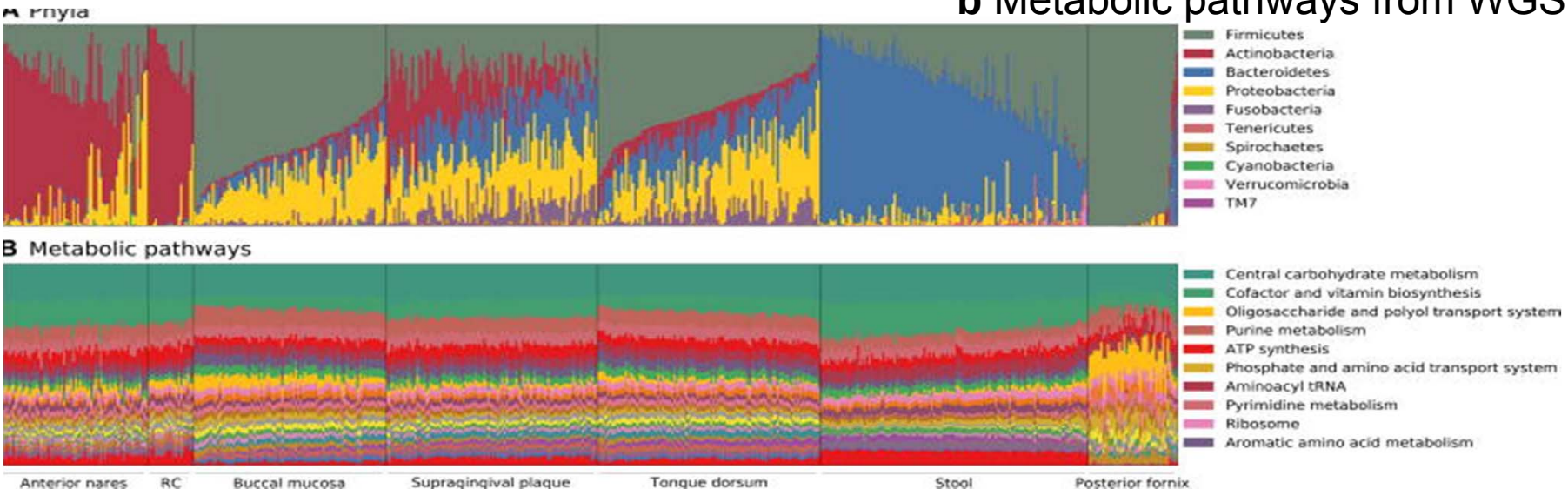
"Structure, Function and Diversity of the Human Microbiome in an Adult Reference Population" The Human Microbiome Consortium. **HMP Consortium (2012)**

HMP Consortium, Nature 2012

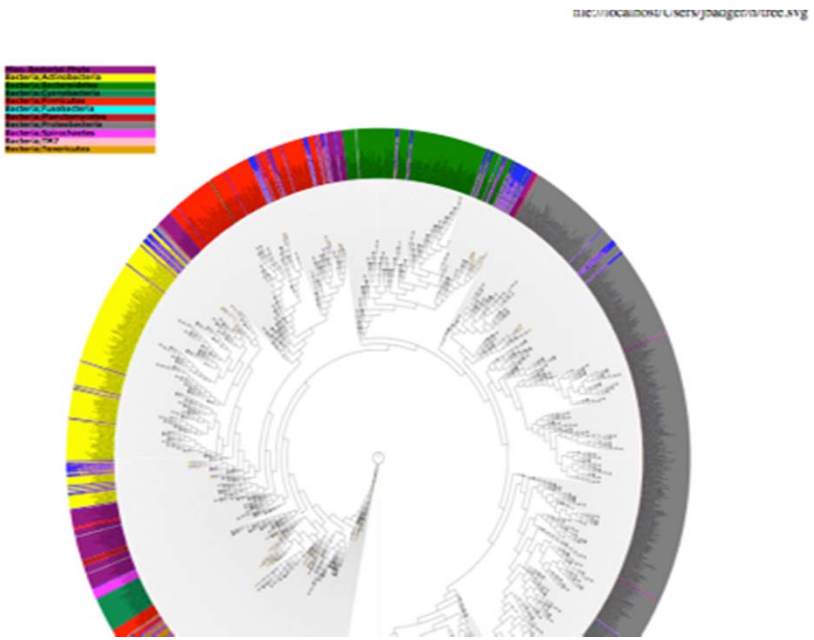
Supplementary Figure 8. Phylum abundances per body site. For each of the body sites studied by both 16S rRNA gene sequencing (A) and whole-genome shotgun sequencing (B) the five most abundant phyla are shown. The small remaining fraction of the data is collapsed and labeled as other phyla (grey).



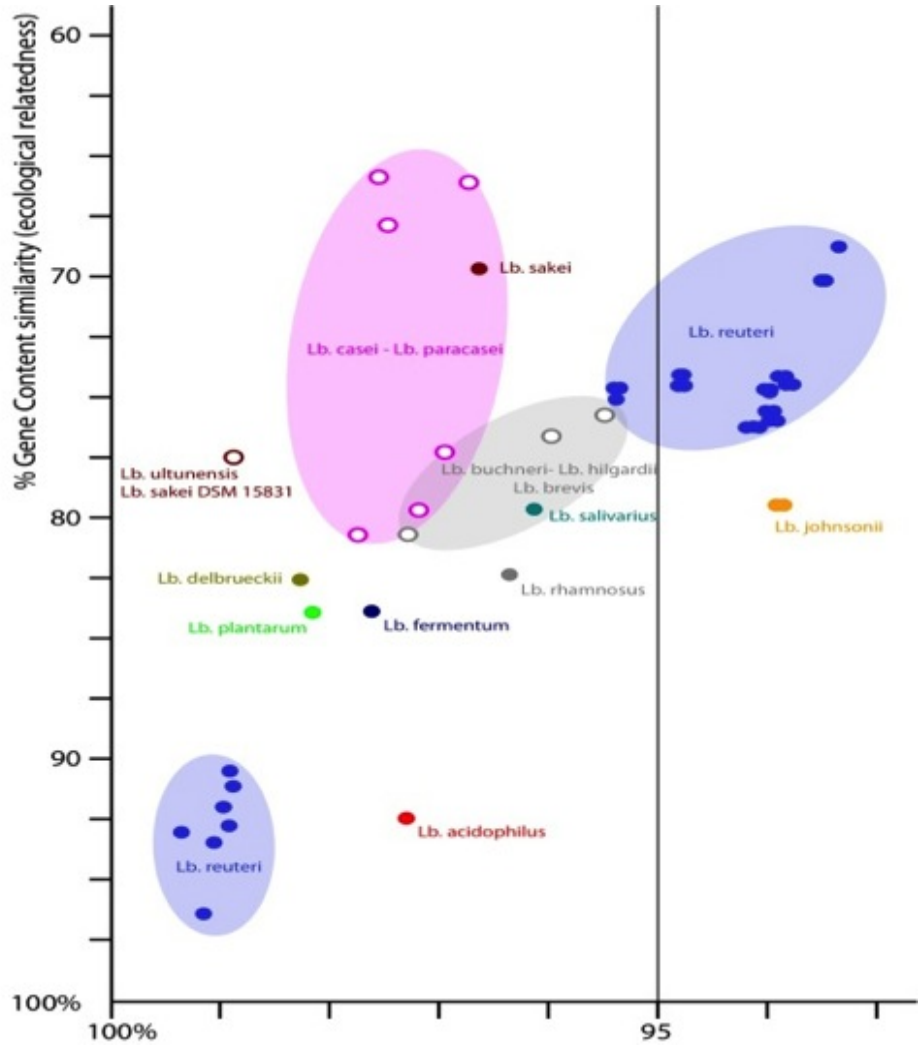
a 16S-based OTU bins
b Metabolic pathways from WGS data



A Catalog of Reference Genomes from the Human Microbiome



178 genomes
~550,000 genes
Nelson et al., Science
May 21, 2010



JCVI Microbiome Space

J. Craig Venter™
INSTITUTE

Ongoing Microbiome/Disease Studies at JCVI

/

- ✓ **Type 1 Diabetes**
- ✓Progression of **esophageal cancer**
- ✓Bacterial vaginosis and **pre-term babies**
- ✓Nasopharynx microbiome and **infant pneumonia**
- ✓Skin microbiome, **acne and psoriasis**
- ✓Oral diseases including **periodontitis**
- ✓**Colon cancer**
- ✓Mouse models of **alcoholism**
- ✓**Urinary Tract Infections**
- ✓**Spinal cord injuries**
- ✓Upper **Respiratory tract infections** in animal models
- ✓**Febrile illnesses** in Children
- ✓Molecular Hallmarks of Naturally Acquired Immunity to **Malaria**
- ✓**Rotaviral infections**
- ✓Astronaut microbiome (NASA)
- ✓**Chronic Wound** microbiomes
- ✓Animal models of **stress**

http://www.jcvi.org/cms/research/projects/hms/overview

J. Craig Venter INSTITUTE

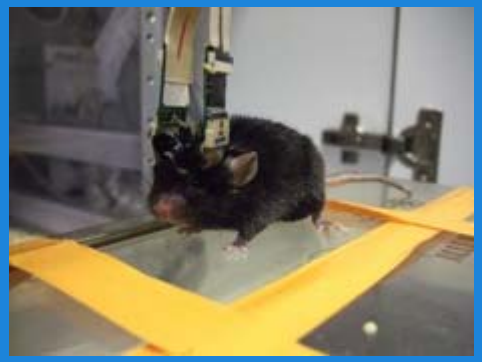
Oral Metagenomes - Health - and the Power of Twins

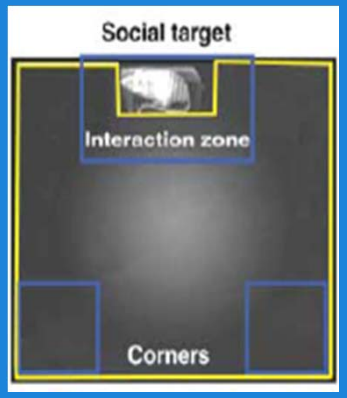
- JCVI – Universities of Melbourne and Adelaide

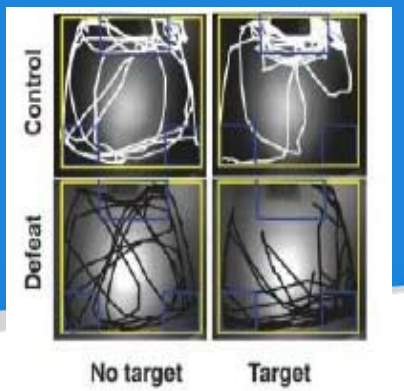
- Human Variability – Enrolled 600 twin pairs
- Dizygotic ~300 Monozygotic ~300
- Genetic background, diet, lifestyle, geography, socio-economic status, etc.

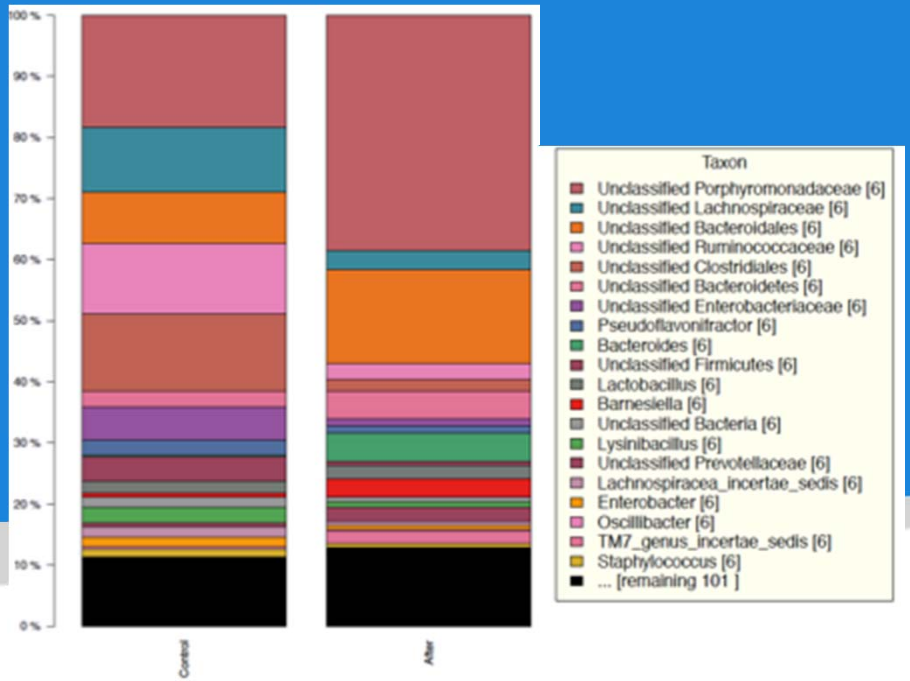
J. Craig Venter™
INSTITUTE

Impact of Stress on the Murine Gut Microbiota





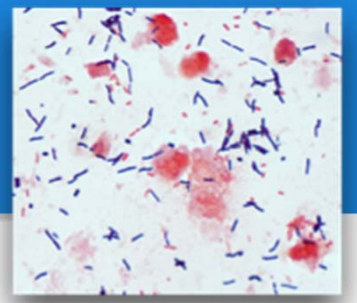




Dr. Marcus Jones in collaboration with Kafui Dzirasa @ Duke

J. Craig Venter INSTITUTE

Clostridium difficile



Gram negative, sporulating, obligate anaerobe

Member of the normal human gut flora

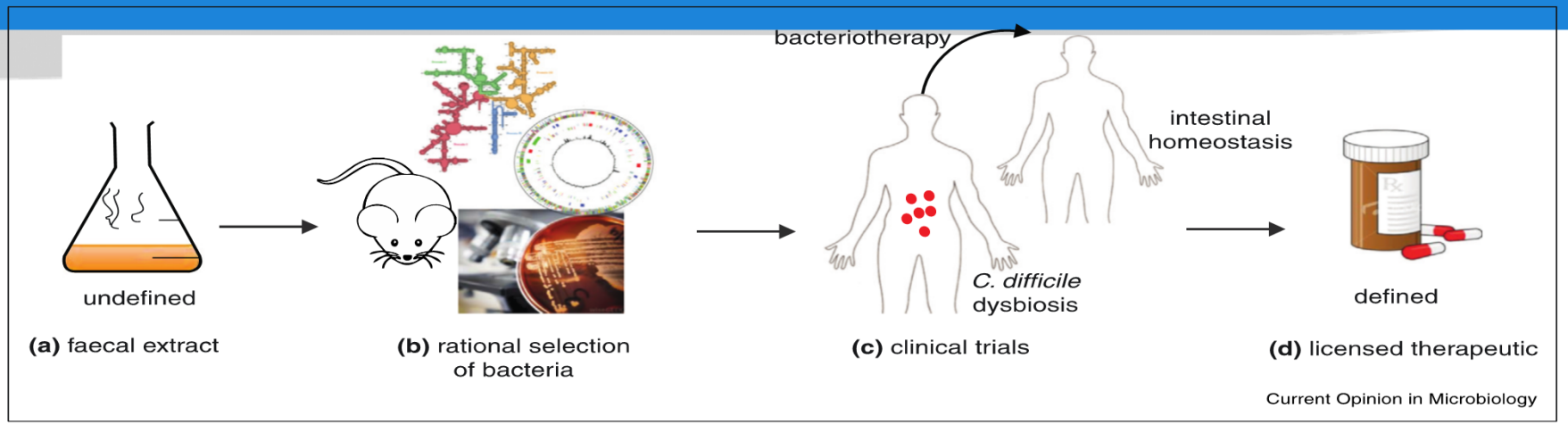
Strains produce cytotoxins, TcdA and TcdB

Major cause of antibiotic-associated diarrhea (>500,000 cases/yr), especially in the elderly, hospitalized, nursing homes

Antibiotic-resistant recurrence increasingly common

J. Craig Venter™ INSTITUTE

First Report of Fecal Bacteriotherapy – return to early studies



THE LANCET, MAY 27, 1989

BACTERIOTHERAPY FOR CHRONIC RELAPSING CLOSTRIDIUM DIFFICILE DIARRHOEA IN SIX PATIENTS

M. TVEDE[1] J. RASK-MADSEN[2]

Department of Clinical Microbiology, Rigshospitalet, Statens Seruminstitut;[1] and Section of Gastroenterology, Department of Medicine G, Bispebjerg Hospital, University of Copenhagen, Denmark[2]

- ✓ *Blautia producta*
- ✓ *Clostridium bifermentans*
- ✓ *Clostridium innocuum*
- ✓ *Clostridium ramosum*
- ✓ *Enterococcus faecalis*
- ✓ *Bacteroides ovatus*
- ✓ *Bacteroides thetaiotaomicron*
- ✓ *Bacteroides vulgatus*
- ✓ *Escherichia coli* (2 strains)

- ✓ 5 patients, >59 yo with relapsing CDI
- ✓ treated, by enema, with mixture of 10 bacterial strains
- ✓ normal bowel function within 24 h, *C. difficile* negative within 7 d

Slide courtesy of Dr. Sarah Highlander, JCVI

J. Craig Venter™ INSTITUTE

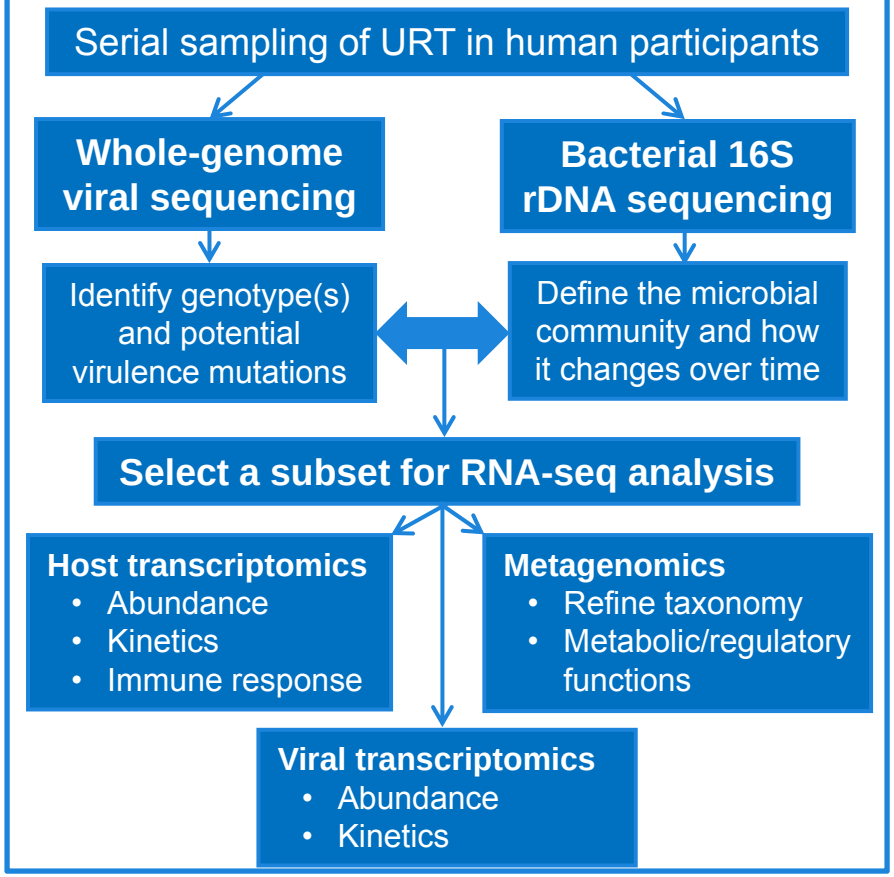
Exploiting Viral Genomics and Metagenomics to Understand Disease

J. Craig Venter™
INSTITUTE

Virus Host Microbiome Interactions

Elucidating viral-host-microbiome determinants that influence viral pathogenesis.

Study design



Hypotheses and preliminary data

- Acute viral infections alter the URT microbiome directly and via host responses to infection
  - Ferrets inoculated with H1N1 IAV have higher URT abundances of *Pseudomonas* and *Acinetobacter* than uninfected controls.
  - URT microbial community structure was perturbed by intra-nasal live attenuated influenza virus vaccination in humans.
- Microbial community structure may influence URT disease severity during acute viral infection and, in the case of RSV, may play a role in the development of subsequent childhood asthma.

Collaborators

Drs. Adolfo Garcia-Sastre and Rafael Medina, Icahn School of Medicine at Mount Sinai, NY (IAV);
Dr. Tina Hartert, Vanderbilt University School of Medicine, TN (RSV)

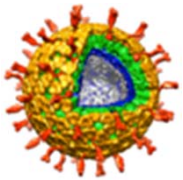
Image Credits

http://www.cdc.gov/flu/images.htm (virus); Janice Haney Carr, CDC (bacteria)

J. Craig Venter INSTITUTE

Enteric Microbiome -Pediatric Diarrheal Disease – Rotaviral Infections in Africa

To elucidate pathogen-microbiome relationships involved in acute pediatric gastrointestinal rotavirus infections and vaccine efficiency



Generalized study design

~150 pediatric diarrheal stool samples collected before and after introducing national rotavirus vaccination campaigns in sub-Saharan Africa

Whole-genome viral sequencing

Bacterial 16S rDNA sequencing

Identify genotype(s)

Define the microbial community

Select a subset for DNA-based metagenomics

- Identify other pathogens (viral, bacterial, fungal, protozoan, helminth, etc.)
- Refine bacterial taxonomy
- Evaluate metabolic/regulatory functions

J. Craig Venter INSTITUTE

Microbiome Correlations with Infectious Diseases

- ✓ Continues to be a major source of **morbidity and mortality** with differences in developed and developing nations.
- ✓ Major problems exist in **diagnosis** and **therapeutics**.
- ✓ **Host-pathogen interactions** and the **human microbiome** present new opportunities
  - ✓ better diagnostics
  - ✓ therapeutics
  - ✓ disease management approaches
- ✓ Can aspects of microbiome be used to supplement during an episode (eg. refined fecal transplantation)

NIH recently awarded JCVI ~$25 million for continued research in this space; Gates ~ $4 million.

J. Craig Venter INSTITUTE

MICROBIOMES and CANCER

J. Craig Venter™
INSTITUTE

Microbial-Cancer Link Established

Viral

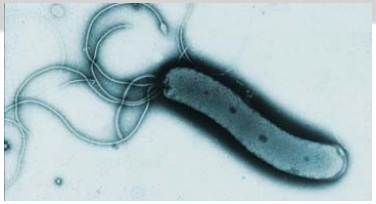
Papillomavirus – Cervical carcinoma
Hepatitis B and C viruses – Hepatocellular carcinoma
Epstein Barr virus – Lymphomas, nasopharyngeal carcinomas
Herpes virus 8 – Kaposi sarcoma
Human T lymphotrophic virus type 1 - Leukemia

Bacterial

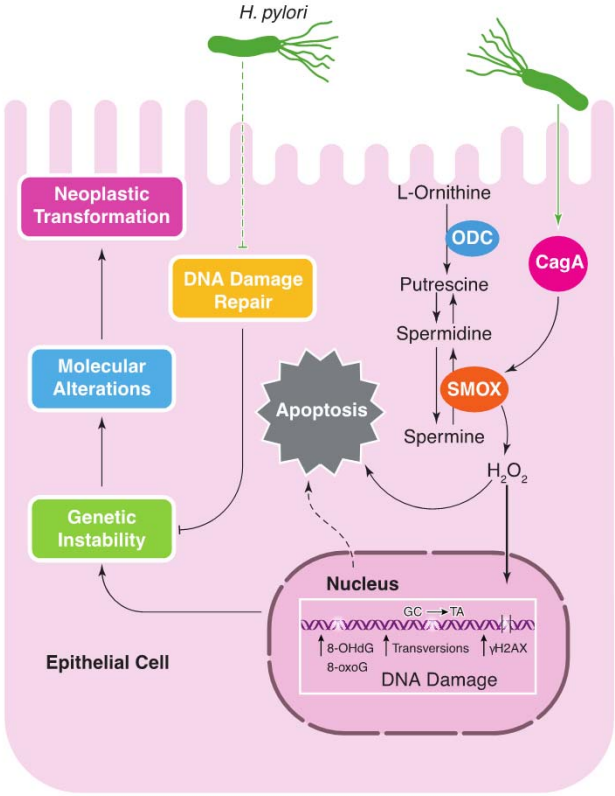
Helicobacter pylori

J. Craig Venter™ INSTITUTE

Helicobacter pylori Gastric Cancer

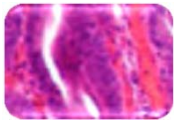


- Gram-negative bacterium, colonizes 50% of world population
- Bacterial urease permits survival in low pH of the stomach
- Organism elaborates several virulence factors that promote inflammation, DNA damage and leads to epithelial cell transformation

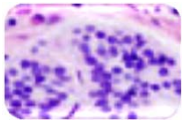


J. Craig Venter™ INSTITUTE

JCVI Microbiome-Cancer Projects



Progression of Esophageal Cancer
Karen E. Nelson, PI
Zhiheng Pei, NYU School of Medicine



Colon Cancer Pilot Study
Shibu Yooseph, PI
Hassan Brim & Hassan Ashktorab, Howard University

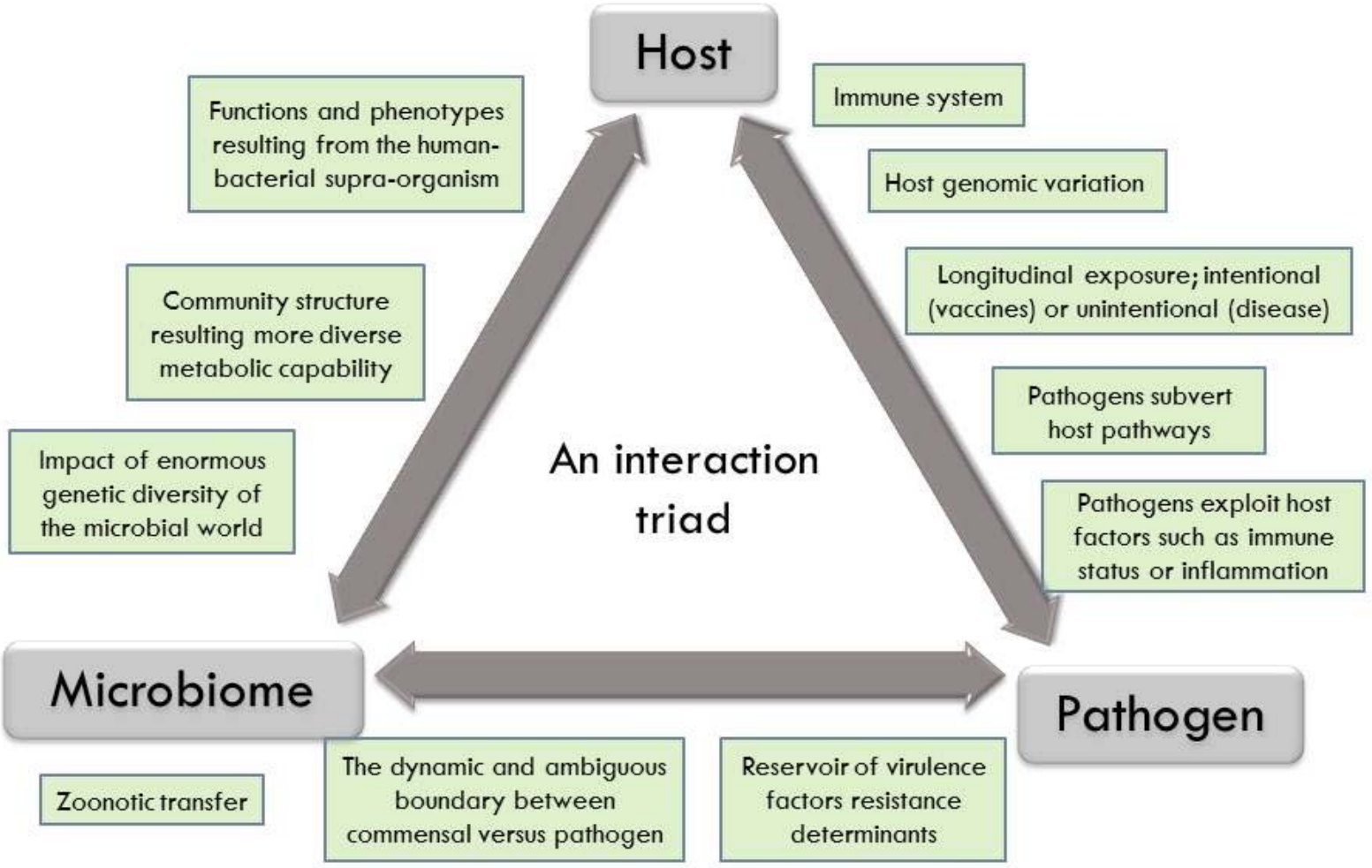
Microbial Inflammation in Pancreatic Cancer
Agnes Chang, PI
Tom Fishbein & Michael Zasloff, Georgetown University

J. Craig Venter™
INSTITUTE

The Human Respiratory Tract Microbiome in Health and Disease

J. Craig Venter™
INSTITUTE

The host-pathogen-microbiome disease paradigm

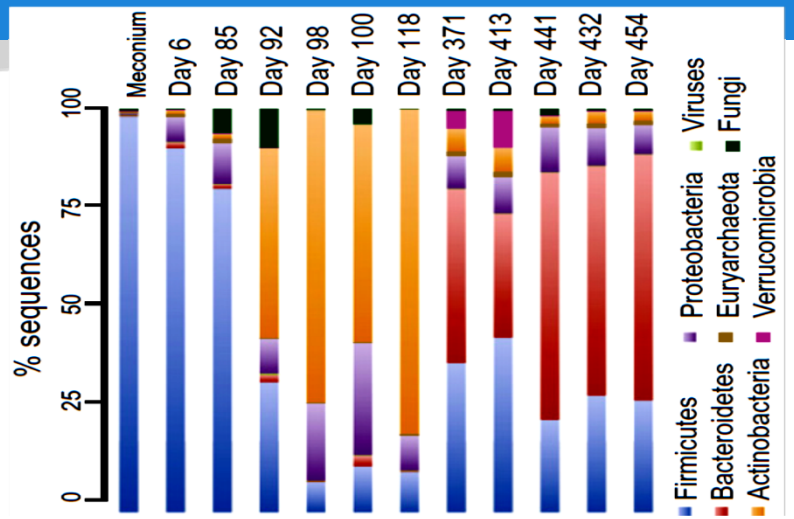


Slide provided by Claire Fraser

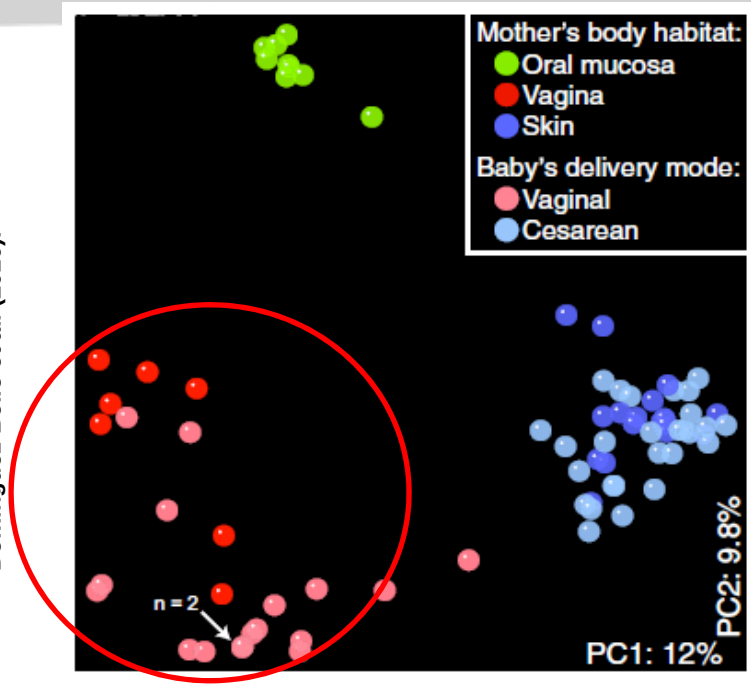
J. Craig Venter™ INSTITUTE

Microbiota are acquired from immediate environment at birth

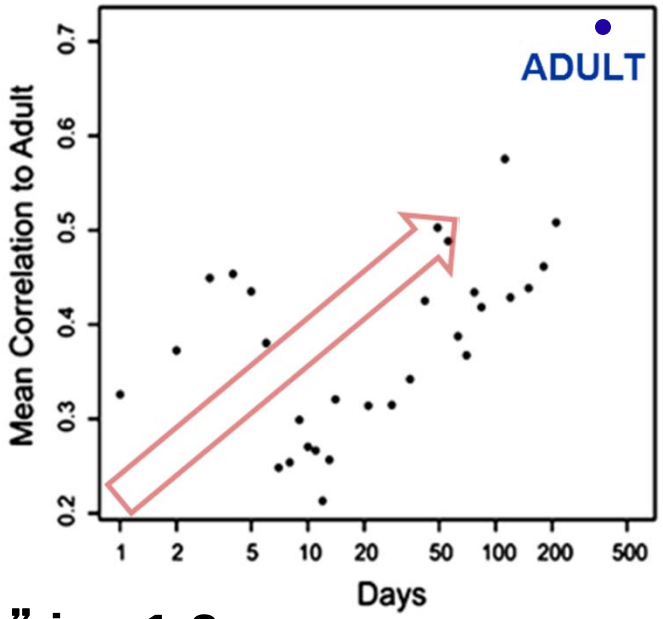
2) Microbial succession over ~1-2 yrs.



Koenig et al. (2010)



Dominguez-Bello et al. (2010).



Palmer et al. (2007)

1) Infants obtain microbes from mother or environment – mode of delivery impacts population structure

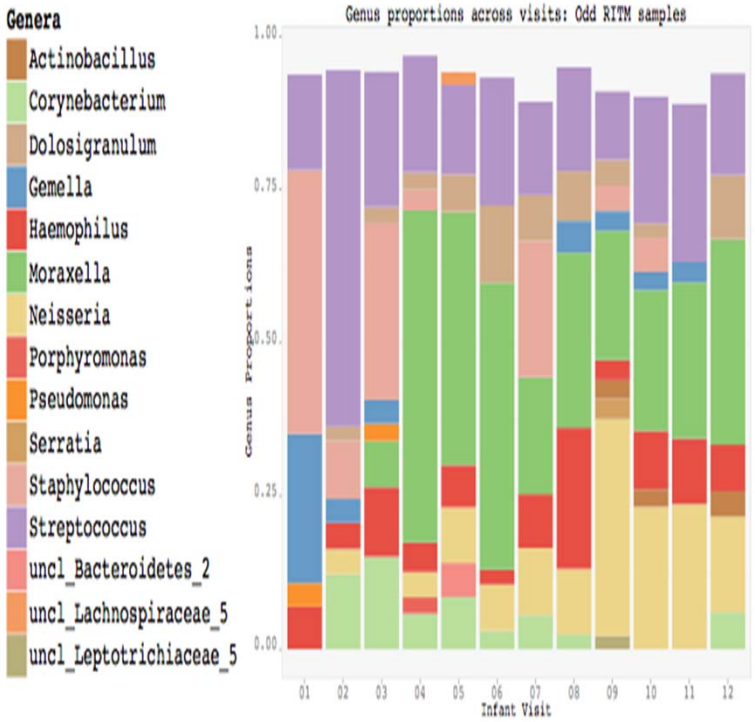
3) Microbiome becomes "adult-like" in ~1-2 yrs.

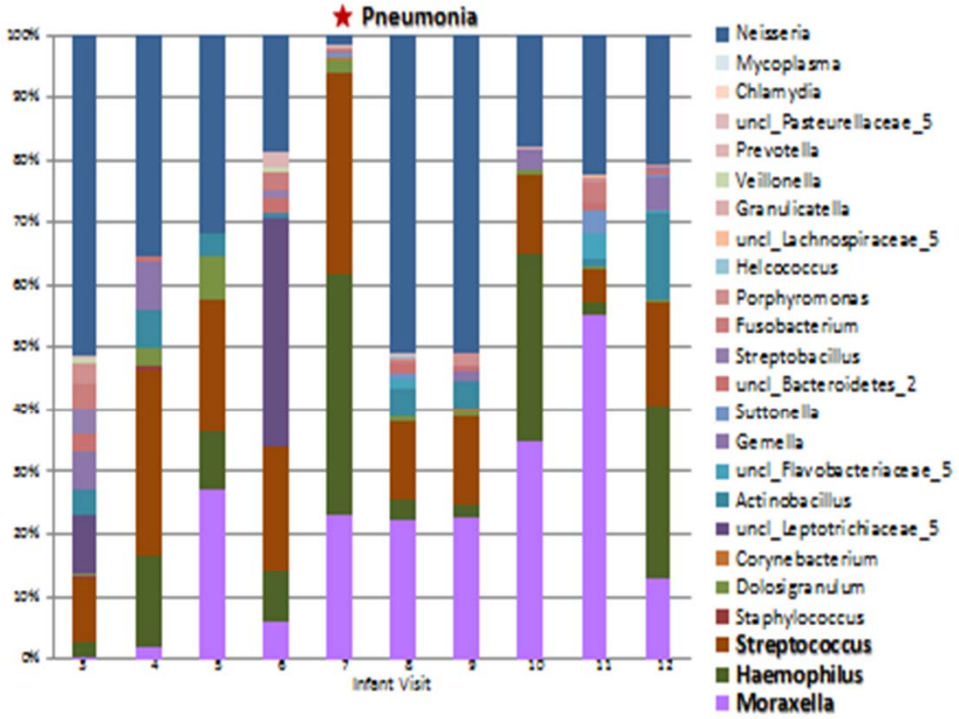
J. Craig Venter INSTITUTE

Respiratory Tract Microbiome

to better understand the effect of pneumonia vaccine on the infant nasopharyngeal microbiome

BILL & MELINDA GATES foundation



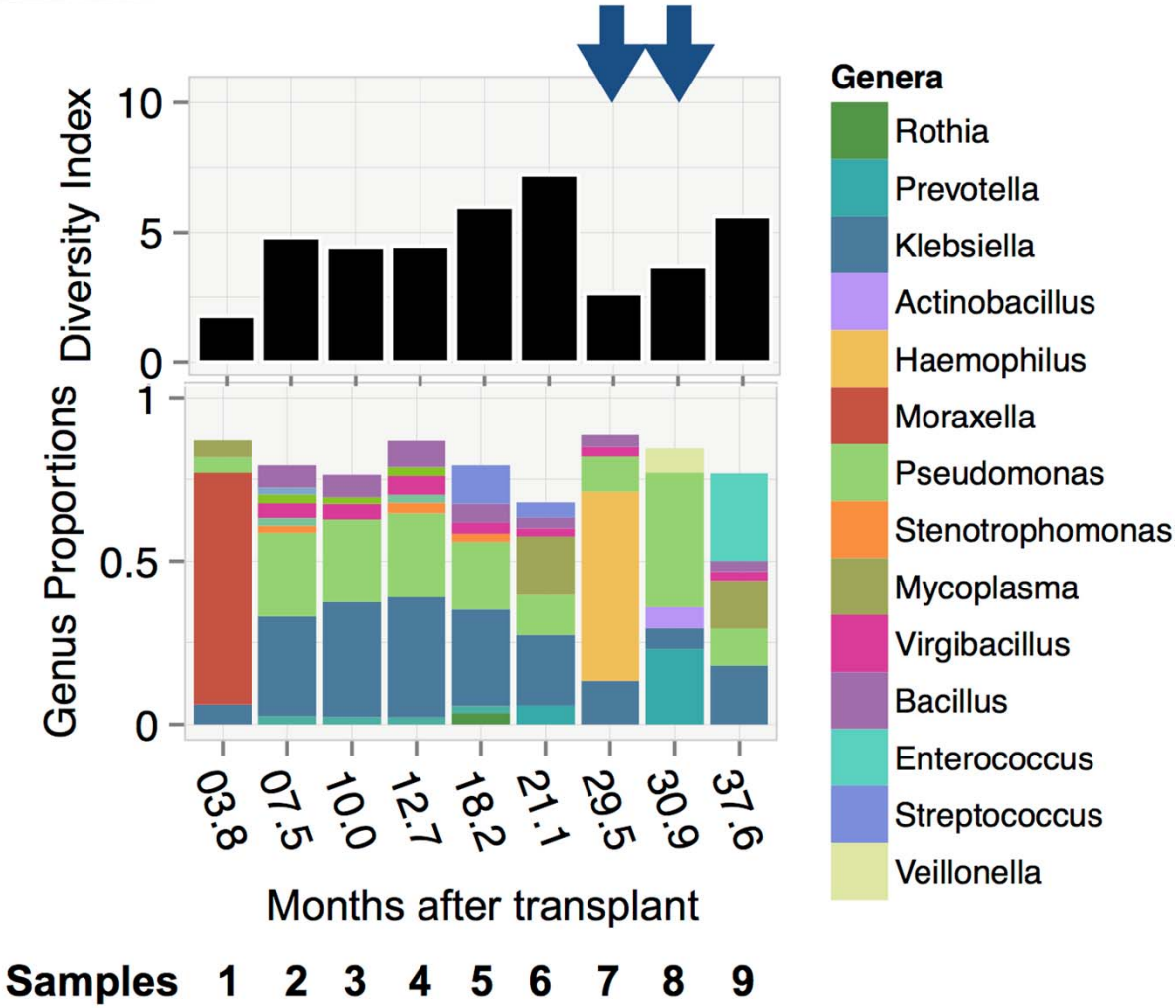


Infant with Pneumonia Episode at 7 Months of Age

Streptococcus pneumoniae in vaccinated infants

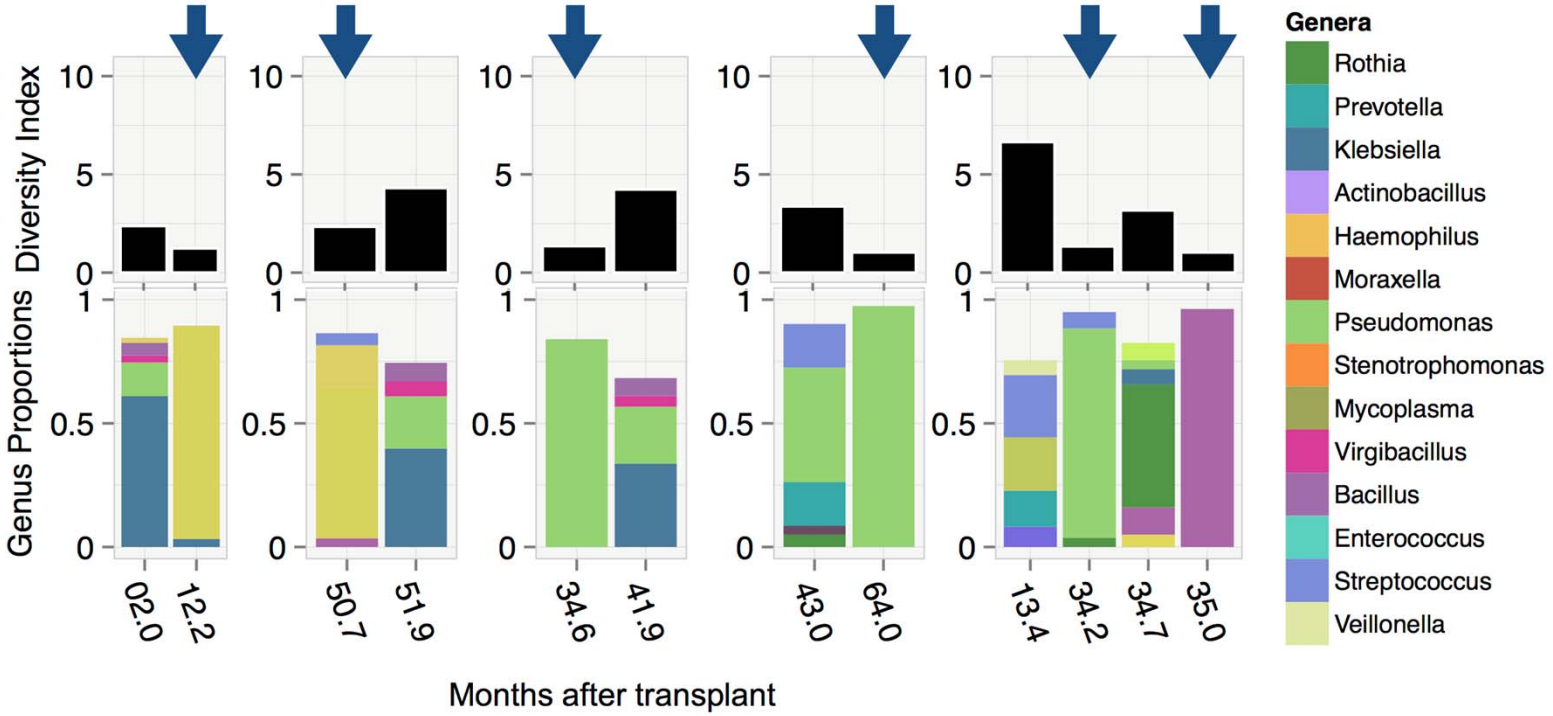
J. Craig Venter INSTITUTE

Lung Transplantation



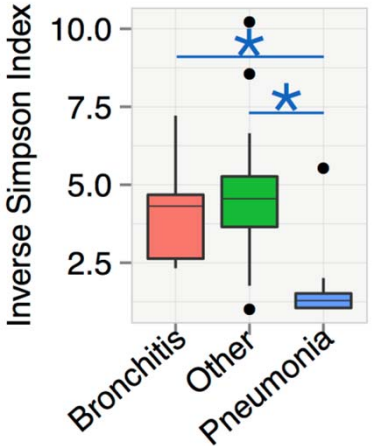
J. Craig Venter INSTITUTE

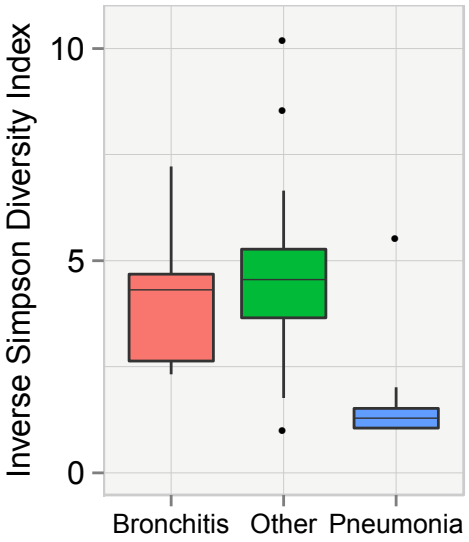
Lung Transplantation

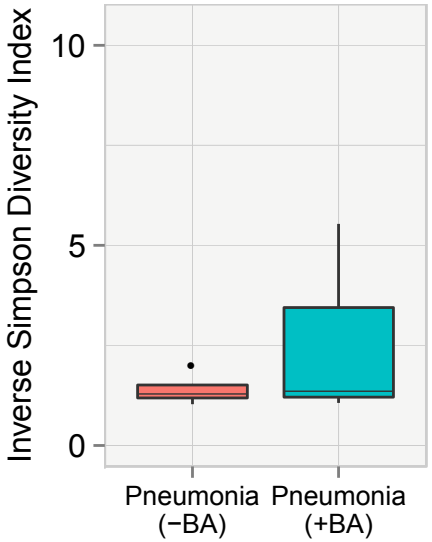


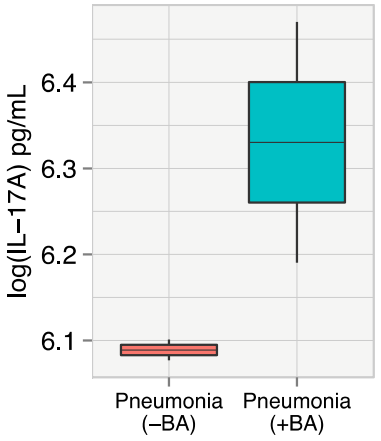
J. Craig Venter™
INSTITUTE

Lung Transplantation



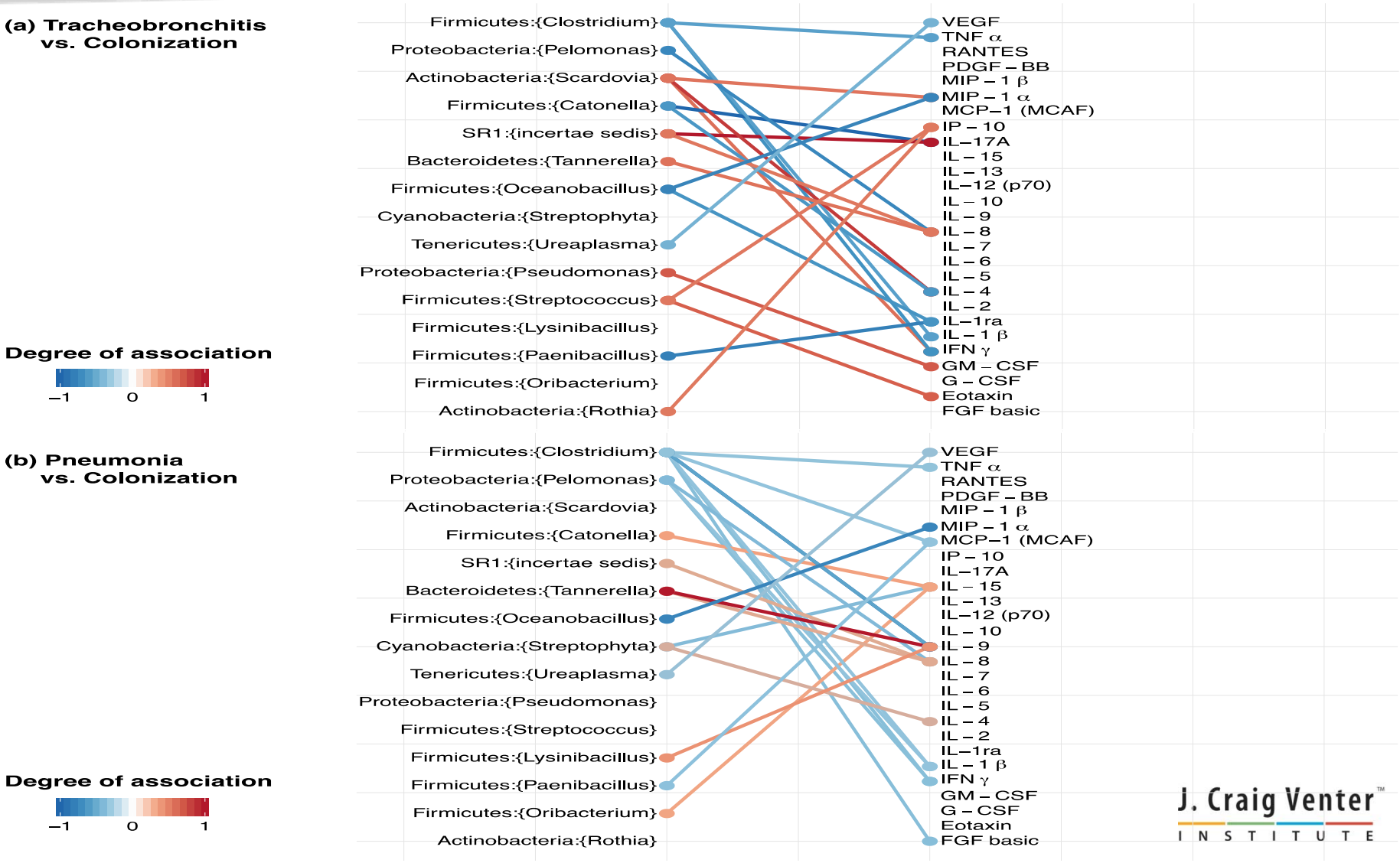






J. Craig Venter™ INSTITUTE

Cytokine-Microbiome Associations (BMA)



J. Craig Venter™ INSTITUTE

Respiratory Tract Microbiome Observations

- The adult respiratory tract microbiome is **stable and homogeneous** between the upper and lower tract.
- The infant and lung transplant respiratory tract microbiomes are **unstable** sample to sample compared to healthy adults.
- Infants in the first year of life in the Philippines are most susceptible to respiratory infections, particularly **pneumonia**.
- **Transplanted lungs** are susceptible to frequent infections.
- **The respiratory tract microbiome may have a role in protecting from or predisposing to respiratory infections**.

J. Craig Venter™ INSTITUTE

A Case Study in Microbiome Analysis

Translational Target

Candidiasis

Surrogate response

Risk Factors for GI Colonization

"High-dimensional" measurements

Known ▶ Antibiotics, Immunosuppression, Invasive Procedures, Colonization

Novel ▶ **Bacterial & Fungal Microbiome**

Key Questions

1. **Do all** GI bacteria, fungi and cytokines influence *C. albicans* colonization?

2. Are some **more important** for colonization than others?

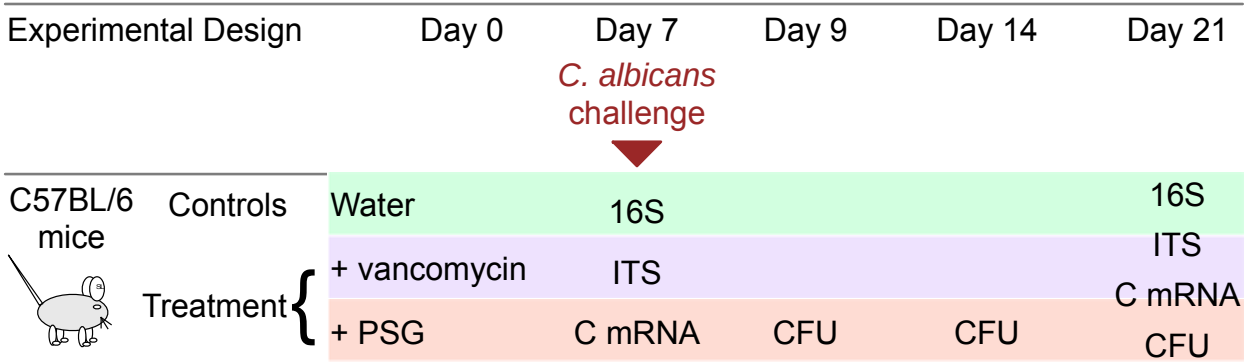
3. What are the **magnitudes and directions** of effects?

4. **How consistent** are these effects?

Are these patterns going to be **replicable, given a different slice** of the population?

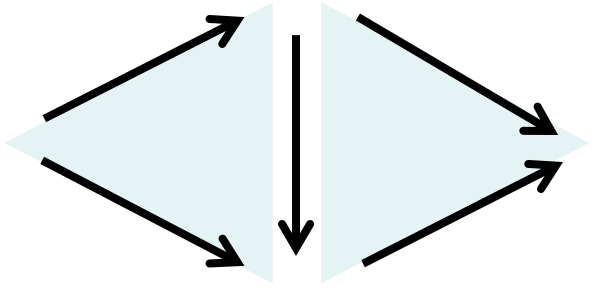
Study Design → Analysis Design

Preclinical mouse study



Bacterial & Fungal Microbiome

Antibiotic Treatment



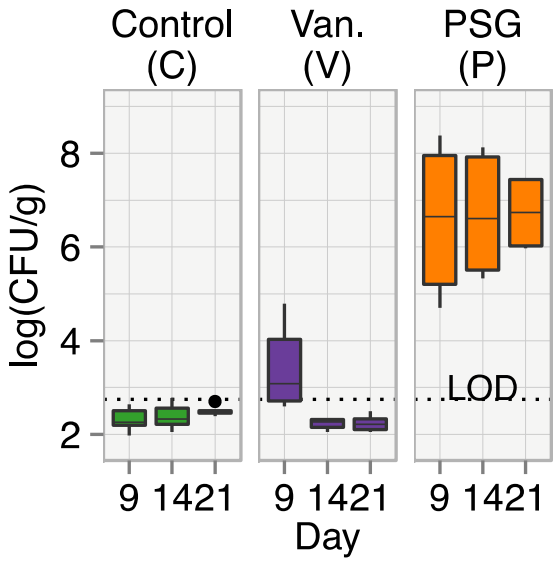
***Candida* Colonization**

Host Immune Response

In collaboration with Scott Filler

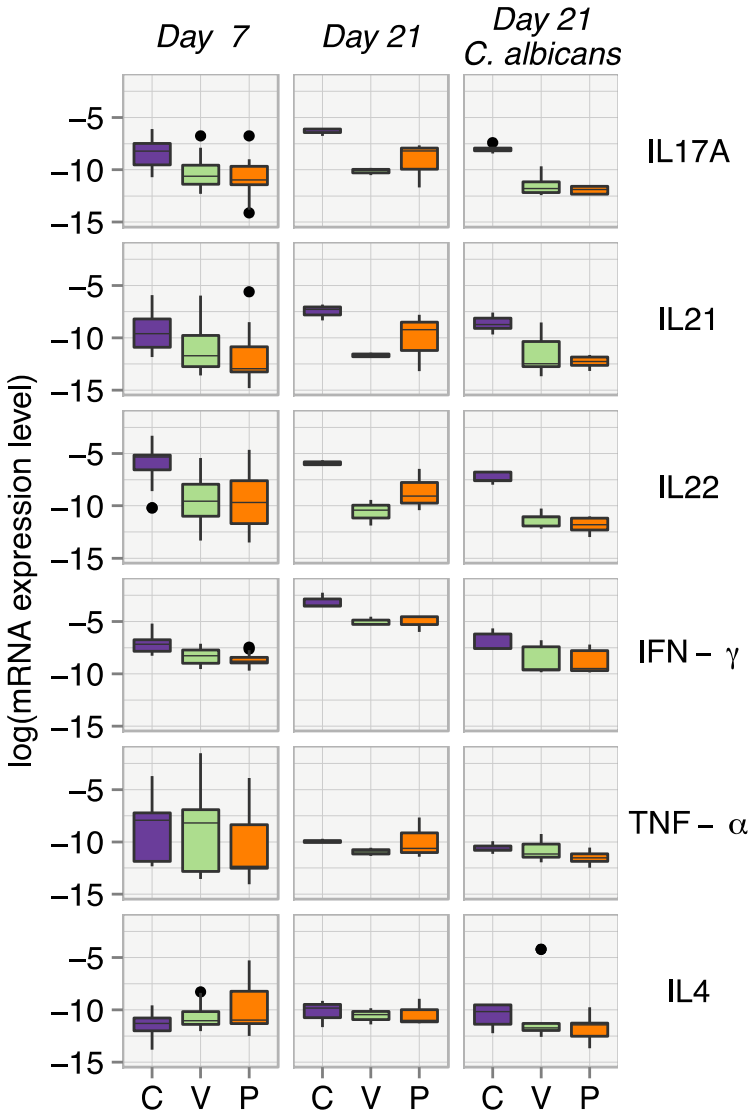
Exploring known facets: Antibiotics.

Antibiotic Treatment → ***Candida* Colonization**



PSG induces higher level of colonization than vancomycin

Exploring known facets: Immune response

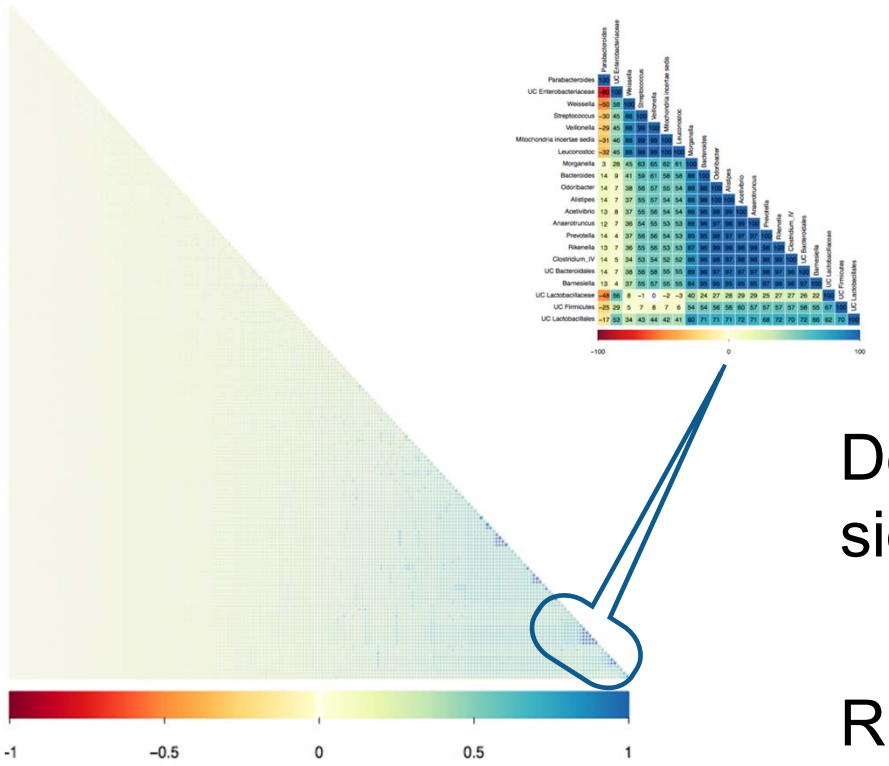


Antibiotic Treatment

Host Immune Response

Both PSG and vancomycin suppress Th17 and Th1 immune response

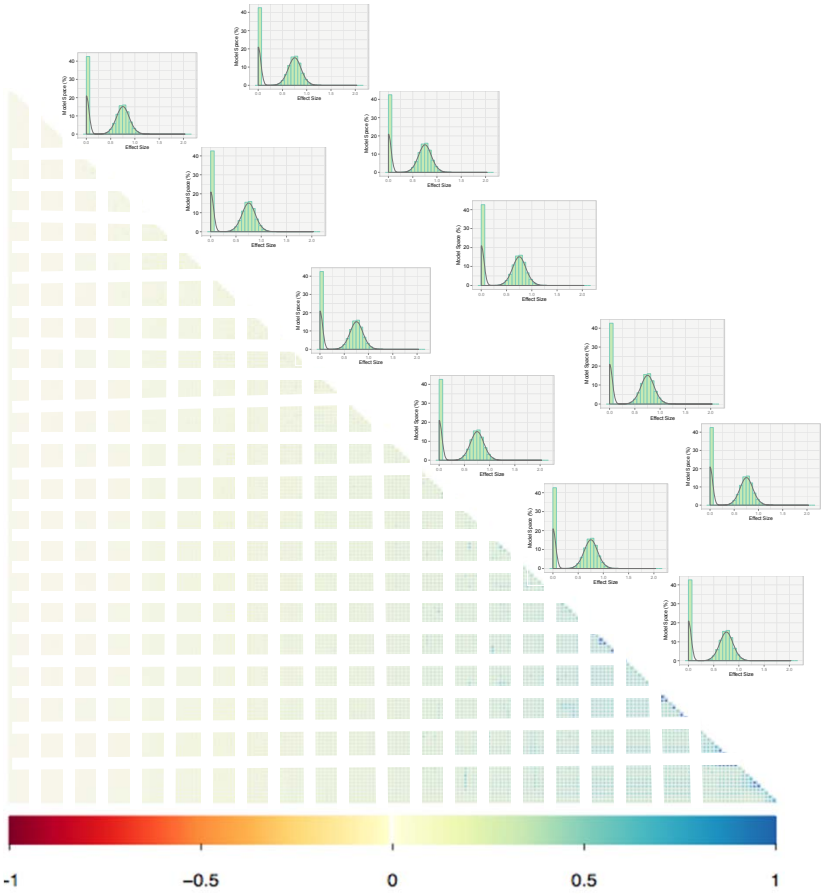
Microbiome Facet. Complex. High Dimensional



Detecting true and consistent signals is non-trivial.

Running univariate tests for every microbe leads to
-- Error accumulation
-- False signals

Ensemble Models: Explore Efficiently



Build ensemble of models

Each model learns from all the other models

Assign probabilities to signals

Rank these probabilities

Higher ranks: more likely true signals

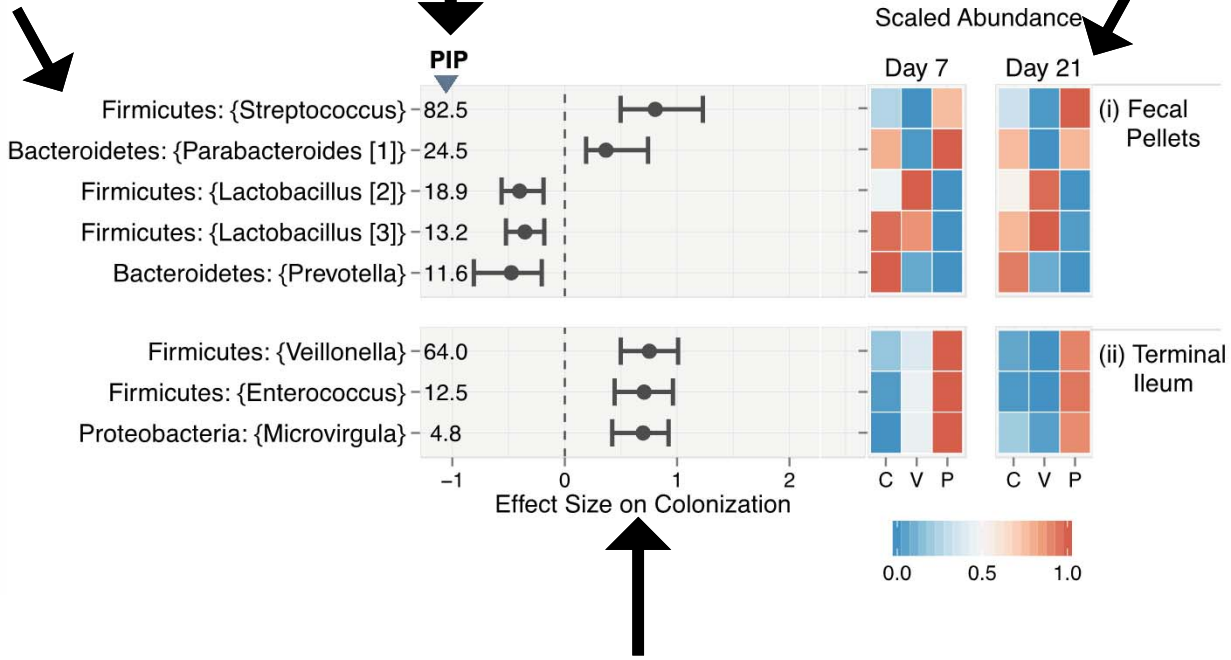
And do all of this simultaneously to avoid accumulating errors!

Ensemble Model Summaries

1. List of most influential microbes

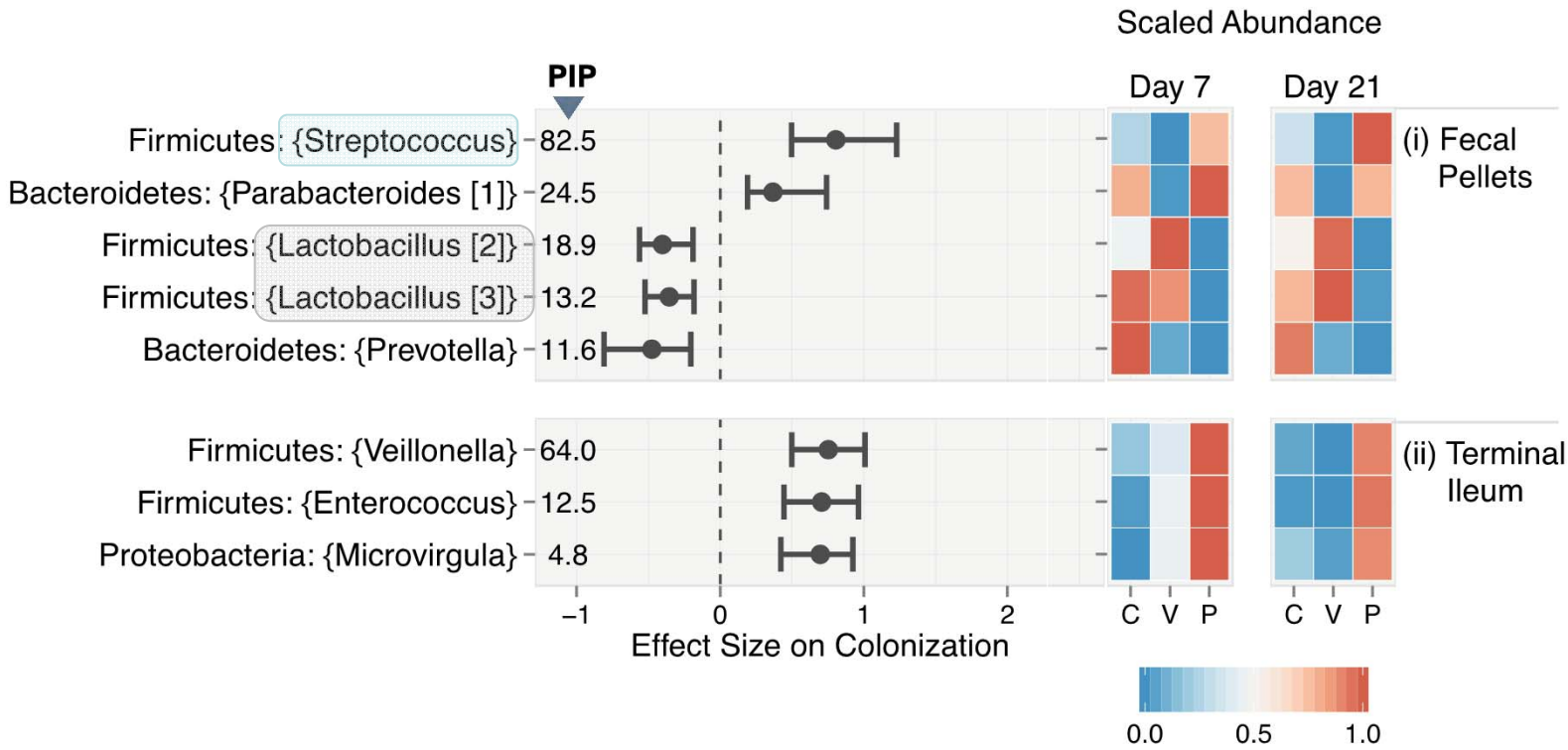
2. Ranked probabilities

4. How these influential microbes changed under PSG, vancomycin.



3. Magnitudes of effect on colonization

But what about biological relevance?



Innocent until proven guilty: mechanisms and roles of *Streptococcus–Candida* interactions in oral health and disease

H. Xu[1], H.F. Jenkinson[2] and A. Dongari-Bagtzoglou[1]

1 Division of Periodontology, School of Dental Medicine, University of Connecticut, Farmington, CT, USA
2 School of Oral and Dental Sciences, University of Bristol, Bristol, UK

Lactobacillus crispatus modulates epithelial cell defense against *Candida albicans* through Toll-like receptors 2 and 4, interleukin 8 and human β-defensins 2 and 3

Antonietta Rizzo, Antonio Losacco, Caterina Romano Carratelli*

Department of Experimental Medicine, Section of Microbiology and Clinical Microbiology, Faculty of Medicine and Surgery, Second University of Naples, Naples, Italy

Acknowledgments

NIH

GSC award from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900007C.

HMP project supported by Award Number U54 AI-084844-01 administered by the National Institute of Allergy and Infectious Diseases on behalf of the NIH Roadmap Human Microbiome Project.

NIDDK T1D: Investigation the Gut Microbiome, Urinary proteome and Metabolome under Grant Number: 1 DP3 DK 94343 – 01

Bill and Melinda Gates Foundation, Vaccination and the Infant Microbiome

JCVI
Karen Nelson
Manolito Torralba
Derrick Fouts
Barbara Methe
Shibu Yooseph
Sarah Highlander
Granger Sutton
Marcus Jones
Sarah Lucas
Indresh Singh
Derek Harkins
Jason Inman
Andrey T.
Rembert Pieper
John Glass
Sanjay Vashee
Hernan Lorenzi
Karla Stucker
Suman Das
Reed Shabman
Liliana Losada
Jyoti Shankar
Stephanie Mounaud

Collaborators
Bryan White (Illinois)
Brenda Wilson (Illinois)
David Brenner (UCSD)
Bernd Schnabl (UCSD)
Zhipeng Pei (NYU)
Scott Filler (UCLA Biomed)
Neal Clancy (U Pitt)
Hong Nguyen (U Pitt)

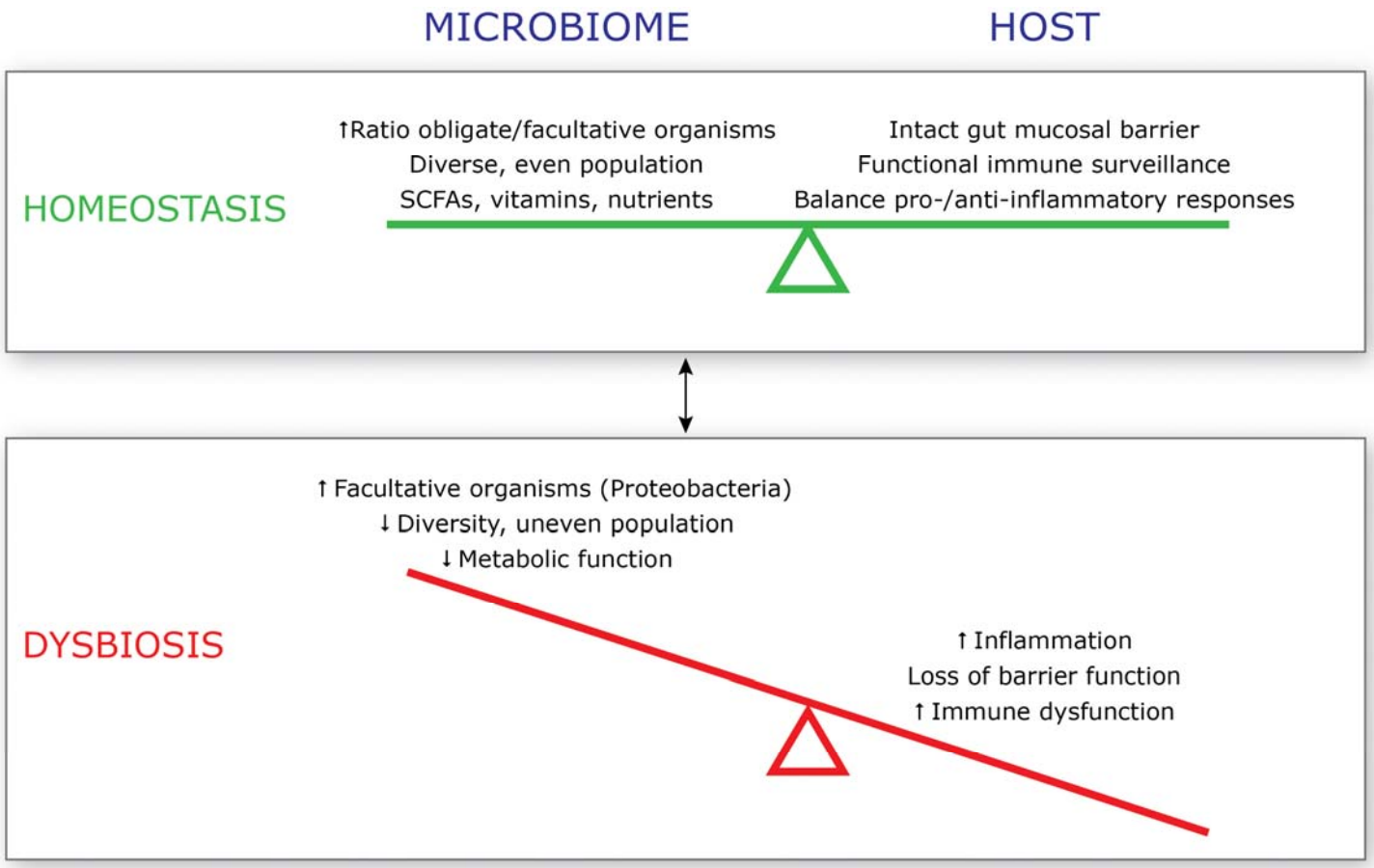
Others
Maria Giovanni (NIAID)

J. Craig Venter™
INSTITUTE

Dysbioses of the GI Tract

J. Craig Venter™
INSTITUTE

Gut Dysbiosis



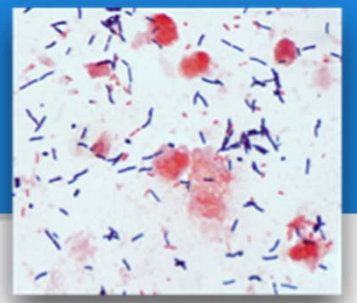
J. Craig Venter™ INSTITUTE

Restoration of a Disrupted (Dysbiotic) GI "Ecosystem"

Lozupone, *et al*. Nature *489* 2012

J. Craig Venter INSTITUTE

Clostridium difficile



Gram negative, sporulating, obligate anaerobe

Member of the normal human gut flora

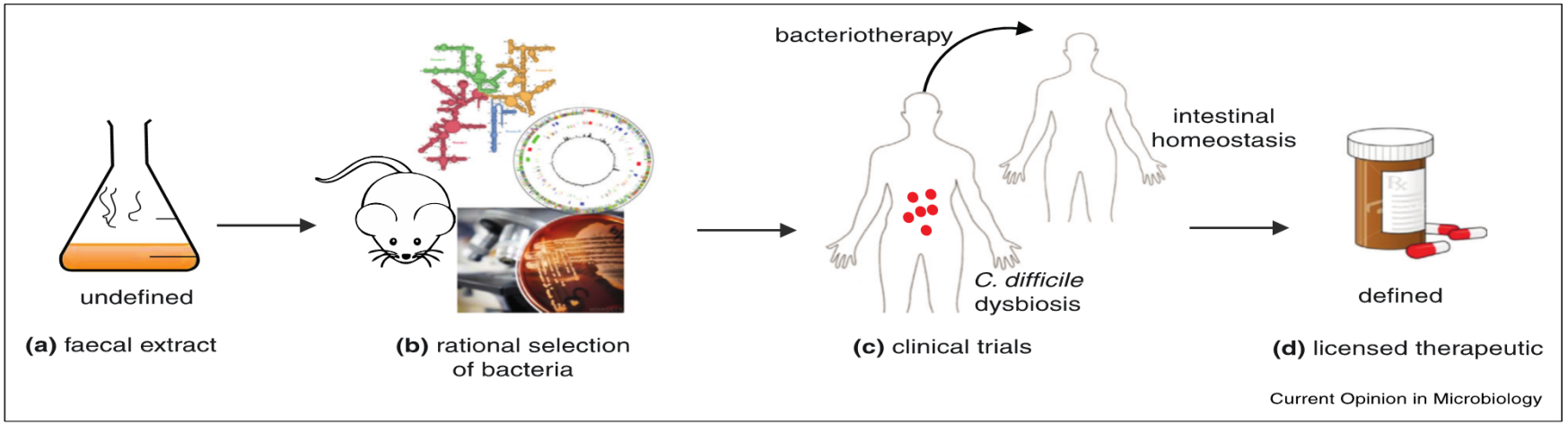
Strains produce cytotoxins, TcdA and TcdB

Major cause of antibiotic-associated diarrhea (>500,000 cases/yr), especially in the elderly, hospitalized, nursing homes

Antibiotic-resistant recurrence increasingly common

J. Craig Venter™ INSTITUTE

First Report of Fecal Bacteriotherapy – return to early studies



THE LANCET, MAY 27, 1989

BACTERIOTHERAPY FOR CHRONIC RELAPSING CLOSTRIDIUM DIFFICILE DIARRHOEA IN SIX PATIENTS

M. TVEDE[1] J. RASK-MADSEN[2]

Department of Clinical Microbiology, Rigshospitalet, Statens Seruminstitut;[1] and Section of Gastroenterology, Department of Medicine G, Bispebjerg Hospital, University of Copenhagen, Denmark[2]

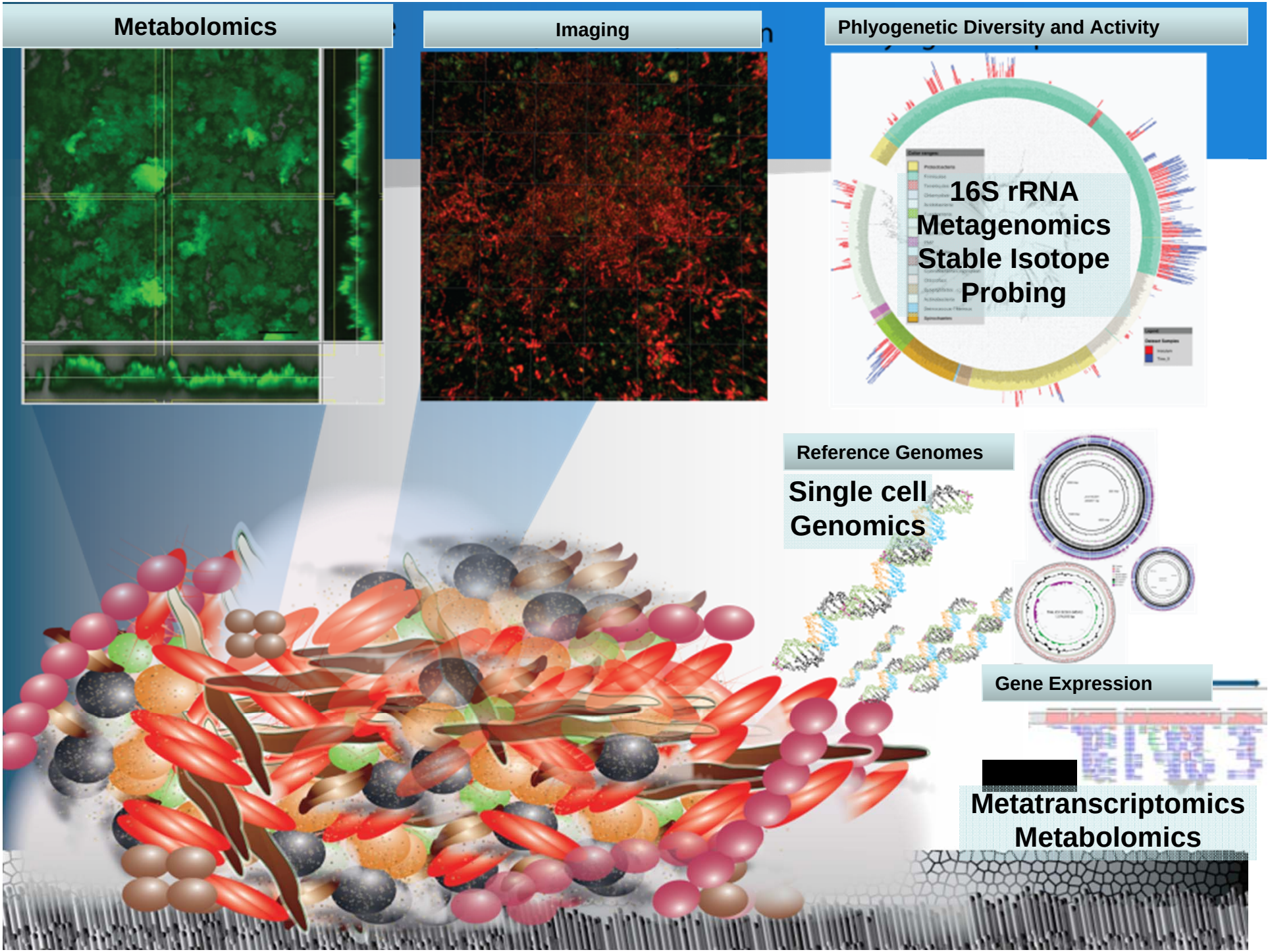
- ✓ *Blautia producta*
- ✓ *Clostridium bifermentans*
- ✓ *Clostridium innocuum*
- ✓ *Clostridium ramosum*
- ✓ *Enterococcus faecalis*
- ✓ *Bacteroides ovatus*
- ✓ *Bacteroides thetaiotaomicron*
- ✓ *Bacteroides vulgatus*
- ✓ *Escherichia coli* (2 strains)

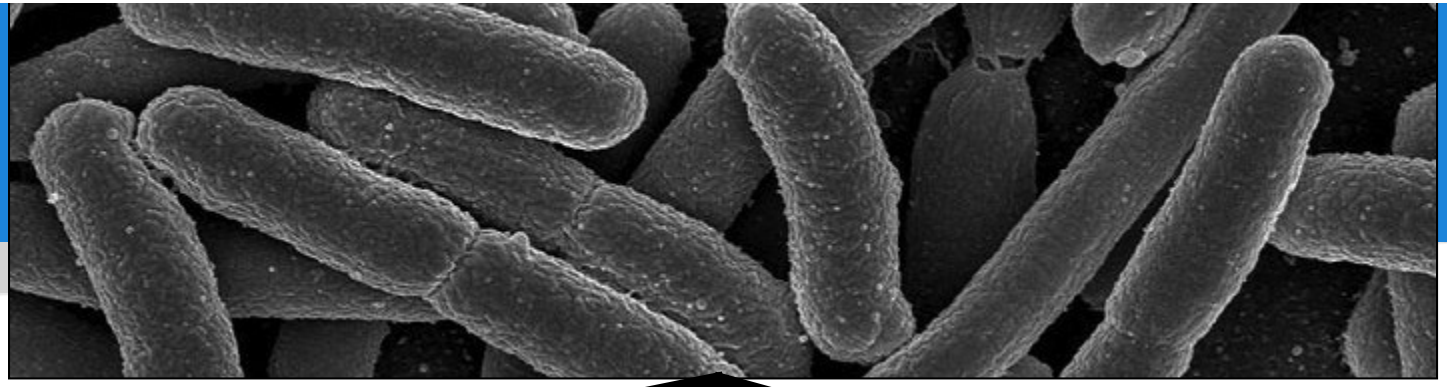
- ✓ 5 patients, >59 yo with relapsing CDI
- ✓ treated, by enema, with mixture of 10 bacterial strains
- ✓ normal bowel function within 24 h, *C. difficile* negative within 7 d

Slide courtesy of Dr. Sarah Highlander, JCVI

Understanding Microbial Communities

J. Craig Venter™
INSTITUTE


Metabolomics
Imaging
Phlyogenetic Diversity and Activity
16S rRNA
Metagenomics
Stable Isotope
Probing
Reference Genomes
Single cell
Genomics
Gene Expression
Metatranscriptomics
Metabolomics



16S rRNA sequencing

Taxonomic Content

Metagenomic sequencing

Functional predictions based on gene content

Long Reads + Better Assembly = group genes by genome

Metatranscriptomic sequencing

Functional predictions based on gene expression

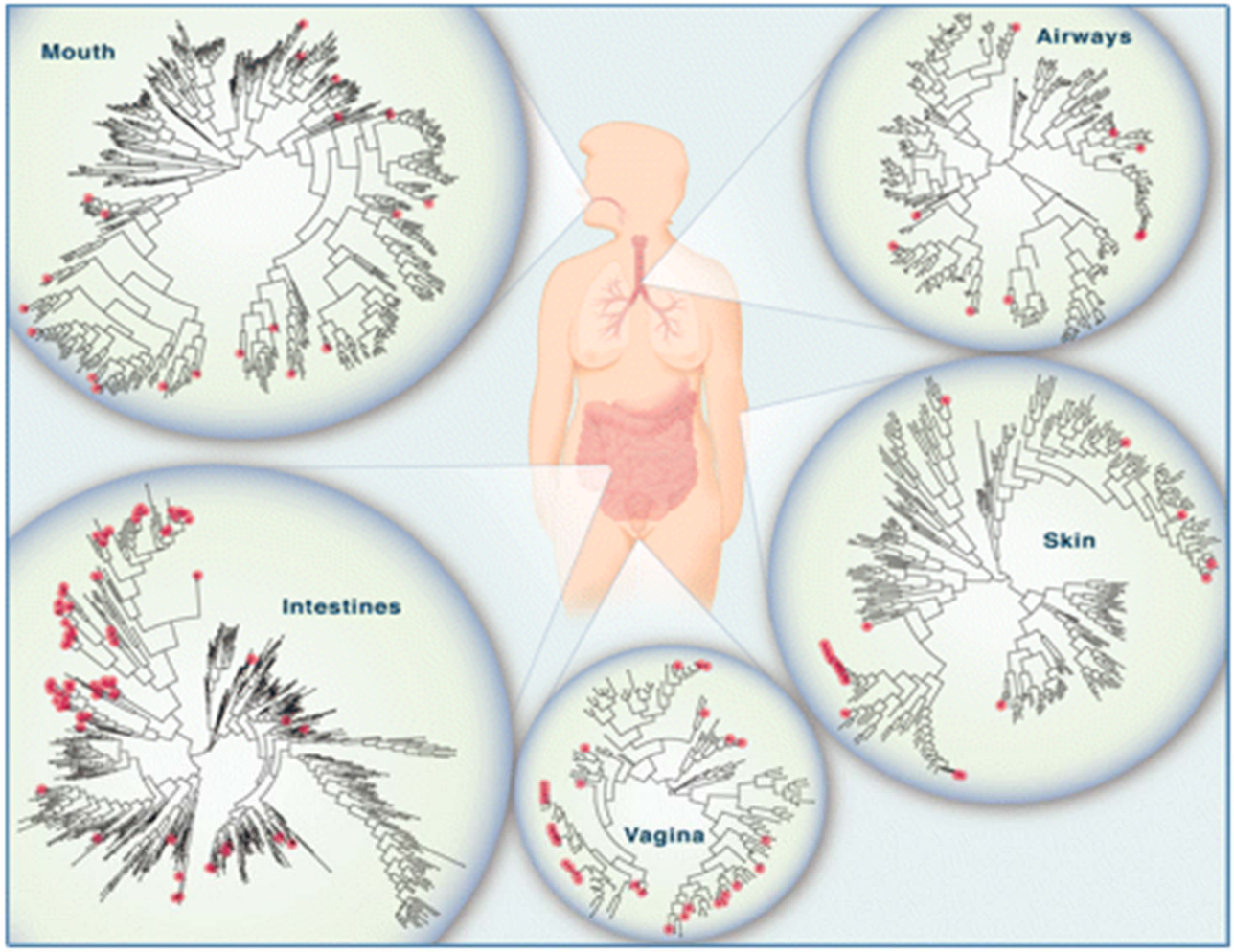
Other 'Omics
Proteome
Glycome
etc

J. Craig Venter™ INSTITUTE

WHAT DO WE KNOW?

J. Craig Venter™
INSTITUTE

Each body site has site-specific microbiome



J. Craig Venter™
INSTITUTE

Few Mechanistic Correlations

Most at the single species level

Gastric cancer microbiome:

✓Pertubation of the colonic epithelium by toxin-producing strains may increase the risk of developing malignancies.

✓Value associated with monitoring microbiota, circulating metabolites and host biomarkers (genome) in healthy and diseased individuals.

✓Microbial populations can be used as indicators of disease progression (based on 16S rRNA analysis). See J. P. Zackular, *Cancer Prevention Research*, 2014.

J. Craig Venter™
INSTITUTE

Several Animal Model studies of Microbiome Associations

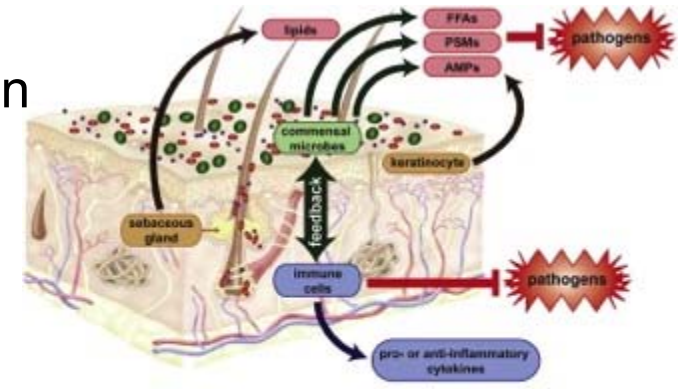
Mouse models

- liver cirrhosis
- obesity
- autism
- stress
- non-alcoholic fatty liver disease

J. Craig Venter™ INSTITUTE

For Example Skin

- ✓ Healthy skin microbiome - 1 million bacteria inhabit each cm^2 of skin
- ✓ ~100-200 different species reside on skin surface and in deep layers of skin
- ✓ Imbalances in the normal ecosystem cause skin pathology
  - ✓ Contribute to non infectious diseases including atopic dermatitis, psoriasis, rosacea (impacts 3% worldwide), and acne (affects 85% of teenagers) – *P. acnes* population structures are different healthy vs. acne cohorts
- ✓ Immunocompromised individuals have altered skin microbiomes with increased colonization by pathogenic bacteria and fungi
- ✓ Novel approaches and therapies can utilize this information



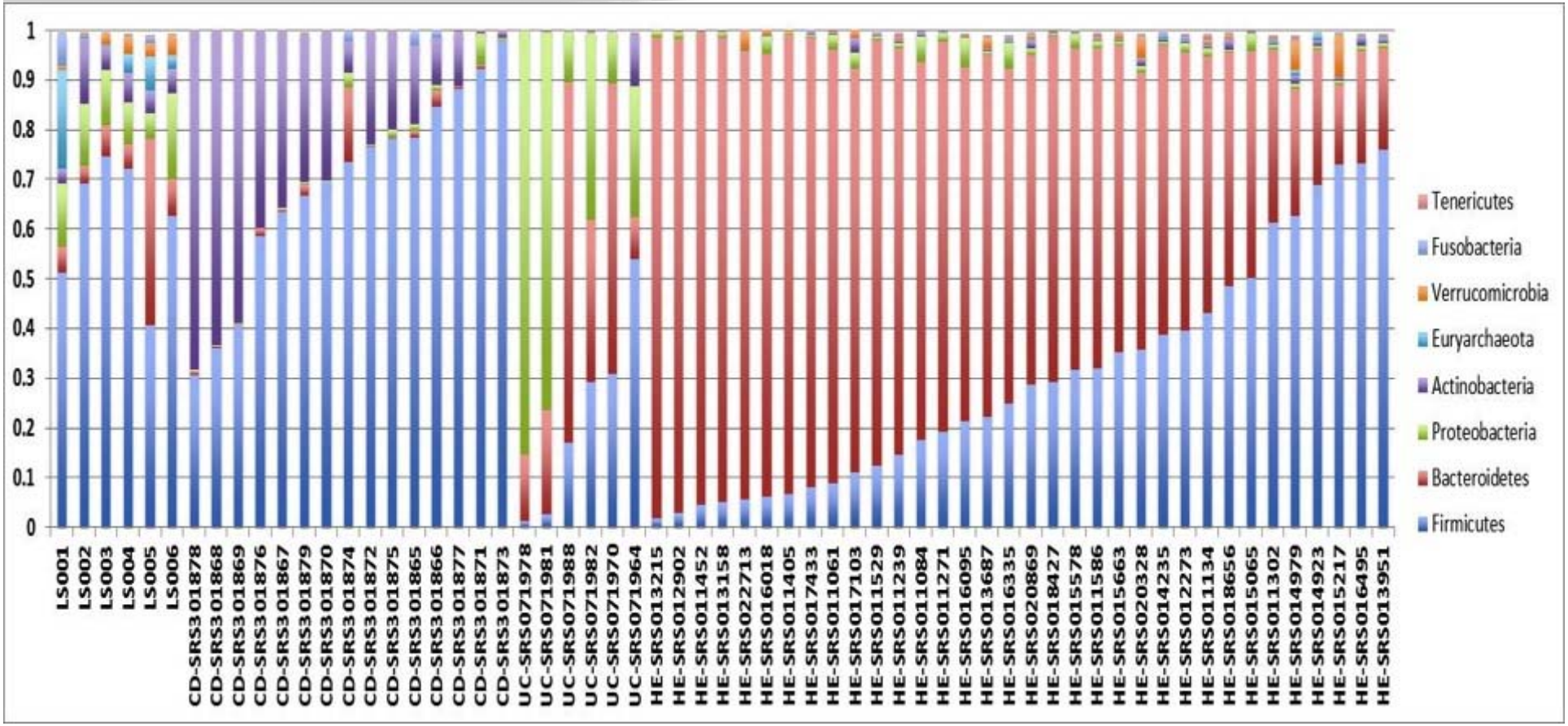
J. Craig Venter INSTITUTE

Chen, et al. J Am Acad Dermatol;69(1):143-155.e3.

Opportunities for Individualized Medicine

J. Craig Venter™
INSTITUTE

Phyla Gut Microbial Abundance LS, Crohn's, Ulcerative Colitis, and Healthy Subjects



Toward Noninvasive Microbial Ecology Diagnostics

J. Craig Venter™ INSTITUTE

Bottom Line

Significant Opportunities in Microbiome Space

J. Craig Venter™
INSTITUTE

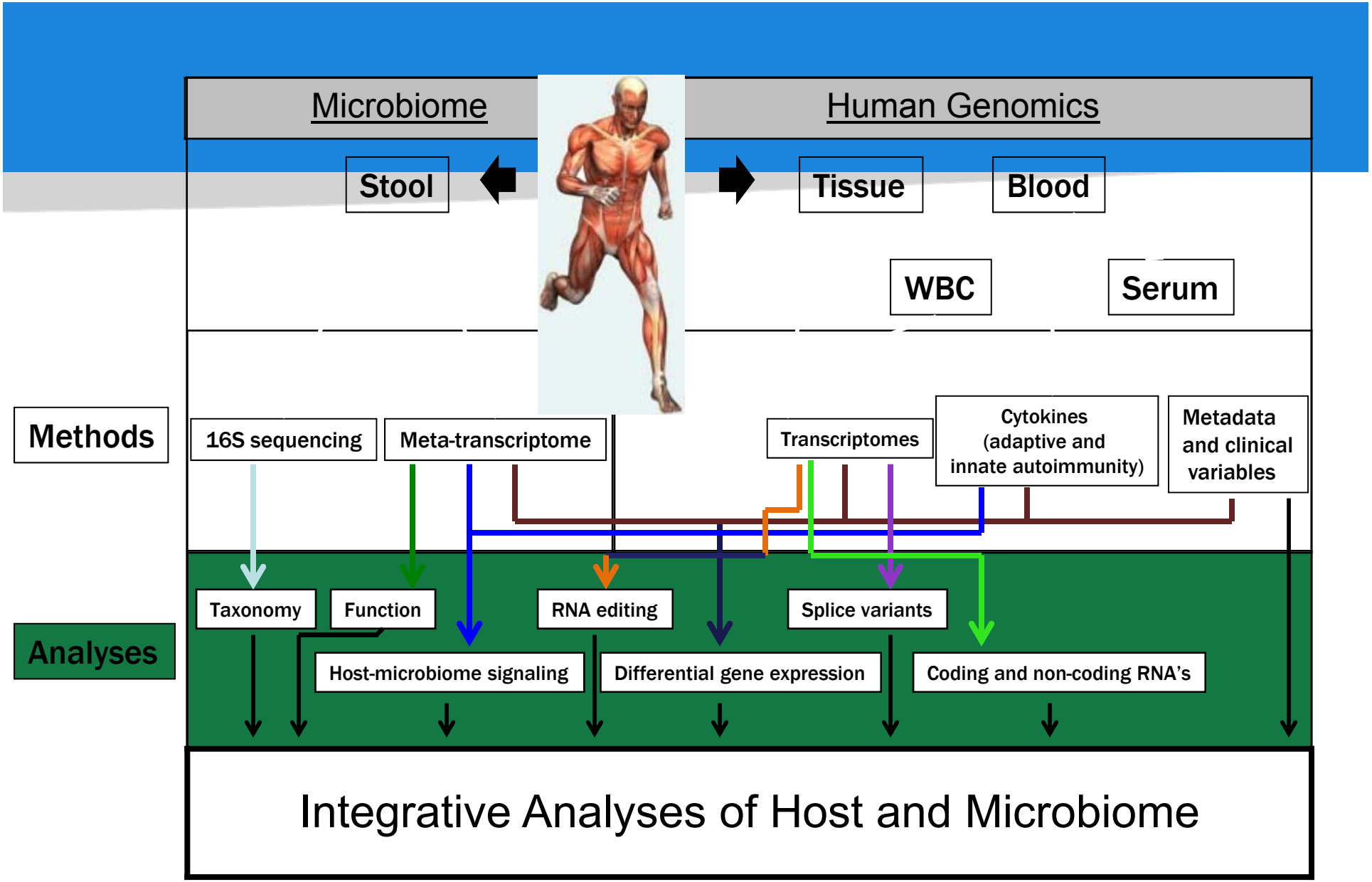
- ✓ ***Microbiome Therapeutics*** (formerly NuMe) – providing specialized nutrition that augments the growth of targeted desirable bacteria.
- ✓ ***Second Genome*** - discovery and development of therapeutic products – sequencing based
- ✓ ***Enterome*** – Microbiome biomarkers for IBD and NAFLD
- ✓ ***Metabiomics***
- ✓ ***Osel*** – women's health and gastrointestinal disorders
- ✓ ***Rebiotix*** - FMT for *C. difficile*
- ✓ ***CIPAC*** – as above
- ✓ ***Redanta*** – focused on microbial based modulation of pathways of interaction between the human microbiome and the host immune system
- ✓ ***ViThera Pharmaceuticals*** – bacteria therapeutic molecules for IBD
- ✓ ***Enterologics*** - Probiotics as live biotherapeutics
- ✓ ***AOBiome*** – science of Ammonia Oxidizing bacteria
- ✓ ***Synlogic*** – Synthetic microbiomes
- ✓ ***Ceres Health***

enterome
bioscience

Metabiomics

µBiome

qb3 California Institute for Quantitative Biomedical Research

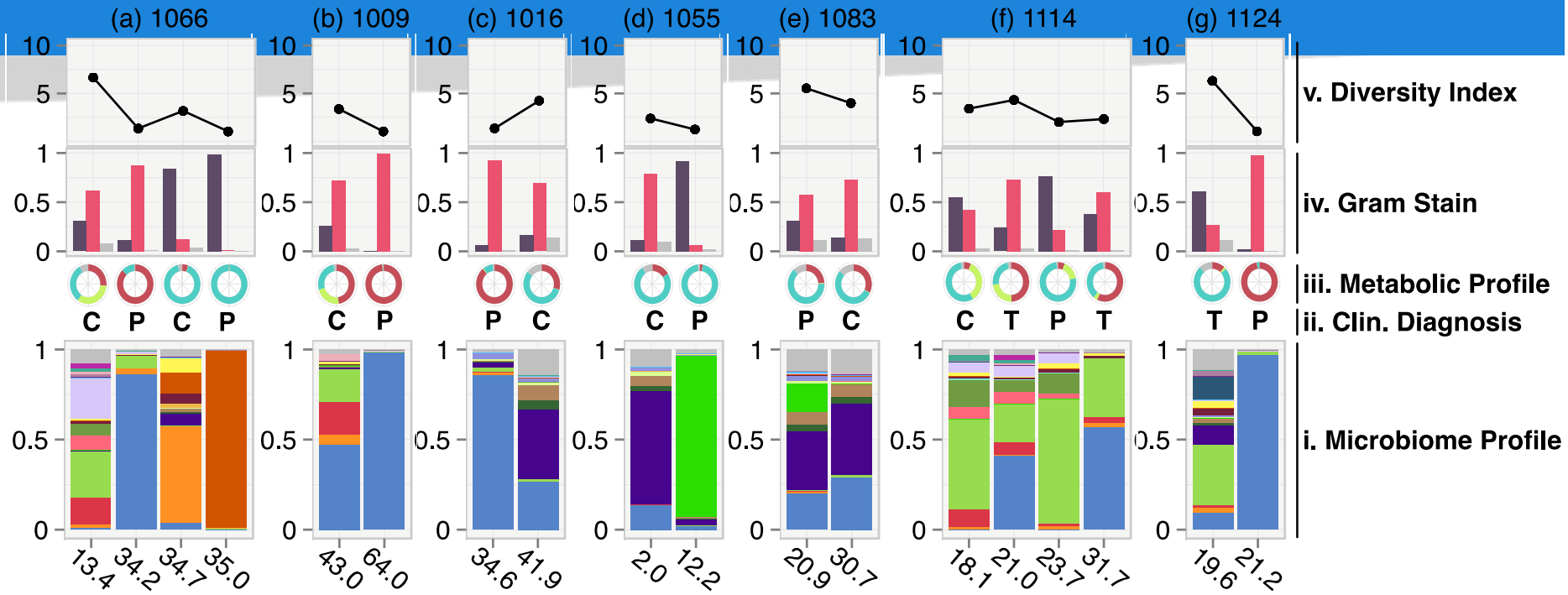


J. Craig Venter INSTITUTE

Characterizing acute infectious events post lung-transplantation

J. Craig Venter™
INSTITUTE

Patients with at least one diagnosis of Pneumonia



Genera

- *Actinobacteria:{Actinomyces}*
- *Actinobacteria:{Corynebacterium}*
- *Actinobacteria:{Rothia}*
- *Bacteroidetes:{Paraprevotella}*
- *Bacteroidetes:{Porphyromonas}*
- *Bacteroidetes:{Prevotella}*
- *Firmicutes:{Bacillus}*
- *Firmicutes:{Enterococcus}*
- *Firmicutes:{Granulicatella}*
- *Firmicutes:{Lactobacillus}*
- *Firmicutes:{Staphylococcus}*
- *Firmicutes:{Streptococcus}*
- *Firmicutes:{Veillonella}*
- *Firmicutes:{Virgibacillus}*
- *Fusobacteria:{Fusobacterium}*
- *Fusobacteria:{Leptotrichia}*
- *Other taxa (<5%)*
- *Proteobacteria:{Actinobacillus}*
- *Proteobacteria:{Campylobacter}*
- *Proteobacteria:{Haemophilus}*
- *Proteobacteria:{Klebsiella}*
- *Proteobacteria:{Moraxella}*
- *Proteobacteria:{Neisseria}*
- *Proteobacteria:{Pseudomonas}*
- *Proteobacteria:{Stenotrophomonas}*
- *Tenericutes:{Mycoplasma}*
- *Tenericutes:{Ureaplasma}*

Gram Stain

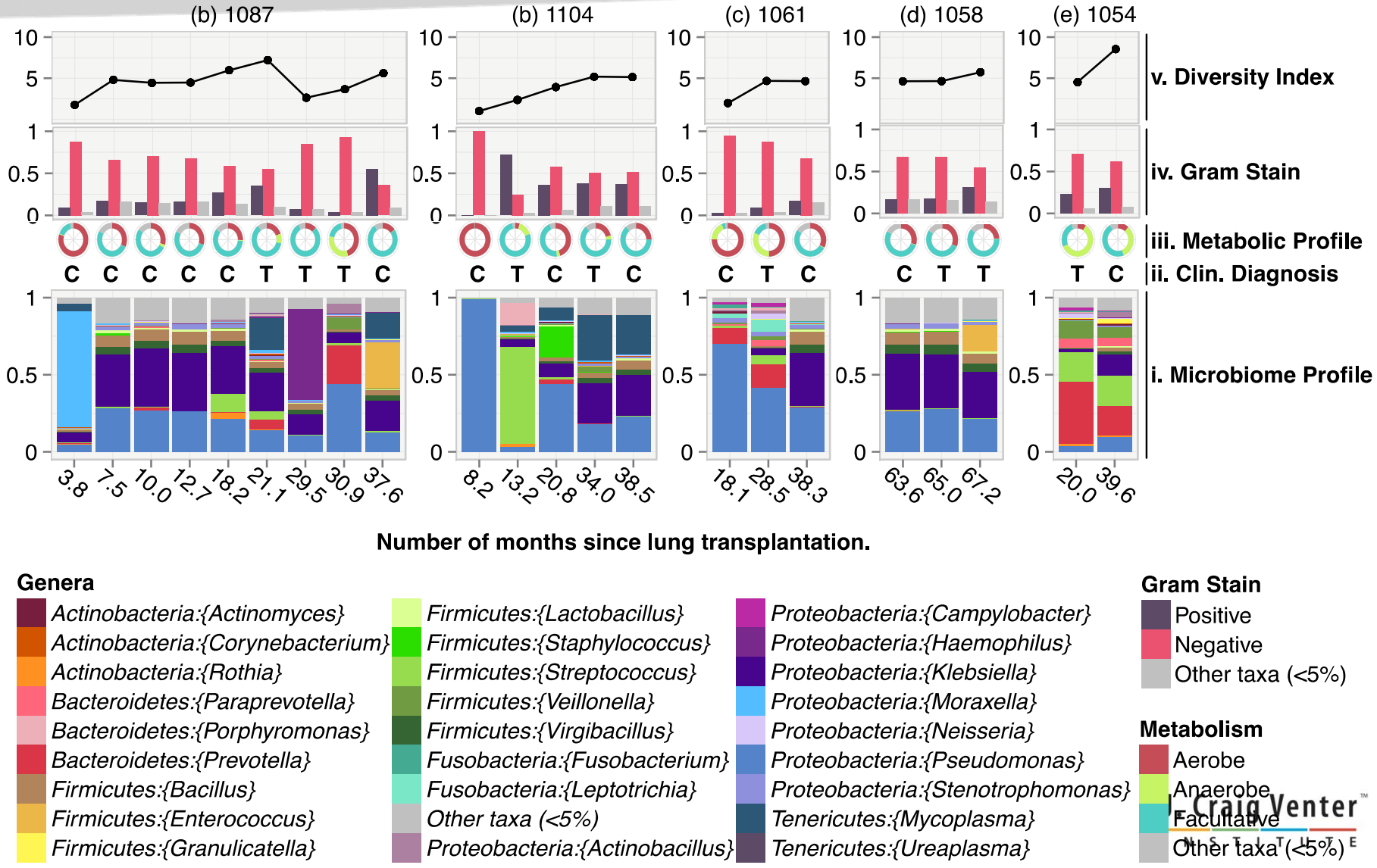
- Positive
- Negative
- Other taxa (<5%)

Metabolism

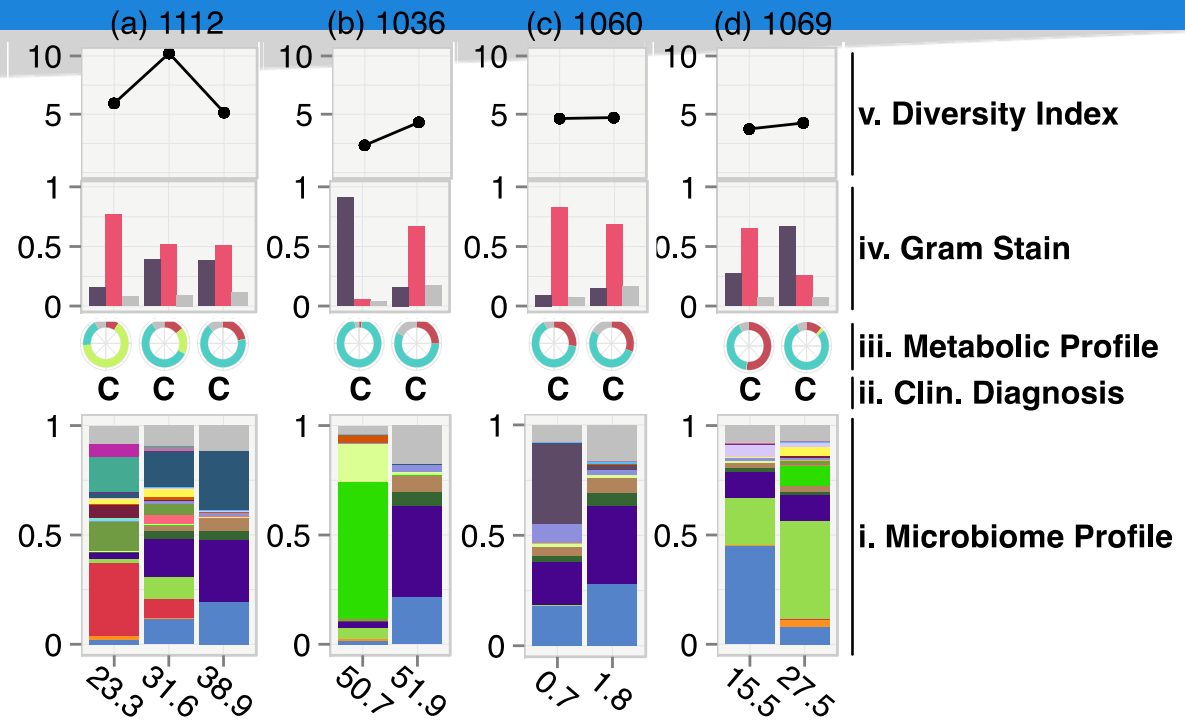
- Aerobe
- Anaerobe
- Facultative
- Other taxa (<5%)

J. Craig Venter INSTITUTE

Patients with at least one diagnosis of Tracheobronchitis



Patients without a concomitant acute infectious event.



Number of months since lung transplantation.

Genera

- *Actinobacteria:{Actinomyces}*
- *Actinobacteria:{Corynebacterium}*
- *Actinobacteria:{Rothia}*
- *Bacteroidetes:{Paraprevotella}*
- *Bacteroidetes:{Porphyromonas}*
- *Bacteroidetes:{Prevotella}*
- *Firmicutes:{Bacillus}*
- *Firmicutes:{Enterococcus}*
- *Firmicutes:{Granulicatella}*
- *Firmicutes:{Lactobacillus}*
- *Firmicutes:{Staphylococcus}*
- *Firmicutes:{Streptococcus}*
- *Firmicutes:{Veillonella}*
- *Firmicutes:{Virgibacillus}*
- *Fusobacteria:{Fusobacterium}*
- *Fusobacteria:{Leptotrichia}*
- *Other taxa (<5%)*
- *Proteobacteria:{Actinobacillus}*
- *Proteobacteria:{Campylobacter}*
- *Proteobacteria:{Haemophilus}*
- *Proteobacteria:{Klebsiella}*
- *Proteobacteria:{Moraxella}*
- *Proteobacteria:{Neisseria}*
- *Proteobacteria:{Pseudomonas}*
- *Proteobacteria:{Stenotrophomonas}*
- *Tenericutes:{Mycoplasma}*
- *Tenericutes:{Ureaplasma}*

Gram Stain

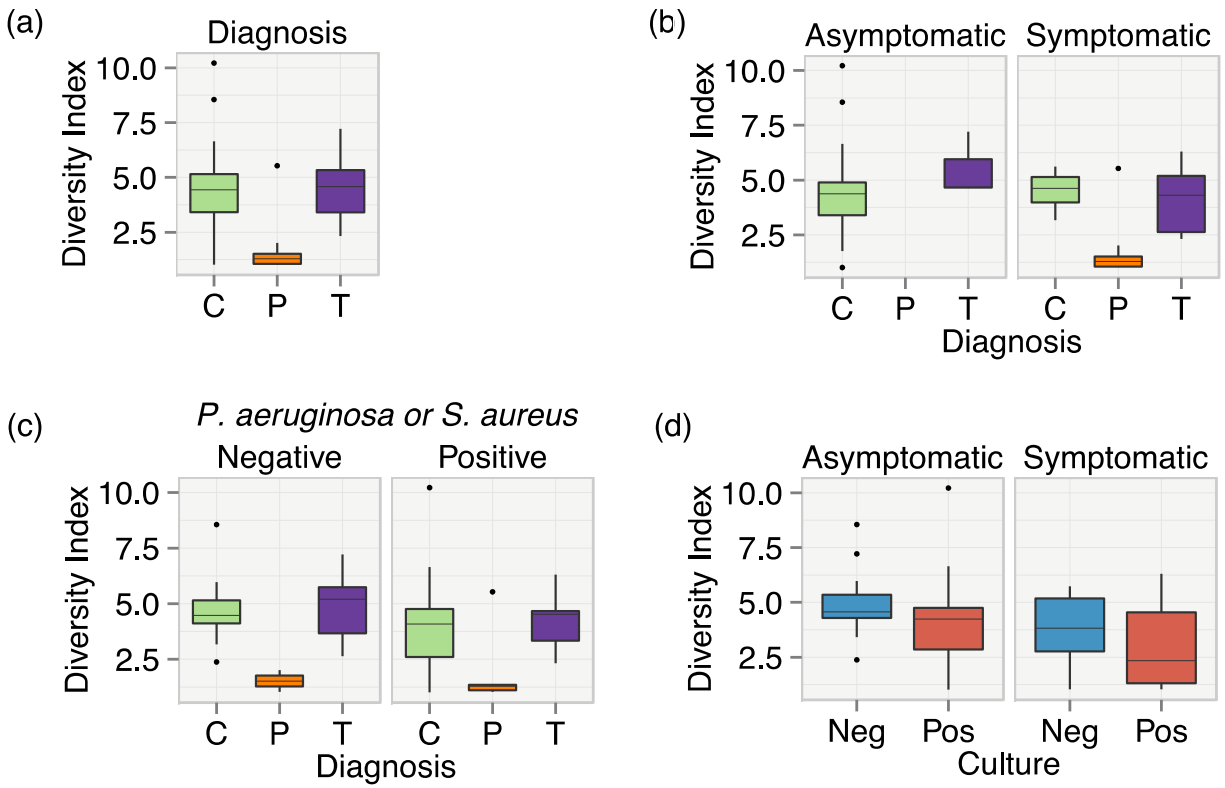
- Positive
- Negative
- Other taxa (<5%)

Metabolism

- Aerobe
- Anaerobe
- Facultative
- Other taxa (<5%)

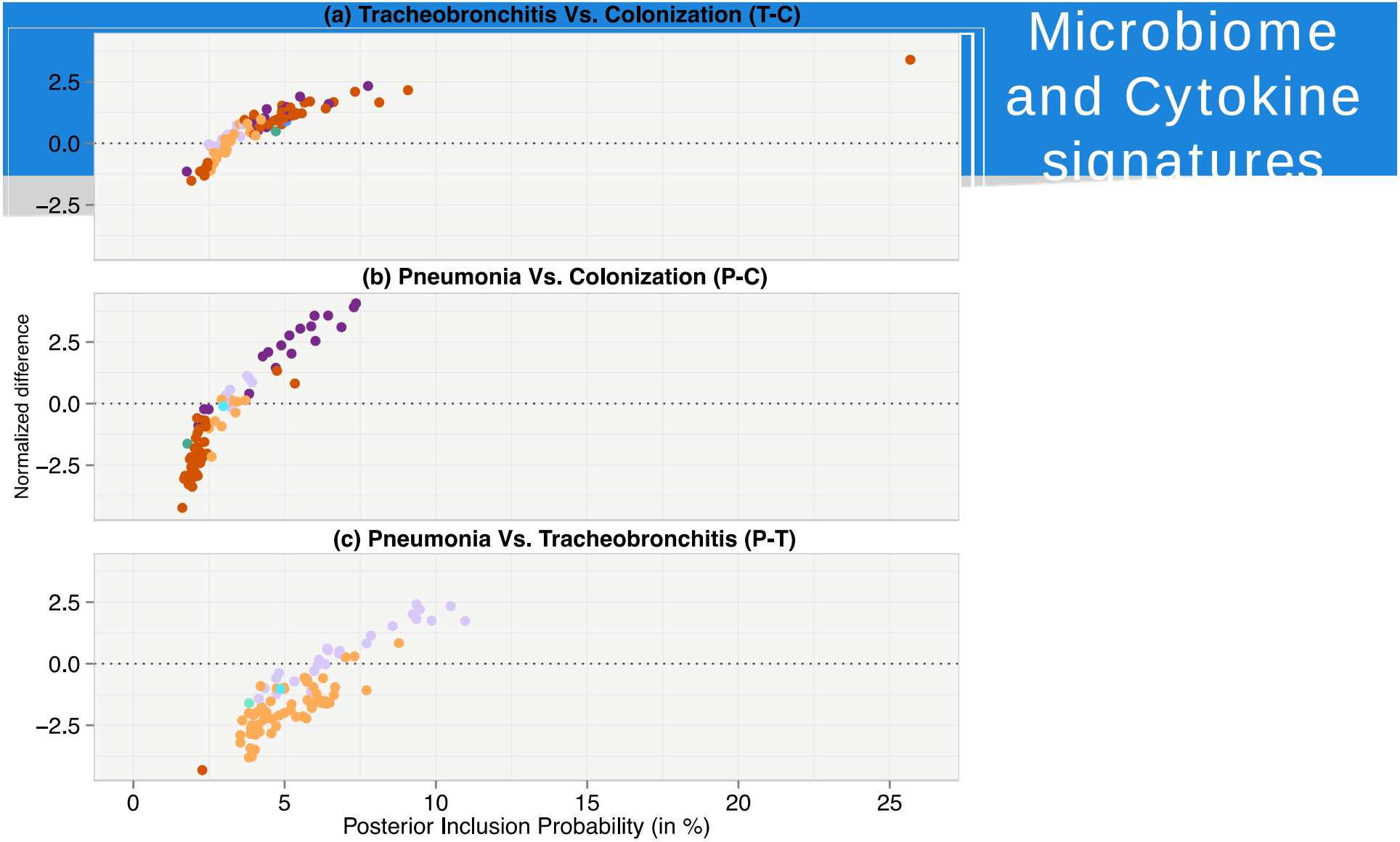
J. Craig Venter INSTITUTE

Descriptive analysis: Microbiome Diversity



J. Craig Venter™ INSTITUTE

Microbiome and Cytokine signatures

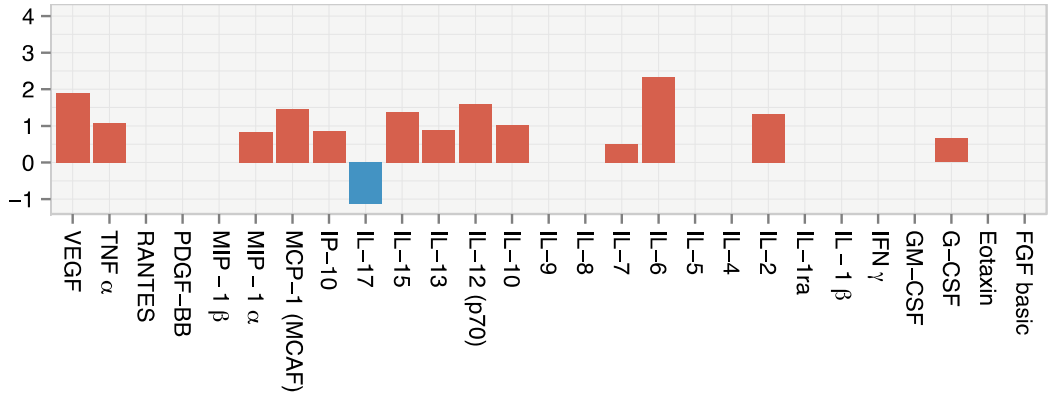


J. Craig Venter™ INSTITUTE

Cytokine signatures (BMA)

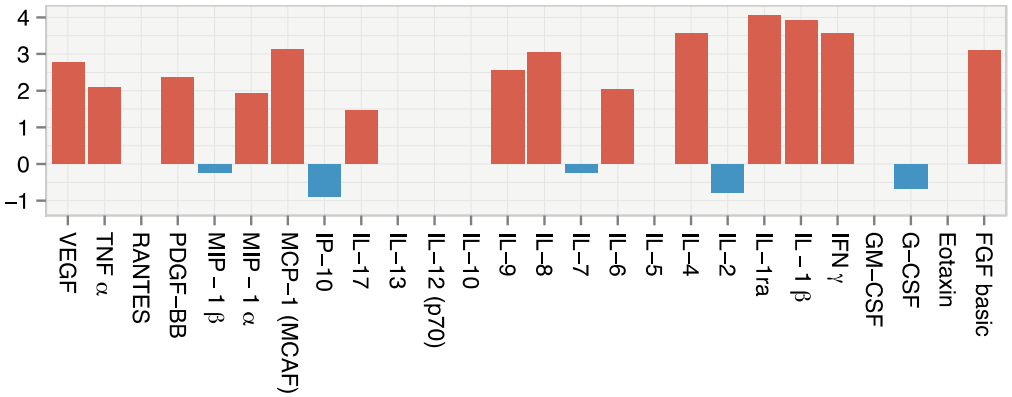
(a) Tracheobronchitis vs. Colonization

(i) ⬆Prob. (TBr)

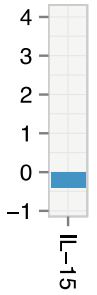


(b) Pneumonia vs. Colonization

(i) ⬆Prob. (PNA)



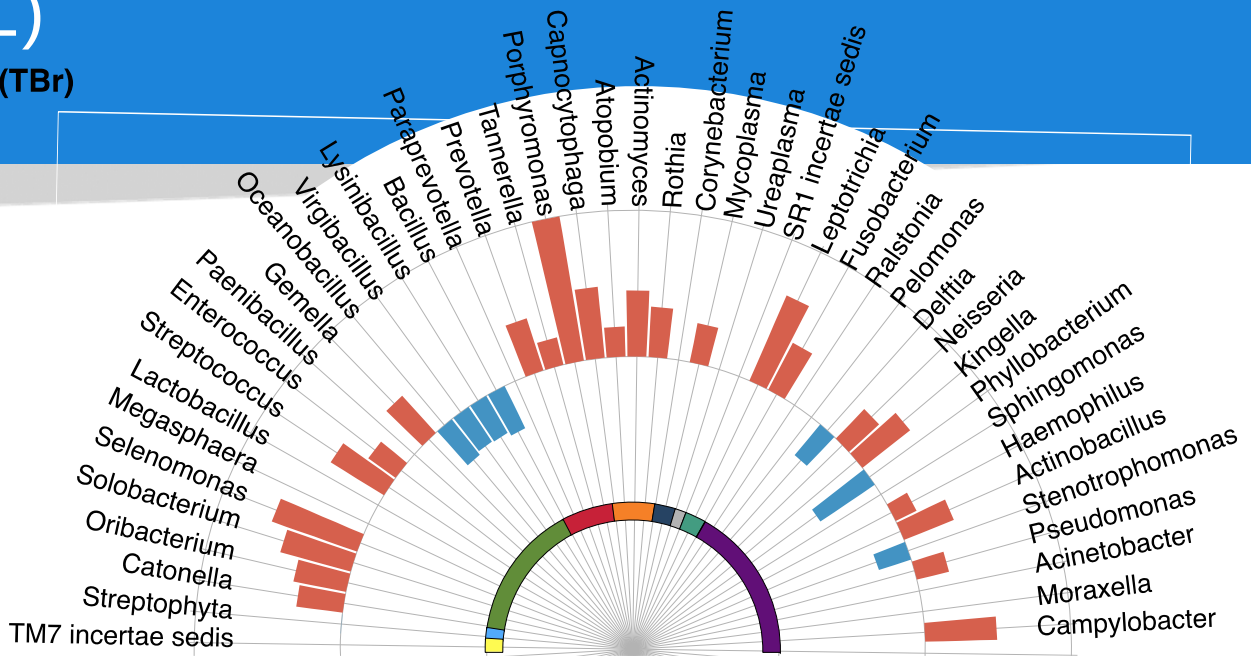
(ii) ⬇Prob. (PNA)



J. Craig Venter™ INSTITUTE

Microbiome signatures (TBr vs. COL)

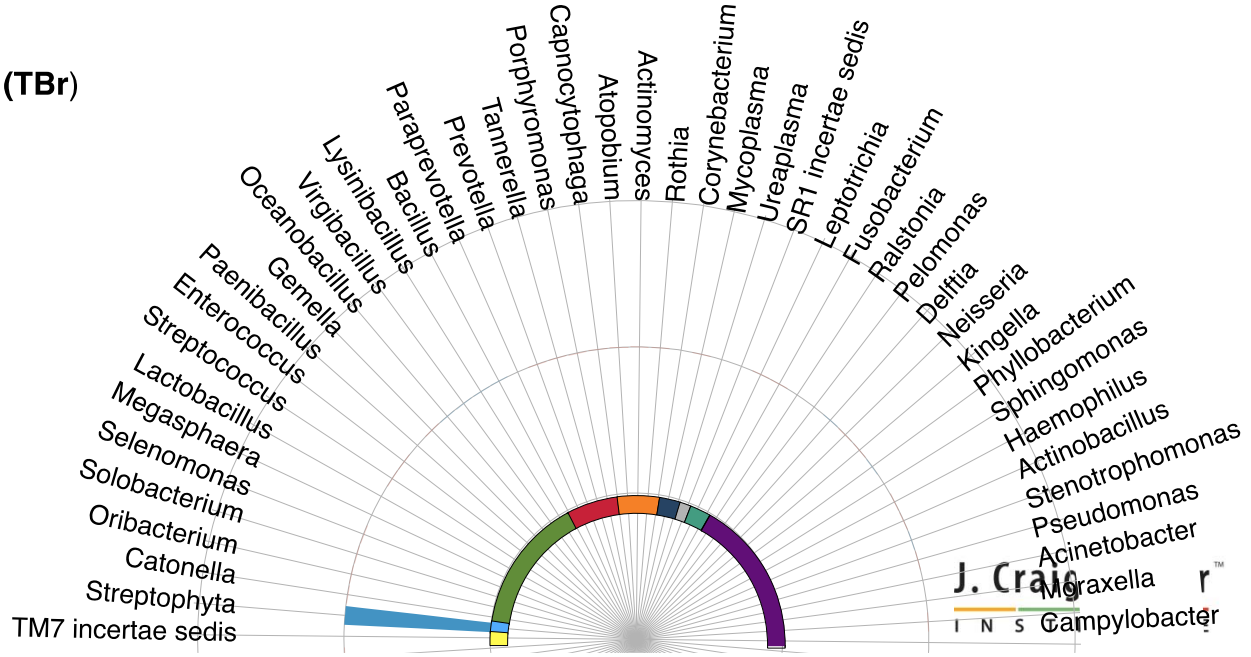
a. ⬆Prob. (TBr)



Phyla

- Proteobacteria
- Fusobacteria
- SR1
- Tenericutes
- Actinobacteria
- Bacteroidetes
- Firmicutes
- Cyanobacteria
- TM7

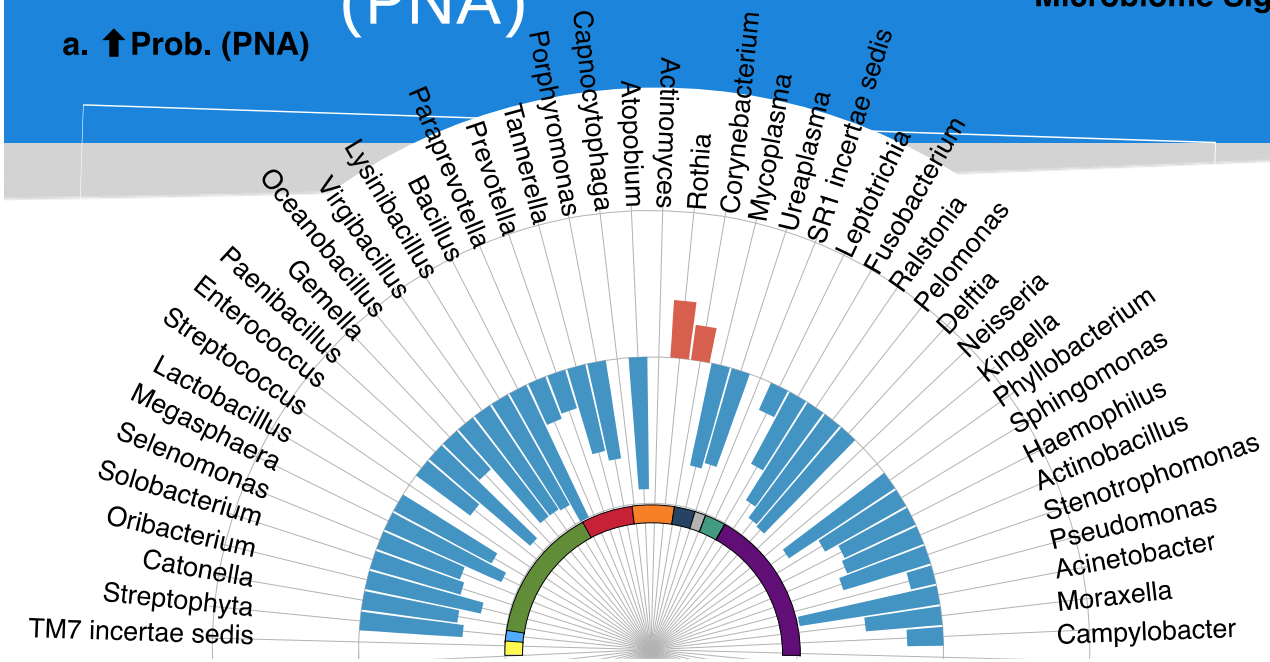
b. ⬇Prob. (TBr)



Microbiome signatures (PNA)

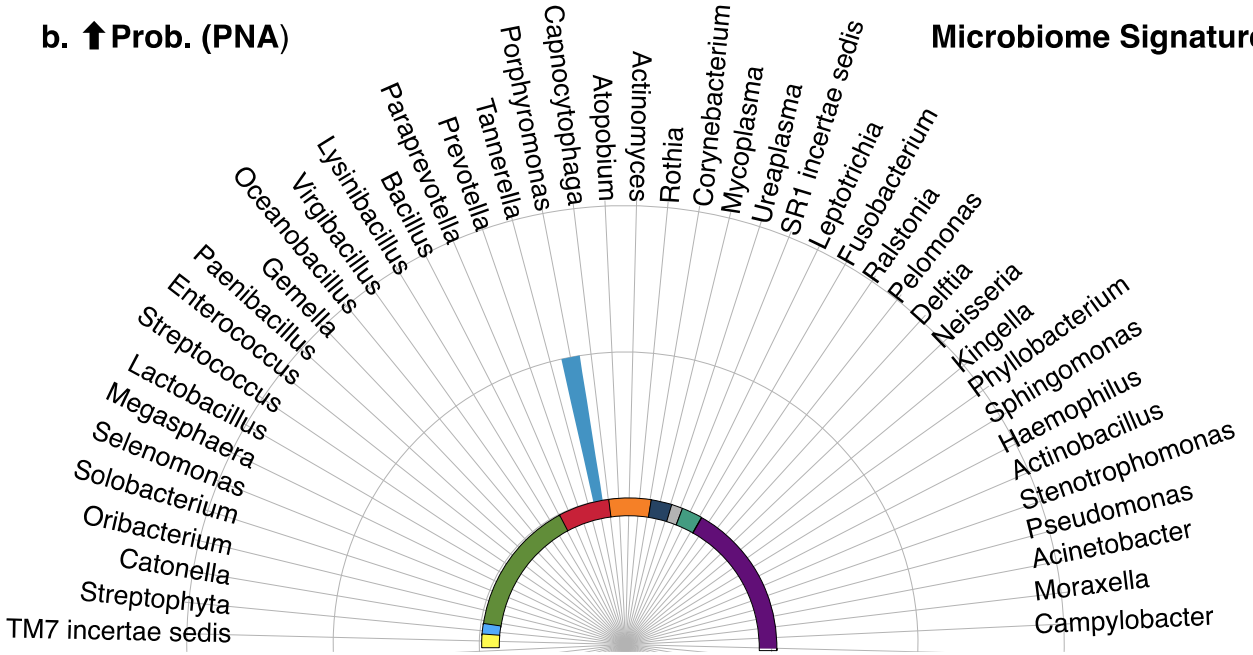
Microbiome Signatures: Pneumonia vs. Colonization

a. ⬆Prob. (PNA)

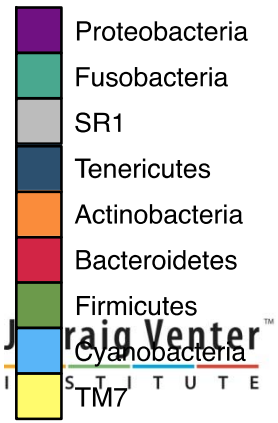


b. ⬆Prob. (PNA)

Microbiome Signatures: Pneumonia vs. Tracheobronchitis



Phyla



J. Craig Venter™ INSTITUTE

Dr. Adesh Ramsubhag

Metagenomic Study of the Bacterial Diversity of the Nariva Swamp, Trinidad

Department of Life Sciences, UWI, St. Augustine:

Integrating genomics and metagenomics in microbiology research

Department of Life Science (DLS)

- A leading department at UWI, St. Augustine

- Several teaching and research units:
  - Environmental and Ecological Sciences
  - Zoology
  - Plant Sciences
  - Biochemistry
  - Marine Biology
  - Biotechnology and Microbiology

Microbiology research in DLS

- Integrated with other disciplines

- Collaborations with other departments at UWI and international partners (e.g. JCVI, CIRAD, UF, GSCU, UP, UG)
- Major focal areas:
  - Environmental microbiology
  - Microbial Ecology
  - Plant microbiology (pathology, biofertilizers and biological control)
  - Microbial natural products

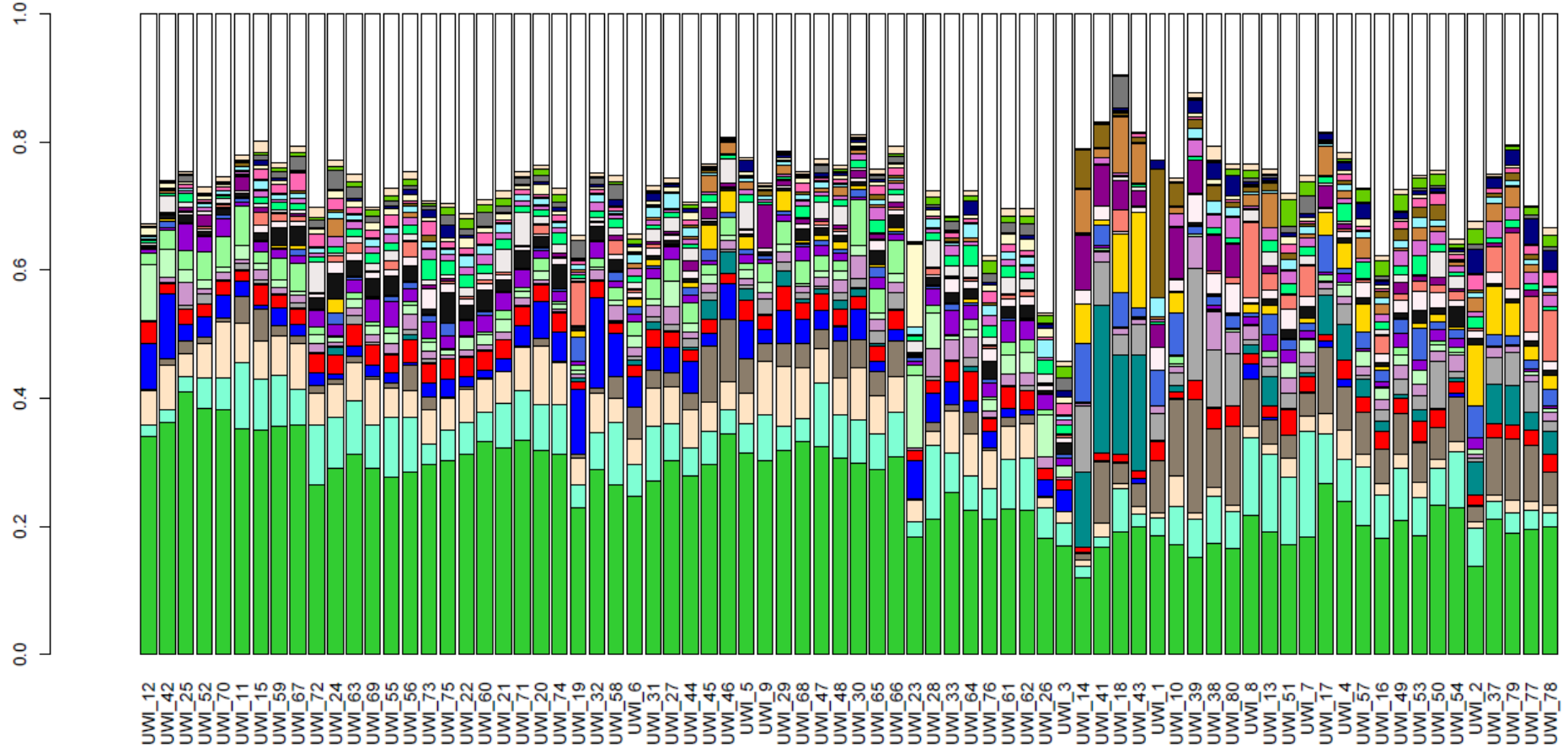
Microbiological research tools and techniques in DLS

- Conventional culture based
- Traditional molecular methods
- Recently started using next generation sequencing technology
  - Powerful tool for achieving old and new research goals

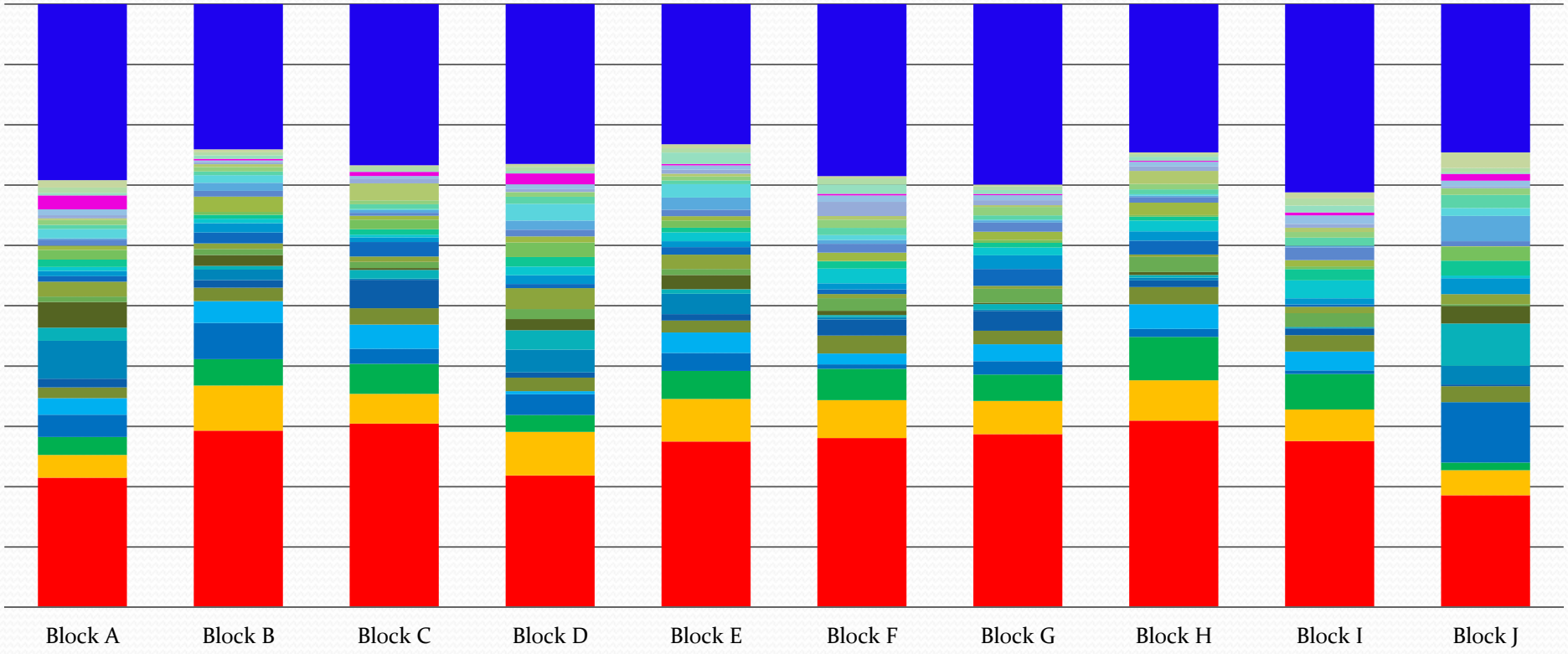
Environmental and Ecological studies

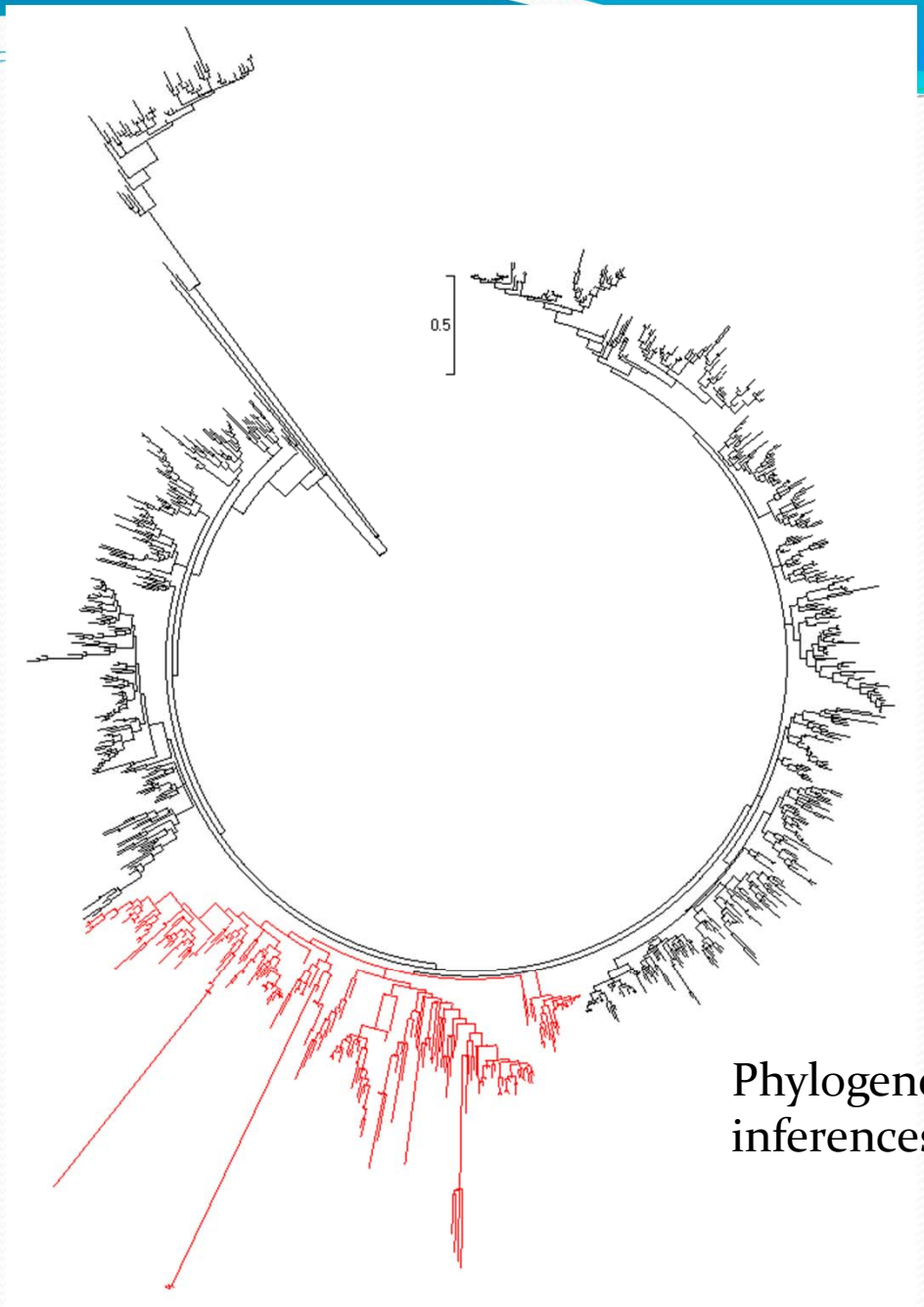
- 16 S bacterial community analysis of 120 environmental samples
  - Sequencing V1 –V3 region of 16S gene using IlluminaMiSeq.
  - >14M reads
  - After preliminary filtering, ~1.2M reads used in analysiis
  - >1.2M bacterial 16 Sequences identified
  - ~230000 representative sequences
  - >1000 genera or higher levels of taxa identidied
  - Downstream processing of data

16S community analysis to understand microbial dynamics associated with carbon cycling in Nariva swamp

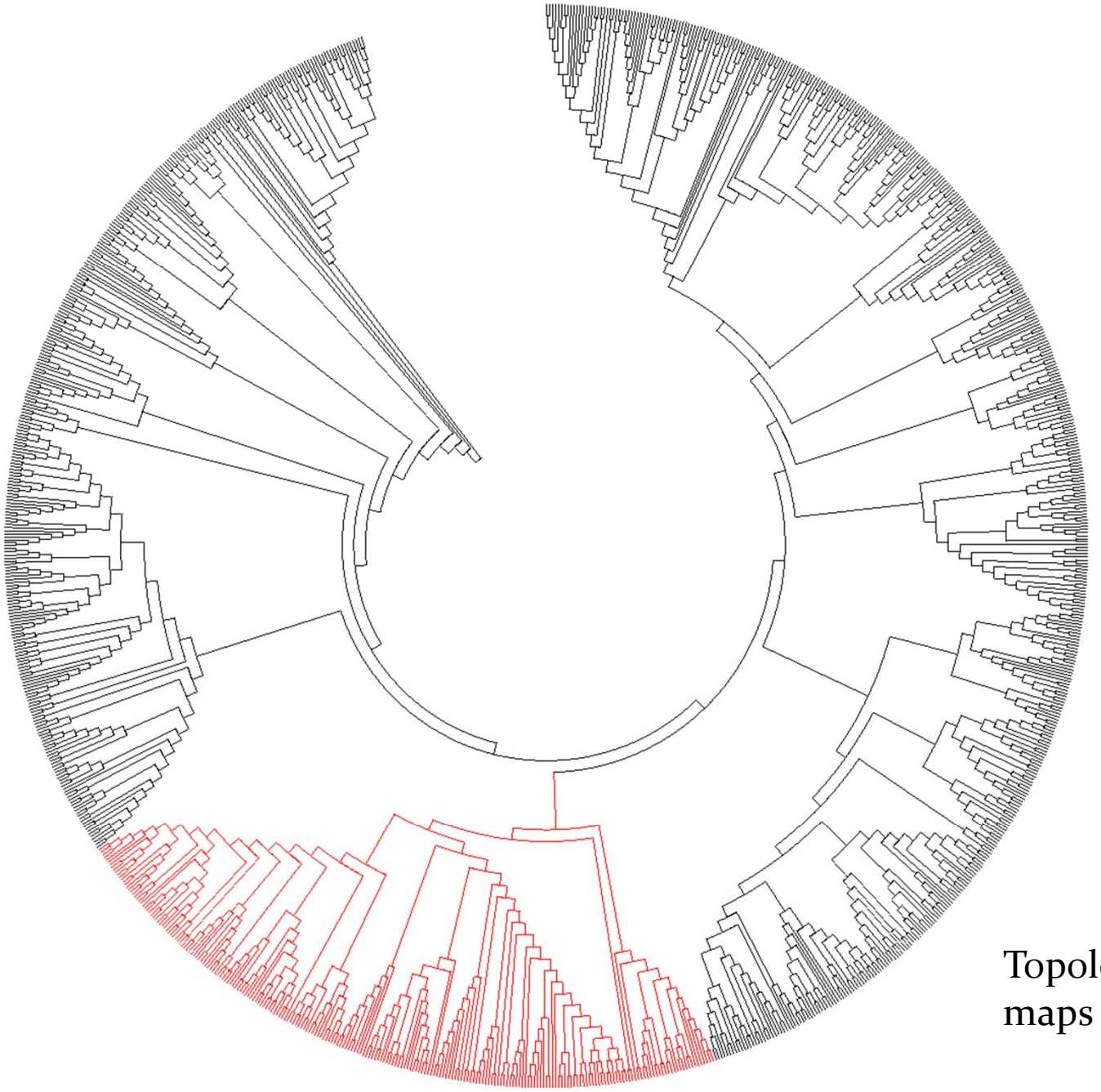


16S community analysis to understand microbial dynamics associated with carbon cycling in Nariva swamp

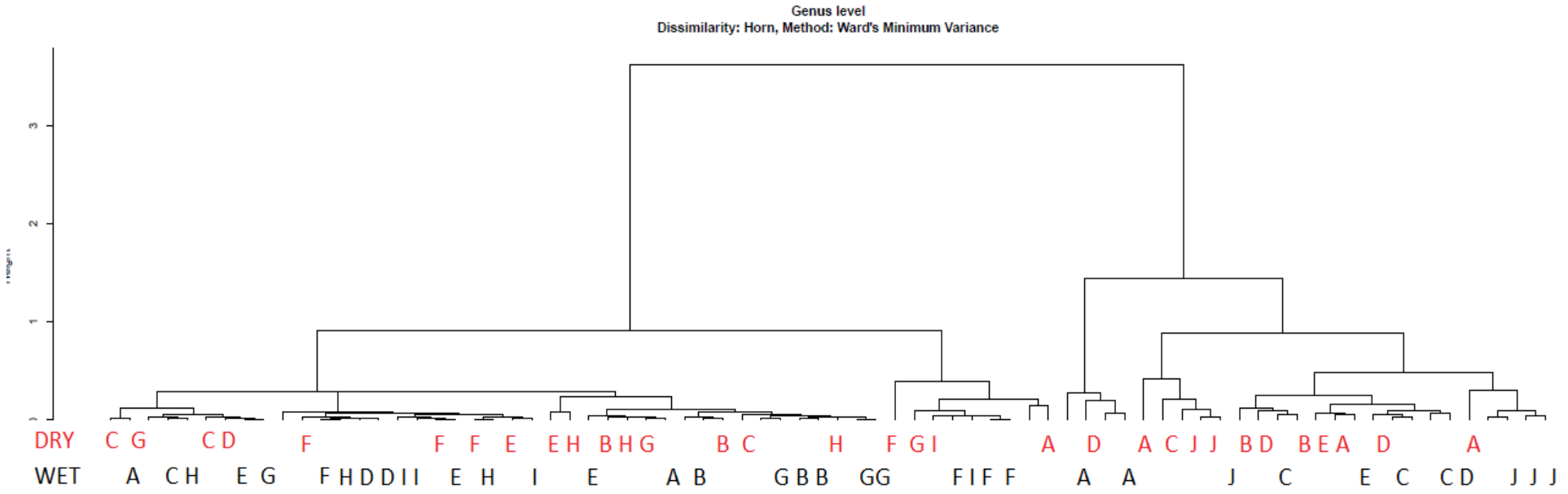


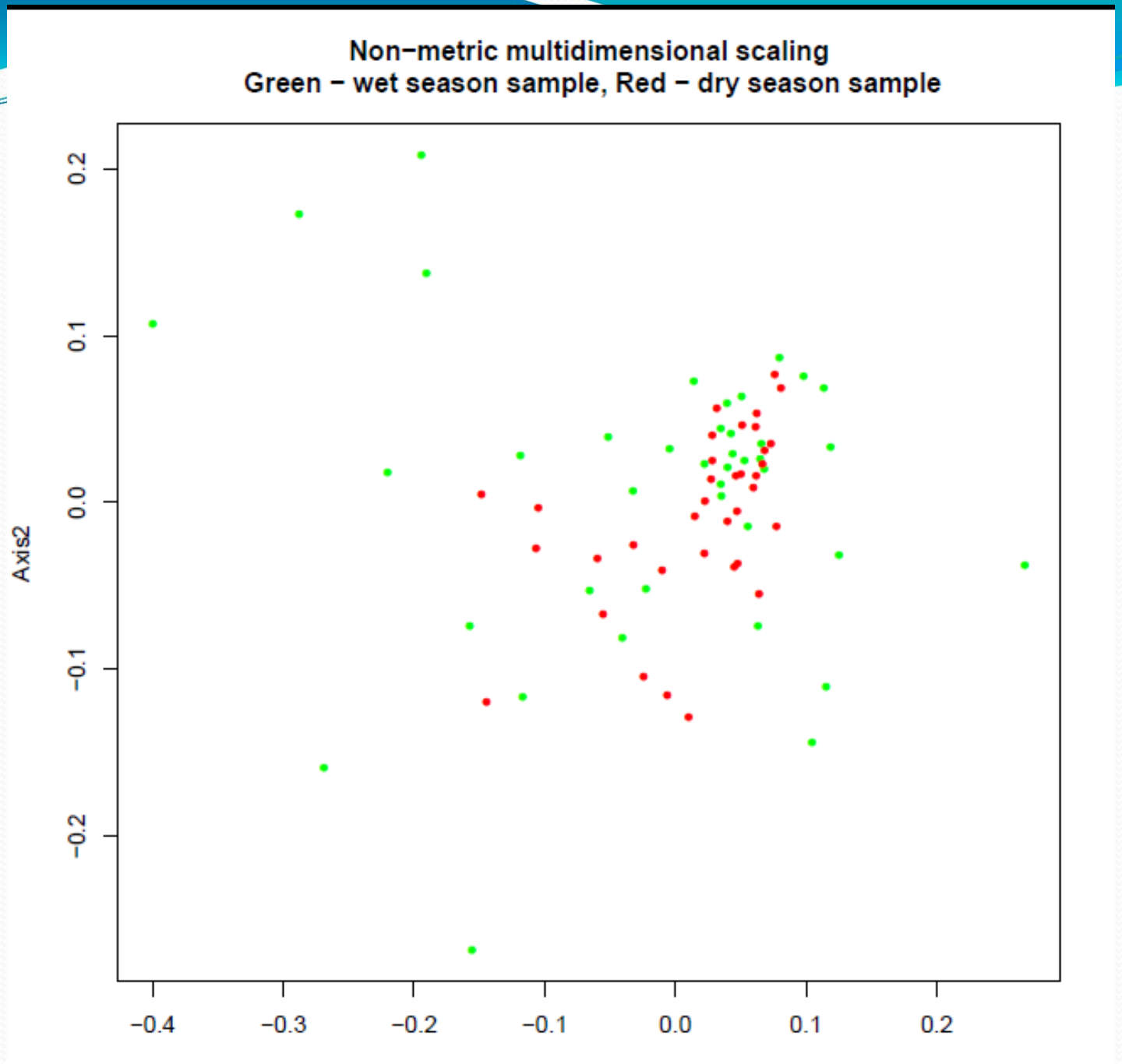


Phylogenetic inferences

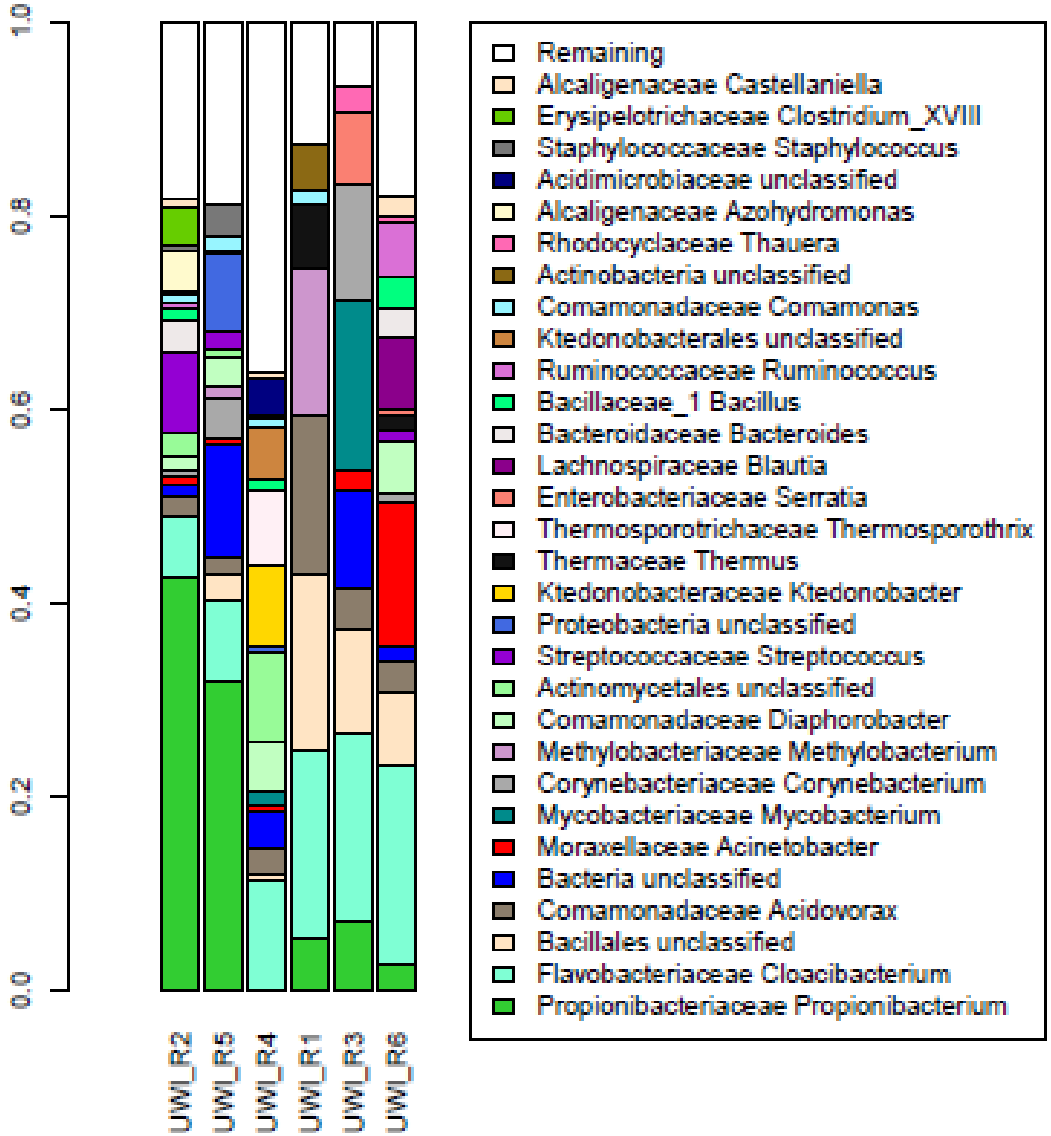


Topology maps

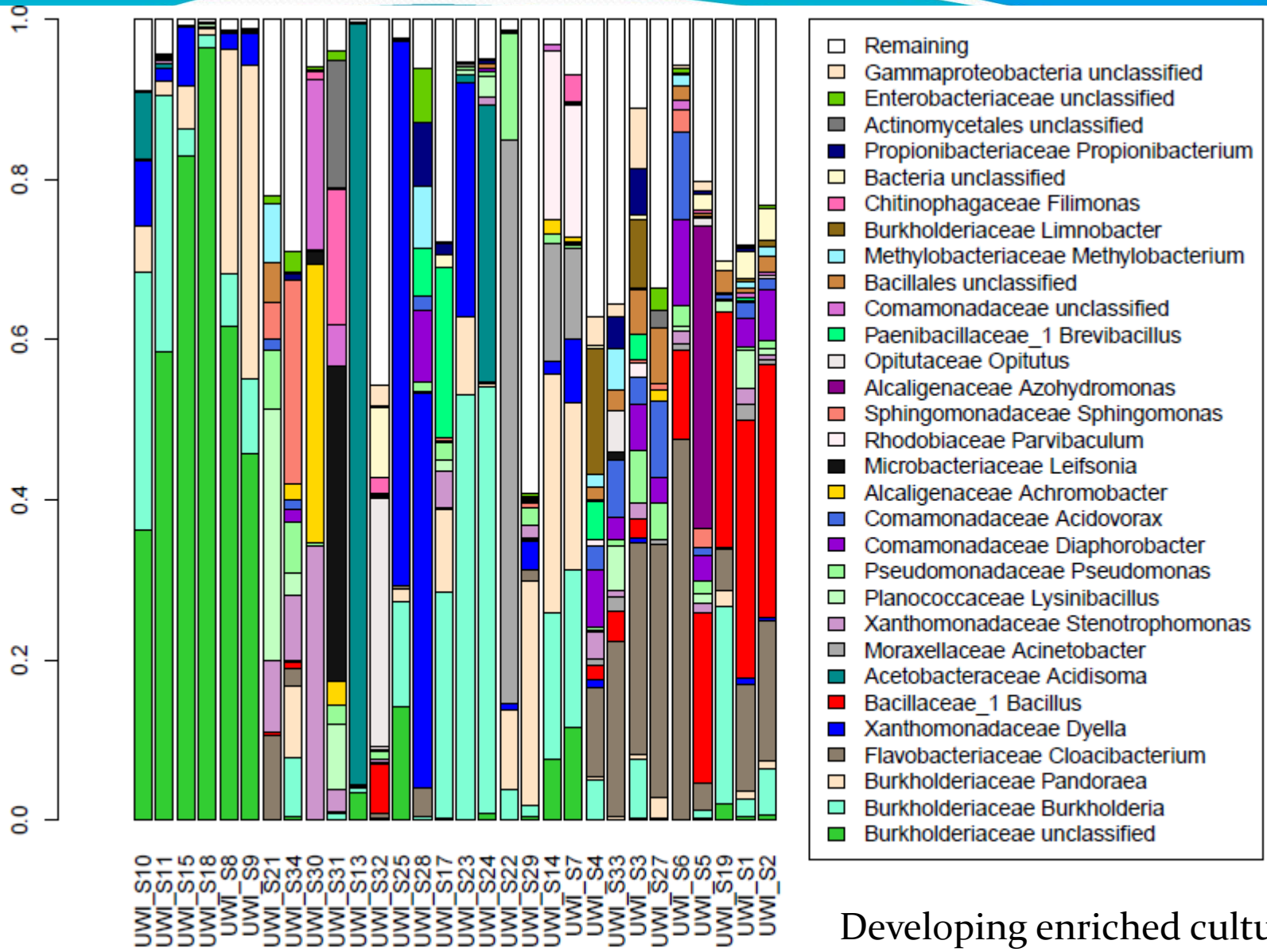

Genus level
Dissimilarity: Horn, Method: Ward's Minimum Variance
Height
0
1
2
3
DRY C G C D F F F E E H B H G B C H F G I A D A C J J B D B E A D A
WET A C H E G F H D D I I E H I E A B G B B G G F I F F A A J C E C C D J J J


Non−metric multidimensional scaling
Green − wet season sample, Red − dry season sample
Axis2
0.2
0.1
0.0
−0.1
−0.2
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2

Bioremediation of petroleum hydrocarbons

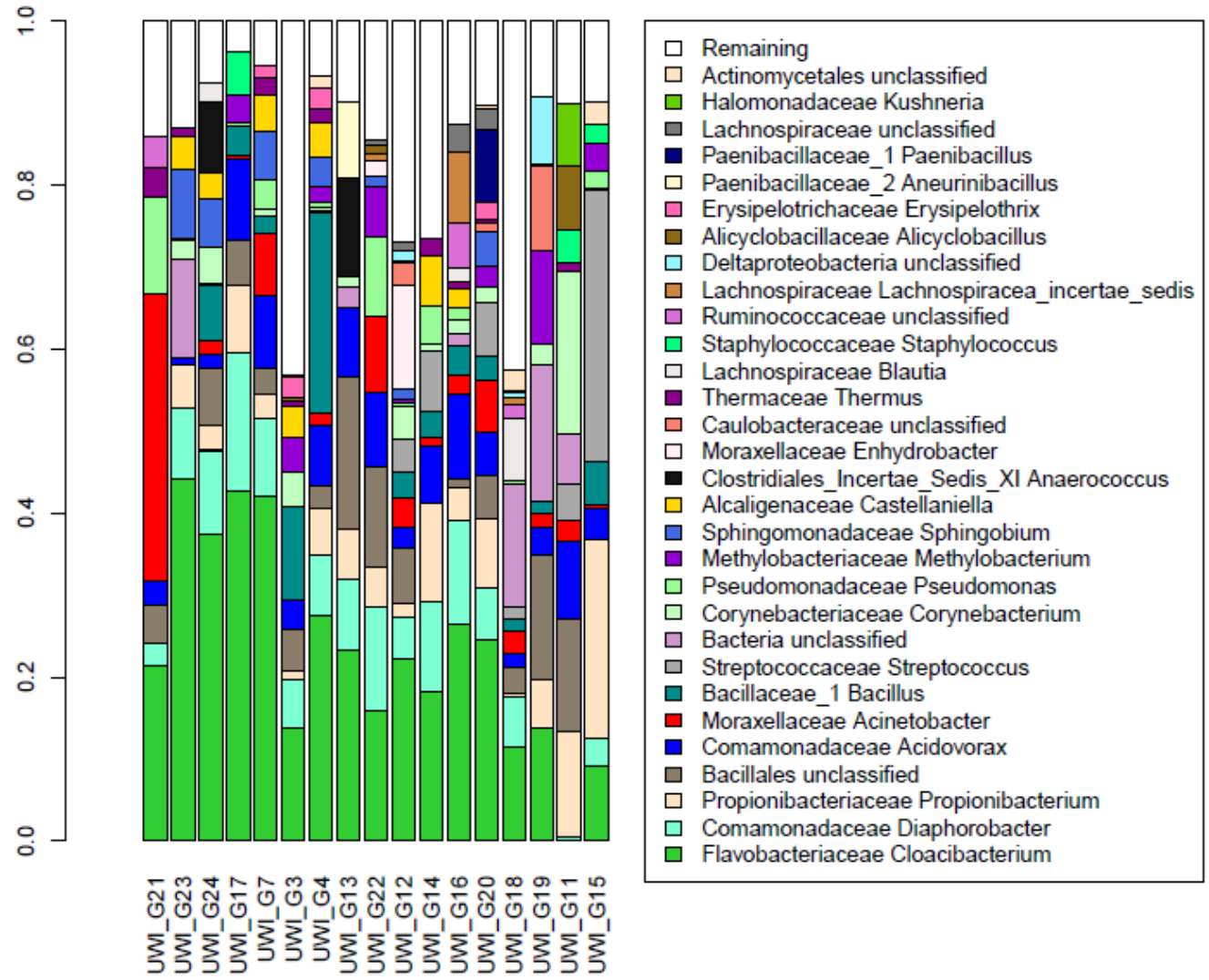


Optimizing DNA extraction methods



Developing enriched cultures for bioremediation

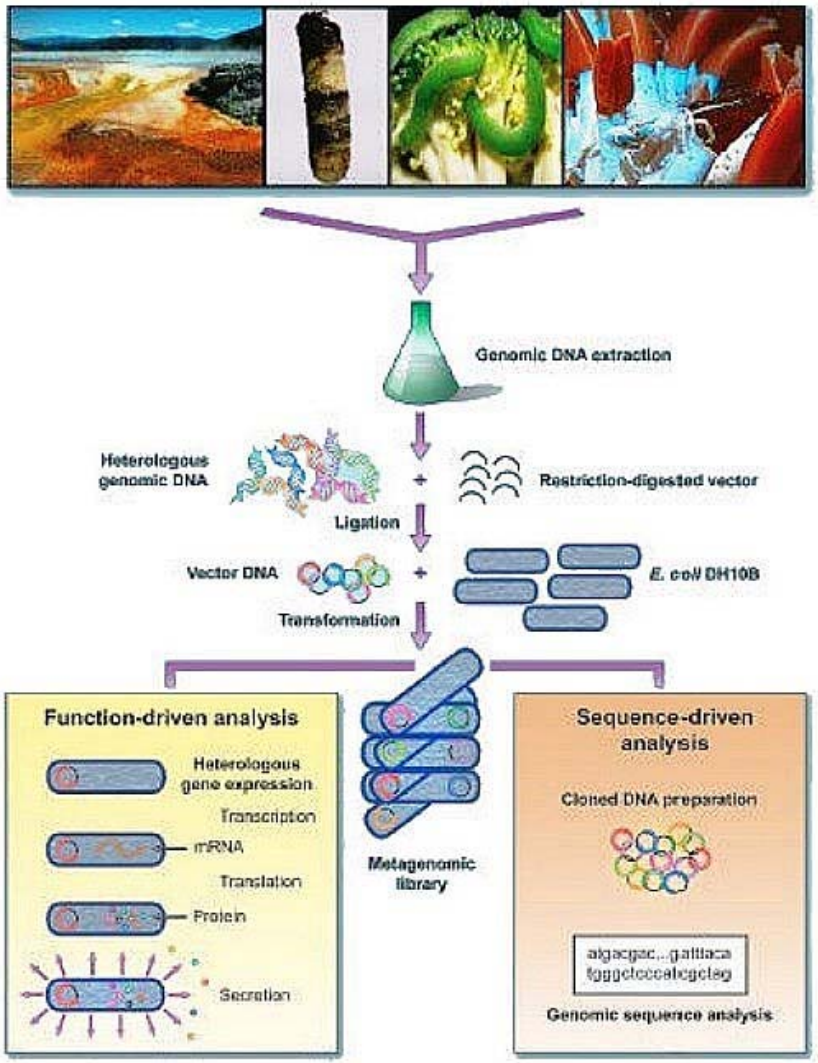
Microbial community dynamics associated with agricultural cropping systems



Further work

- Investigate specific groups of interesting organisms in more detail
- More detailed community analysis
- Understanding role of microbes in different processes using transcriptomic and proteomic analyses
- Other related environmental based projects on-going

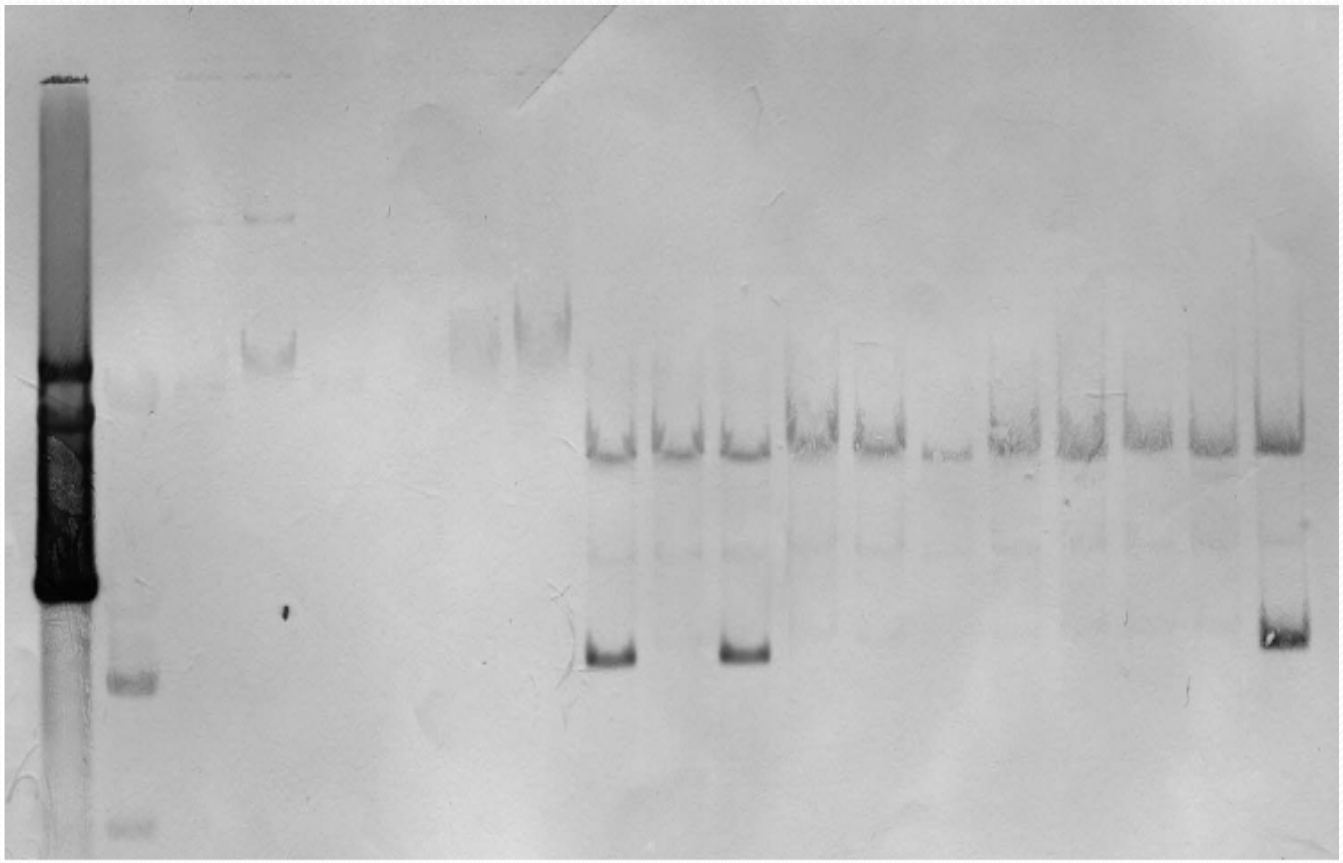
Natural products microbiology



Metagenomic approach to discovering novel bioactive compounds

Plant microbiology

- Copper resistance in Xanthomonas campestris



Plant microbiology

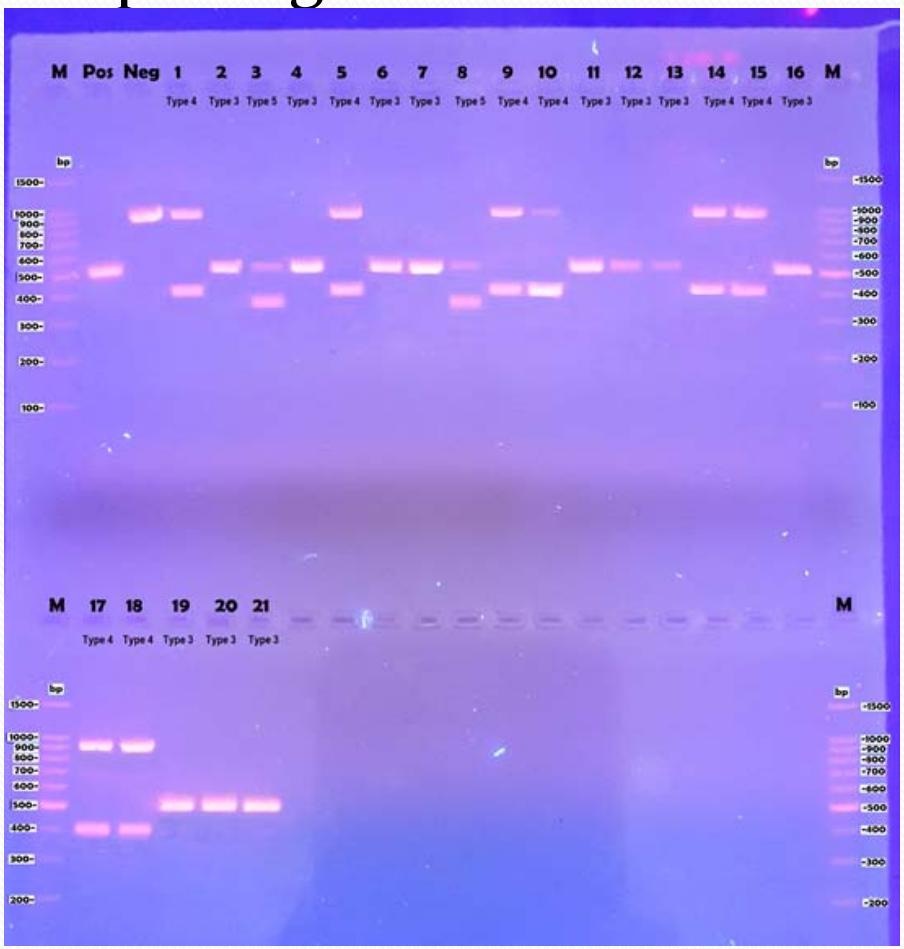
- Additional on-going projects
  - Plant pathology-
    - Epidemiology and evolution of pathogenic bacteria in the Caribbean

    disease diagnostics
    - mixed virus infections

  - Biofertilizers

Infectious diseases

- Antimicrobial resistance in infectious common pathogenic bacteria



END

Dr. Shibu Yooseph
Computational Analysis of Meta-omics Datasets

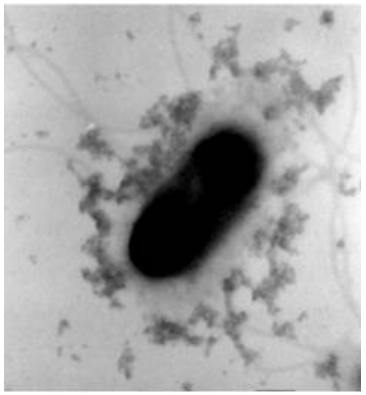
Computational tools for the analysis of "omics" data to study microbial communities

Shibu Yooseph
Professor, Informatics

J. Craig Venter™
INSTITUTE

Microbes

- Small organisms (typically <100μm) not visible to the naked eye
- Either
  - Prokaryotic: cells lacking a true nucleus
    - Bacteria, Archaea
  - Eukaryotic: cells with true nucleus
    - Fungi, Algae, Protists
- *Our focus here is on Prokaryotes*



Microscope image of a bacterium
(by Jeffrey McLean)

- Existed on Earth for 3-4 billion years
- Extremely abundant
  - "Unseen majority" (Whitman et al., 1998)
  - Estimate of 4-6x10^{30} cells
  - Constitute more than half of the biomass of the Earth
- Incredible diversity
- Found almost everywhere, including in extreme environments
  - Temperature (-15°C to 121°C)
  - pH (0 to 11)
  - High pressure (1300 atmospheres)

J. Craig Venter™
INSTITUTE

Microbes and their importance

- Symbiotic relationships with other life forms
- Diseases and human health
- Biogeochemical cycling
- Regulate atmosphere and weather
- Carry out transformations of matter essential for life
- Engineering microbes:
  - Producing chemicals of major industrial importance
  - Producing useful enzymes
  - Food production and preservation
- .. and many many more reasons!

J. Craig Venter™
INSTITUTE

Studying microbes

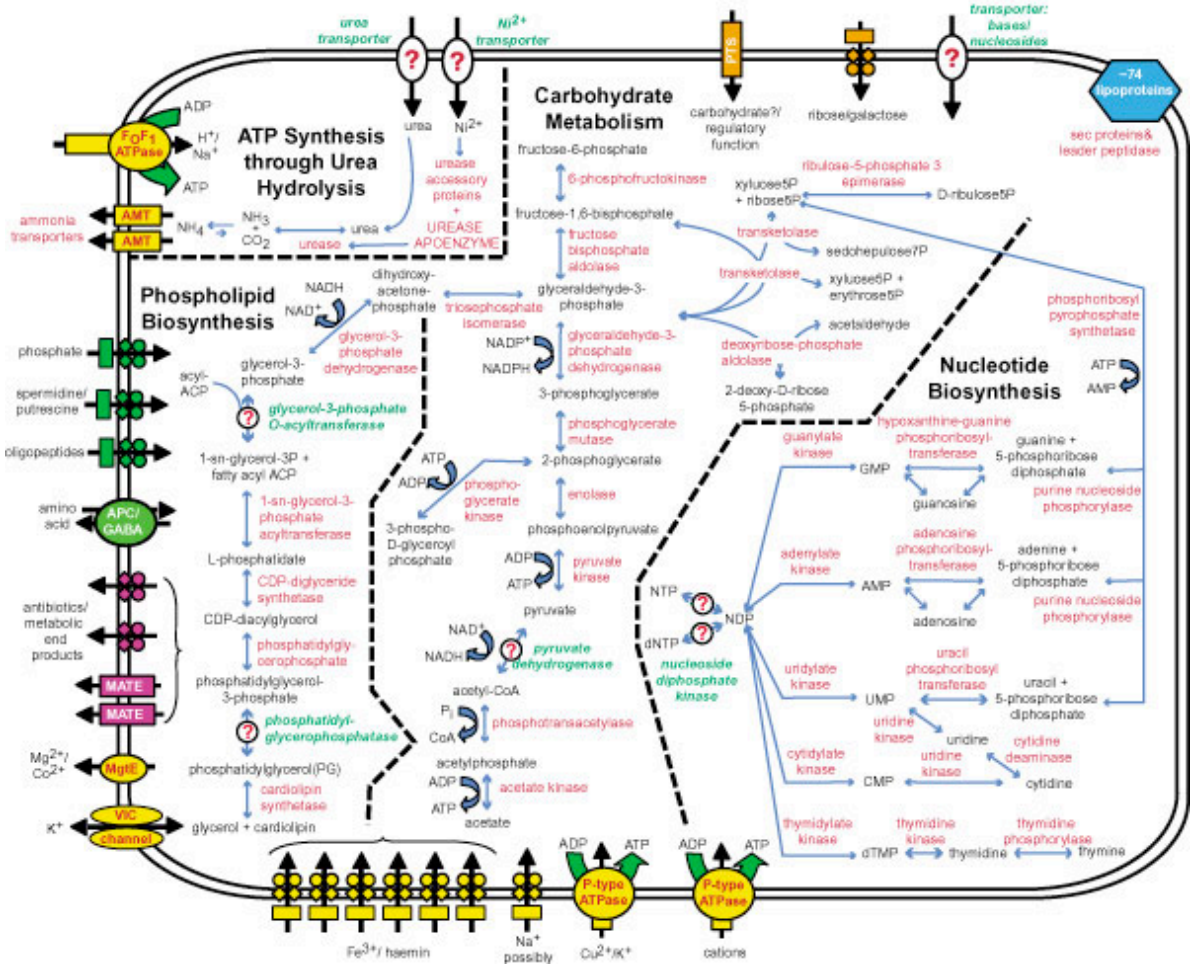
- Cultivation or culturing: grow them in the laboratory
  - Agarose gel
  - Liquid culture
- Plate count anomaly (Stanley and Konopka 1985)
  - Cells counts obtained by cultivation using agar media are very often orders of magnitude smaller than cell counts under a microscope
- Current estimate: in many environments, we can only cultivate <1% of microbes
  - Do not know right conditions and nutrients
  - Could be co-existing with other microbes

Molecular sequencing

- 16S rRNA gene has been used as a phylogenetic marker to produce taxonomic classification of prokaryotes (Woese 1977)
- Sequencing of microbial genomes
  - *Haemophilus influenzae* was first bacterium to be sequenced (Fleischmann et al., 1995)
  - Currently over 16,000 prokaryotic genomes sequenced (complete or draft)

J. Craig Venter™ INSTITUTE

Map of metabolic pathway and substrate transport



Glass et al, 2000

J. Craig Venter INSTITUTE

Microbial community assays: "omics" data

- Marker gene based (using PCR)
  - Taxonomic: 16S rRNA gene, RecA gene
  - Functional genes
- Metagenomics
  - Genomic content
  - Infer metabolic potential
- Metatranscriptomics
  - Gene expression: transcriptomic content
- Metaproteomics
  - Protein expression
- Metabolomics
  - Profile of small molecule metabolites

J. Craig Venter™ INSTITUTE

Microbial community studies

- Driven by a combination of advances in experimental methodology, laboratory techniques, sequencing technology, and informatics

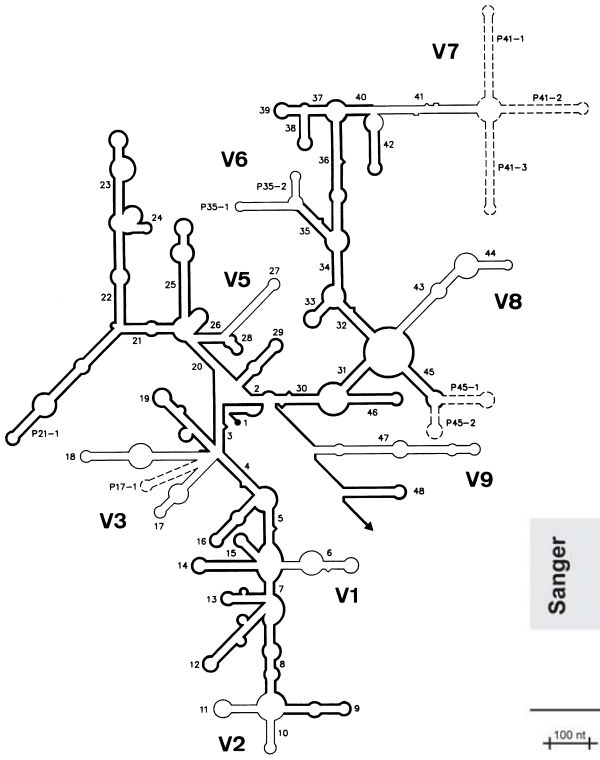
- Next-generation sequencing
  - Quality and volume of data
  - Error rates and sequence lengths
  - Illumina, IonTorrent, PacBio, etc.

- Computational challenges
  - Big Data challenge: data rich field!
  - ~1 Tbp can be generated per run (Illumina HiSeq2500)
  - High-throughput computing
  - Design of efficient algorithms

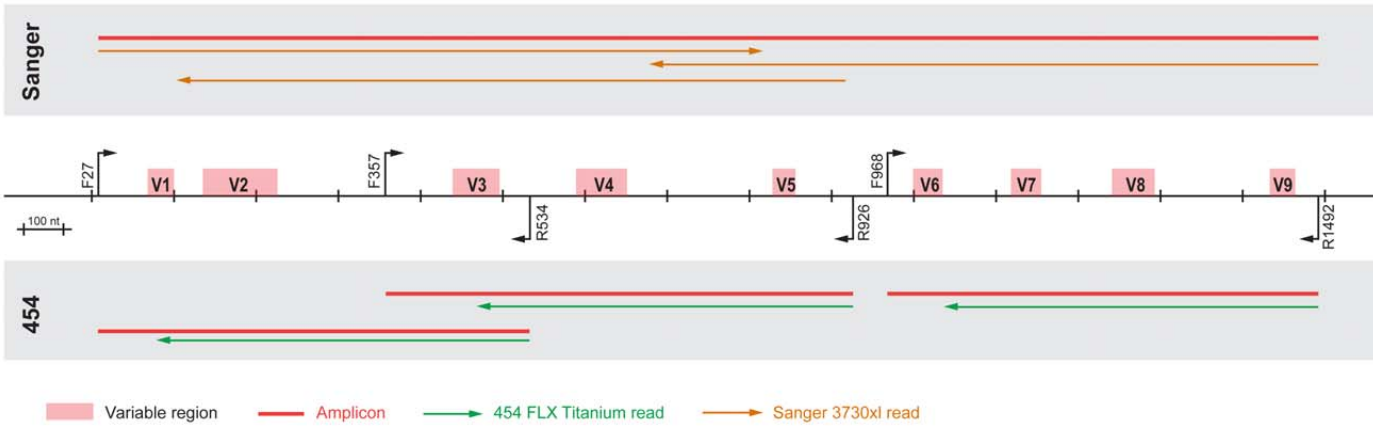
J. Craig Venter™
INSTITUTE

16S rRNA gene as a taxonomic marker

Neefs et al., *Nucleic Acids Res.* 1990



Sequencing the variable regions of the 16S rRNA gene



(Jumpstart Consortium HMP, *PLoS One* 2012)

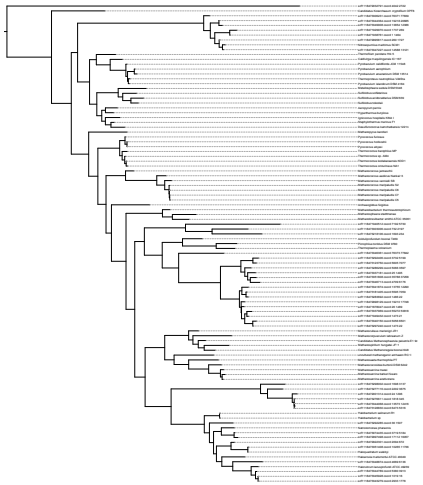
J. Craig Venter™ INSTITUTE

16S sequence data analysis

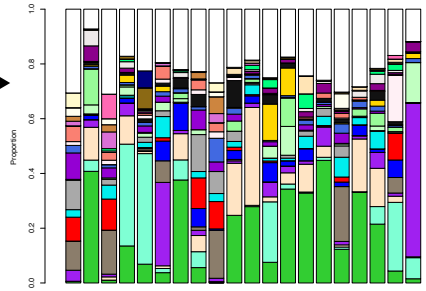
Raw sequences

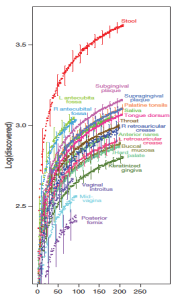
- *Deconvolution*
- *Quality Checking*
- *Trimming*
- *Sequence stitching (if paired end)*
- *Alignment*

Taxonomic assignments

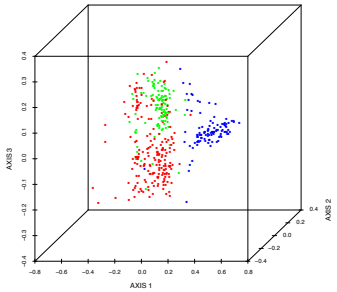


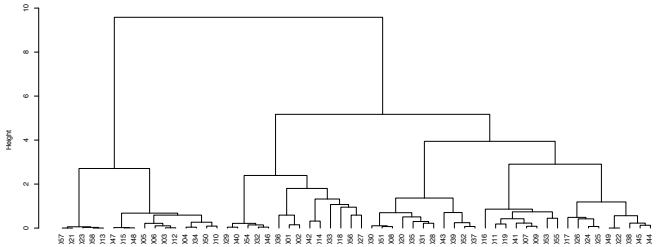
Sample composition and diversity



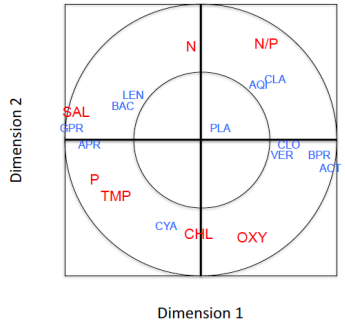


Sample ordination and clustering





Multivariate analysis with sample metadata



J. Craig Venter INSTITUTE

16S rRNA sequence analysis

- Taxonomic assignment of sequences
  - Phylogeny based
    - Any tree building program
  - Non-phylogeny based
    - Example: Naïve Bayesian Classifier (Wang et al, *AEM* 2007) used by the Ribosomal Database project
- Operational Taxonomic Unit (OTU)
  - Sequence clustering
  - Species otus: 97% sequence identity
- Alpha and beta diversity of samples
  - Diversity: richness, evenness
- Sample comparison, ordination, and clustering
  - PCA, NMDS
  - K-means, PAM, hierarchical clustering, model-based clustering
  - Several implementations available (Example: in the R package)
- Software packages that are one-stop shop
  - Mothur (Schloss et al., *AEM* 2009)
  - Qiime (Caporaso et al., *Nature Methods 2010)*

J. Craig Venter™ INSTITUTE

Understanding limitations of 16S rRNA gene based analysis

- Many genomes have multiple and nearly-identical copies of 16S rRNA operons
- Not always possible to accurately normalize 16S data
- Cannot always obtain accurate estimates of diversity and abundance
- Strain-level resolution of the microbial community is not always possible
- Strains of the same species can have different phenotypes and functional capabilities
  - Not possible to accurately infer or reconstruct functional and metabolic capabilities
- Conservation of 16S gene sequence can mask sequence diversity in rest of the genome

J. Craig Venter™ INSTITUTE

Metagenomics

- Examining genomic content of organisms in community to better understand
  - Diversity of organisms
  - Their roles and interactions in the ecosystem

- Cultivation independent approach to study microbial communities
  - DNA is directly isolated from sample and sequenced

J. Craig Venter™
INSTITUTE

Metagenomics

- Who's there?
  - Taxonomic composition
  - Accurate strain level resolution and abundance estimates possible
- What are they doing?
  - Functional composition and metabolic potential

J. Craig Venter™ INSTITUTE

Microbial communities

- Collection of organisms (taxonomically distinct)
- Varying abundances
- (Possibly) different %GC content and codon usage biases
- Strain variants, genome rearrangements, etc.

- Community complexity is a function of carbon, nutrient, and energy sources, and environmental variables like temperature, pH, salinity, etc.

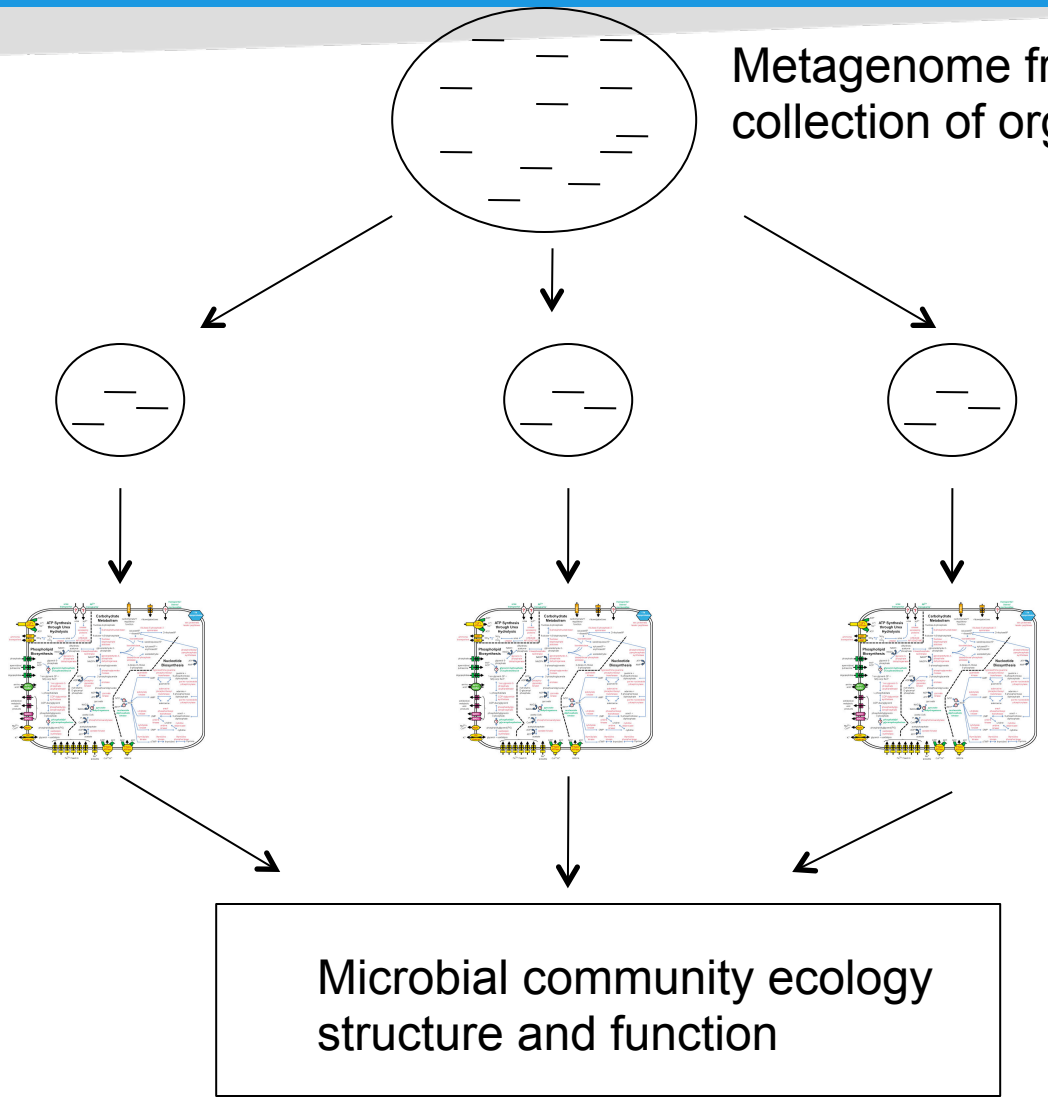
J. Craig Venter™
INSTITUTE

A few examples

- Acid mine drainage study (Tyson et al., *Science* 2004)
  - Low diversity environment
- Soil metagenomics (Vogel et al., *Nature Reviews* 2009)
  - High diversity environment
- Sargasso Sea (Venter et al, *Science* 2004) and Global ocean sampling (GOS) expedition (Rusch et al, *PLoS Biol* 2007; Yooseph et al, *PLoS Biol* 2007)
  - Medium diversity environment
- Human body (Gill et al., *Science* 2006; Qin et al., *Nature* 2010; HMP Consortium, *Nature 2012)*
  - Variable diversity

J. Craig Venter™
INSTITUTE

Metagenomics



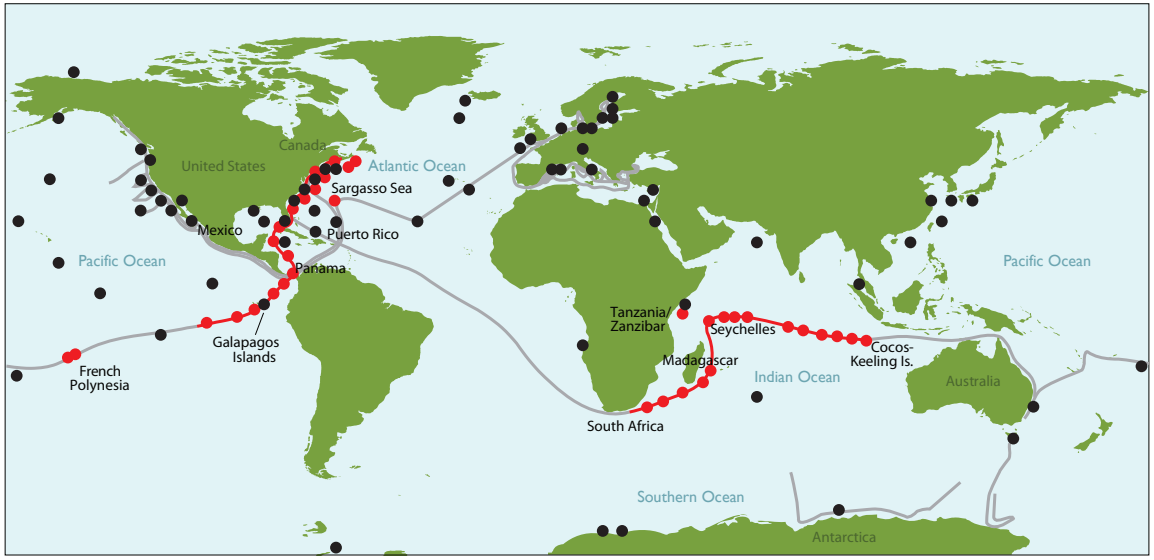
J. Craig Venter™ INSTITUTE

Computational analysis of metagenomic data

- Taxonomic classification or binning
  - Reference genome based: read mapping
  - Classification of assembled contigs
- Metagenomic assembly
  - Reconstruction of genomes of organisms
- Gene finding
  - Prediction of protein coding genes
- Gene annotation
  - Assigning function to predicted genes
- Pathway reconstruction
  - Inference of metabolic pathways from gene annotation data
- Choice of DNA preparation protocols and sequencing technology affects complexity of computational inference

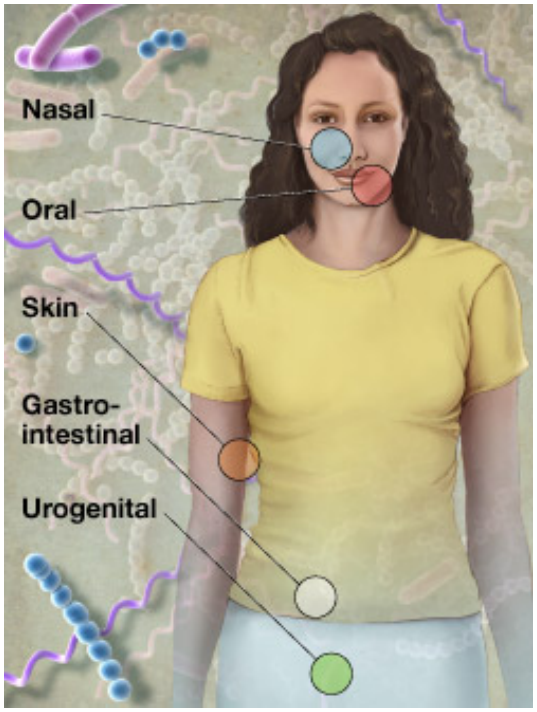
J. Craig Venter™ INSTITUTE

Computational infrastructure developed for large scale metagenomic projects at JCVI



J. Craig Venter INSTITUTE

JCVI's Global Ocean Sampling (GOS) Expedition



NIH funded Human Microbiome Project

J. Craig Venter INSTITUTE

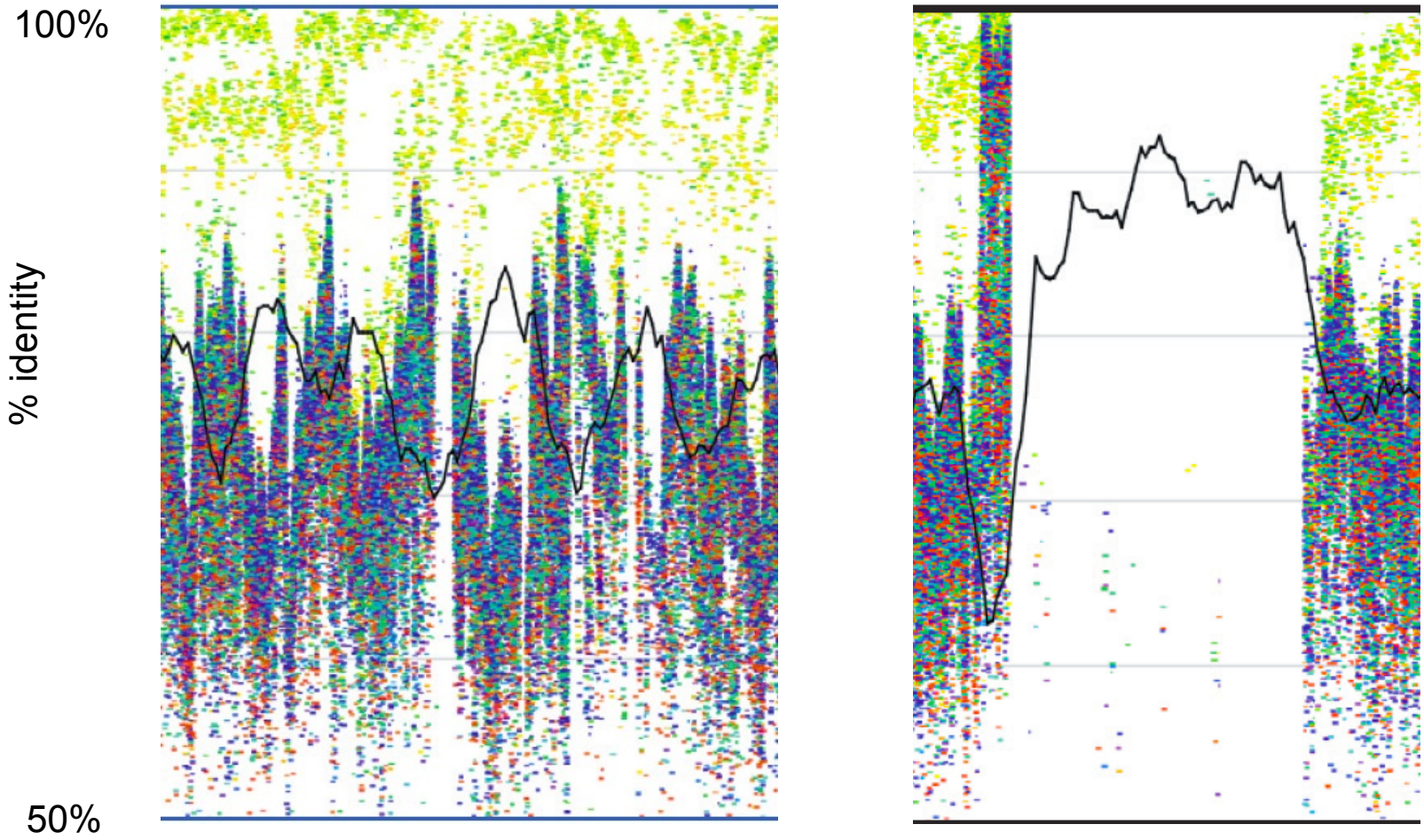
Taxonomic classification via mapping of metagenomic reads

- Use of reference genome database
- Read mapping tools include
  - BWA (Li and Durbin, *Bioinformatics* 2009)
  - Bowtie2 (Langmead and Salzberg, *Nature Methods* 2012)
  - BLAST (Altschul et al, *Nucleic Acids Res.* 1997)
  - BLAT (Kent, *Genome Res.* 2002)

- Inference of relative abundances of organisms; examples include
  - MEGAN (Huson et al, *Genome Res.* 2007)
  - MetaPhlAn (Segata et al, *Nature Methods* 2011)
  - PhymmBL (Brady and Salzberg, *Nature Methods* 2009)
  - GRAMMy (Xia et al, *PLoS One* 2011)

J. Craig Venter™ INSTITUTE

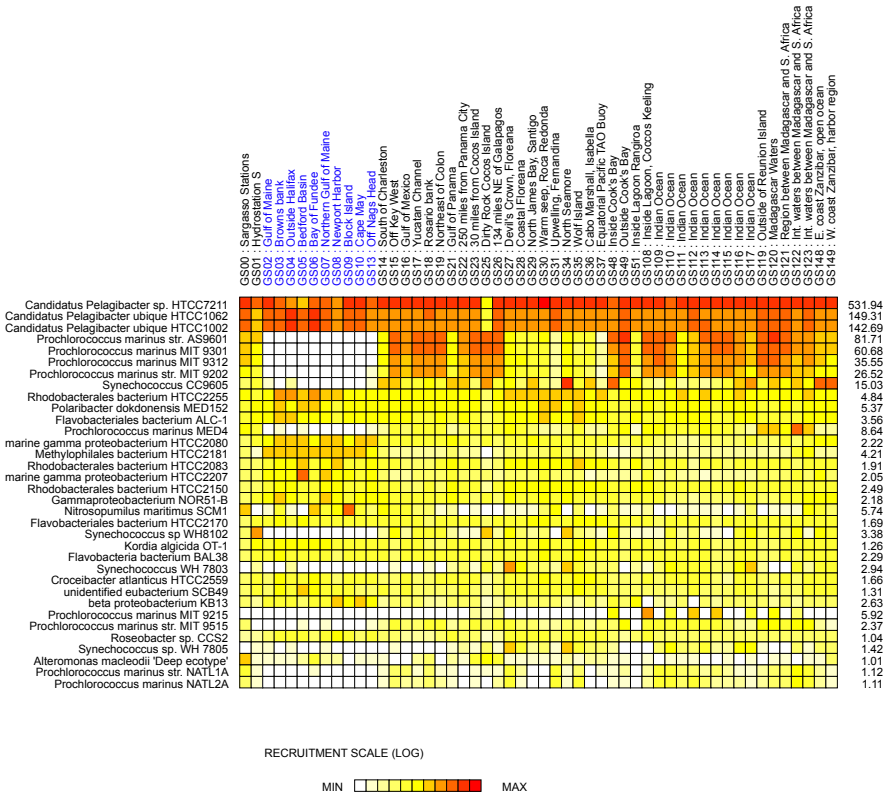
GOS project
Understanding genome diversity using read mapping
Pelagibacter Ubique *HTCC1062* as reference genome



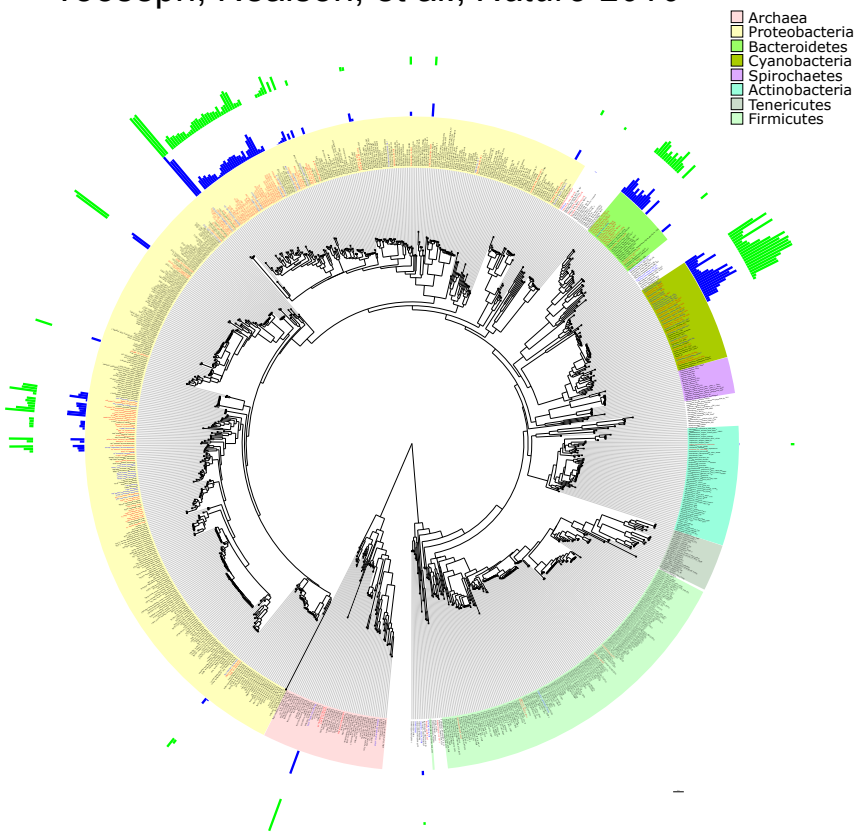
(Rusch et al, *PLoS Biol* 2007)

J. Craig Venter™ INSTITUTE

Studying marine microbial ecology using read mapping



Yooseph, Nealson, et al., *Nature 2010*



High recruiting genomes and low recruiting genomes have distinct functional and metabolic capabilities

J. Craig Venter™ INSTITUTE

Taxonomic classification using marker genes

- Use of protein coding genes as taxonomic markers to ascertain sample composition; genes selection criteria include
  - Single copy
  - Universal (or near universal) presence
  - No support for horizontal gene transfer

  (Ciccarelli et al., *Science* 2006)

- Software packages include
  - Amphora (Wu and Eisen, *Genome Res* 2008)

J. Craig Venter™ INSTITUTE

Metagenomic assembly

- Goal: Inference of genomes of constituent organisms in a metagenomic sample

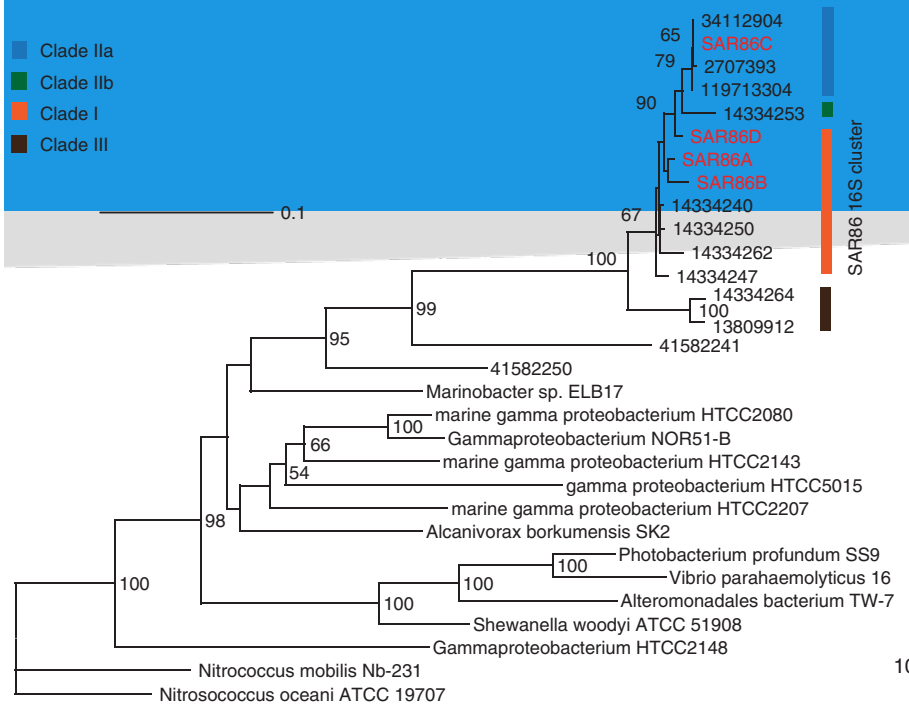
- Assembly quality is dependent on many criteria, including
  - Diversity and complexity of microbial community being sampled: strain variation, genome rearrangement, etc.
  - Sequencing depth of sample: sequence coverage of organisms
  - For a given sequencing depth, better assemblies possible from lower diversity communities

Metagenomic assemblers

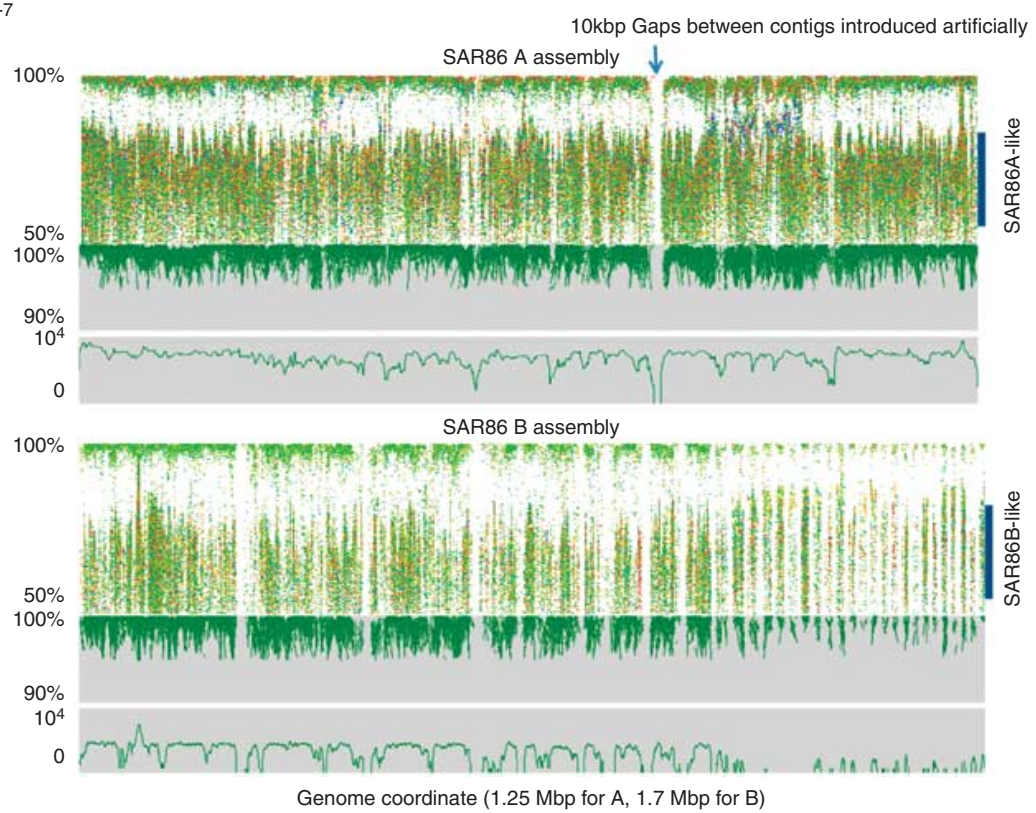
- Short-read assemblers
- Construct some variant of a read overlap graph: usually *k-mer* graph (*de Bruijn* graph)
- Assembly contigs constructed by traversing this graph
- Metagenomic assemblers include
  - MetaVelvet (Namiki et al., *ACM BCBB* 2011)
  - IDBA-UD (Peng et al., *Bioinformatics* 2012)
  - Ray Meta (Boisvert et al., *Genome Biol.* 2012)
- Assemblers for single genomes also used frequently
  - Celera Assembler (Adams et al., *Science* 2000)
  - SOAPdenovo (Li et al., *Genome Res.* 2010)

Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage

Dupont, Rusch et al, 2012 *ISME J*



Feasibility and utility of genome assembly



Gene prediction from metagenomic data

- Accurate and fast *de novo* gene finders are available that can predict protein coding genes from either reads or assembled contigs

- Examples
  - MetaGeneAnnotator (Noguchi et al., *DNA Res.* 2008)
  - FragGeneScan (Rho et al., *Nucleic Acids Res.* 2010)
  - Glimmer-MG (Kelley et al., *Nucleic Acids Res.* 2012)
  - MetaGeneMark (Zhu et al., *Nucleic Acids Res.* 2010)

Functional annotation
Assigning name and function to a predicted protein

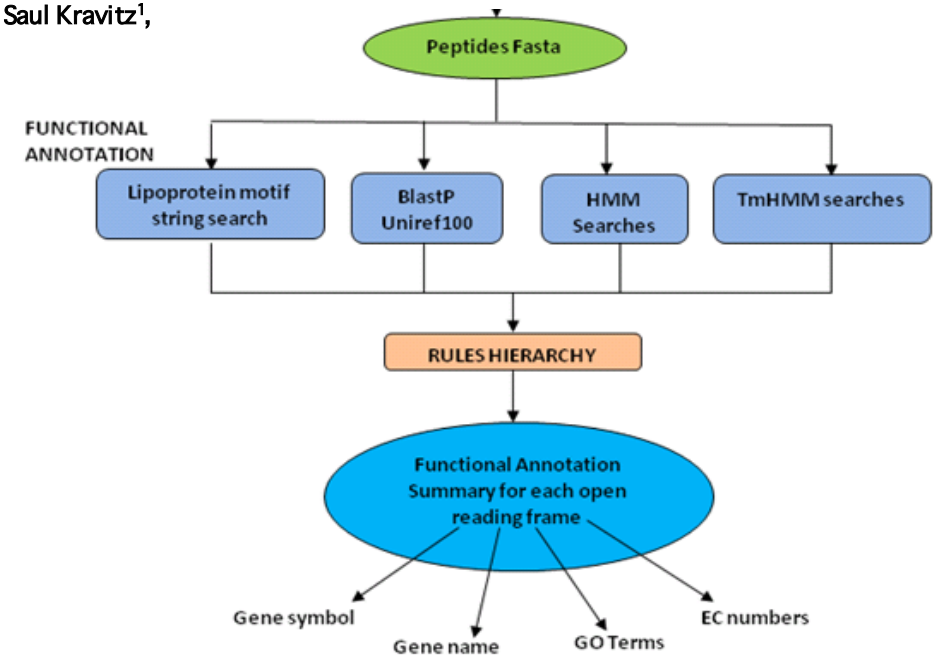
Standards in Genomic Sciences (2010) 2:229-237

The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data

David M. Tanenbaum[1], Johannes Goll[1], Sean Murphy[1], Prateek Kumar[2], Nikhat Zafar[1], Mathangi Thiagarajan[1], Ramana Madupu[1], Tanja Davidsen[1], Leonid Kagan[1], Saul Kravitz[1], Douglas B. Rusch[1], Shibu Yooseph[2]*

[1] J. Craig Venter Institute, Rockville, MD 20850
[2] J. Craig Venter Institute, San Diego, CA 92121

*Corresponding author: Shibu Yooseph (syooseph@jcvi.org)



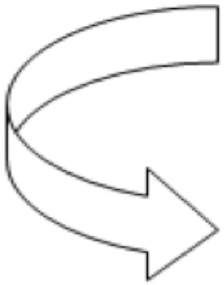
JCVI's functional annotation pipeline was used by the NIH funded Human Microbiome Project for annotating the metagenomic data generated by the project

J. Craig Venter INSTITUTE

Annotation Rules Hierarchy

Evidences

HMM Hits PFAM TIGRFAM **above trusted cut off**	**BlastP Hits** 1e-5; 35% ident; 80% cov **CHAR** database	**Rps PRIAM Hits** 1e-10 cutoff EC Numbers	**TmHMM**	**Lipoprotein Motif**



Annotation Rules

1	TIGRFAM/PFAM	(Equivalog)
2	Characterized (CHAR) BlastP Hit	
3	TIGRFAM/PFAM	(Non-Equivalog)
4	CDP (conserved domain protein) blastp hit	
5	TmHMM hit: "membrane protein"	
6	Lipoprotein motif: "lipoprotein"	
7	"hypthetical protein"	

Common Names, Gene Symbols, EC Numbers, GO Terms, TIGR Role ids

J. Craig Venter™ INSTITUTE

Assembly in amino acid space

Inference of complete protein sequences from metagenomic data sets can provide a more accurate picture of the functional and metabolic potential of the microbial community

Nucleic Acids Research, 2013, ***1–10***
doi:10.1093/nar/gkt118

SPA: a short peptide assembler for metagenomic data

Youngik Yang and Shibu Yooseph*

Informatics Department, J. Craig Venter Institute, San Diego, CA 92121, USA

Received October 2, 2012; Revised February 1, 2013; Accepted February 5, 2013

J. Craig Venter™
INSTITUTE

Visualization and Data Access

METAREP
JCVI Metagenomics Reports

website www.jcvi.org/metarep
source code http://github.com/jcvi/METAREP
blog http://blogs.jcvi.org/tag/metarep
contact metarep-support@jcvi.org

This work was supported by the US Department of Energy [#DEFG02-02ER63453, #DE-FC02-02ER63446], National Institute of Allergy and Infectious Diseases [1U54AI084844], National Cancer Institute [UH2CA14023], Sloan Foundation [#2004-5-46EG], University of Illinois, and the Department of Primary Industries, Victoria, Australia.

BIOINFORMATICS

METAREP: JCVI Metagenomics Reports - an open source tool for high-performance comparative metagenomics

Johannes Goll [1], Doug Rusch [1], David M. Tanenbaum [1],Mathangi Thiagarajan [1], Kelvin Li [1], Barbara A. Methé [1], Shibu Yooseph [1*]

[1]The J. Craig Venter Institute, Rockville, MD 20850, USA

J. Craig Venter™
INSTITUTE

METAREP – Browse

[Browse Blast Taxonomy selected node: Bacteria]

JCVI Metagenomics Reports

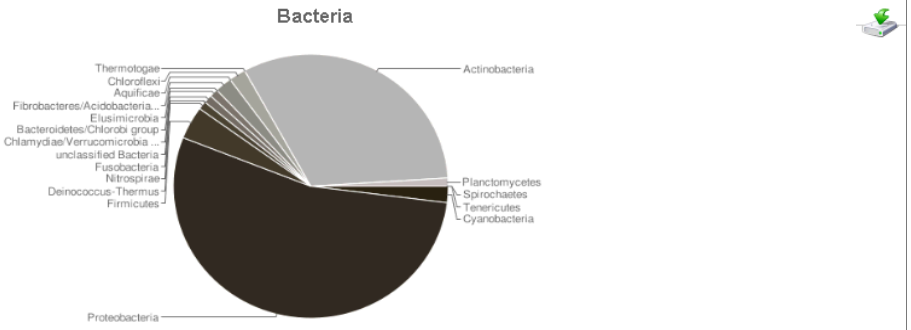
QUICK NAVIGATION | NEW PROJECT | LIST PROJECTS | LIST POPULATIONS | LIST LIBRARIES | PIPELINE PROGRESS LOG

Browse Taxonomy (Blast) F16ZRB301 (Australian Soil)

NCBI Taxonomy Tree

Help

- Viruses (no rank) [135 peptides]
- other sequences (no rank) [3 peptides]
- cellular organisms (no rank) [209,232 peptides]
  - Bacteria (superkingdom) [205,749 peptides]
    - Cyanobacteria (phylum) [4,206 peptides]
    - Proteobacteria (phylum) [110,229 peptides]
    - Firmicutes (phylum) [8,840 peptides]
    - Deinococcus-Thermus (phylum) [1,214 peptides]
    - unclassified Bacteria (no rank) [13 peptides]
    - Fusobacteria (phylum) [42 peptides]
    - Nitrospirae (phylum) [4 peptides]
    - Chlamydiae/Verrucomicrobia group (superphylum) [1,085 peptic
    - Bacteroidetes/Chlorobi group (superphylum) [2,600 peptides]
    - Elusimicrobia (phylum) [43 peptides]
    - Fibrobacteres/Acidobacteria group (superphylum) [4,246 peptid
    - Aquificae (phylum) [109 peptides]
    - Chloroflexi (phylum) [4,487 peptides]
    - Thermotogae (phylum) [426 peptides]
    - Actinobacteria (phylum) [65,161 peptides]
    - Planctomycetes (phylum) [2,672 peptides]
    - Spirochaetes (phylum) [334 peptides]
    - Tenericutes (phylum) [38 peptides]
  - Archaea (superkingdom) [1,942 peptides]
  - Eukaryota (superkingdom) [1,541 peptides]



Top Ten Functional Classifications

Blast Species
1. Nocardioides sp. JS614 (16179) (7.86%)
2. Sphingomonas wittichii RW1 (12364) (6.01%)
3. Sphingopyxis alaskensis RB2256 (6125) (2.98%)
4. Novosphingobium aromaticivorans DSM 12444 (5918) (2.88%)
5. Erythrobacter litoralis HTCC2594 (5258) (2.56%)
6. Rubrobacter xylanophilus DSM 9941 (4105) (2%)
7. Bradyrhizobium japonicum USDA 110 (3623) (1.76%)
8. Mesorhizobium loti MAFF303099 (3264) (1.59%)
9. Saccharopolyspora erythraea NRRL 2338 (2990) (1.45%)
10. Kineococcus radiotolerans SRS30216 (2893) (1.41%)

Common Name
1. hypothetical protein (33741) (16.4%)
2. protein of unknown function (2998) (1.46%)
3. ABC transporter related (1644) (0.8%)
4. major facilitator superfamily MFS_1 (967) (0.47%)
5. TonB-dependent receptor (874) (0.42%)
6. NADH dehydrogenase (quinone) (867) (0.42%)
7. short-chain dehydrogenase/reductase SDR (793) (0.39%)
8. DNA-directed DNA polymerase (790) (0.38%)
9. binding-protein-dependent transport systems inner membrane component (707) (0.34%)
10. histidine kinase (677) (0.33%)

Go Term
1. unassigned (191760) (93.2%)
2. GO:0006355 | regulation of transcription, DNA-dependent (2807) (1.36%)
3. GO:0003700 | transcription factor activity (1829) (0.89%)
4. GO:0003677 | DNA binding (1142) (0.56%)
5. GO:0000160 | two-component signal transduction system (phosphorelay) (979) (0.48%)
6. GO:0006412 | translation (874) (0.42%)
7. GO:0008152 | metabolic process (865) (0.42%)
8. GO:0003735 | structural constituent of ribosome (844) (0.41%)
9. GO:0055114 | oxidation reduction (790) (0.38%)
10. GO:0005524 | ATP binding (716) (0.35%)

Ec number
1. unassigned (147365) (71.62%)
2. 2.7.7.7 | DNA-directed DNA polymerase. (966) (0.47%)
3. 1.6.99.5 | NADH dehydrogenase (quinone). (910) (0.44%)
4. 2.7.7.6 | DNA-directed RNA polymerase. (669) (0.33%)
5. 2.7.13.3 | Histidine kinase. (575) (0.28%)
6. 1.1.1.100 | 3-oxoacyl-[acyl-carrier-protein] reductase. (537) (0.26%)
7. 3.6.3.17 | Monosaccharide-transporting ATPase. (491) (0.24%)
8. 1.9.3.1 | Cytochrome-c oxidase. (458) (0.22%)
9. 5.99.1.3 | DNA topoisomerase (ATP-hydrolyzing). (402) (0.2%)
10. 3.6.3.31 | Polyamine-transporting ATPase. (352) (0.17%)

HMM
1. unassigned (178117) (86.57%)
2. PF00072 | response regulator receiver domain (463) (0.23%)
3. PF02518 | ATPase, histidine kinase-, DNA gyrase B-, and HSP90-like domain protein (411) (0.2%)
4. PF00353 | type I secretion target GGXGXDXX repeat (2 copies) (286) (0.14%)
5. PF04542 | Sigma-70 region 2 (268) (0.13%)
6. PF08240 | Alcohol dehydrogenase GroES-like domain (262) (0.13%)
7. PF00512 | His Kinase A (phosphoacceptor) domain (250) (0.12%)
8. PF00196 | transcriptional regulator, LuxR family (225) (0.11%)
9. PF00440 | transcriptional regulator, TetR family (217) (0.11%)
10. PF08281 | Sigma-70, region 4 (207) (0.1%)

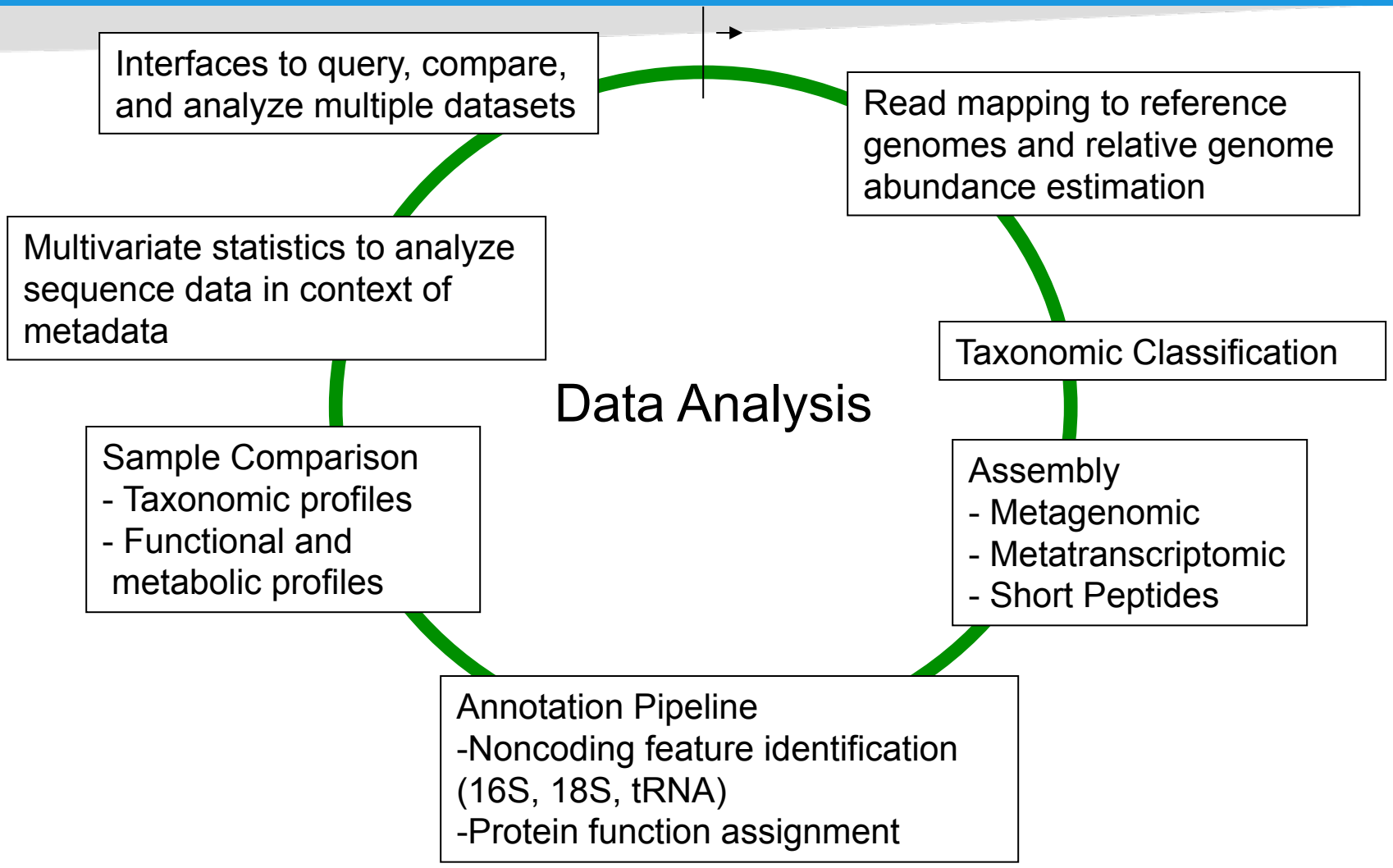
Top Ten Functional Pie Charts

Metatranscriptomic data analysis

- Simultaneous RNAseq for multiple organisms!
- Initial processing has similar flavor as metagenomic data
  - Combination of read mapping and assembly
- Obtain expression levels based on reads mapping to genes
- Differential abundance analyses can be done using tools developed for RNAseq, including
  - DESeq (Anders and Huber, *Genome Biology* 2010)
  - EdgeR (Robinson et al., *Bioinformatics* 2010)

J. Craig Venter™
INSTITUTE

High-throughput Data Analysis



J. Craig Venter™ INSTITUTE

Dr. William Nierman
Overview of Emerging Infectious Diseases

Emerging Infectious Diseases

William C. Nierman
Professor
Infectious Diseases Program Director
J. Craig Venter Institute

J. Craig Venter™
INSTITUTE

Emerging Infectious Diseases
What Are They

- A disease that newly appears in a human population.
- An established disease that has a rapidly increase in incidence or an expansion of geographic range.

J. Craig Venter™
INSTITUTE

The Process of Emerging

1. A disease agent is introduced to a new host.

2. The disease becomes established and further disseminated within the population.

J. Craig Venter™
INSTITUTE

Factors That Promote Disease Emergence

- Ecological change and agricultural development
- Human demographics or behavior
- International travel and commerce
- Technology and industrial activity
- Microbial adaptation and changes
- Breakdown/deficiencies in public health infrastructure

J. Craig Venter™
INSTITUTE

The Process of Emerging Human Behavior

Measles Cases and Outbreaks

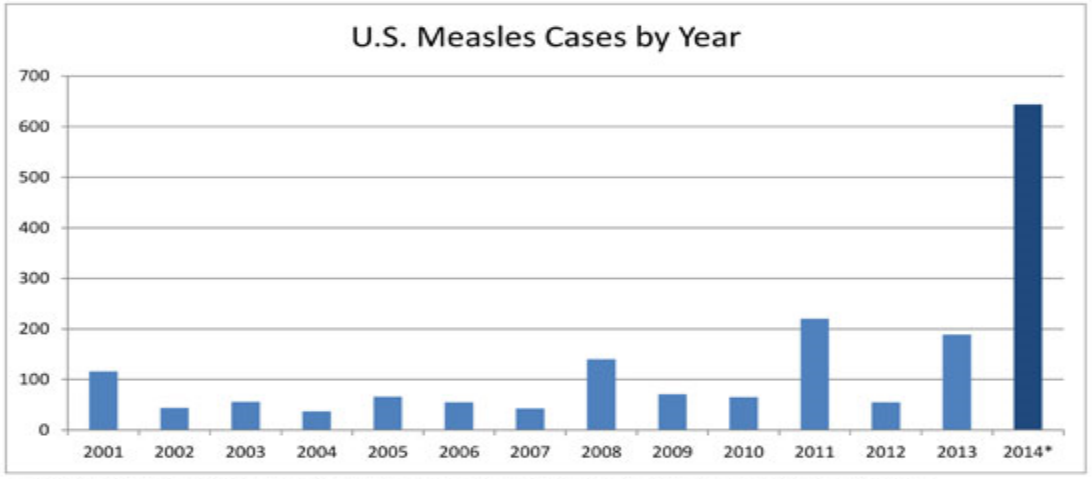
During 2014*

644 Cases

23 Outbreaks

reported in 27 states: Alabama, California, Colorado, Connecticut, Hawaii, Illinois, Indiana, Kansas, Massachusetts, Michigan, Minnesota, Missouri, Nebraska, New Jersey, New Mexico, New York, North Carolina, Ohio, Oregon, Pennsylvania, South Dakota, Tennessee, Texas, Utah, Virginia, Washington, Wisconsin

representing 89% of reported cases this year



*Provisional data reported to CDC's National Center for Immunization and Respiratory Diseases

CDC

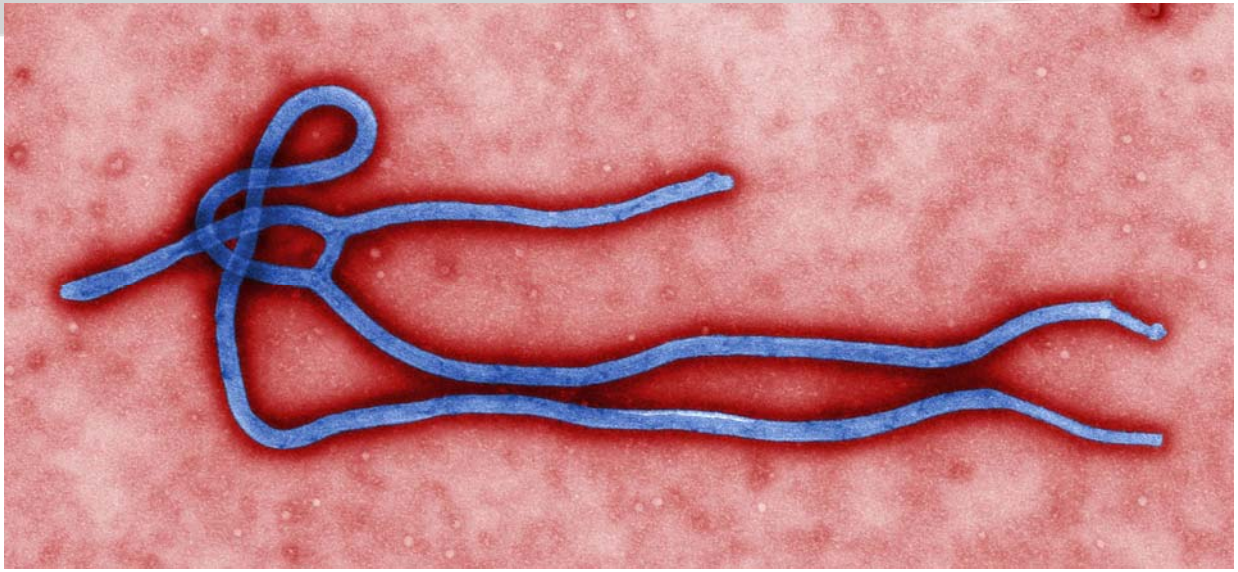
J. Craig Venter™ INSTITUTE

2014 Outbreaks and Notable Infectious Diseases

- Ebola (West Africa)
- MERS (Saudi Arabia)
- Avian Influenza H7N9 (China)
- Whooping Cough (US)
- Carbapenem-resistant *Klebsiella pneumoniae* (US)
- Chikungunya (Caribbean, Dr. Carrington to present)

J. Craig Venter™ INSTITUTE

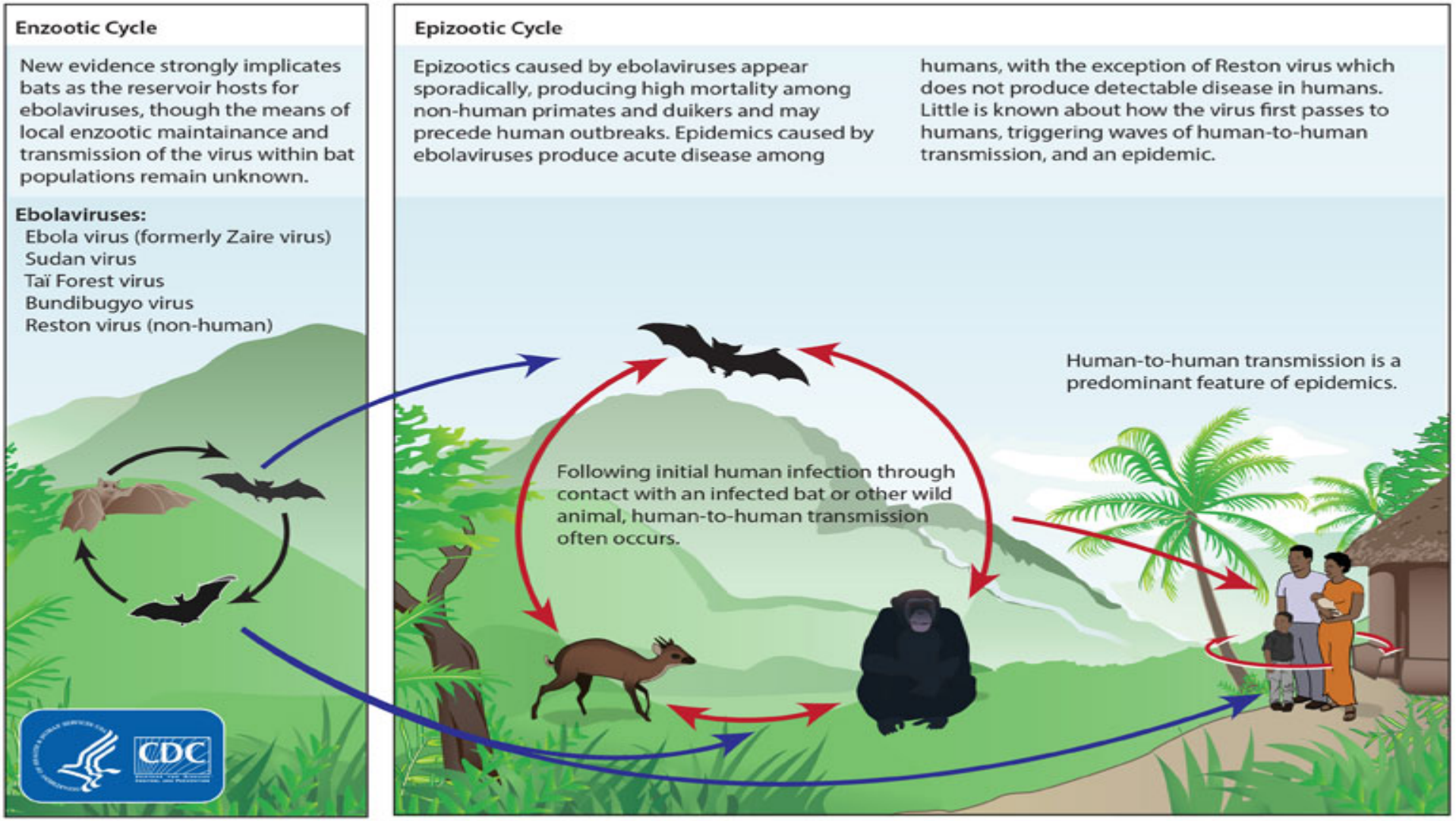
Ebola



- Negative strand RNA genome
- Five closely related Marburg viruses
- Spread by direct contact

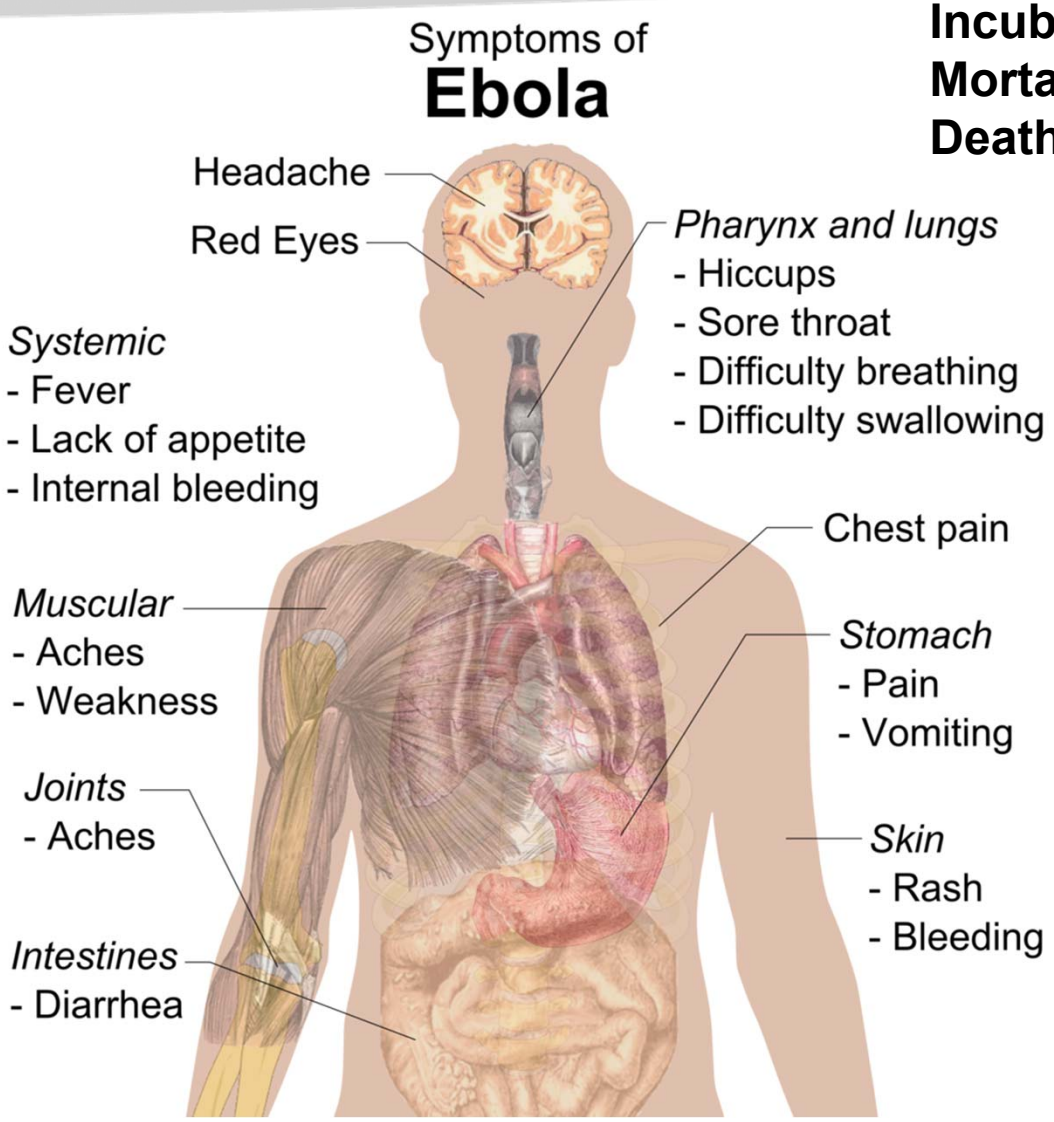
J. Craig Venter™
INSTITUTE

Ebola


Enzootic Cycle
New evidence strongly implicates bats as the reservoir hosts for ebolaviruses, though the means of local enzootic maintainance and transmission of the virus within bat populations remain unknown.
Ebolaviruses:
Ebola virus (formerly Zaire virus)
Sudan virus
Taï Forest virus
Bundibugyo virus
Reston virus (non-human)
Epizootic Cycle
Epizootics caused by ebolaviruses appear sporadically, producing high mortality among non-human primates and duikers and may precede human outbreaks. Epidemics caused by ebolaviruses produce acute disease among humans, with the exception of Reston virus which does not produce detectable disease in humans. Little is known about how the virus first passes to humans, triggering waves of human-to-human transmission, and an epidemic.
Human-to-human transmission is a predominant feature of epidemics.
Following initial human infection through contact with an infected bat or other wild animal, human-to-human transmission often occurs.
CDC

Ebola

Incubation period 2das – 3 wks
Mortality 50%
Death 6-16 das post symptoms



J. Craig Venter™ INSTITUTE

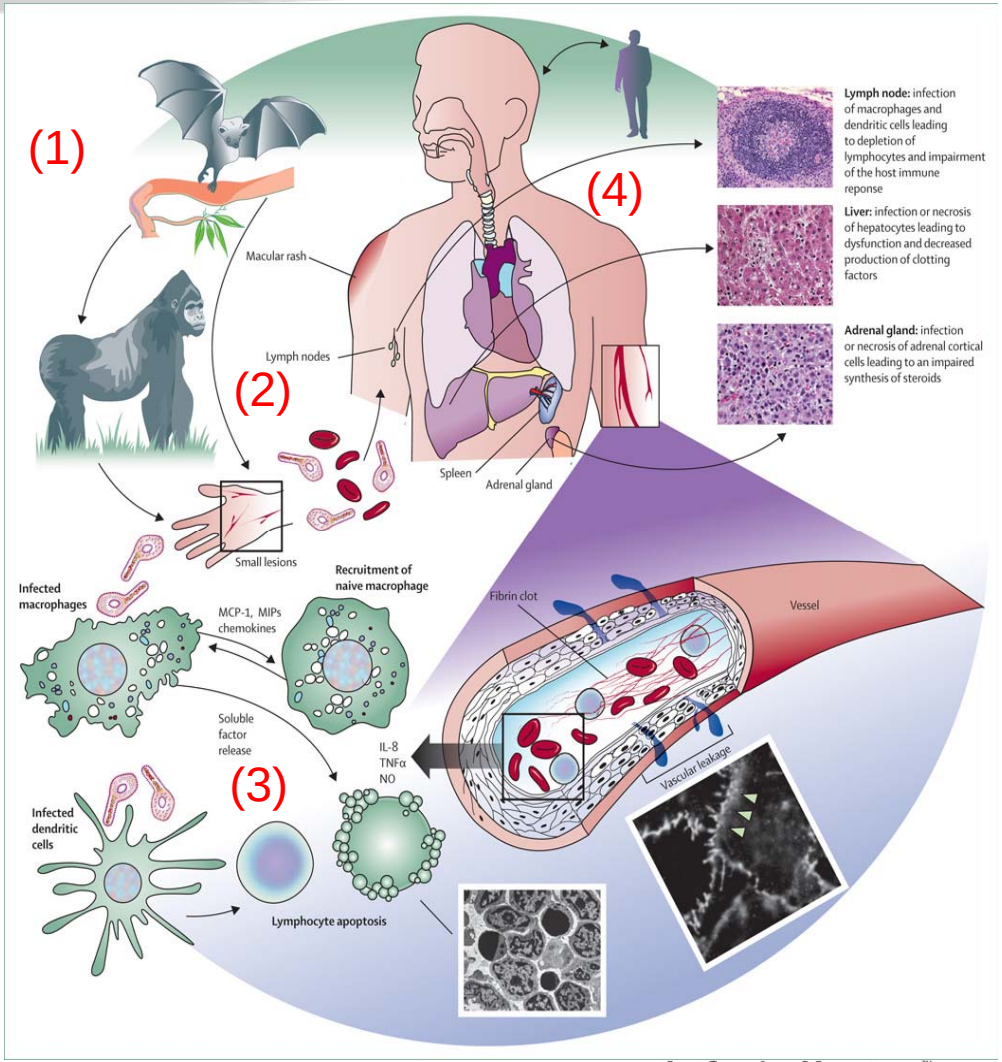
A model of Ebolavirus Pathogenesis

(1) Bat reservoirs spill infect either humans or non-human primates

(2) Ebola spreads from the initial infection site to regional lymph nodes

(3) Innate and adaptive immune suppression mediated by virally encoded proteins allows for systemic viral spread

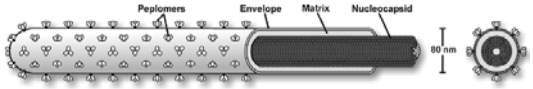
(4) Systemic viral spread and immune dysregulation leads to multi-organ failure and severe disease

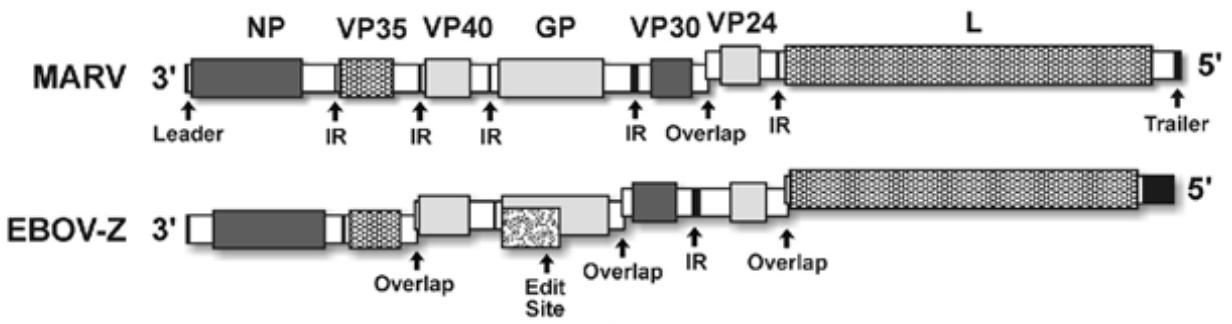


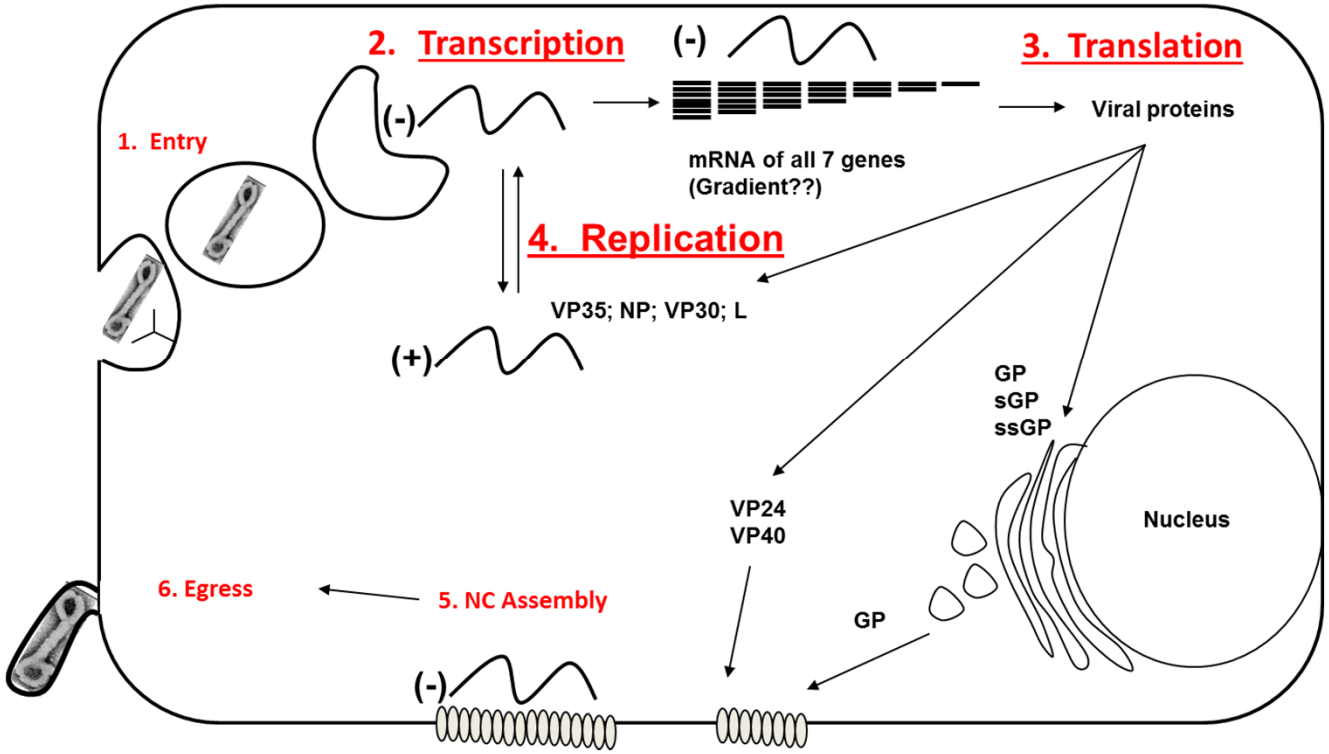
Modified from: Lancet. 2011 Mar 5;377(9768):849-62

J. Craig Venter INSTITUTE

Filovirus Genomic Organization and Replication Cycle

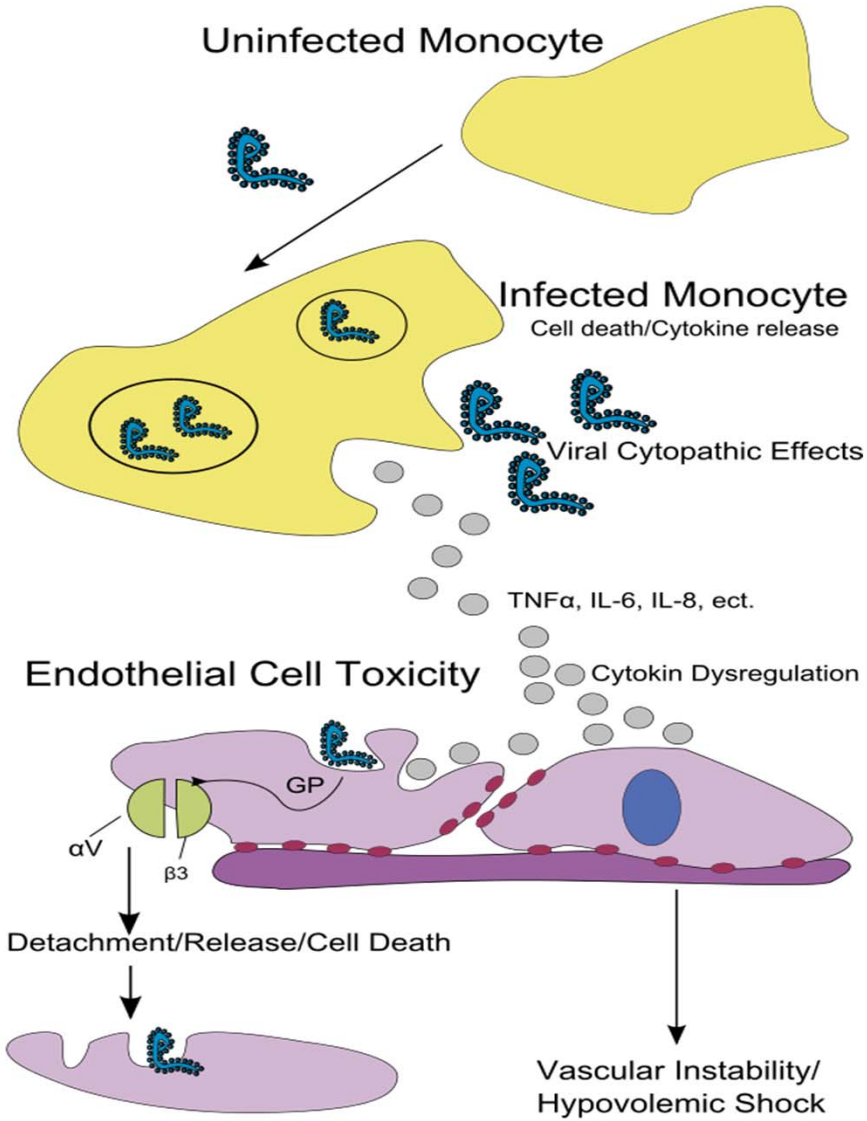






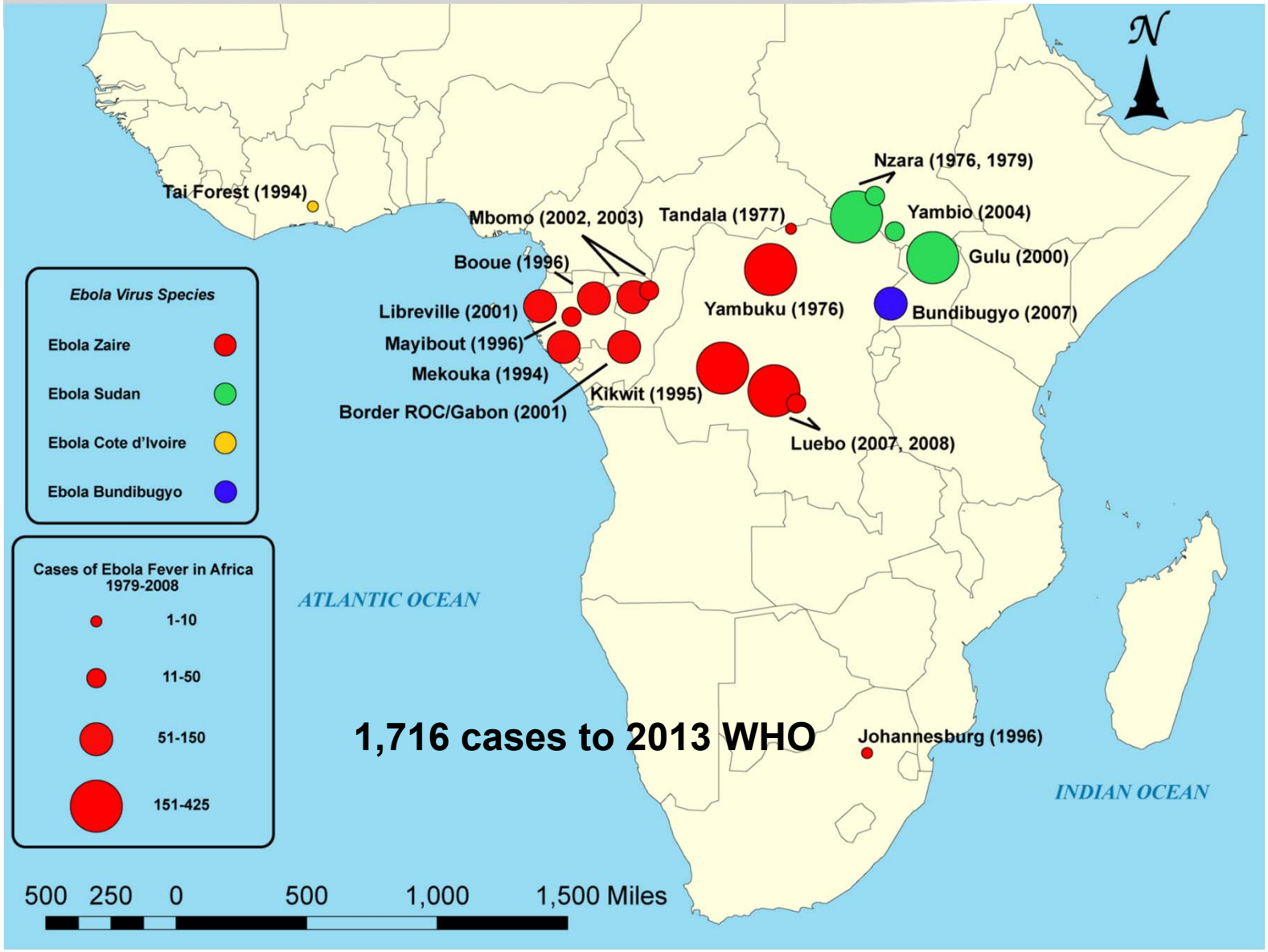
J. Craig Venter INSTITUTE

Ebola



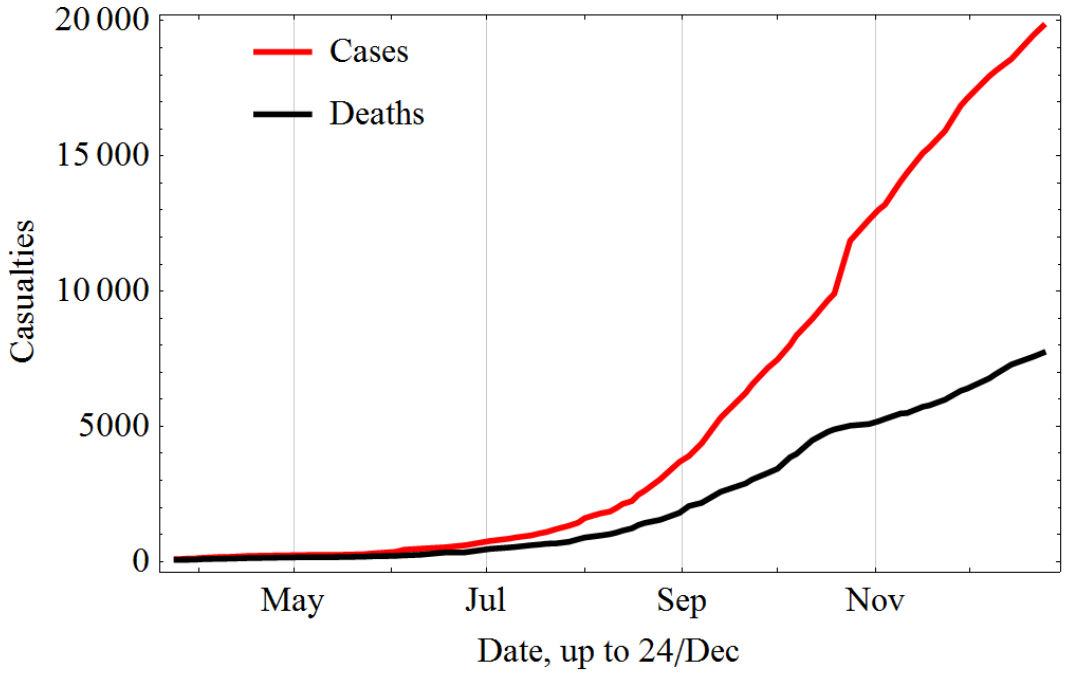
J. Craig Venter™
INSTITUTE

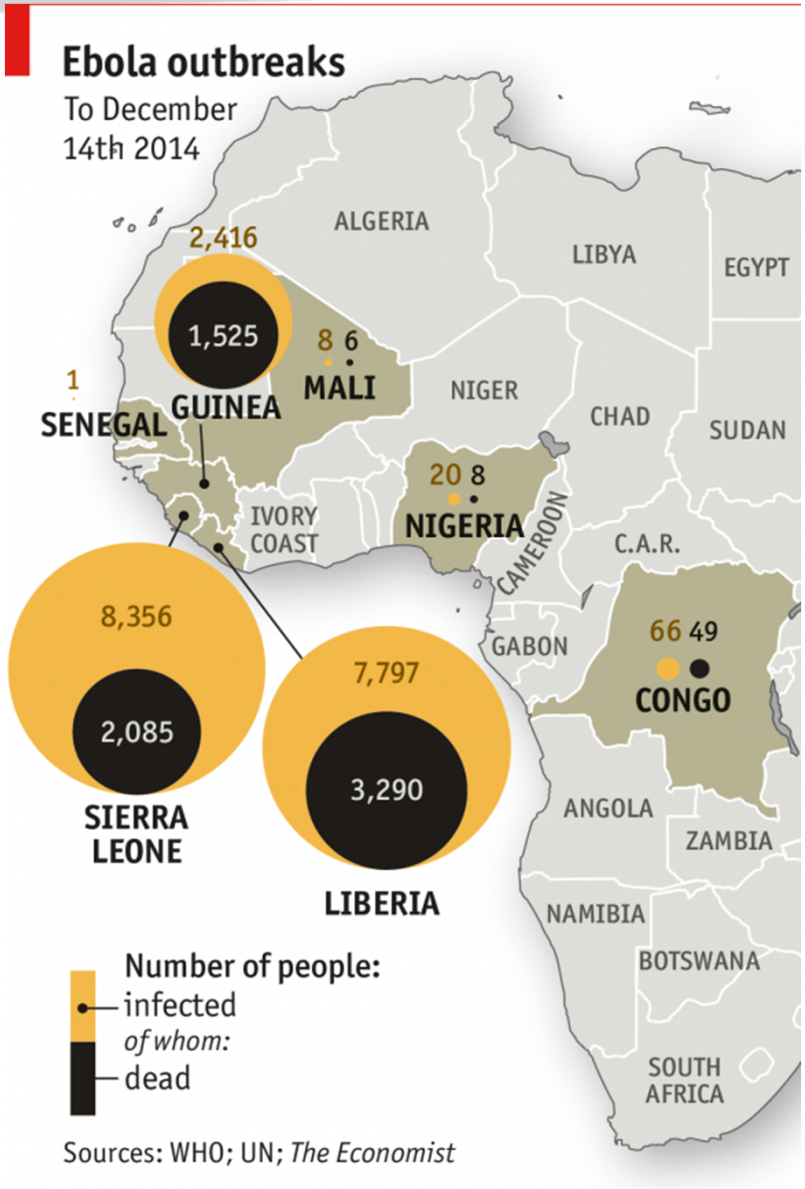
Ebola



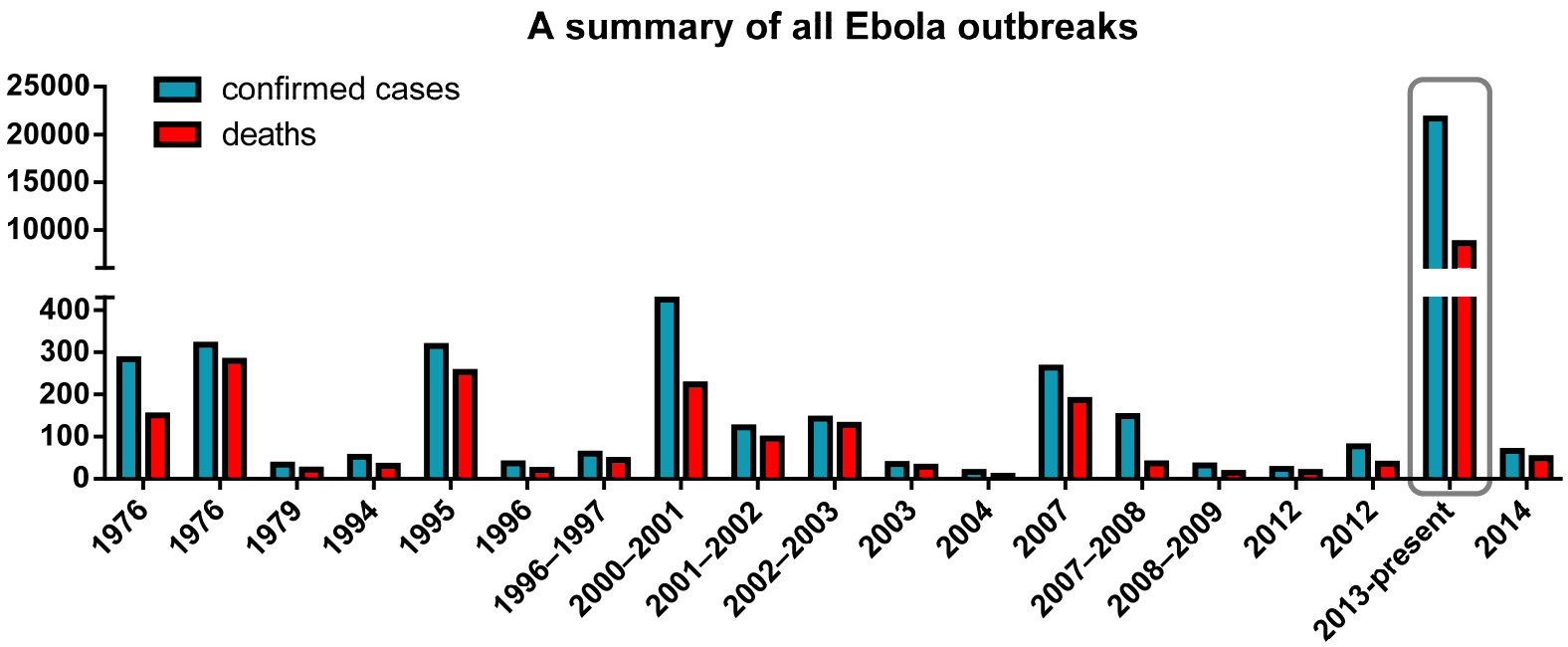
J. Craig Venter™ INSTITUTE

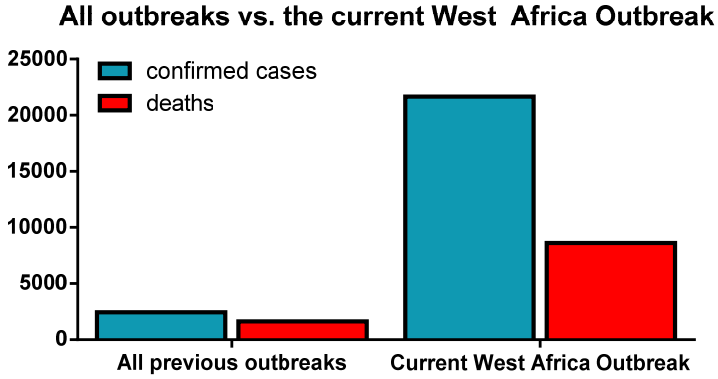
Ebola


20 000
15 000
10 000
5000
0
Casualties
Cases
Deaths
May
Jul
Sep
Nov
Date, up to 24/Dec


Ebola outbreaks
To December 14th 2014
ALGERIA
LIBYA
EGYPT
2,416
1,525
8 6
MALI
1
SENEGAL
GUINEA
NIGER
CHAD
SUDAN
20 8
NIGERIA
IVORY COAST
CAMEROON
C.A.R.
8,356
2,085
SIERRA LEONE
7,797
3,290
LIBERIA
GABON
66 49
CONGO
ANGOLA
ZAMBIA
NAMIBIA
BOTSWANA
SOUTH AFRICA
Number of people:
infected
of whom:
dead
Sources: WHO; UN; The Economist

The Size of the Current Ebola Outbreak in West Africa is Unprecedented





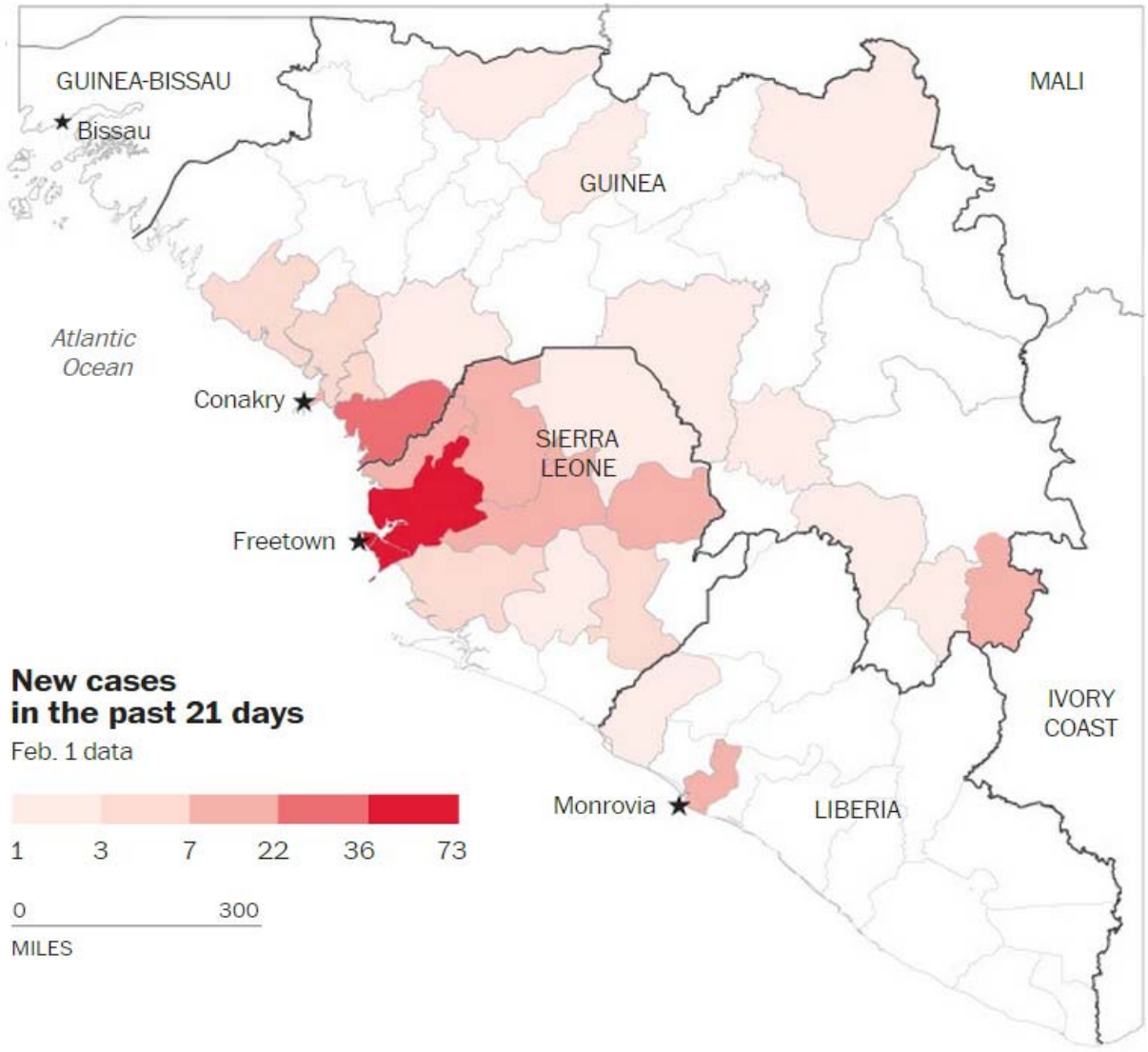
J. Craig Venter™
INSTITUTE

Ebola
Factors Promoting Outbreak

- Population density – rural vs urban
- Breakdown/deficiencies in public health infrastructure

J. Craig Venter™
INSTITUTE

Ebola
Recent Status



J. Craig Venter INSTITUTE

Ebola
Lessons Learned – Disrupt Transmission

- Simple changes to disrupt transmission can provide significant results – hospitals and funerals not a source of transmission
  - Early detection of infected patients (village monitors)
  - Early isolation of patients even in rudimentary facilities
  - 82% of cases from the community, 72% of cases from family members

J. Craig Venter™
INSTITUTE

Ebola
Lessons Learned

- Rely on local leadership
- Be sensitive to local cultures
- Speed and agility is more important than size
- Ounce of prevention
- Keep fear in check
- Prepare for rapid world response to outbreaks

J. Craig Venter™
INSTITUTE

Ebola
Future Management

- Vaccine development consortia produced a vaccine now in phase 1 trials (Johnson & Johnson, Bavarian Nordic A/S).
- US FDA being flexible in allowing for vaccination of at risk workers after a successful phase 1 trial.

J. Craig Venter™
INSTITUTE

Ebola and Marburg virus sequencing / analysis @ JCVI

1. ***Sequencing and analysis of genomic RNA from Ebola and Marburgvirus stock strains.***

- *Obtain gold standard complete coding genome consensus sequences for each of these stocks.*
- *Determine if the stocks are free of other contaminating pathogens, including other filovirus strains.*
- *Determine the level of single nucleotide polymorphisms (SNP) within the virus populations, and if possible identify other potential contaminants in the sample.*

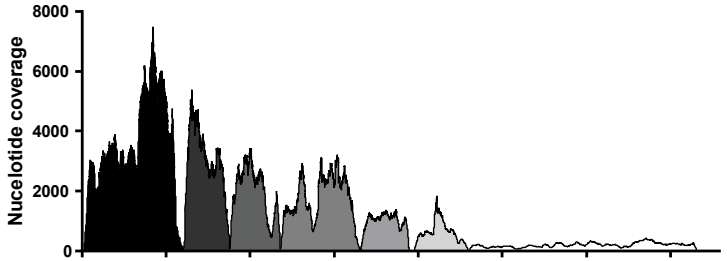
2. ***Analysis of deep sequencing data sets from Ebola and Marburg virus messenger RNAs in infected cells.***

Data recently published: MBio. 2014 Nov 4;5(6):e02011

J. Craig Venter™
INSTITUTE

JCVI Ebola Project Findings

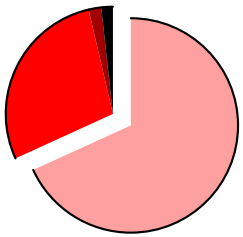
- We identified the presence of a filovirus transcriptional gradient



- We highlight regions of filovirus genomes prone to minor variants, possibly caused by RNA editing via cellular enzymes

Genome position*	Designation	A	C	G	T	Change	% of minor variant	Type of change	Amino acid
2590	NP 3'-UTR	0	687	0	8164	T/C	8%	3'-UTR +398	
2598	NP 3'-UTR	1	674	0	7755	T/C	8%	3'-UTR +406	
2640	NP 3'-UTR	0	704	1	5946	T/C	11%	3'-UTR +448	
2643	NP 3'-UTR	1	699	1	5968	T/C	10%	3'-UTR +451	
2663	NP 3'-UTR	0	800	1	6101	T/C	12%	3'-UTR +471	
2724	NP 3'-UTR	1	926	0	8981	T/C	9%	3'-UTR +532	
2822	NP 3'-UTR	0	566	0	5349	T/C	10%	3'-UTR +630	
2824	NP 3'-UTR	2	568	1	5080	T/C	10%	3'-UTR +632	
2834	NP 3'-UTR	947	0	224	0	A/G	19%	3'-UTR +642	
2836	NP 3'-UTR	0	108	0	350	T/C	24%	3'-UTR +644	
3400	VP35 ORF	9126	6	6	2038	A/T	18%	subst_NONSYNONYMOUS[TTA:L TTt:F c]	152
4283	VP35 3'-UTR	1	346	0	3801	T/C	8%	3'-UTR +348	

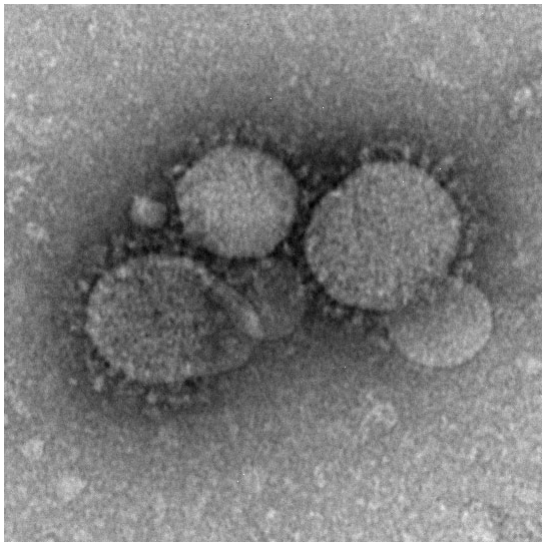
- We further describe the known Ebola glycoprotein editing site during the course of infection

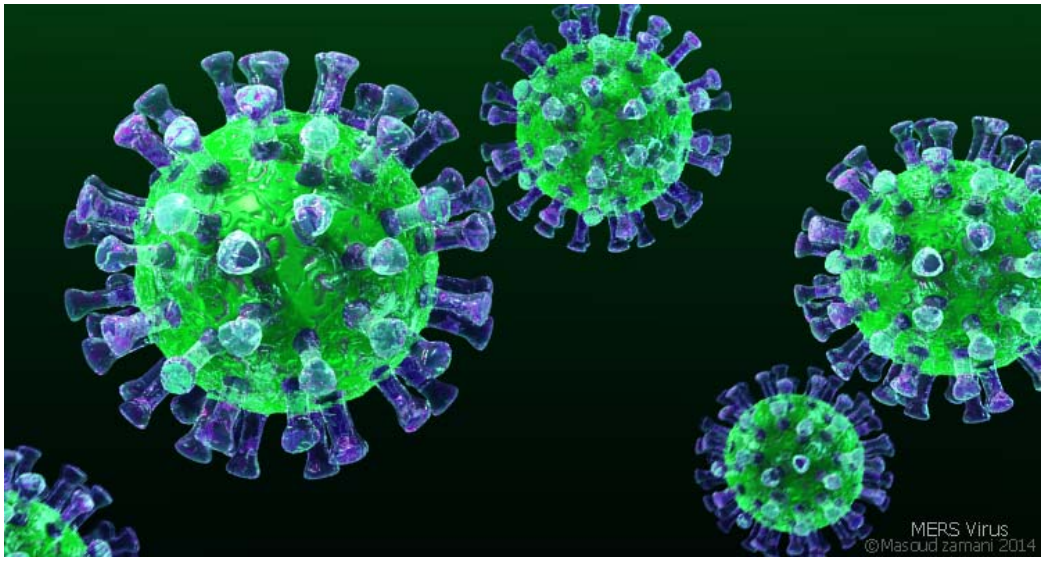


Reed Shabman PI

J. Craig Venter™ INSTITUTE

MERS-CoV

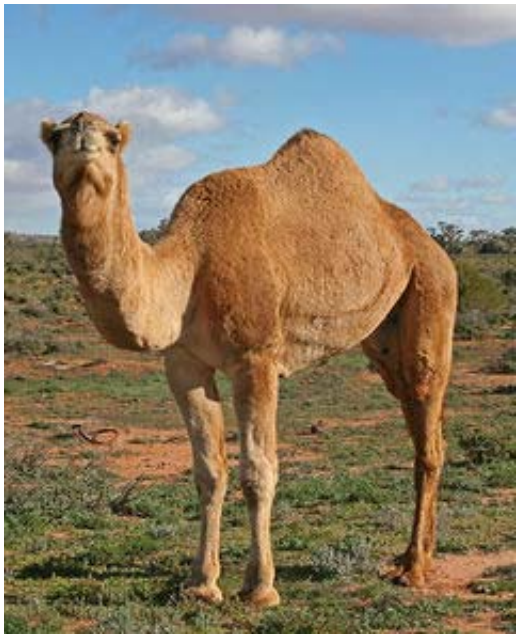




- Large positive strand RNA genome
- Tropism for respiratory epithelium
- Not particularly contagious
- Presentation symptoms: fever, cough, pneumonia

J. Craig Venter™
INSTITUTE

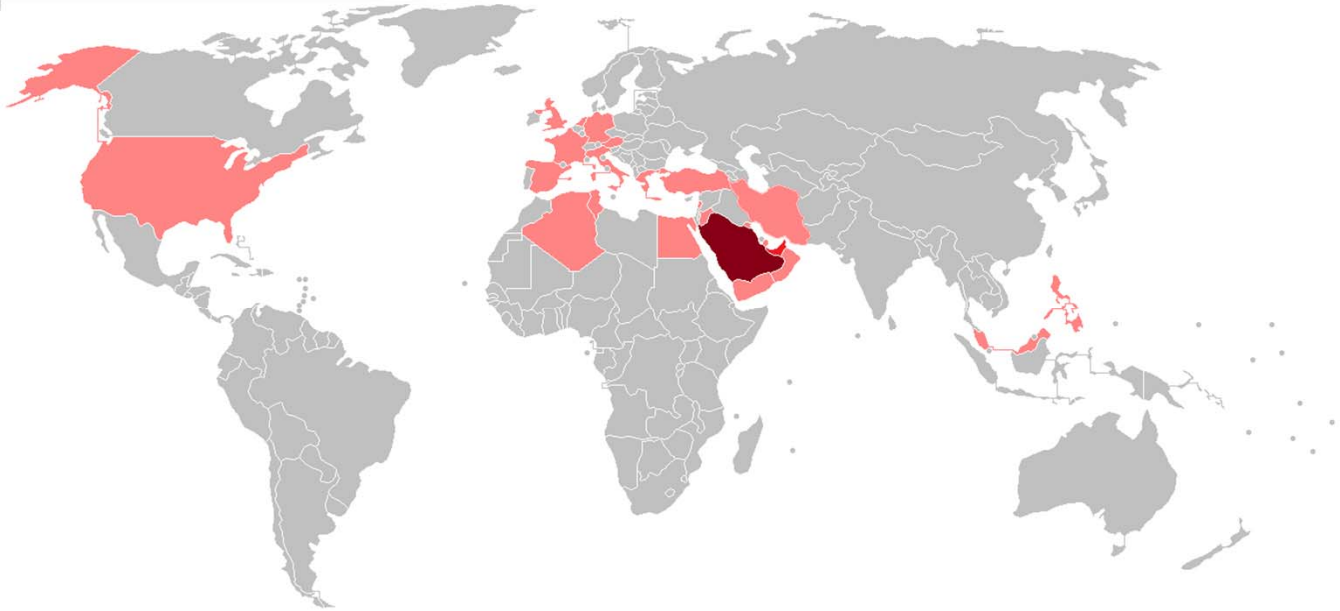
MERS-CoV
Source of Infection



- First patient virus matched one in Egyptian tomb bat.
- High incidence of virus in dromedary camel.
- Camel meat, milk, and urine consumed.

J. Craig Venter™
INSTITUTE

MERS-CoV
Case Geographies



- 938 cases, 343 deaths, nearly all linked to Saudi Arabia.
- Over half of cases are primary cases.
- First case in 2012, surge in spring 2014.

J. Craig Venter™
INSTITUTE

MERS
Factors Promoting/Inhibiting Outbreak

- Low person to person transmission
- Cultural/economic factors as camels are the major source of human infections
- 3.3 to 1 gender disparity
- International cases from travel
- Reason for the outbreak remains undetermined

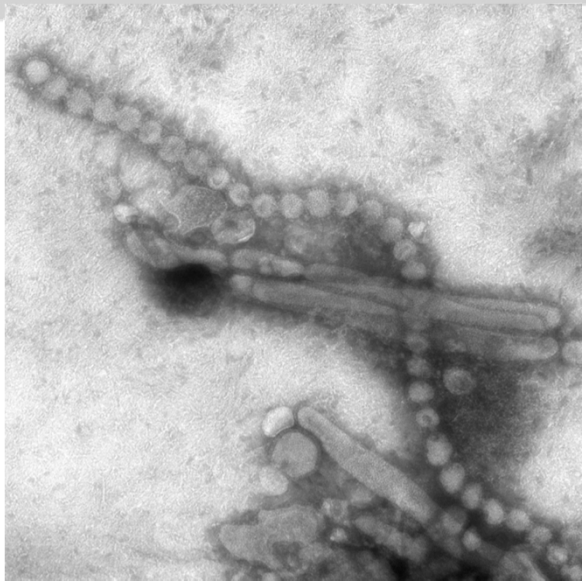
J. Craig Venter™
INSTITUTE

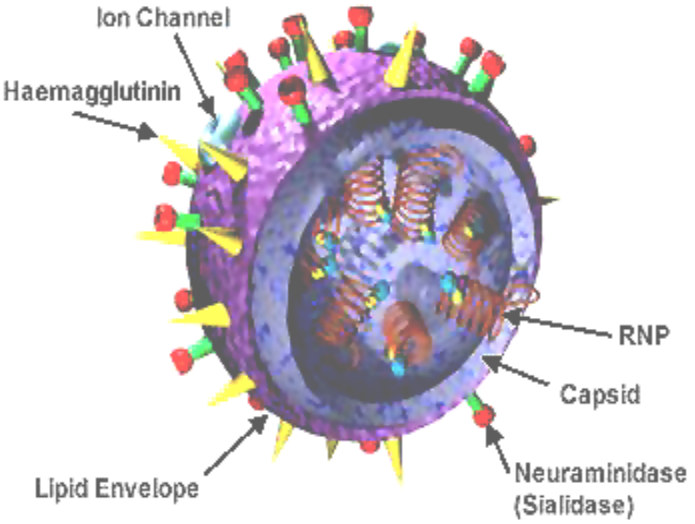
MERS-CoV
Future Management

- Camels get MERS infections but symptoms are as a cold, cleared in weeks
- Virus vented through nostrils during and after infection
- Camel vaccine is in testing
- Project is collaborative undertaking between NIAID (Vincent Munster) and Colorado State University (Richard Bowen)

J. Craig Venter™
INSTITUTE

Influenza

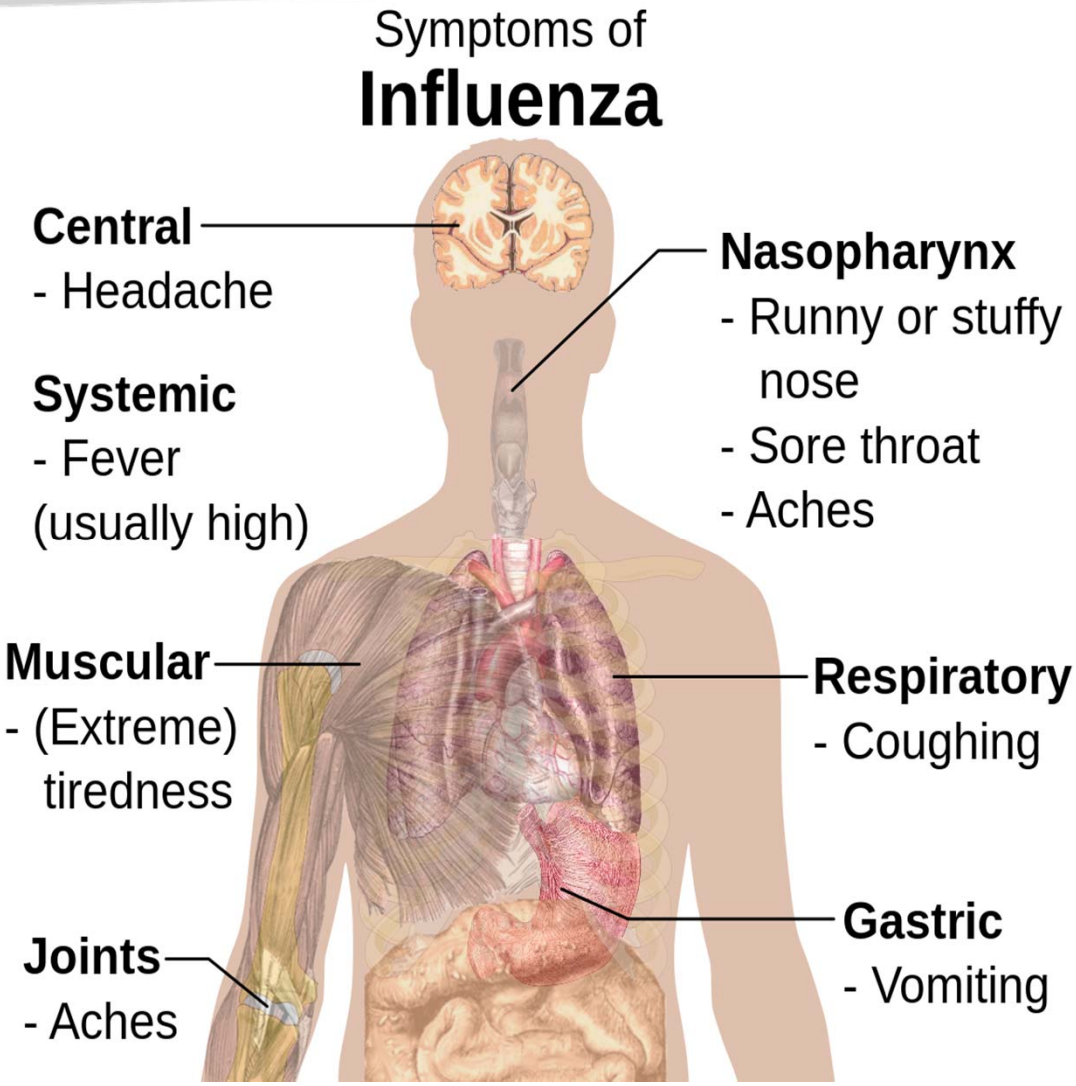




- Negative strand segmented SS RNA genome, 11 genes in influenza virus A
- Three species

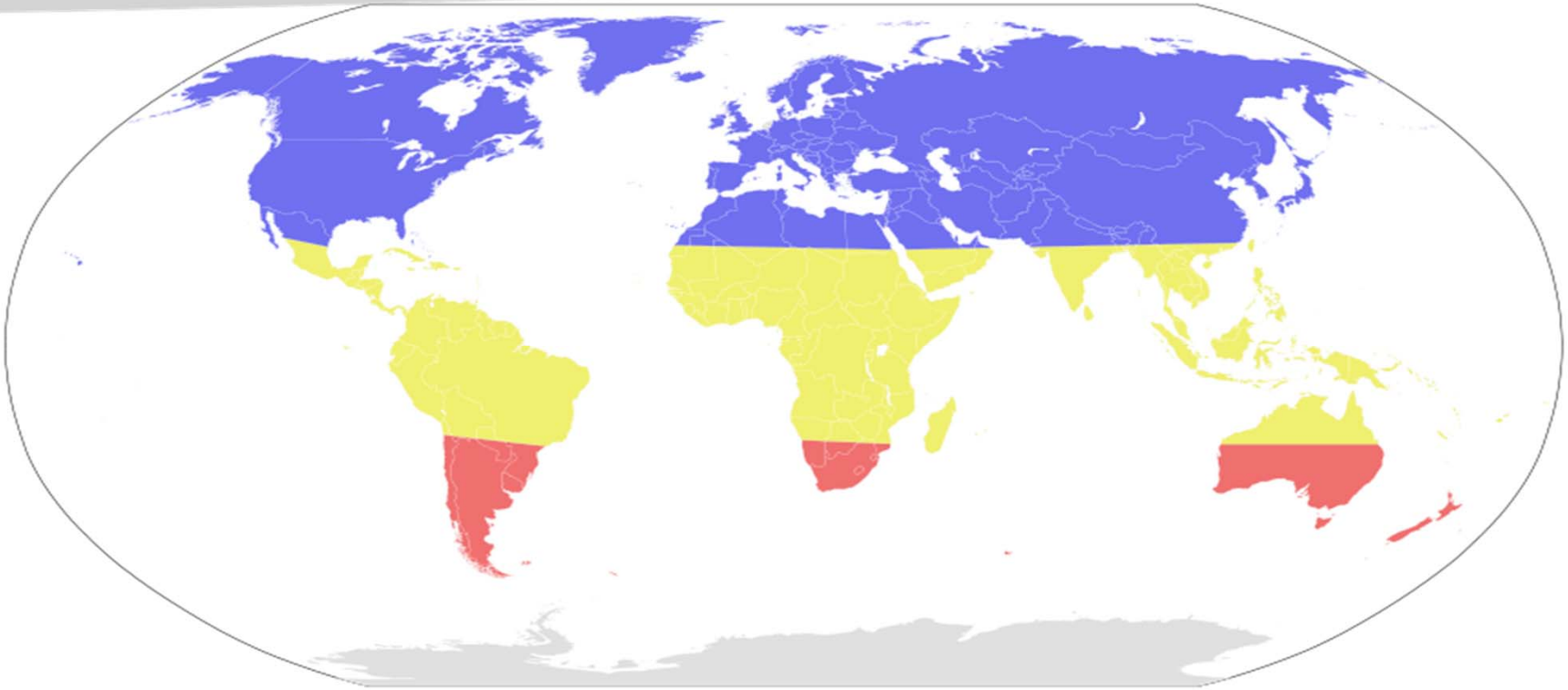
J. Craig Venter™
INSTITUTE

Influenza



J. Craig Venter™ INSTITUTE

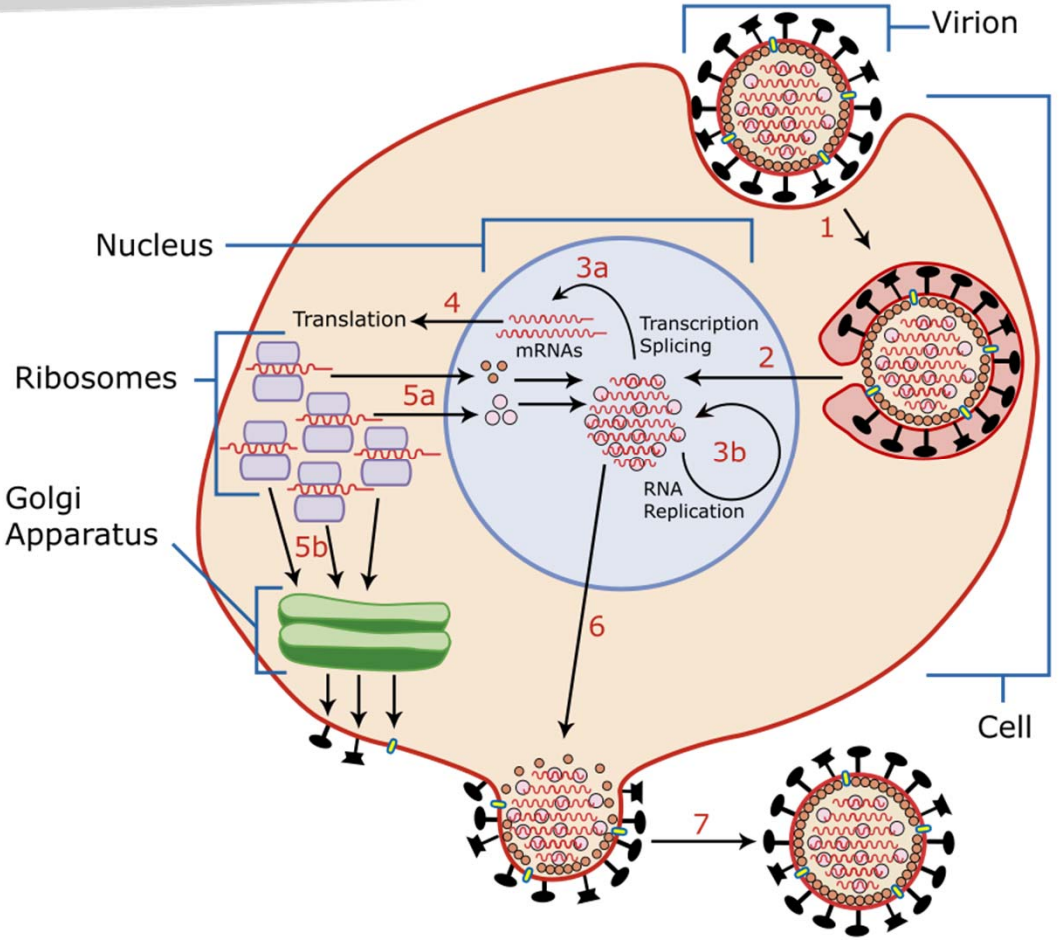
Influenza Seasonality



- Seasonal outbreaks in winter
- 3-5 million cases annually, 250-500K deaths
- Deaths occur mostly in young and old
- Person to person transmission

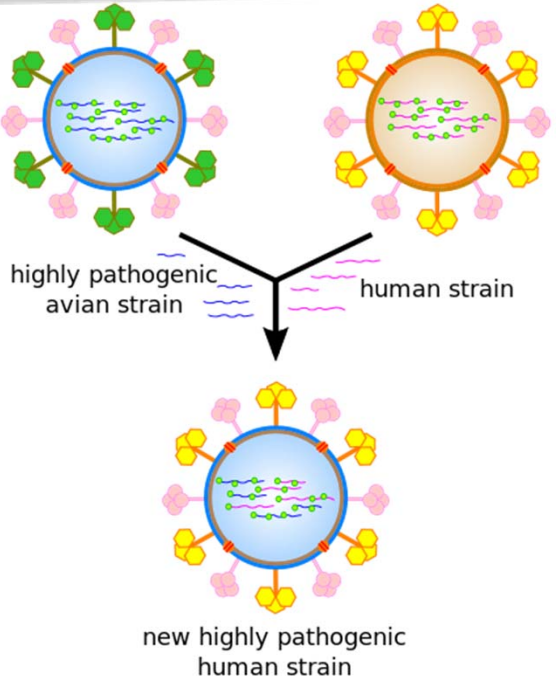
J. Craig Venter™ INSTITUTE

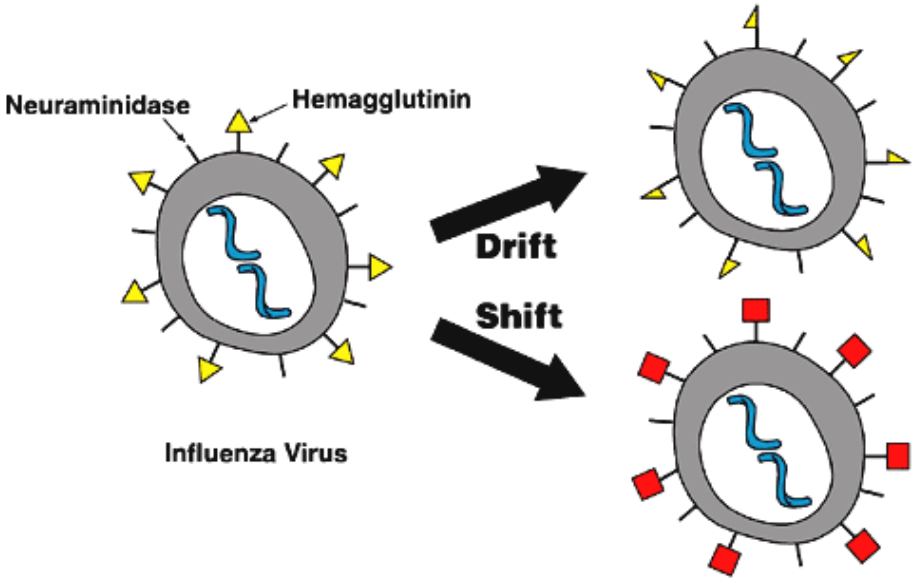
Influenza



J. Craig Venter™
INSTITUTE

Influenza
Genetic Drift and Shift

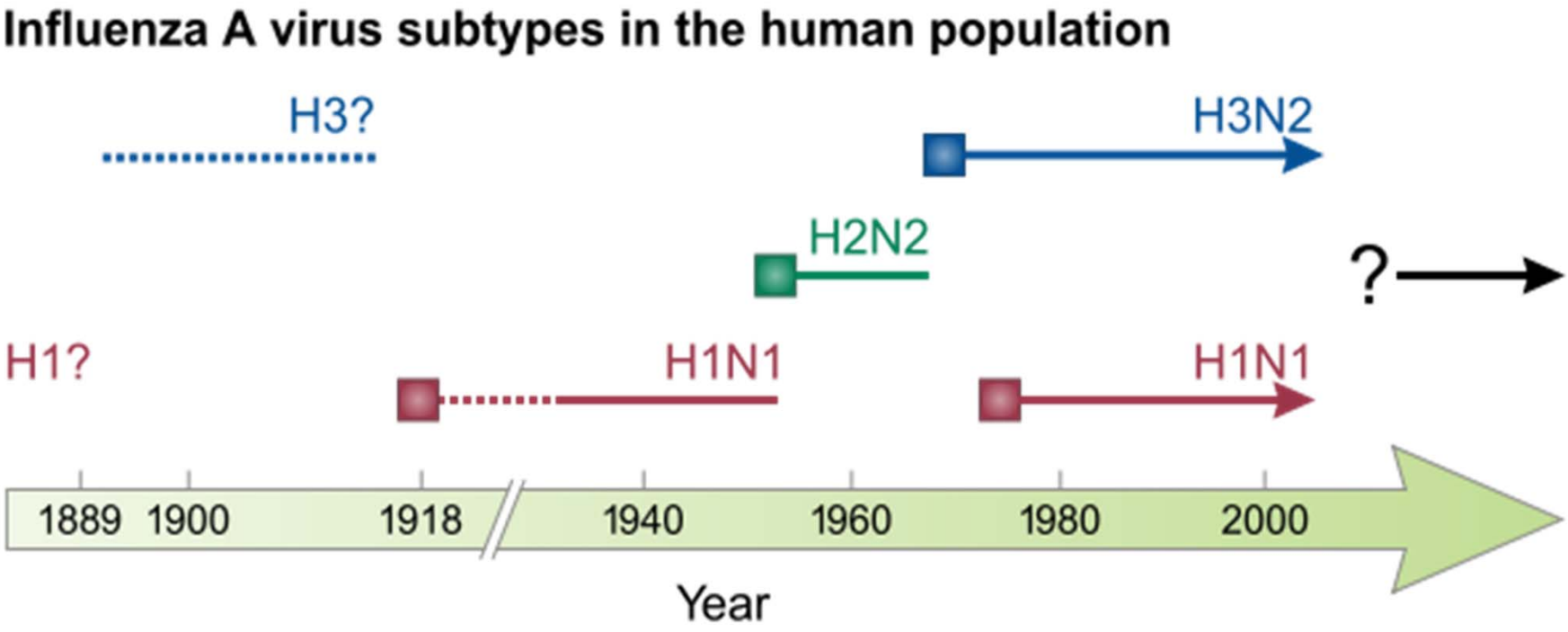




- Error prone replication
- Genomic segment shuffling

J. Craig Venter™
INSTITUTE

Influenza



J. Craig Venter™ INSTITUTE

Influenza

Known flu pandemics					
Name of pandemic	Date	Deaths	Case fatality rate	Subtype involved	Pandemic Severity Index
1889–1890 flu pandemic (Asiatic or Russian Flu)[175]	1889–1890	1 million	0.15%	possibly H3N8 or H2N2	N/A
1918 flu pandemic (Spanish flu)[176]	1918–1920	20 to 100 million	2%	H1N1	5
Asian Flu	1957–1958	1 to 1.5 million	0.13%	H2N2	2
Hong Kong Flu	1968–1969	0.75 to 1 million	<0.1%	H3N2	2
Russian flu	1977–1978	no accurate count	N/A	H1N1	N/A
2009 flu pandemic[177]	2009–2010	105,700-395,600[178]	0.03%	H1N1	N/A

J. Craig Venter™ INSTITUTE

Influenza

Aircraft arrivals in Beijing during 2009 H1N1 pandemic

J. Craig Venter™
INSTITUTE

Influenza A H7N9 Outbreak China 2013-2014

- 454 confirmed cases, severe symptom rate ~100%, mortality ~30%.
- Older male prevalence bias.
- No evidence of sustained person to person transmission.
- High prevalence in domestic birds without bird disease.
- Genome sequence analysis revealed the outbreak strain to be recombinant of genes between parent viruses in wild birds and poultry but show adaptation to mammals.
- Novel low pathogenicity in birds, high pathogenicity in humans.

J. Craig Venter™
INSTITUTE

Influenza A H7N9 Outbreak China 2013-2014

Transmission to humans at live poultry markets or from poultry market environment.

J. Craig Venter™
INSTITUTE

Influenza A H7N9 Outbreak
Factors Promoting/Inhibiting Outbreak

- Virus evolving in wild and domestic avian population and infected humans
- Low transmission between humans
- Commercial live poultry markets
- Vaccines for poultry and humans being developed in China and US
- Seasonal flu vaccine is ~60% protective

J. Craig Venter™
INSTITUTE

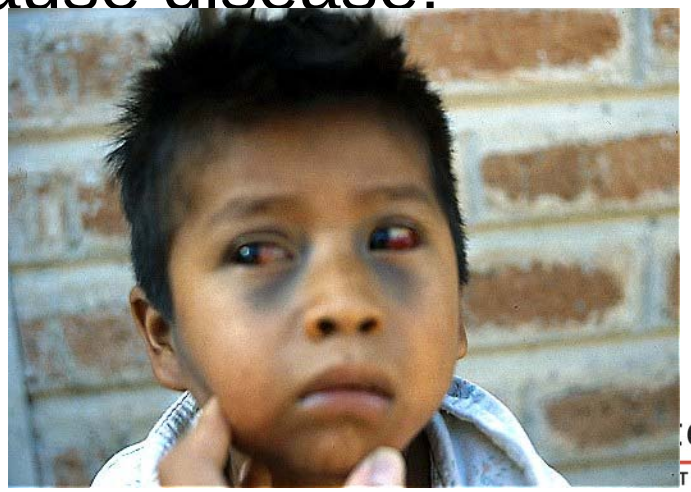
Whooping Cough

J. Craig Venter™
INSTITUTE

Bordetella spp. and Whooping Cough

- Serious respiratory illness caused primarily by *B. pertussis.*
- Disease is characterized by whooping sound after coughing
- Coughing can be so violent that may result in vomiting, burst vessels, bruising
- Other Bordetellae can also cause disease. though usually less severe

Whooping sound

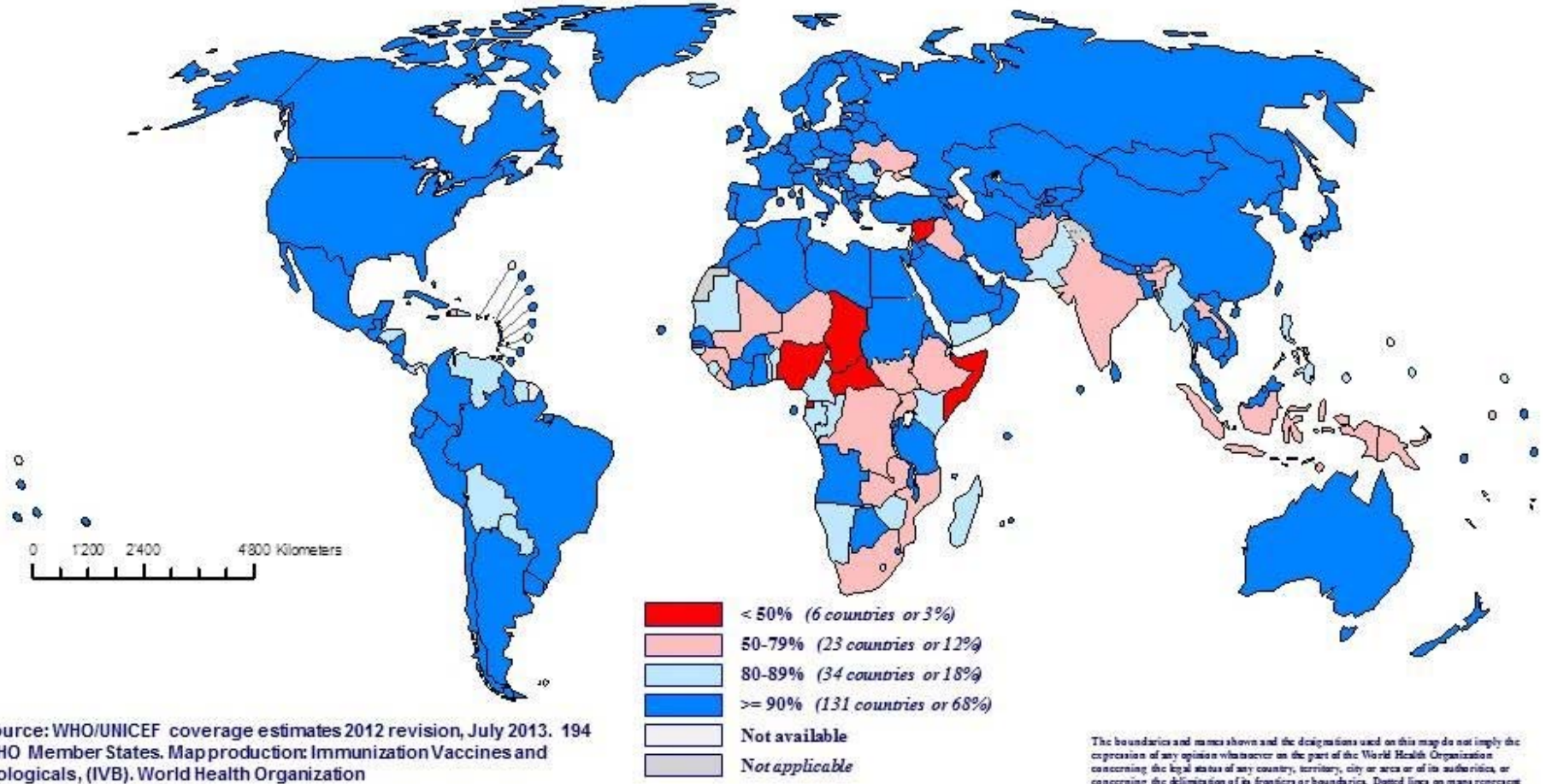


Whooping Cough

- 1 in 2 infants will be hospitalized
- Of hospitalized infants:
  - 1 in 4 (23%) get pneumonia (lung infection)
  - 1 or 2 in 100 (1.6%) will have convulsions (violent, uncontrolled shaking)
  - Two thirds (67%) will have apnea (slowed or stopped breathing)
  - 1 in 300 (0.4%) will have encephalopathy (disease of the brain)
  - 1 or 2 in 100 (1.6%) will die

J. Craig Venter™ INSTITUTE

Immunization coverage with DTP3 vaccines in infants (from <50%), 2012



Source: WHO/UNICEF coverage estimates 2012 revision, July 2013. 194 WHO Member States. Map production: Immunization Vaccines and Biologicals, (IVB). World Health Organization
Date of slide: 16 July 2013

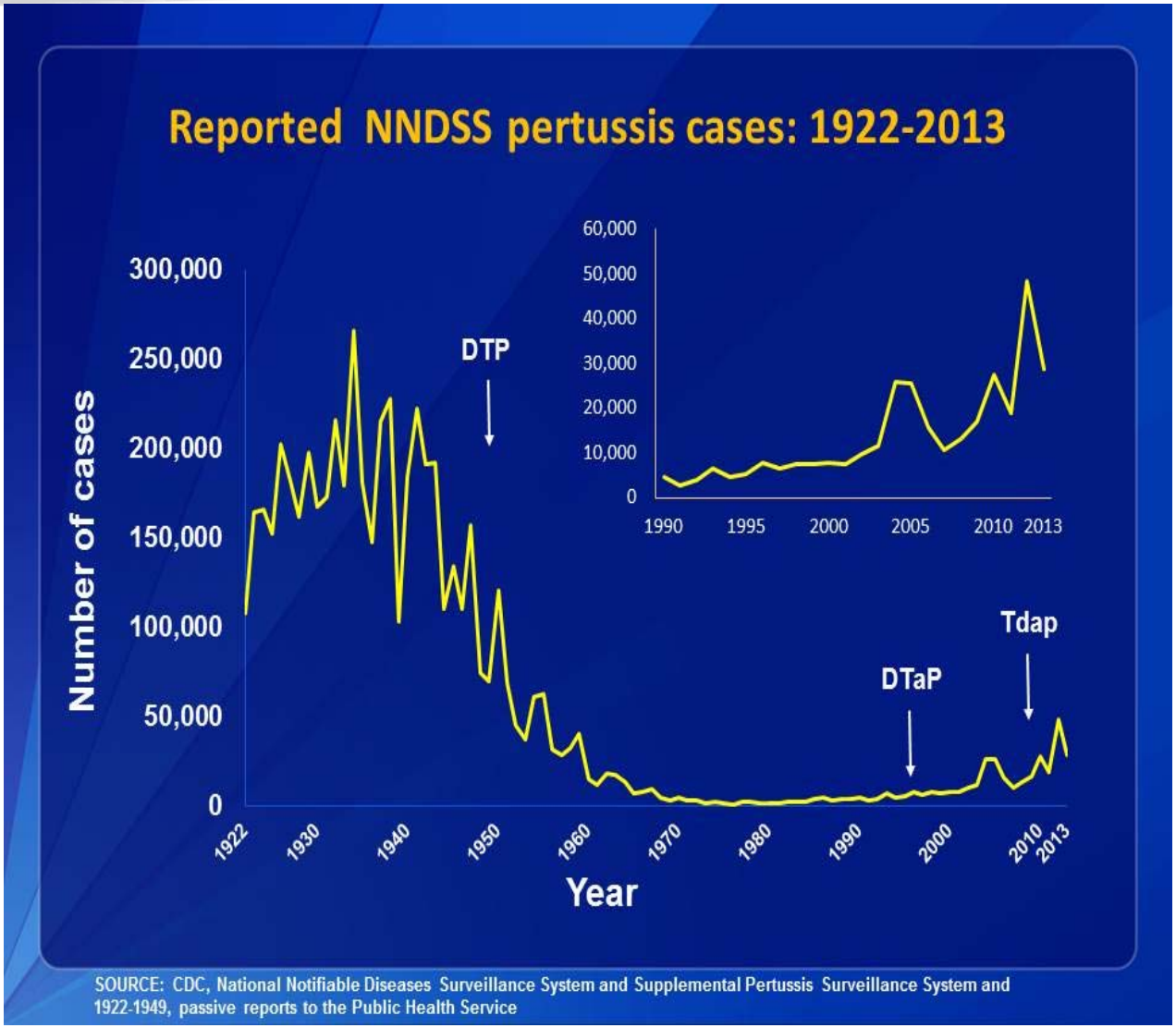
World Health Organization

Pertussis Vaccine

- Killed whole-cell vaccine was introduced in the mid-1940's.
- Replaced by an acellular formulation in the early 1990's.
- Contains one or more antigens:
  - Filamentous hemagglutinins
  - Pertactin
  - Detoxified pertussis toxin
  - Fimbriae

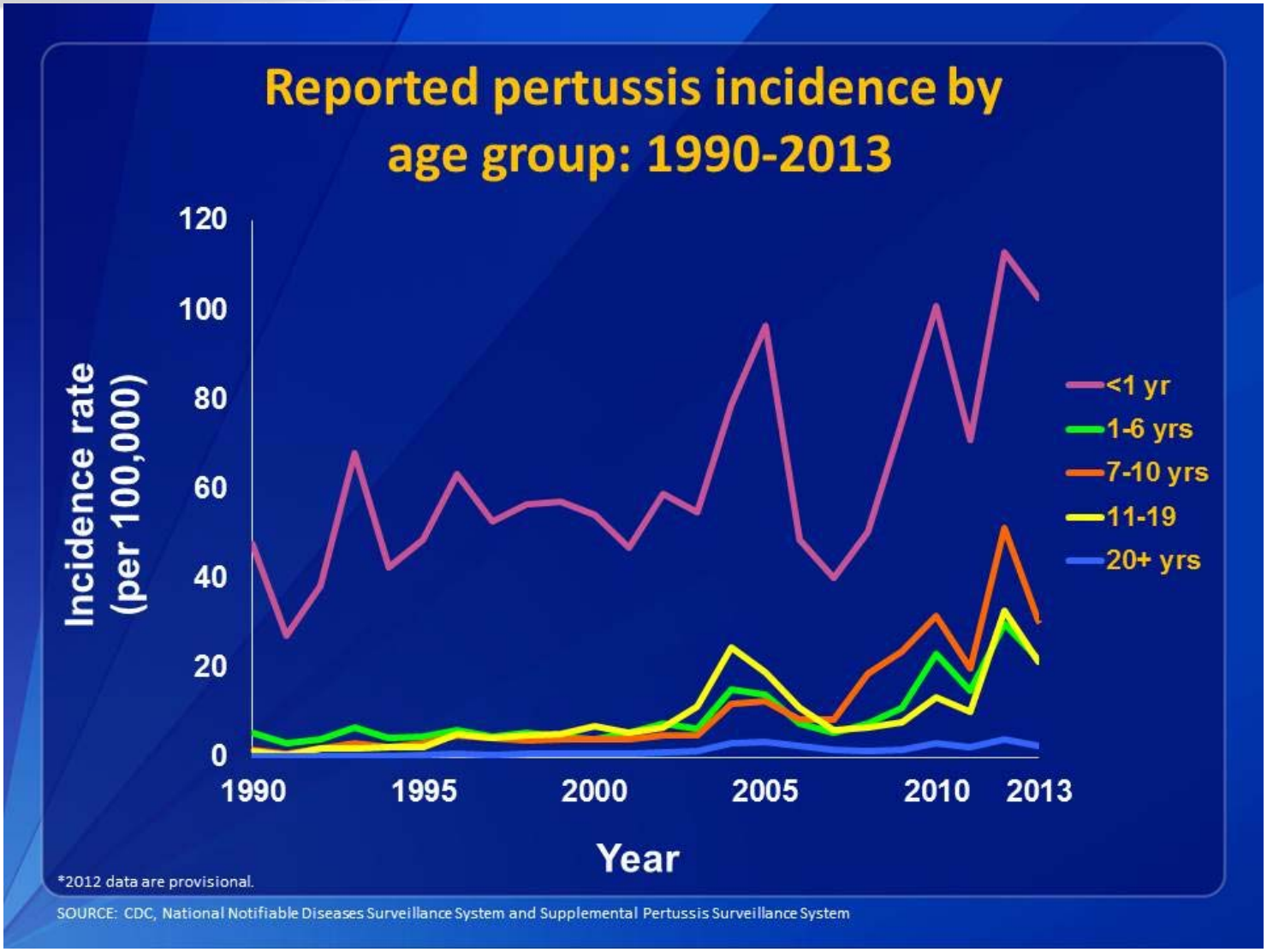
J. Craig Venter™ INSTITUTE

Whooping Cough



J. Craig Venter™ INSTITUTE

Whooping Cough



J. Craig Venter™ INSTITUTE

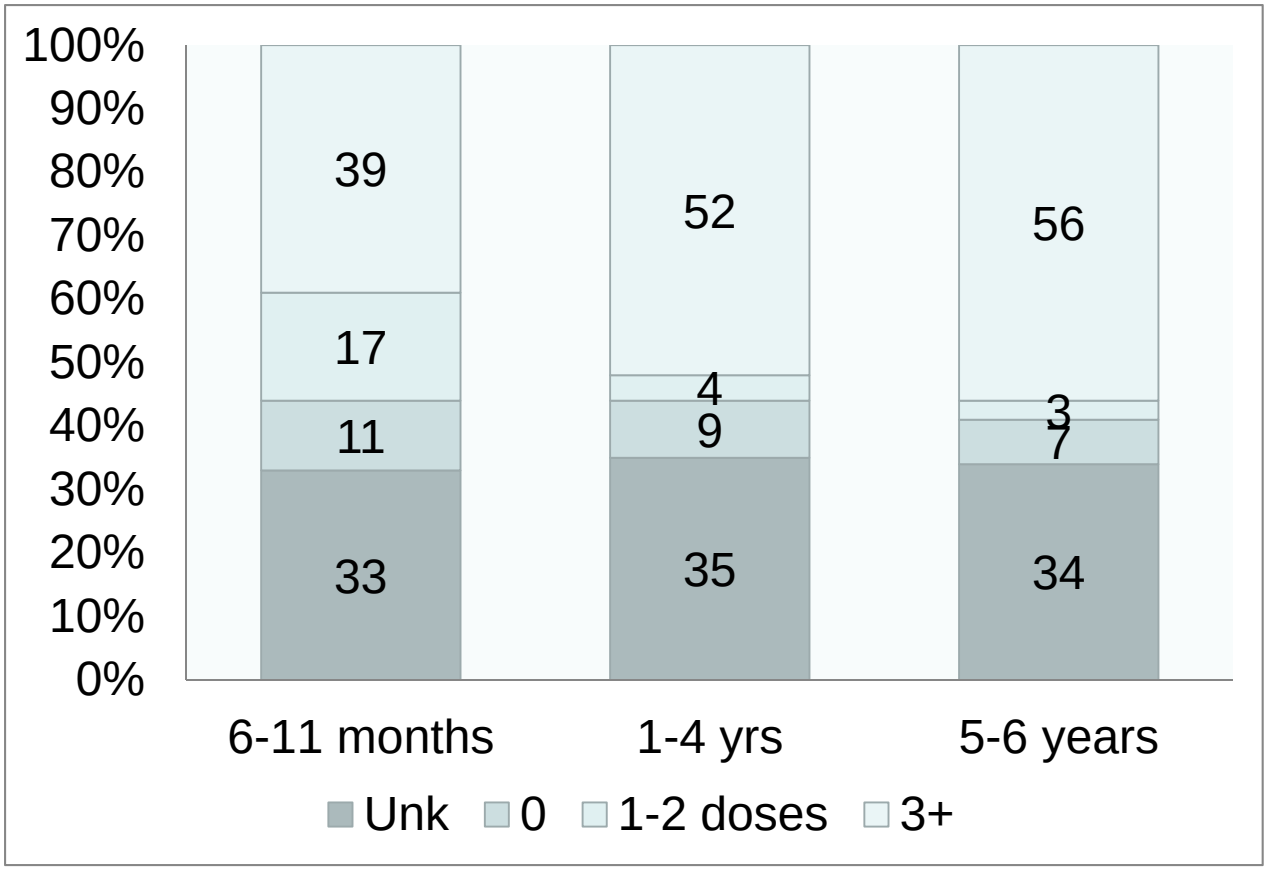
Whooping Cough

- In 2014 , 25,987 US cases. 30% increase from 2013, 26 deaths

- 2014 in California
  - 10,838 cases, 26/100,000 – worst in 70 yrs
  - Infants highest risk group 172/100,000
  - Of 378 hospitalized, 300 under 1 yr age, 233 under 4 mos.
  - Most mothers not vaccinated
  - Spike in teen cases even though vaccinated as infants

J. Craig Venter™
INSTITUTE

Majority of children with whooping cough have had several doses of acellular vaccine



N=8433 cases

J. Craig Venter™ INSTITUTE

Concerns and Questions

- Acellular vaccine protection appears to wane
  - *Is there a need for a booster?*
  - *Is there a need for better vaccines?*
- Despite high vaccination rates, *B. pertussis* still circulates in the population
  - *Are there specific genetic characteristics that allow human disease and vaccine evasion?*
- Non-*B. pertussis* cases are becoming more prevalent
  - *What is the genetic complement of these organisms?*

J. Craig Venter™ INSTITUTE

Sequencing of Classical Bordetellae

- Collaboration with Dr. Eric Harvill and Dr. Karen Register to sequence veterinary and non-pertussis Bordetellae
- Collaboration with hospitals in the Collaborative Pulmonary Critical Care Research Network (CPCCRN) across USA to sequence 2004, 2010, 2012 outbreak pertussis isolates
- Goals:
  - *Characterize the genetic diversity in Bordetella*
  - *Understand how multiple lineages of Bordtella can infect humans*
  - *Characterize evolution and adaptation in Bordetella*

Liliana Losada, JCVI PI

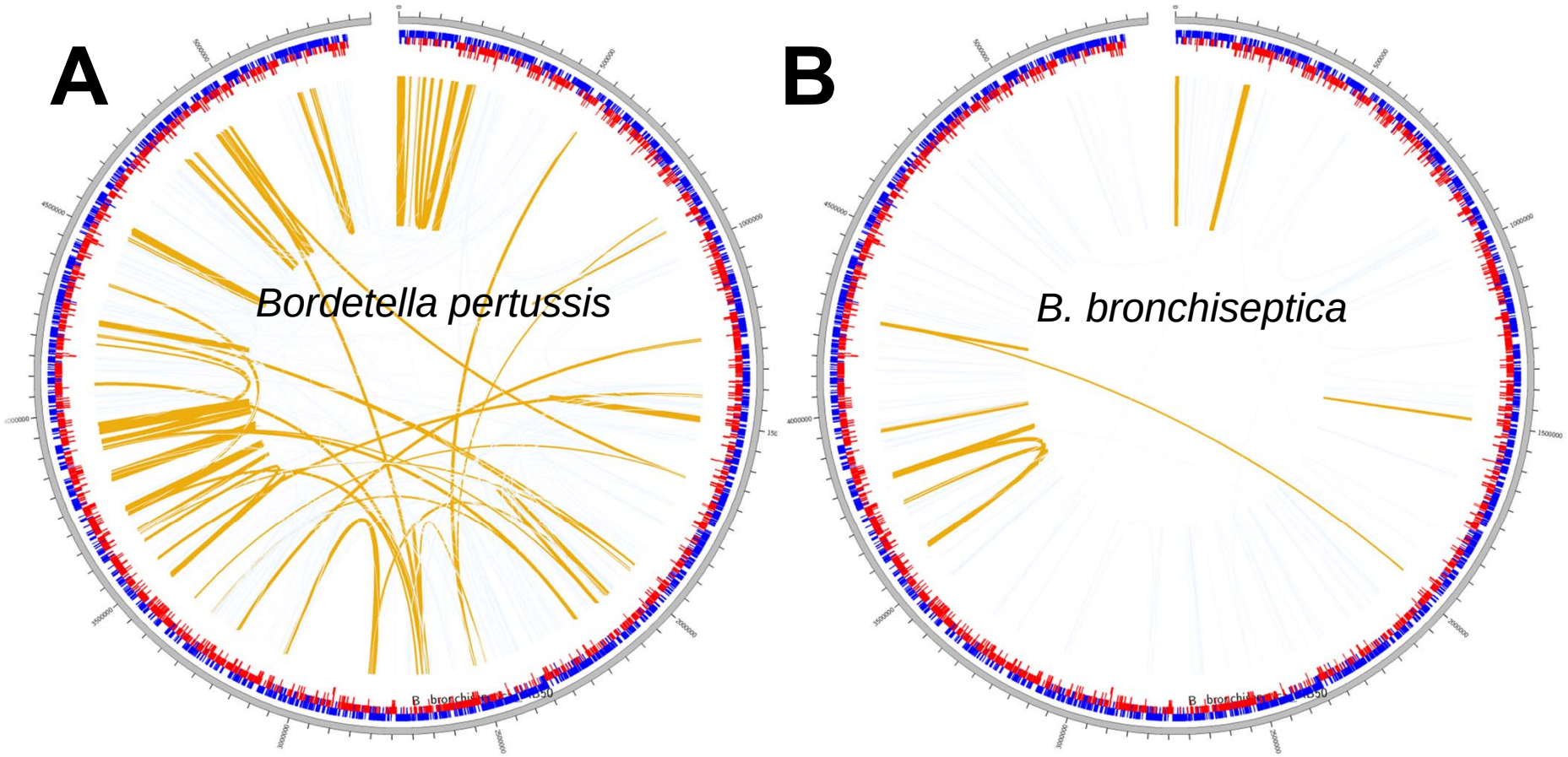
J. Craig Venter™ INSTITUTE

Re-emergence Factors

The “re-emergence” of Bordetella is believed to be due to **vaccine efficiency waning** 10+ years after vaccination which leaves **teenagers** at risk for whooping cough and to **become carriers that then expose a non-vaccinated subpopulation** (the anti-vax movement). This is especially noticeable in the states with the highest number of outbreaks: California, Washington, etc. The currently-used acellular vaccine is still effective against the *B. pertussis* target, and the **genomic data do not support any gain in virulence** by Bpt. While the Bpt strains have undergone significant genomic rearrangements due to IS element expansion, these should not have an effect of increasing virulence or on vaccine “escape”. Bpt is a highly specialized pathogen that cannot survive outside of a host and has very little recourse for gene gain by horizontal transfer. **The circulating lineage of Bpt is an evolutionary dead-end that can only continue to evolve by shedding more and more of its genome.**

J. Craig Venter™
INSTITUTE

Rearrangements in Non-bronchiseptica Genomes

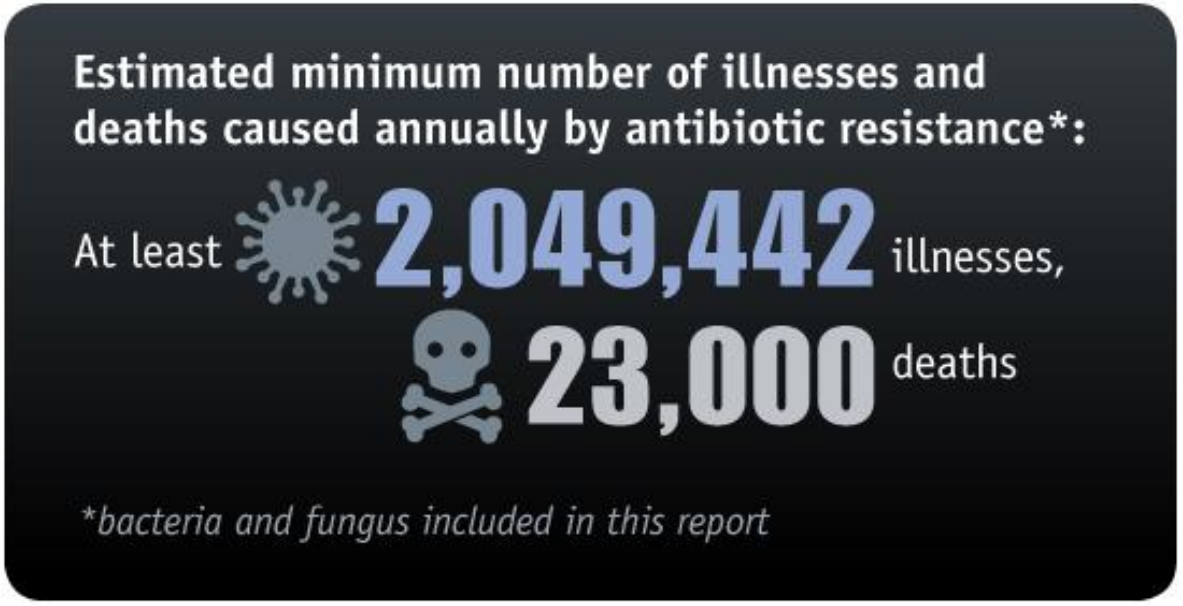


J. Craig Venter™ INSTITUTE

Carbapenem-Resistant *Klebsiella pneumoniae*

J. Craig Venter™
INSTITUTE

Bacterial Drug Resistance



Multi-drug resistance as a critical issue

- Infections in critical-care settings
- Transfer to community-acquired infections

CDC 2013 Threat Report

Headline – Washington Post

2 dead, 7 possibly ill after ‘superbug’ bacteria hits UCLA hospital

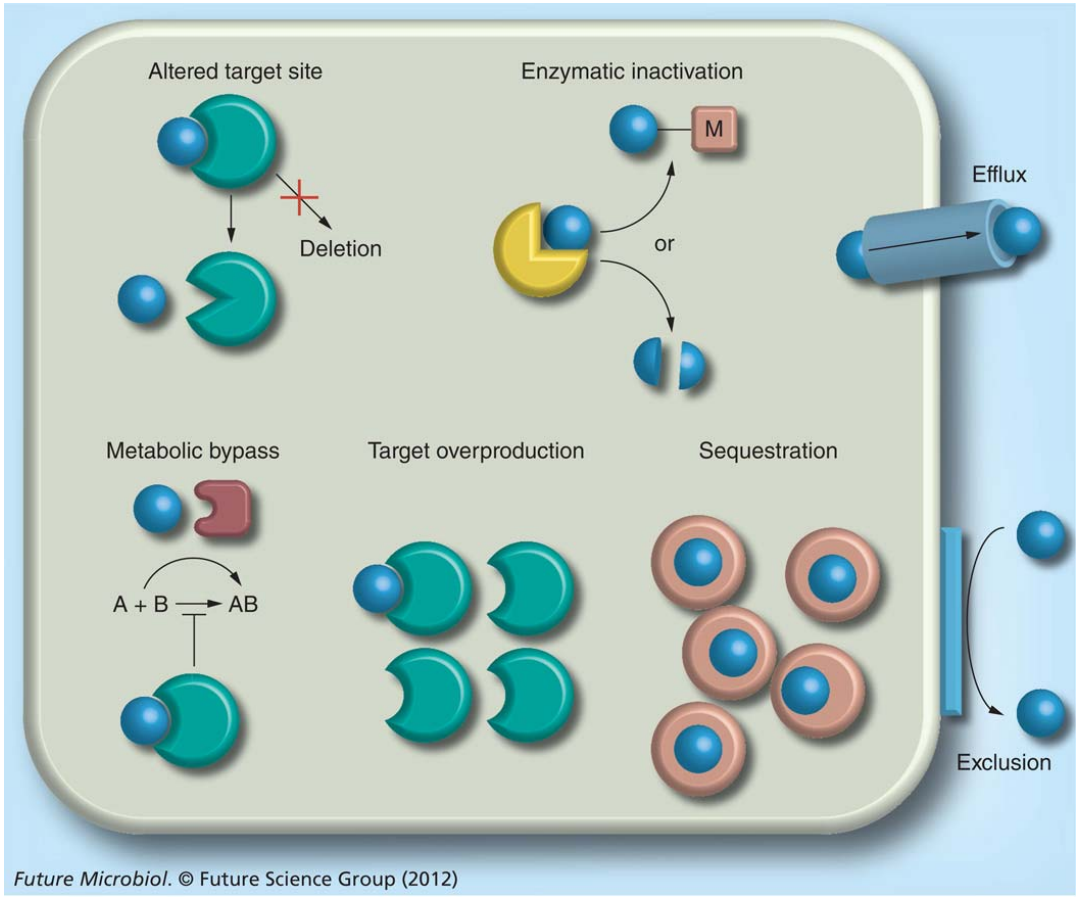
Washington Post, February 19, 2015

J. Craig Venter™
INSTITUTE

Bacterial Drug Resistance and Drug Tolerance

- Drug resistance
  - Genetic alterations to drug target protein, export pumps, regulatory elements, etc.
  - Frequently carried on mobile elements
- Drug tolerance
  - Genetic alteration resulting in slow growth
  - State of organism with low metabolic activity without genetic change (persister state)

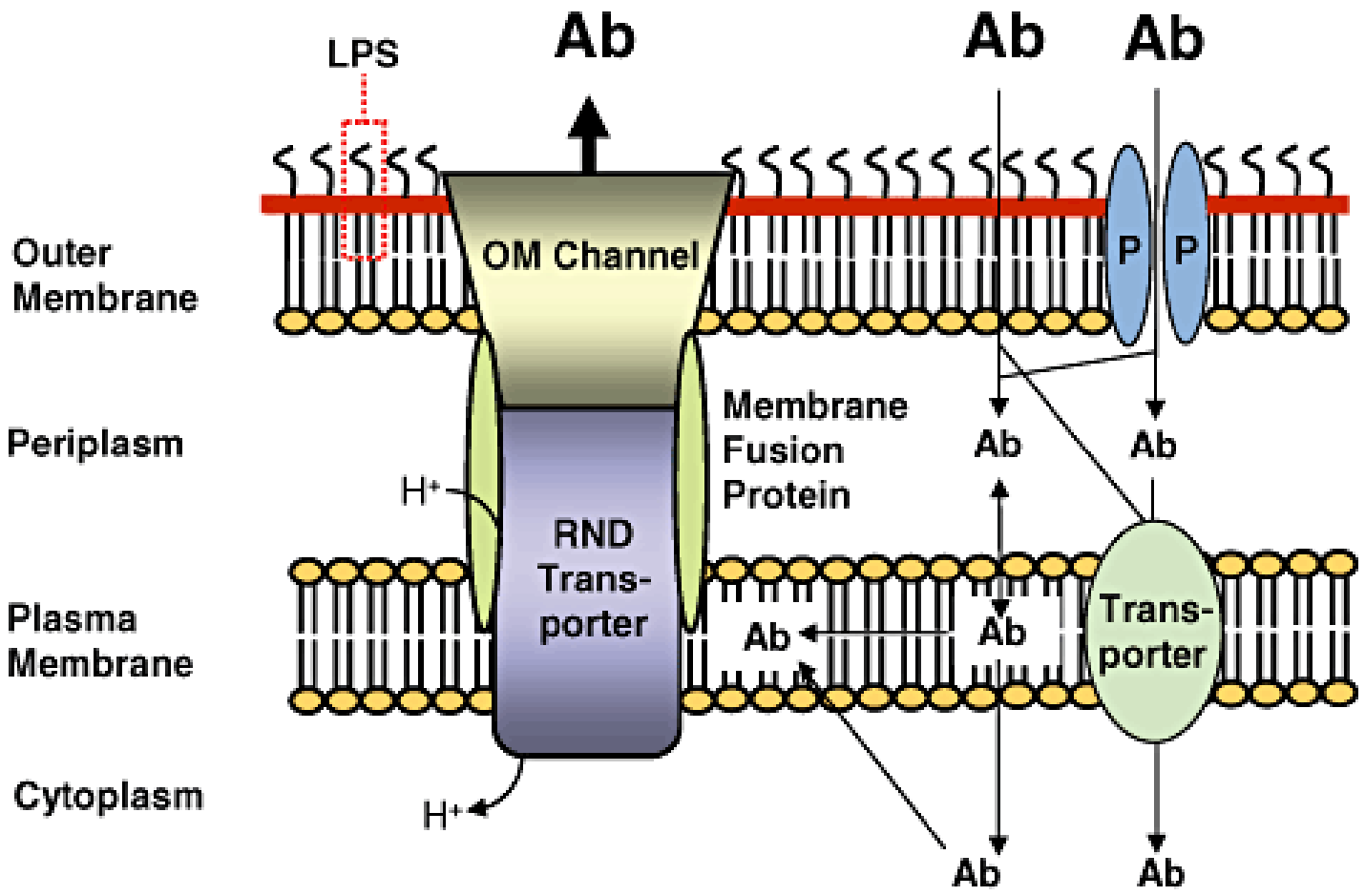
Bacterial Antibiotic Resistance



Schweizer 2012 Future Microbiology 7:1389

J. Craig Venter™ INSTITUTE

Cell Envelop of Gram Negative Bacteria



Schweitzer 2012 Microbial Biotechnology 5: 581

J. Craig Venter INSTITUTE

B. pseudomallei Resistance Mechanisms

***Burkholderia pseudomallei* antibiotic resistance mechanisms.**

Antibiotic or inhibitor class‡	Exclusion	Enzymatic inactivation	Target mutation	Efflux
Aminoglycosides	X			X
β-lactams		X	X	
Chloramphenicol				X
Clavulanic acid			X	
Fluoroquinolones			X	X
Macrolides				X
Polymyxin B	X			
Tetracyclines				X
Trimethoprim				X
Trimethoprim–sulfamethoxazole				X

Schweizer 2012 Future Microbiology 7:1389

J. Craig Venter™ INSTITUTE

Bacterial Persistence

- State of tolerant to ultra-high antibiotic treatment.
- Discovered in 1944. Best characterized in *E. coli.*
- After emergence from this state drug sensitivity is restored.
- May contribute to
  - Antibiotic treatment failure.
  - Disease latency.
  - Reemergent infections.

J. Craig Venter™ INSTITUTE

Persister Assay

- Viability cfu plating before drug challenge
- Megadose drug challenge – 16 hours
  - Cefotaxime 100XMIC (400µg/mL)
  - Ciprofloxacin 10XMIC (20µg/mL)
- Viability cfu plating after drug challenge

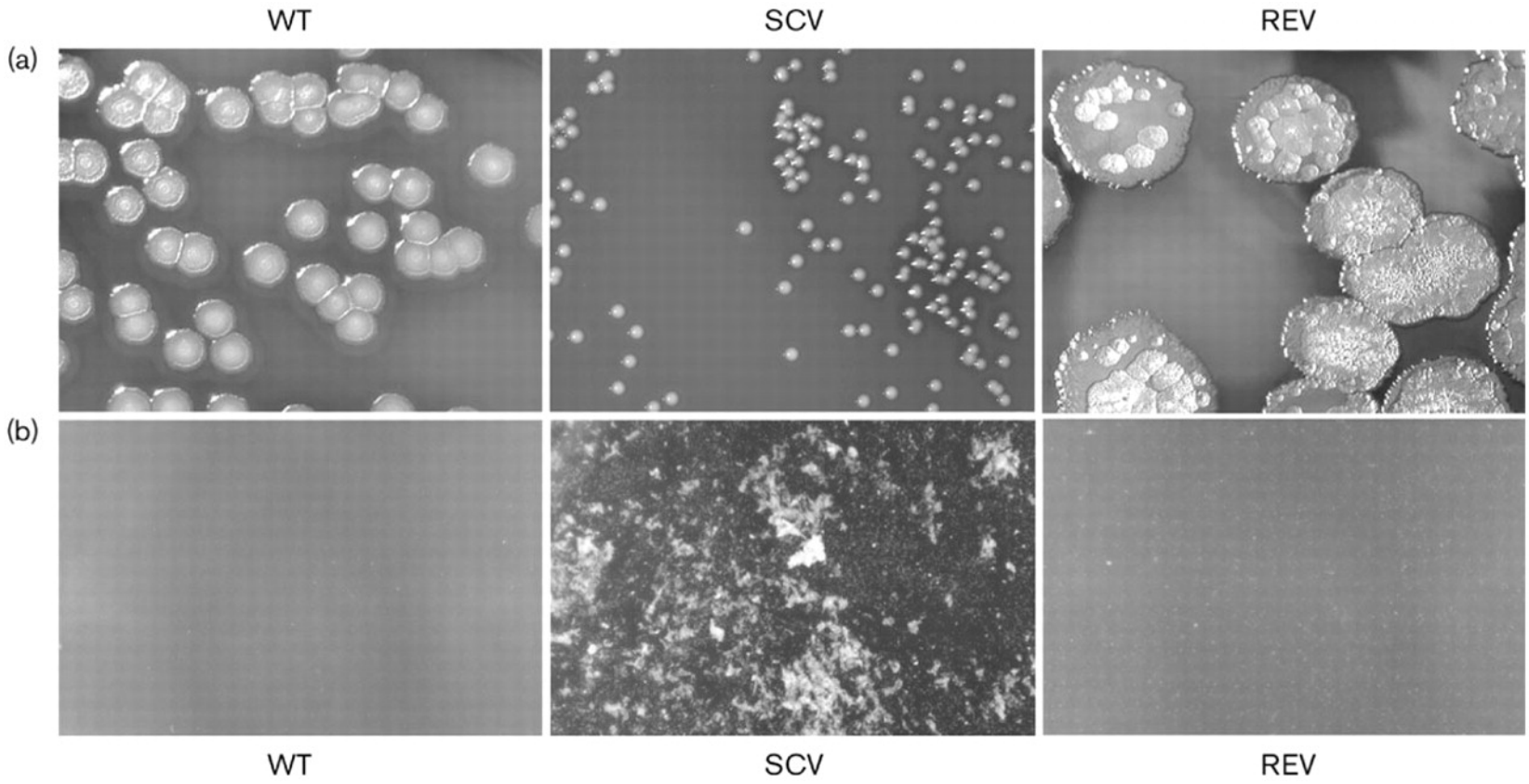
J. Craig Venter™
INSTITUTE

Cystic Fibrosis

- Lungs are chronically colonized by microbes.
- Frequent acute pneumonia episodes.
- Most frequent pathogens *Staphylococcus aureus, Haemophilus influenzae, Pseudomonas aeruginosa*.
- By age 18 80% are colonized by *P. aeruginosa* and 3.5% by *Burkholderia cepacia* complex with more difficulty in managing pneumonia episodes.

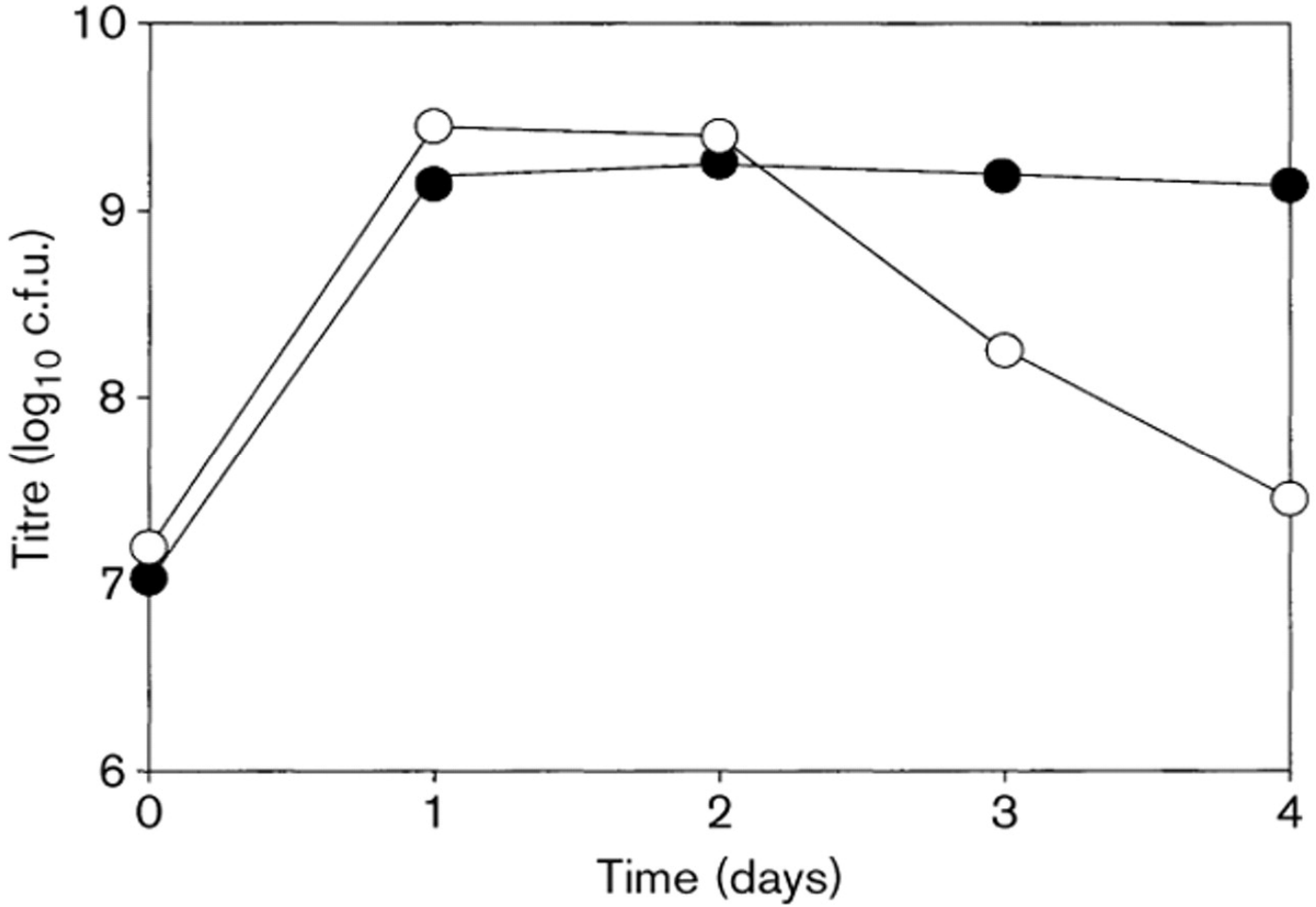
J. Craig Venter™ INSTITUTE

Pseudomonas aeruginosa
Small Colony Variants



Häußler S et al. J Med Microbiol 2003;52:295-301

Pseudomonas aeruginosa
SCVs vs WT Relative Fitness



Häußler S et al. J Med Microbiol 2003;52:295-301

SOCIETY FOR GENERAL MICROBIOLOGY™

In vitro generated *P. aeruginosa* SCVs

- Plate *P. aeruginosa* PAO1at 20X MIC ciprofloxacin. SCVs formed within 10 days.
- Replanted single colony PAO1-Sg1-Sg3
  - SCV phenotype maintained with or without selection.
  - Without selection SCVs appeared by 144 hours (6Da), few larger colonies appeared by 48 hrs. One designated PAO10Sg1-S1.
- Plating Sg-S1 to obtain second round suppressor mutants with wt growth rate.

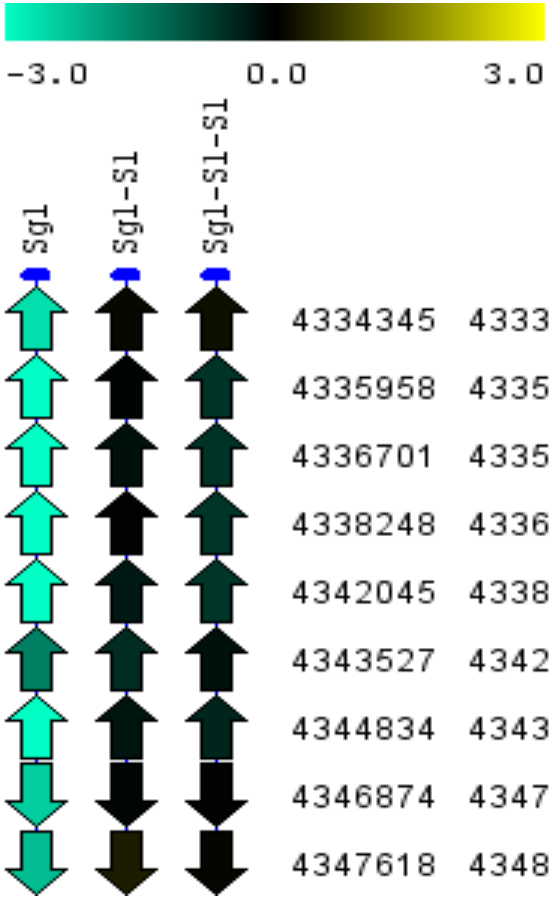
J. Craig Venter INSTITUTE

P. aeruginosa SCVs

- PAO1 doubling time in LB liquid medium
  - 37.3 ± 1.5 min
- For PAO1-Sg1
  - 291.8 + 16.4 min (-8X).
  - In double suppressor isolate wt doubling time restored.
  - Growth rate and MICs for ciprofloxacin and carbenicillin inversely related.

J. Craig Venter™ INSTITUTE

P. aeruginosa SCV Differential Expression



4334345	4333356	**PA3870**	molybdenum cofactor biosynthesis protein A
4335958	4335275	**PA3872**	respiratory nitrate reductase gamma chain
4336701	4335961	**PA3873**	respiratory nitrate reductase delta chain
4338248	4336707	**PA3874**	respiratory nitrate reductase beta chain
4342045	4338260	**PA3875**	respiratory nitrate reductase alpha chain
4343527	4342121	**PA3876**	nitrite extrusion protein 2
4344834	4343539	**PA3877**	nitrite extrusion protein 1
4346874	4347533	**PA3879**	two-component response regulator NarL
4347618	4348013	**PA3880**	hypothetical protein

J. Craig Venter™
INSTITUTE

P. aeruginosa SCV Genomic Analysis

Locus Name	Gene Symbol	Common Name	Position[a]	Ref[b]	Var[c]	PA01	Sg1[d]	Sg1-S1[b]	Sg1-S1-S1[b]
PA2220	No symbol	Probable transcriptional regulator	2442123	T	A	absent	+	+	+
PA0008	glyS	glycyl-tRNA synthetase subunit beta, S	10895	C	T	absent	absent	+	+
PA3790	oprC	Putative copper transport outer membrane porin OprC precursor, NS	4248153	G	A	absent	absent	+	+

[a] Chromosomal coordinate in the *P. aeruginosa* PAO1 genome.
[b] Nucleotide present in the reference genome.
[c] Nucleotide replacing the reference nucleotide in the query strain.
[d] A plus symbol means the variant nucleotide was present in the query strain, while absent means that the query strain had the same nucleotide as the reference genome.
No changes in *gyrA, gyrB, parC, parE* or *nfxB*.

J. Craig Venter™ INSTITUTE

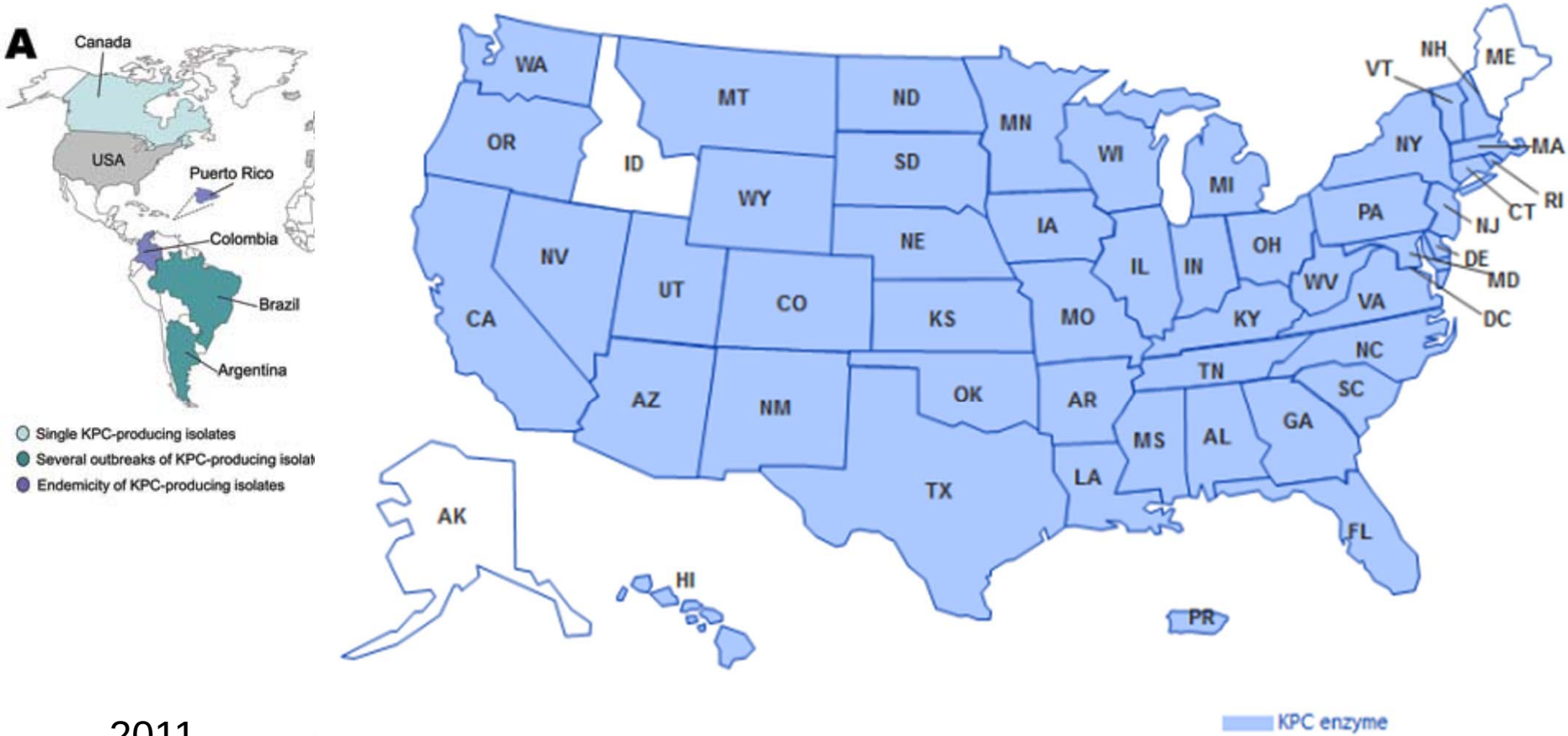
Summary of Major Findings
Pseudomonas aeruginosa

- ***P. aeruginosa*** **PAO1 can form SCVs in response to in vitro antibiotic challenge.**
- **SCVs accumulated tolerance-conferring mutations.**
- **Transcriptional analysis suggested SCV strain sg1 underwent major alterations to nitrate metabolism.**
- **Suppressors mutations restored normal growth rate phenotype.**
- ***P. aeruginosa*** **has a high rate of persoster formation.**

Marcus Jones JCVI PI

J. Craig Venter™
INSTITUTE

Klebsiella pneumoniae Carbapenemase (KPC) in the USA



2011

2014

Nordmann, *et al. Emerg. Infect. Dis.* 2011
http://www.cdc.gov/hai/organisms/cre/TrackingCRE.html

J. Craig Venter INSTITUTE

Genome Architecture of KPC+ ST258

AAC
Journals.ASM.org

Population Structure of KPC-Producing *Klebsiella pneumoniae* Isolates from Midwestern U.S. Hospitals

Meredith S. Wright,[a] Federico Perez,[b] Lauren Brinkac,[a] Michael R. Jacobs,[c] Keith Kaye,[d] Eric Cober,[e] David van Duin,[f] Steven H. Marshall,[b] Andrea M. Hujer,[b,g] Susan D. Rudin,[b,g] Kristine M. Hujer,[b,g] Robert A. Bonomo,[b,g,h,i] Mark D. Adams[a]

PNAS

Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*

Frank R. DeLeo[a,1], Liang Chen[b,1], Stephen F. Porcella[c], Craig A. Martens[c], Scott D. Kobayashi[a], Adeline R. Porter[a], Kalyan D. Chavda[b], Michael R. Jacobs[d], Barun Mathema[b], Randall J. Olsen[e,f], Robert A. Bonomo[g,h], James M. Musser[e,f], and Barry N. Kreiswirth[b,2]

[a]Laboratory of Human Bacterial Pathogenesis and [c]Research Technologies Branch, Rocky Mountain Laboratories, National Institute of Allergy and Infectious

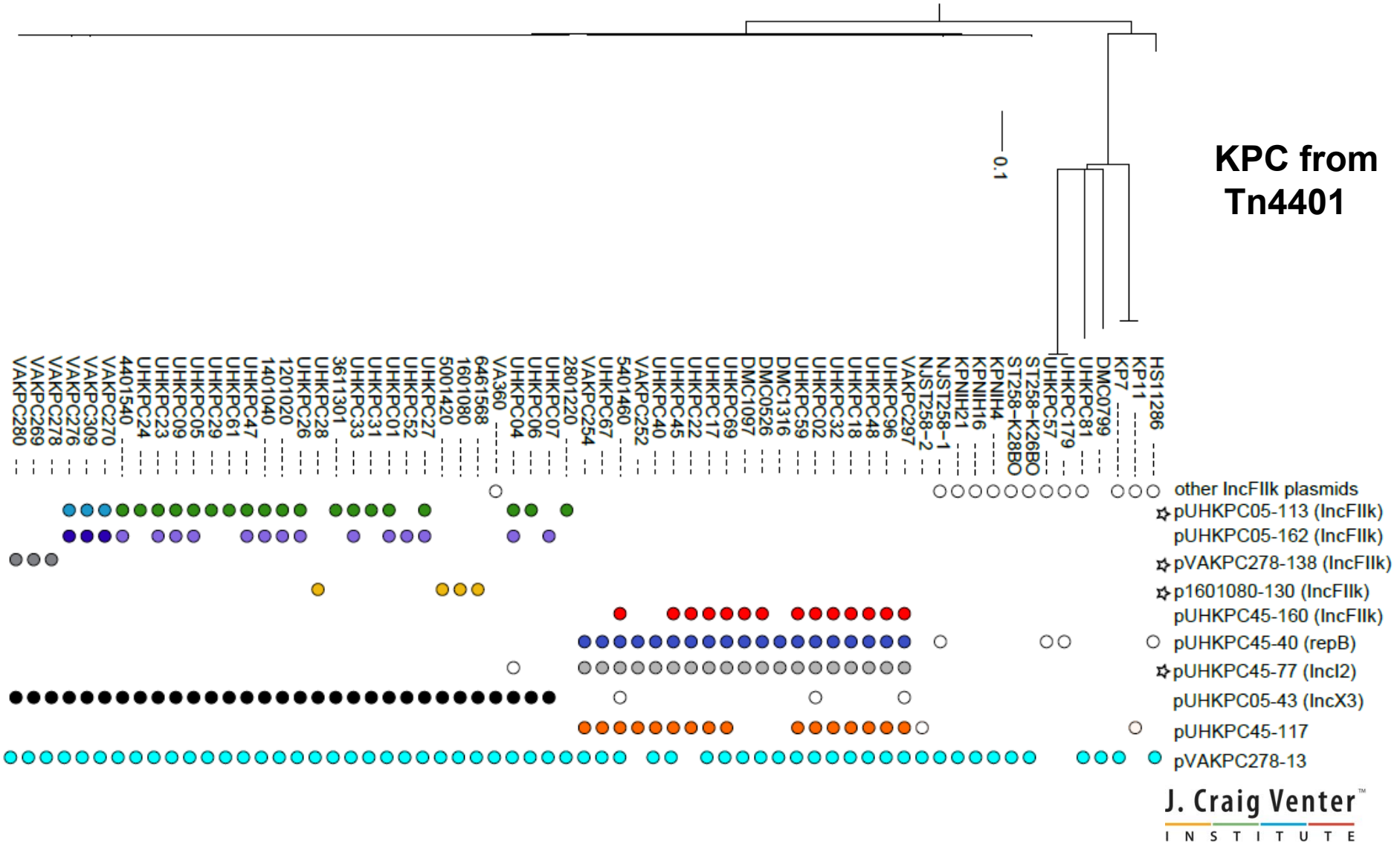
mBio

Epidemic *Klebsiella pneumoniae* ST258 Is a Hybrid Strain

Liang Chen,[a] Barun Mathema,[a,b] Johann D. D. Pitout,[c,d,e] Frank R. DeLeo,[f] Barry N. Kreiswirth[a]

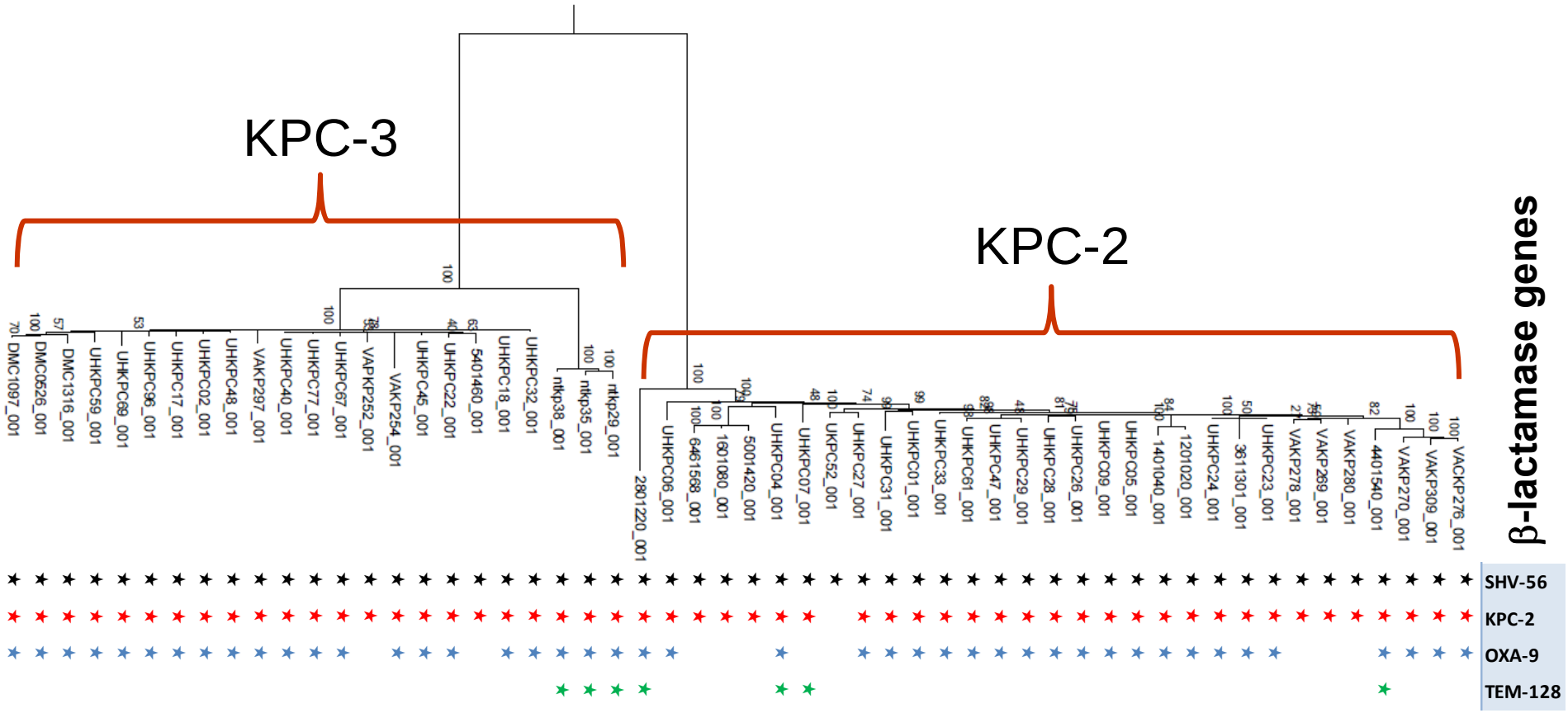
J. Craig Venter™
INSTITUTE

Plasmid Content Varies Across Closely Related Strains



J. Craig Venter INSTITUTE

Two ST258 Sub-clades Redundant Resistance Mechanisms



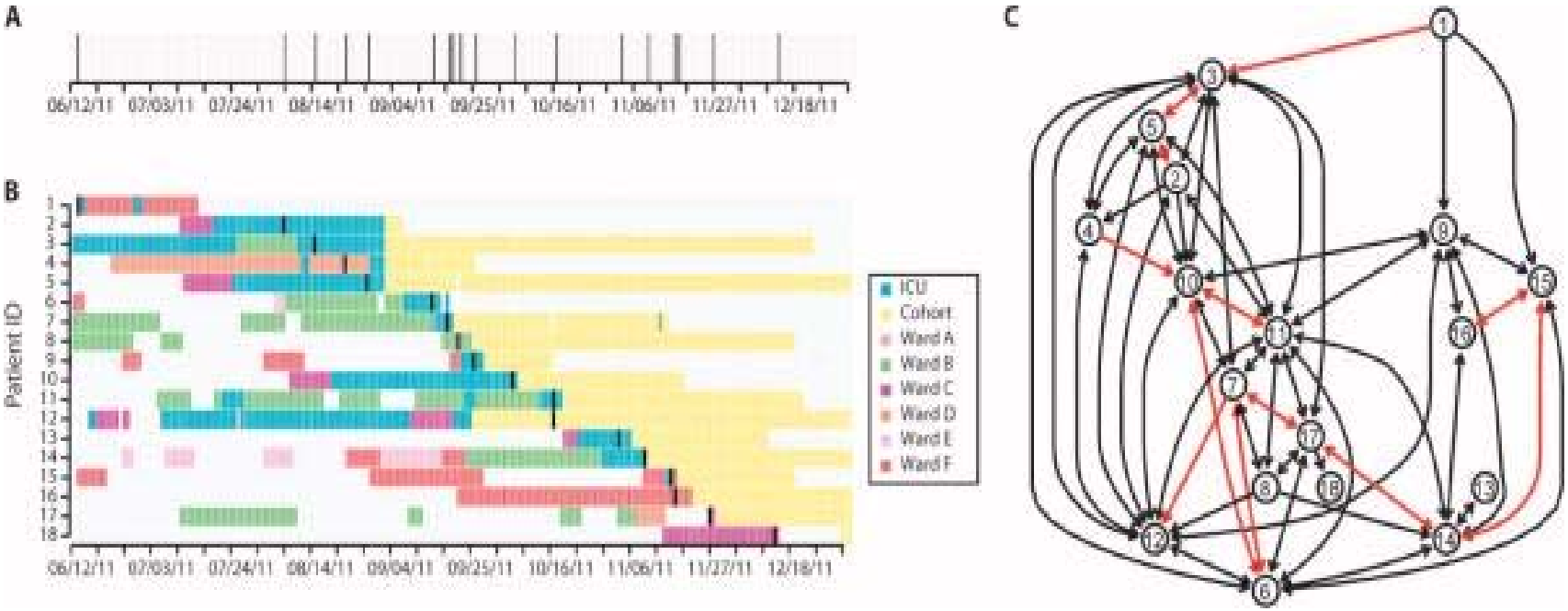
J. Craig Venter INSTITUTE

NIH Clinical Center 2011 - 2012

- *K. pneumoniae* outbreak **case 1** transferred to ICU from NYC 15 July 2011.
- Case 2 diagnosed 25 August, **3 weeks** after case 1 left hospital.
- Total of **18** cases, **11** deaths, **6** directly attributable to *K. pneumoniae* infection.
- Outbreak investigation aggressively undertaken via **genome sequencing** of outbreak strains and **epidemiological data** analysis.
- Findings published
  - Snitkin et al. Sci Trans. Med. (2012) 4:148

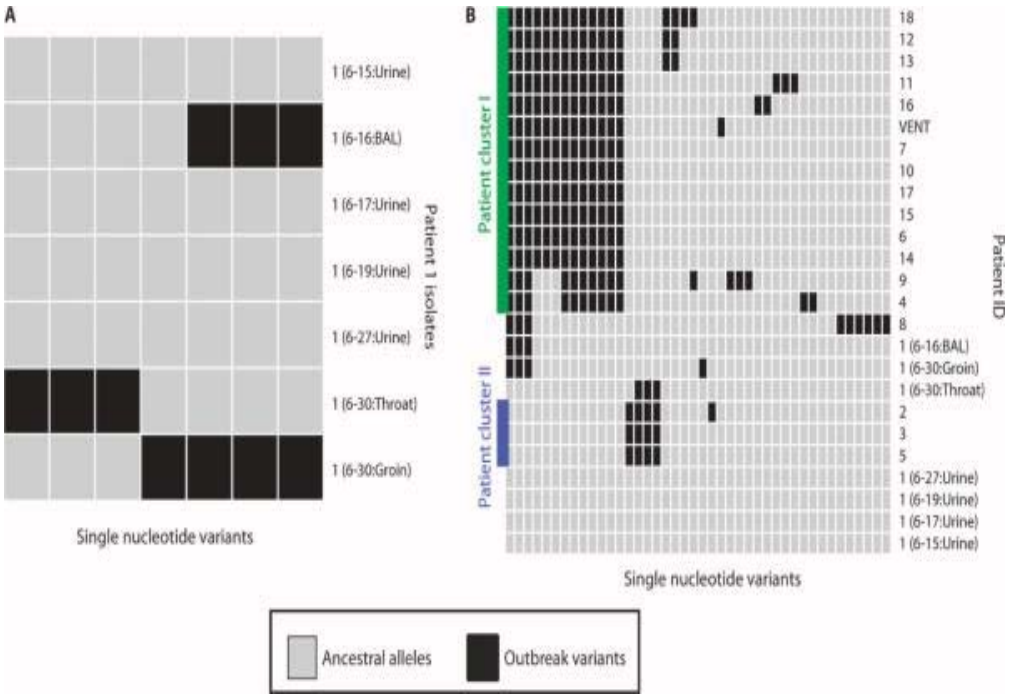
J. Craig Venter INSTITUTE

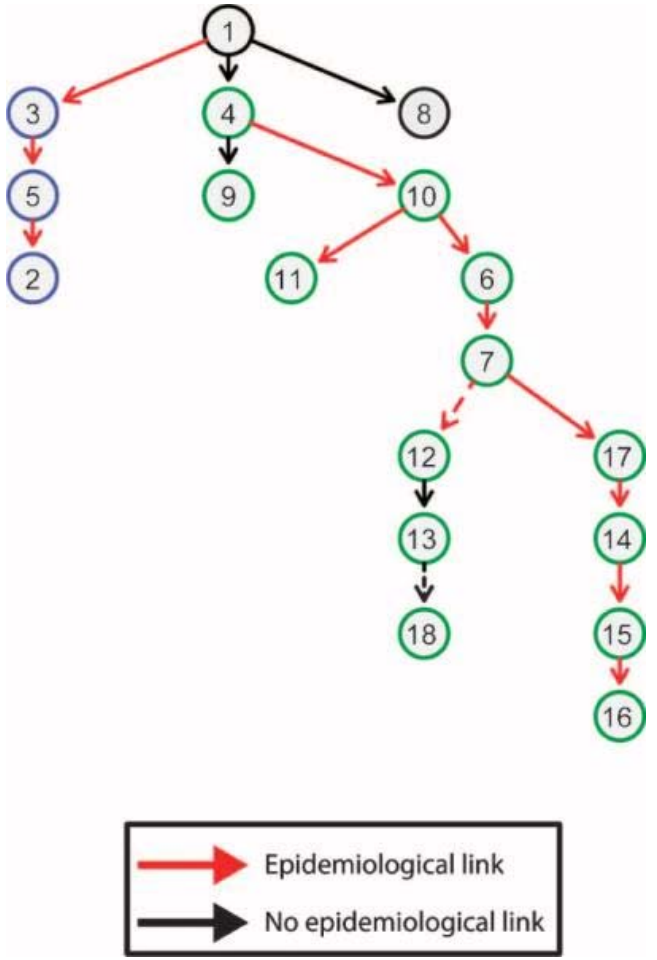
NIH Clinical Center
2011 - 2012



J. Craig Venter™
INSTITUTE

NIH Clinical Center
2011 - 2012





J. Craig Venter™
INSTITUTE

NIH Clinical Center 2011 – 2012 Findings

- *K. pneumoniae* outbreak is propagated by diagnosed patients and **undetected asymptomatic carriers**.
- Isolate detected on respirator after use by patient 6 even **after extensive cleaning**.
- During the course of the outbreak the isolates developed resistance to gentamicin, tigecycline, and colistin, the only antibiotics for which the patient 1 isolates were sensitive.

J. Craig Venter™ INSTITUTE

NIH Clinical Center 2011 – 2012 Lesson Learned

- Do **sequencing in clinical real time** to track transmissions. Reveals silent transmission via asymptomatic carriers.
- Have **effective surveillance protocols** in place before outbreak.
- **Sample multiple sites** from index case.
- **Verify decontamination** after cleaning.

J. Craig Venter™ INSTITUTE

GCID Bacterial Project Specific Aims

- **Specific Aim 1: Assess genetic mechanisms involved in emergence of antibiotic resistance in nosocomial pathogens.**

- **Specific Aim 2: Explore variation in DNA methylation patterns and the role of DNA modification in modulating bacterial gene expression.**

Mark Adams JCVI PI

J. Craig Venter™
INSTITUTE

Acknowledgements

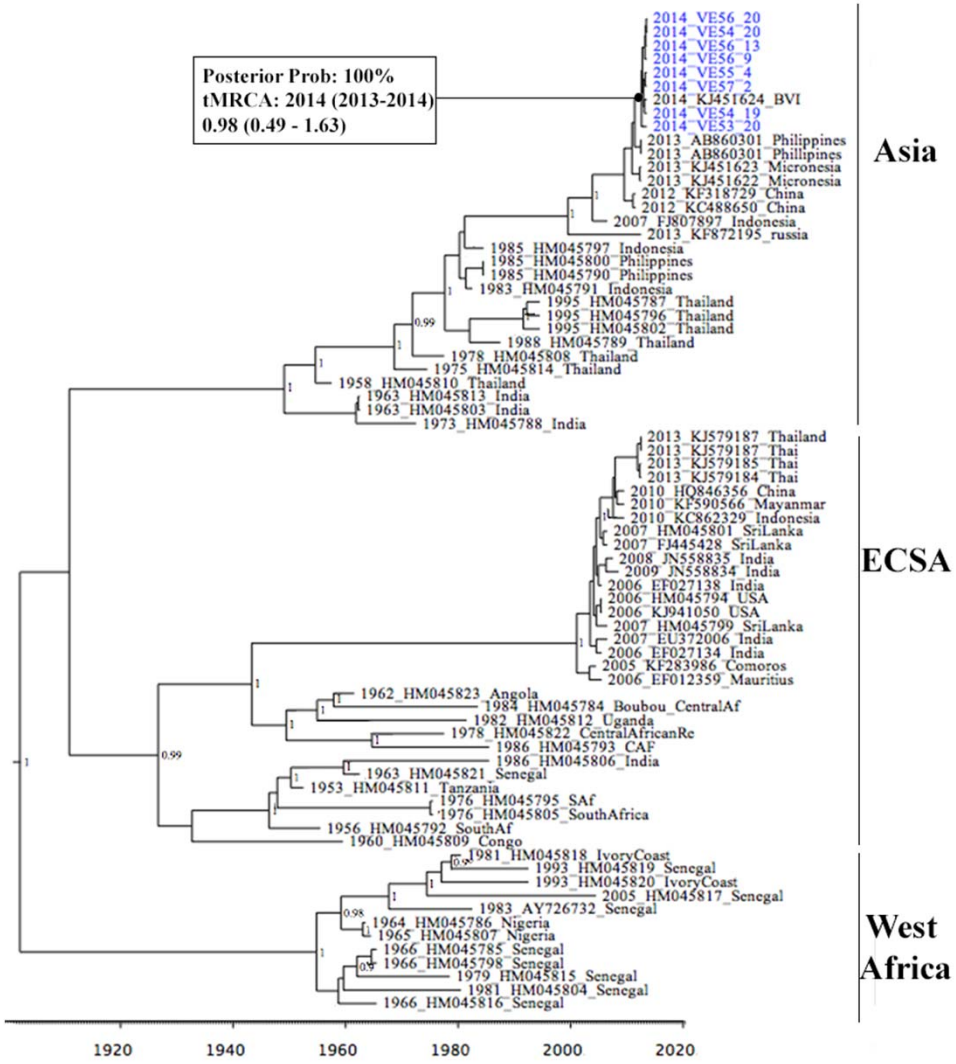
- JCVI Team
  - Granger Sutton
  - Lauren Brinkac
  - Erin Beck
  - Jamison McCorrison
  - Indresh Singh
  - Ravi Sanka
  - Tim Stockwell
  - Brian Bishop
  - Meredith Wright
- Collaborators
  - Robert Bonomo
  - Barry Kreiswirth
  - Joe Vinetz
  - Vance Fowler
- Core Lab
  - Savita Shanker, University of Florida

ARLG
Antibacterial Resistance Leadership Group

J. Craig Venter™
INSTITUTE

Dr. Christine Carrington
Emergence of Chikungunya virus in Trinidad

Emerging and re-emerging viruses in Trinidad and Tobago



Christine Carrington
Department of Preclinical Sciences
Faculty of Medical Sciences
The University of the West Indies

"Omics" at the Faculty of Medical Sciences

- Viral genomics and phylogenetics
  - Dengue
  - Rabies
  - Chikungunya
  - Yellow Fever
- Genome-wide association study
  - National Eye Survey of Trinidad and Tobago (genetics sub-study - Anglia Ruskin University, Duke University, UWI)
- Metagenomics
  - Acute undifferentiated febrile illnesses
  - Oral microbiome – effects of xylitol

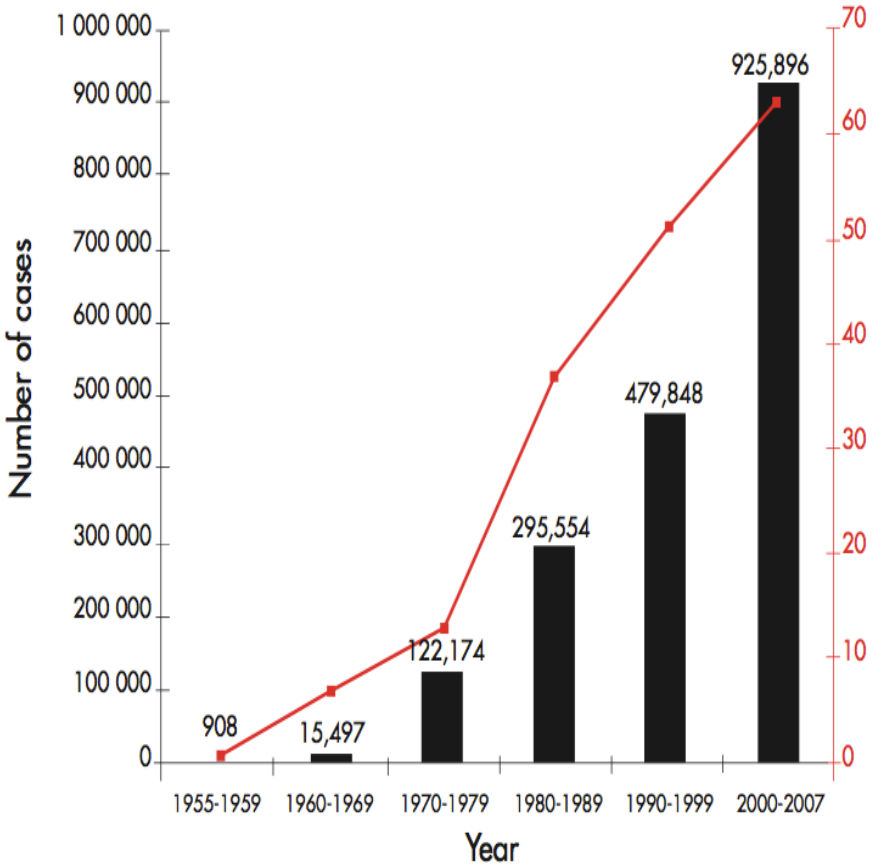
Viral genomics

- Evolutionary and ecological factors underlying the emergence, spread and maintenance of emerging viruses (esp. vector borne RNA viruses)
  - Evolutionary and demographic histories of viruses
  - Spatiotemporal dynamics (phylogeography) of emerging viruses
  - Mechanisms of maintenance (i.e. is there regular reintroduction or low-level endemic transmission between outbreaks?)
  - Animal reservoirs (viral diversity in species that are known or potential sources of emerging viruses)
- Viral diversity (mosquitoes, bats, acute undifferentiated fevers)

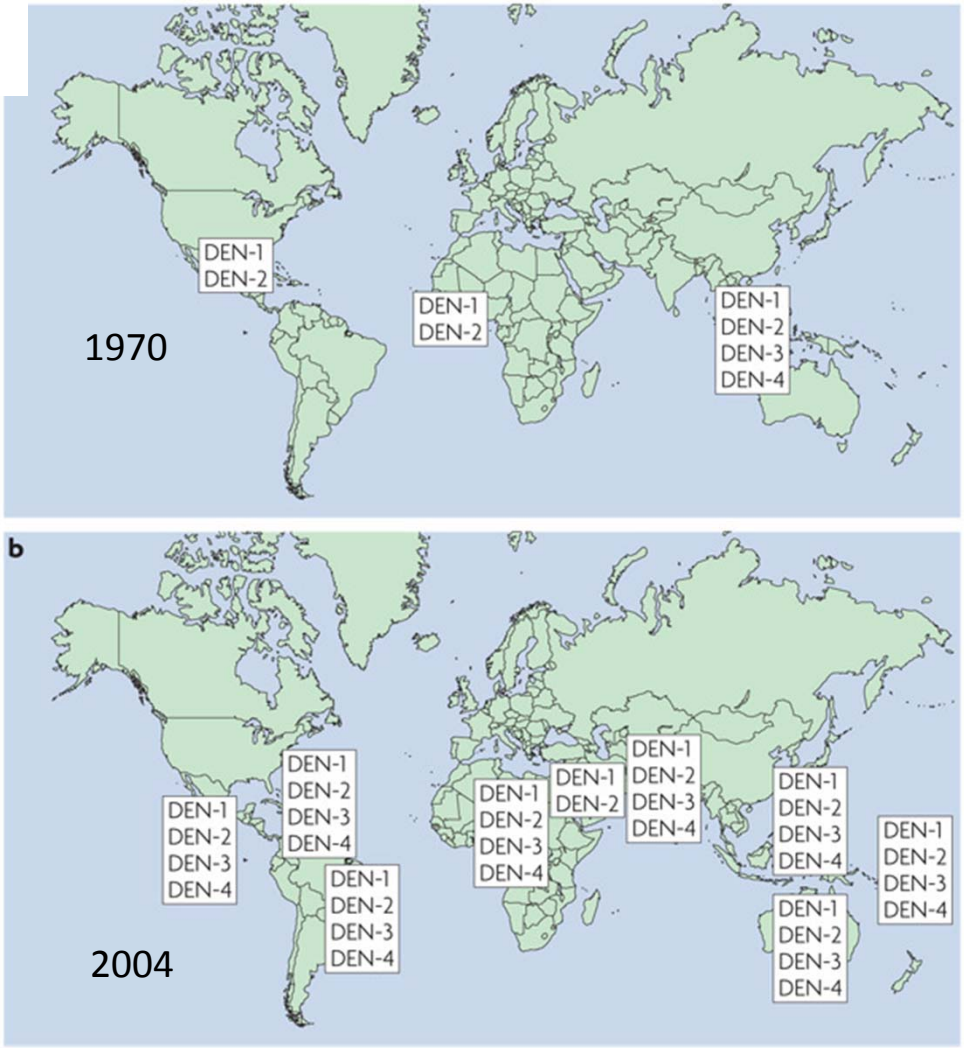
Dengue Virus (DENV)	Chikungunya Virus (CHIKV)
Family *Flaviviridae* Genus *Flavivirus*	Family *Togaviridae* Genus *Alphavirus*
Positive-sense single-stranded RNA genome	Positive-sense single-stranded RNA genome
Genome approx.10.7 kb	Genome approx. 11.6kb
4 distinct serotypes [Several genotypes within each serotype]	1 single serotype [3 genotypes – WAf, ECSA, Asian]
Aedes aegypti *Aedes albopictus*	*Aedes aegypti* *Aedes albopictus*
Dengue (+/- warning signs); Severe Dengue Fever, malaise, arthralgia, myalgia, **low platelet count, plasma leakage,** bleeding manifestations (incl. rash) Full recovery but hospitalization may be necessary, fatalities in severe cases.	Chikungunya Fever Fever, malaise, **severe arthralgia**, **rash** Acute symptoms typically resolve within 7 – 10 days; some patients have persistent joint pains (months – years). Not usually life threatening.

Global Emergence of Dengue

Average annual no. of Dengue cases reported / countries reporting dengue, 1955–2007



WHO 2009. Dengue: Guidelines for diagnosis, treatment, prevention & control



Guzman, M. G. et al. Dengue: A continuing global threat. Nature Reviews Microbiology 8, S7–S16 (2010).

Global emergence of Dengue

WWII troop movements and population displacement

rapid global transport

global population growth
High density human populations
Inadequate infrastructure
unplanned urbanisation

Inadequate water supply
Improper waste disposal

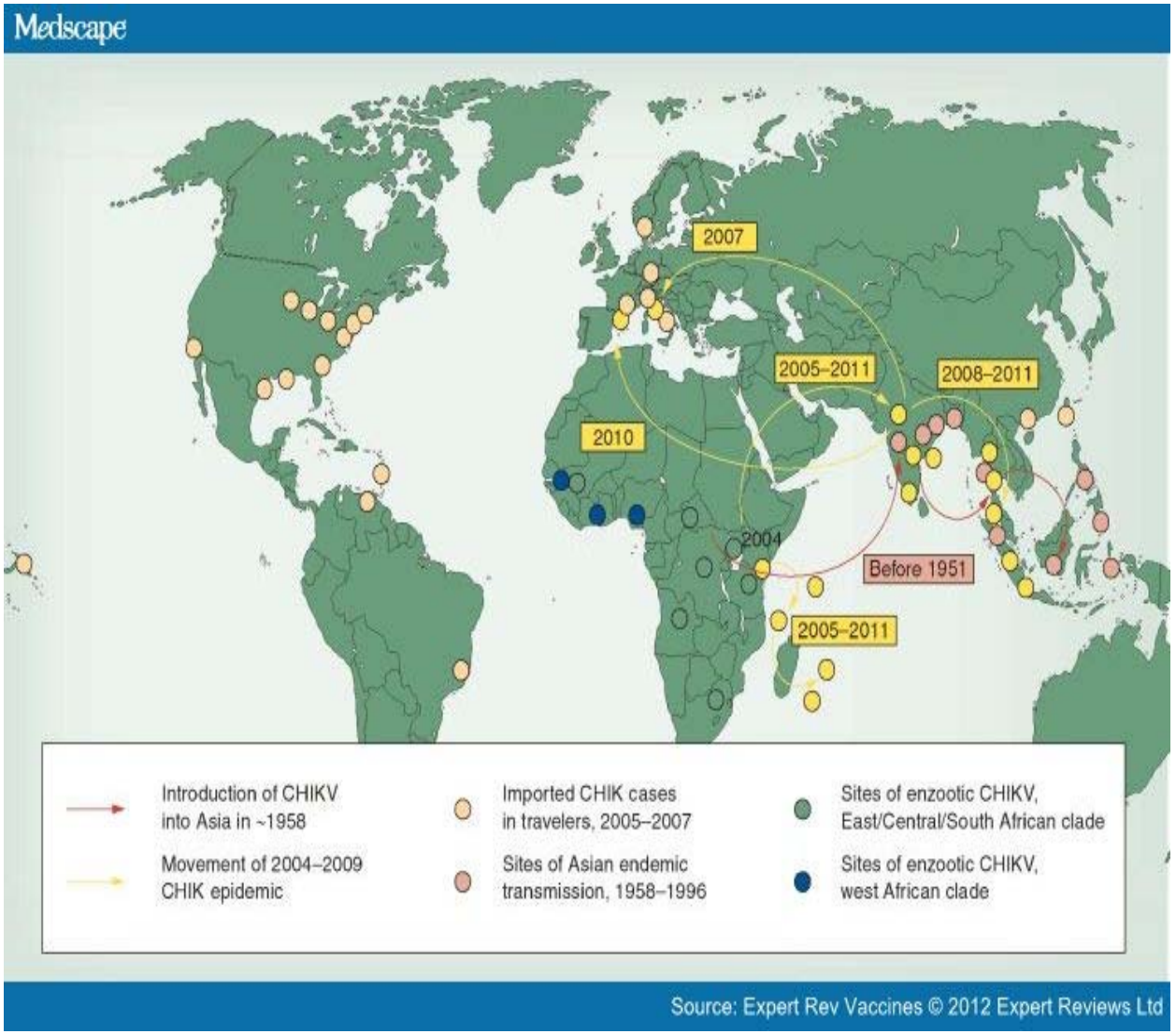
water storage
water collection

mosquito breeding sites

Inadequate vector control + weak implementation of public health policies

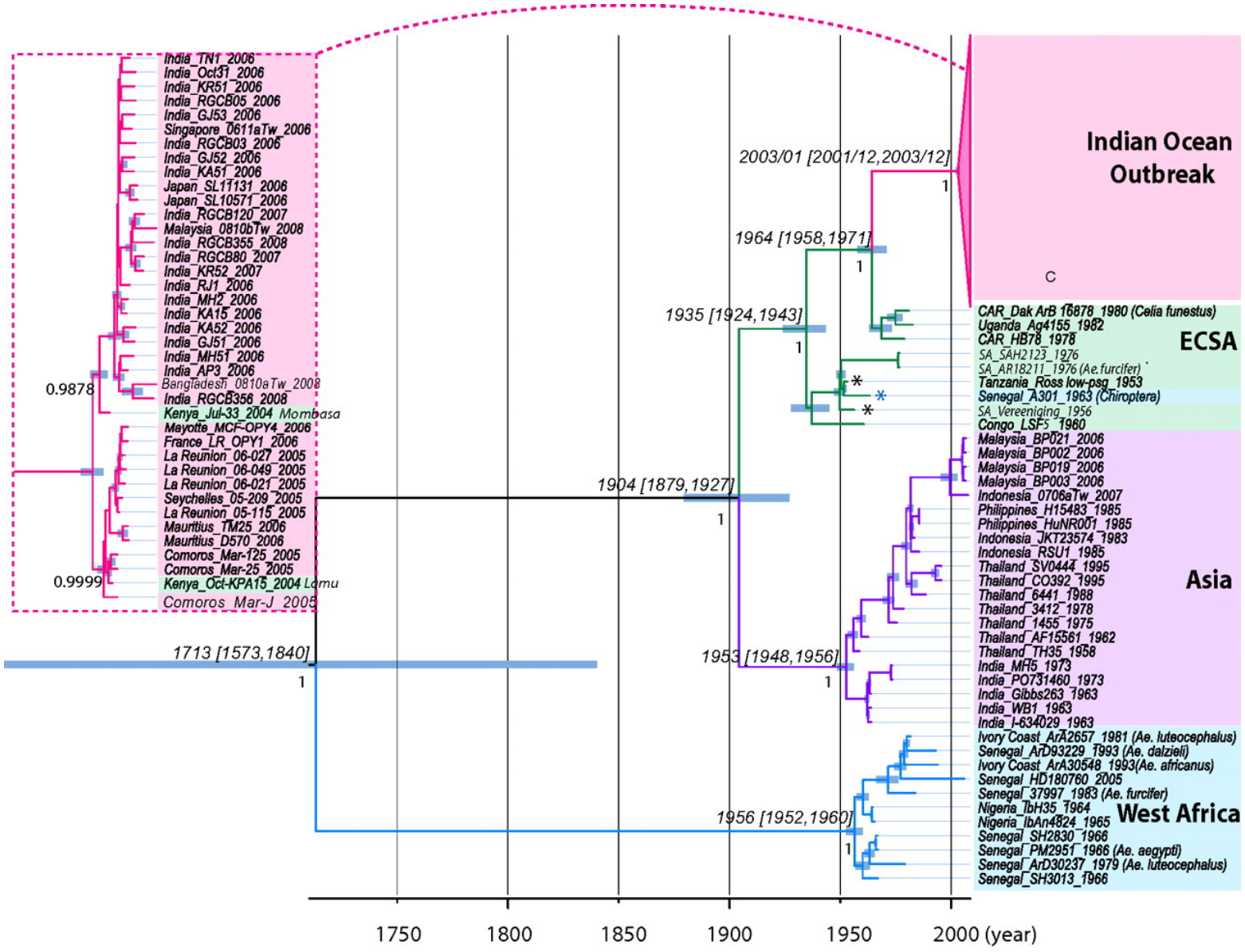
Global expansion of CHIKV (before 2013)

- **Pre 1950s:** Africa and Asia (Central / East Africa origins)
- **1953**: virus isolated (Tanzania)
- **1950s & 60s:** Large outbreaks on Indian subcontinent; disappeared 1970.
- **2005**: Urban epidemic in Indian Ocean; spread to subcontinent.
- **2006 onwards**: imported cases in Europe, USA and Caribbean; outbreak in Italy.



The strain responsible for the Indian Ocean outbreak arose from ECSA strain

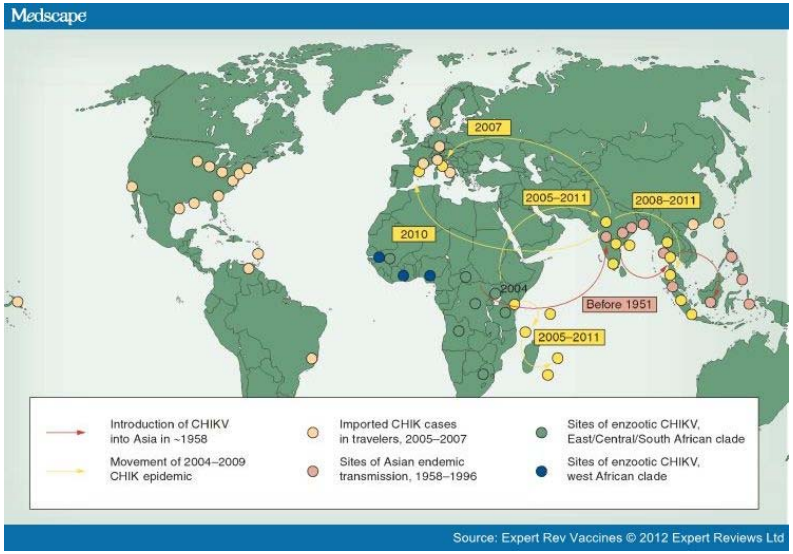
Maximum clade credibility (MCC) tree of 80 CHIKV strains.

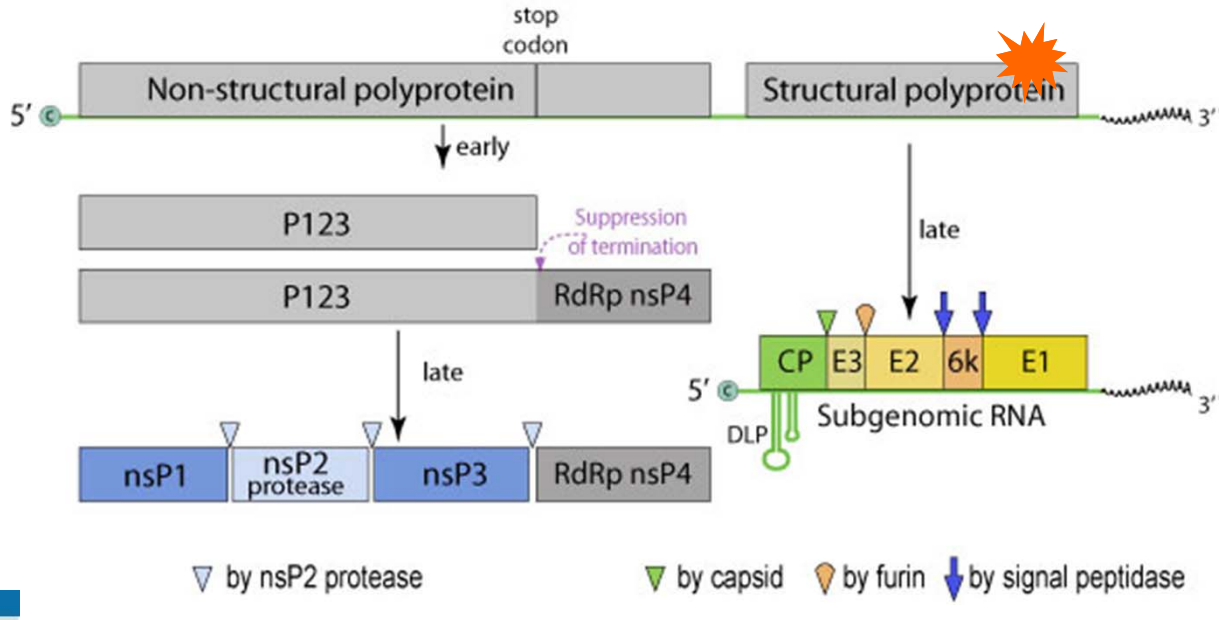


Volk S M et al. J. Virol. 2010;84:6497-6504

Factors underlying the recent global emergence of CHIKV

Increased tourism
Rapid global transport



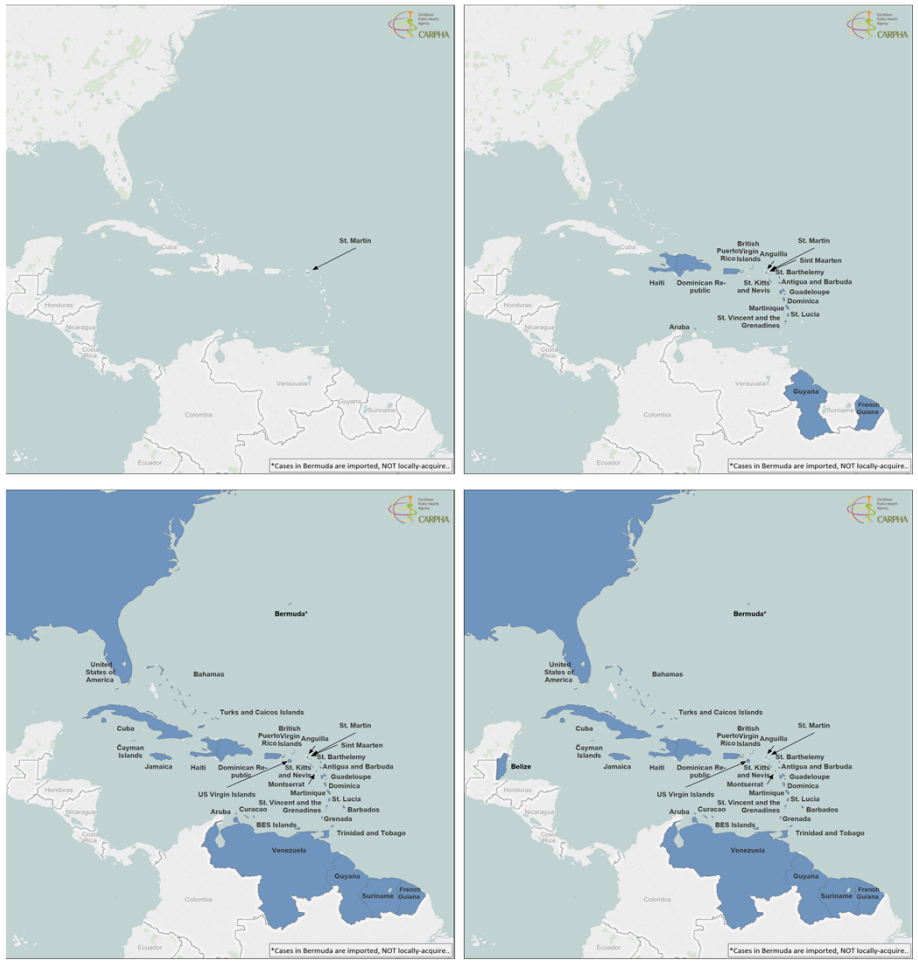


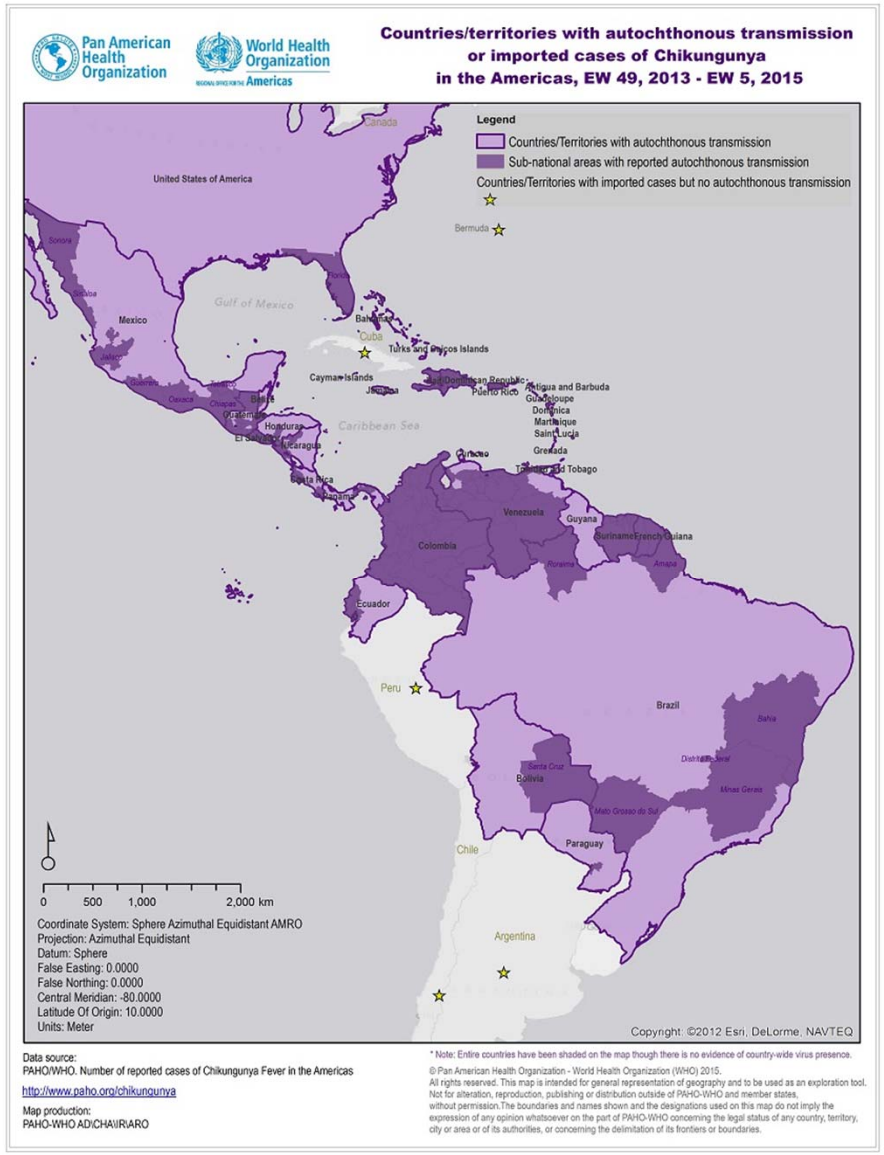
Evolutionary changes
(adaptive mutation in E1 favours replication in *Ae. albopictus;* secondary mutations in E2 improve adaptation)

Immune landscape
(Introduction in naïve populations)

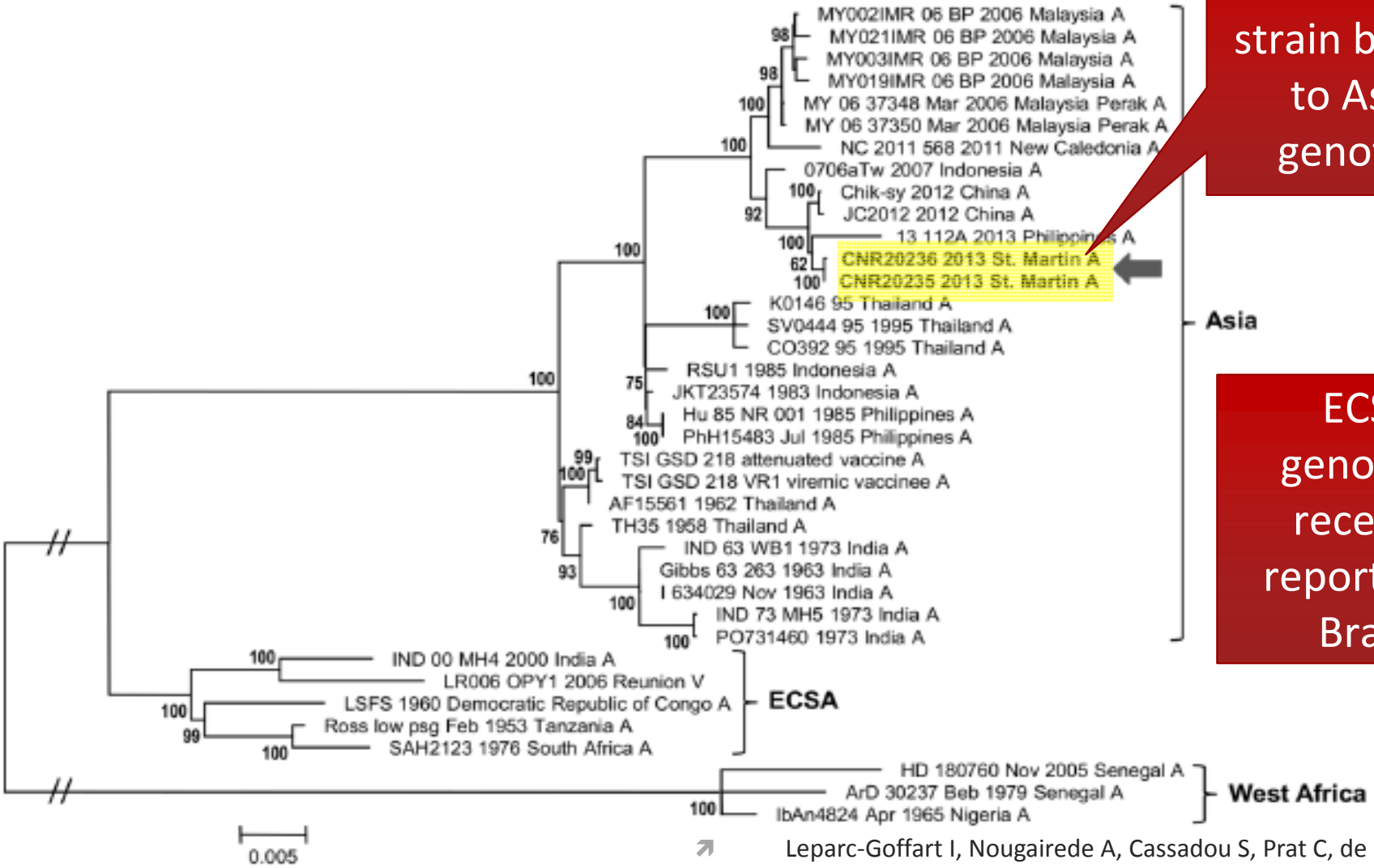
Expansion of CHIKV to the Americas (Dec 2013 – present)

- **Dec 2013:** St. Martin outbreak; spread to other countries in region.





Phylogeny of Chikungunya viruses associated with outbreak in Saint-Martin

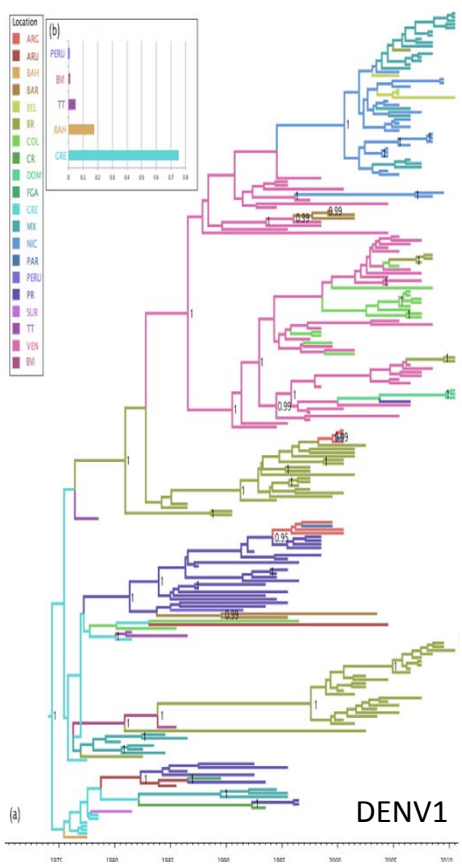


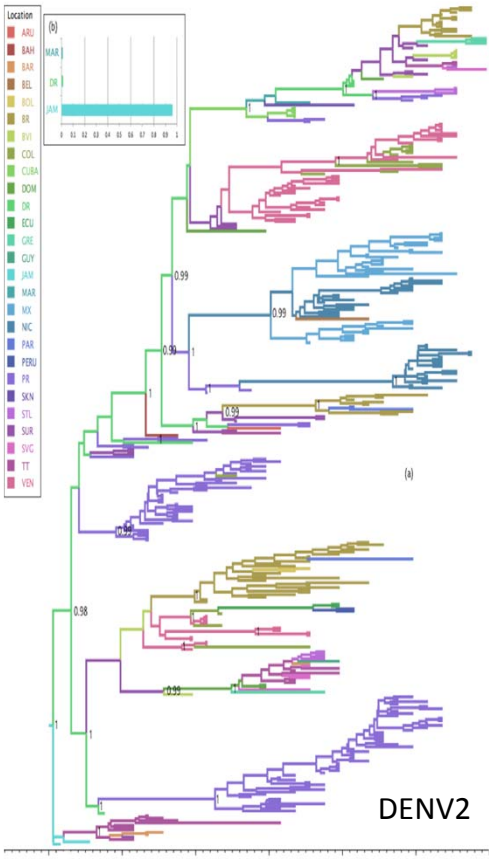
St. Martin strain belongs to Asian genotype

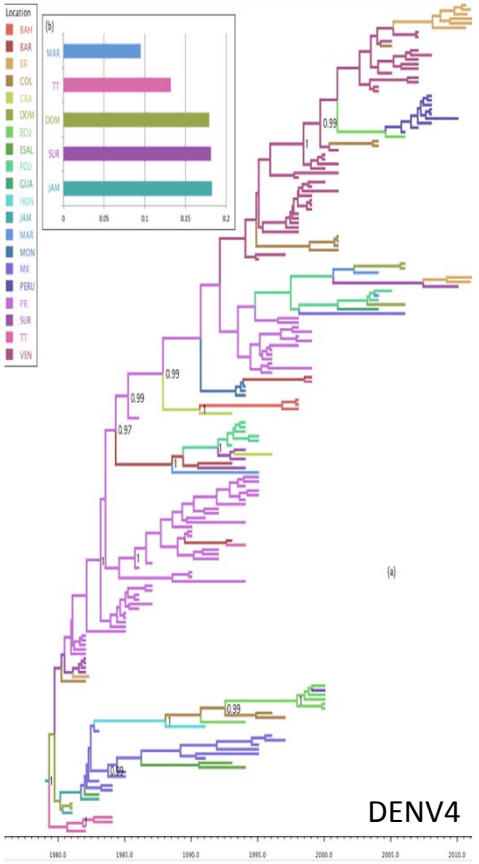
ECSA genotype recently reported in Brazil

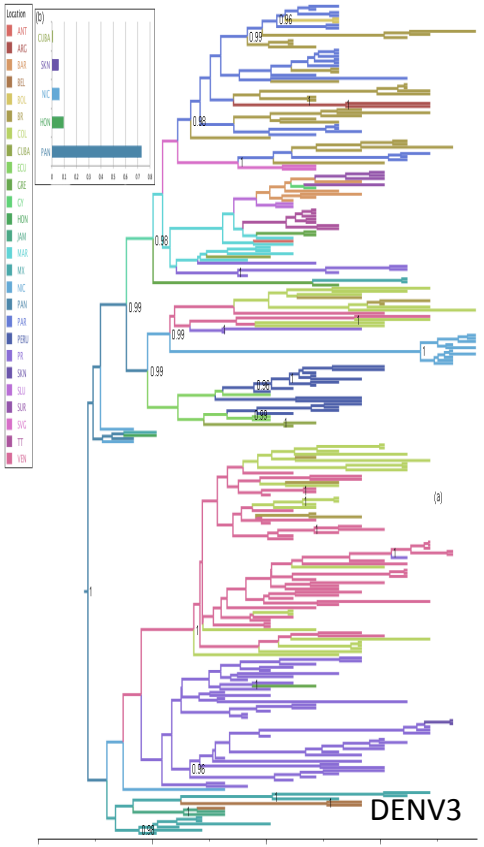
Leparc-Goffart I, Nougairede A, Cassadou S, Prat C, de Lamballerie X. Chikungunya in the Americas. *Lancet* 2014; **383:** 514.

Phylogenies say more than who is related to whom









Strong spatial structure

Clear pattern of lineage extinction and replacement (within and among countries)

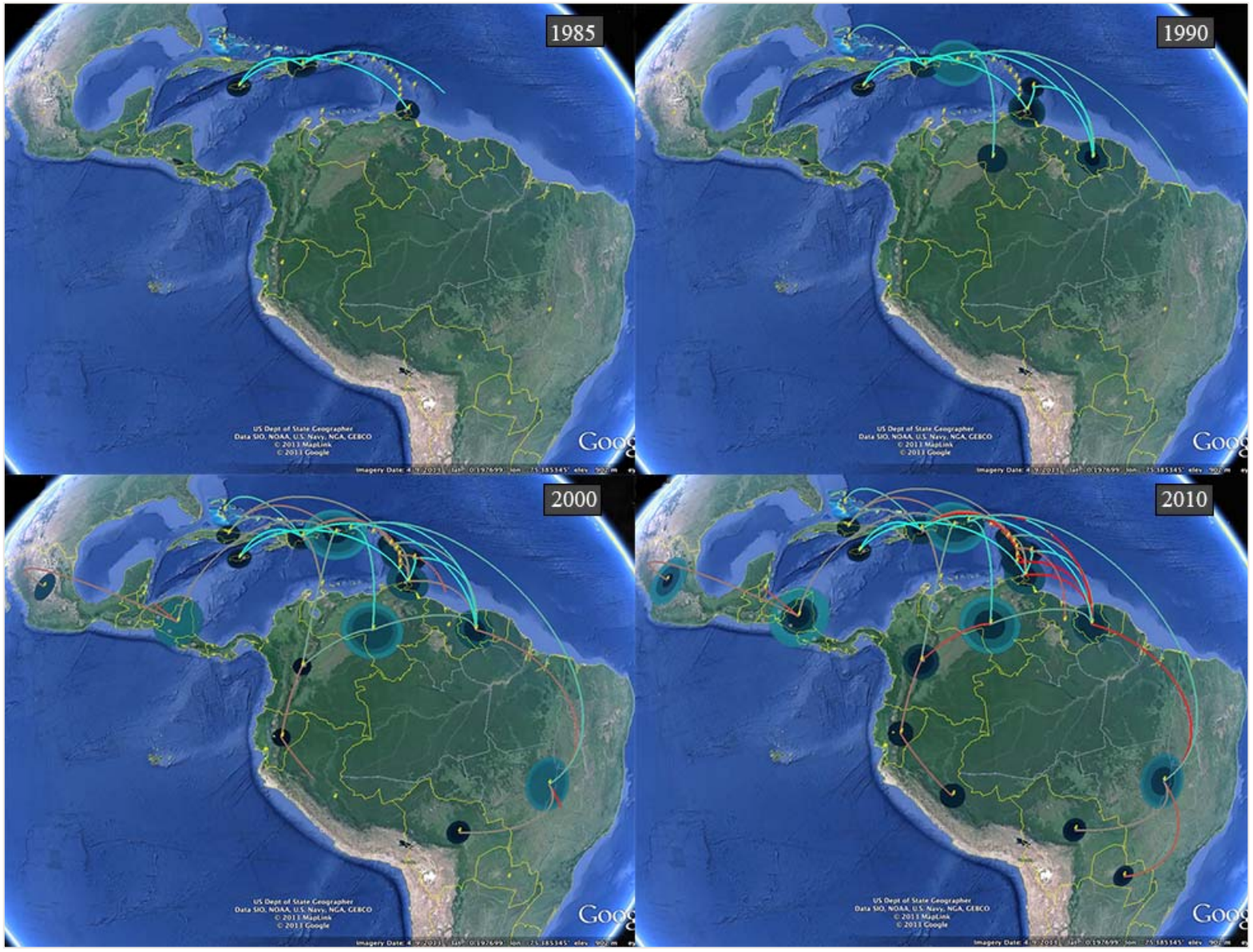
More gene flow among countries

Less lineage turnover

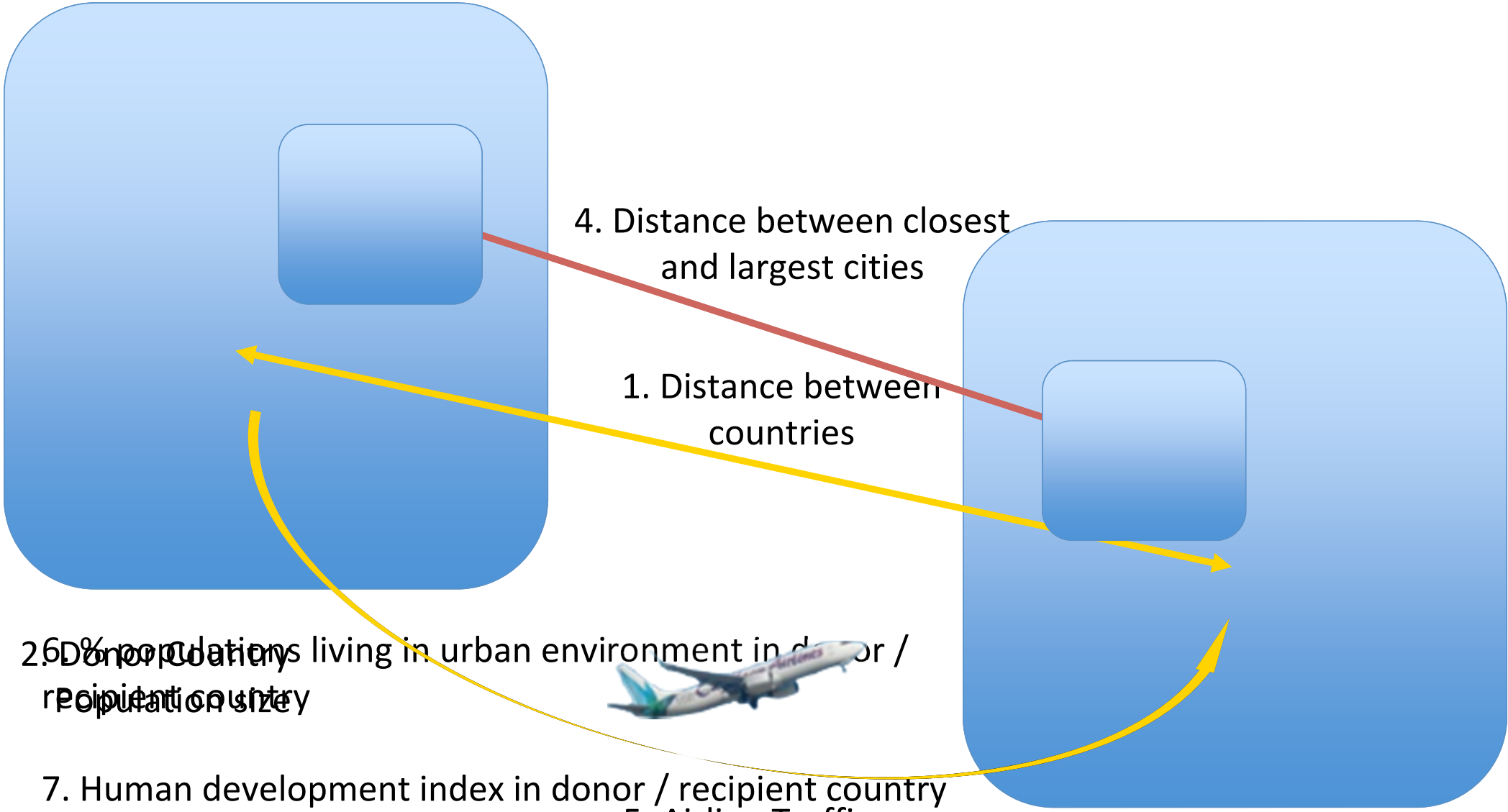
Exponential growth

Spread of DENV 1 inferred from sequence data.

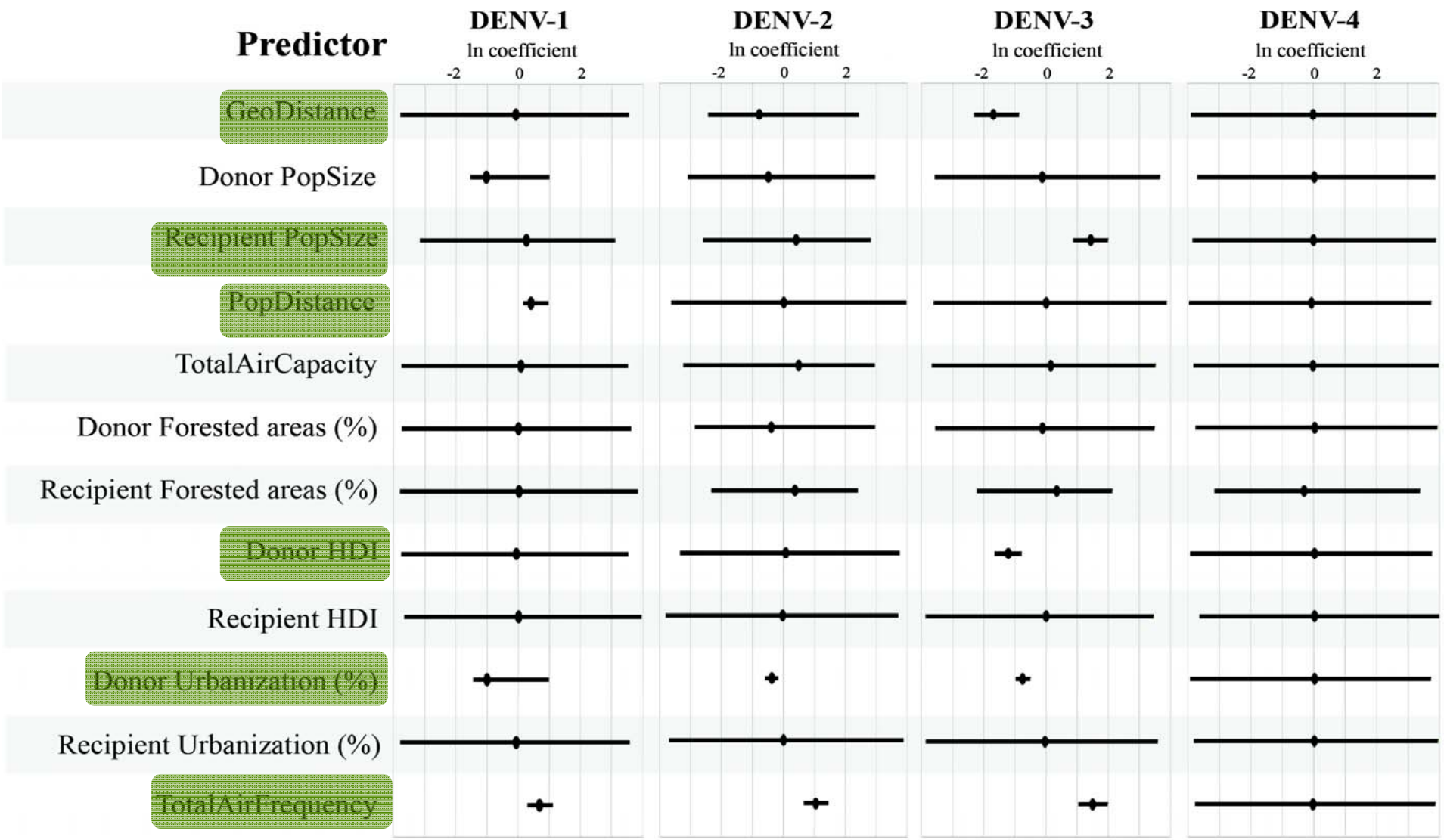
Rapid dispersal followed by more localized maintenance



Testing predictors of dengue virus spread in the Americas



Correlations between predictors & rates of DENV geographic spread within the Americas



Predictors of CHIKV spread

- *"strong spatial signature in the regional epidemic, with the risk of transmission between areas estimated to be inversely proportional to the distance rather than driven by air transportation"*

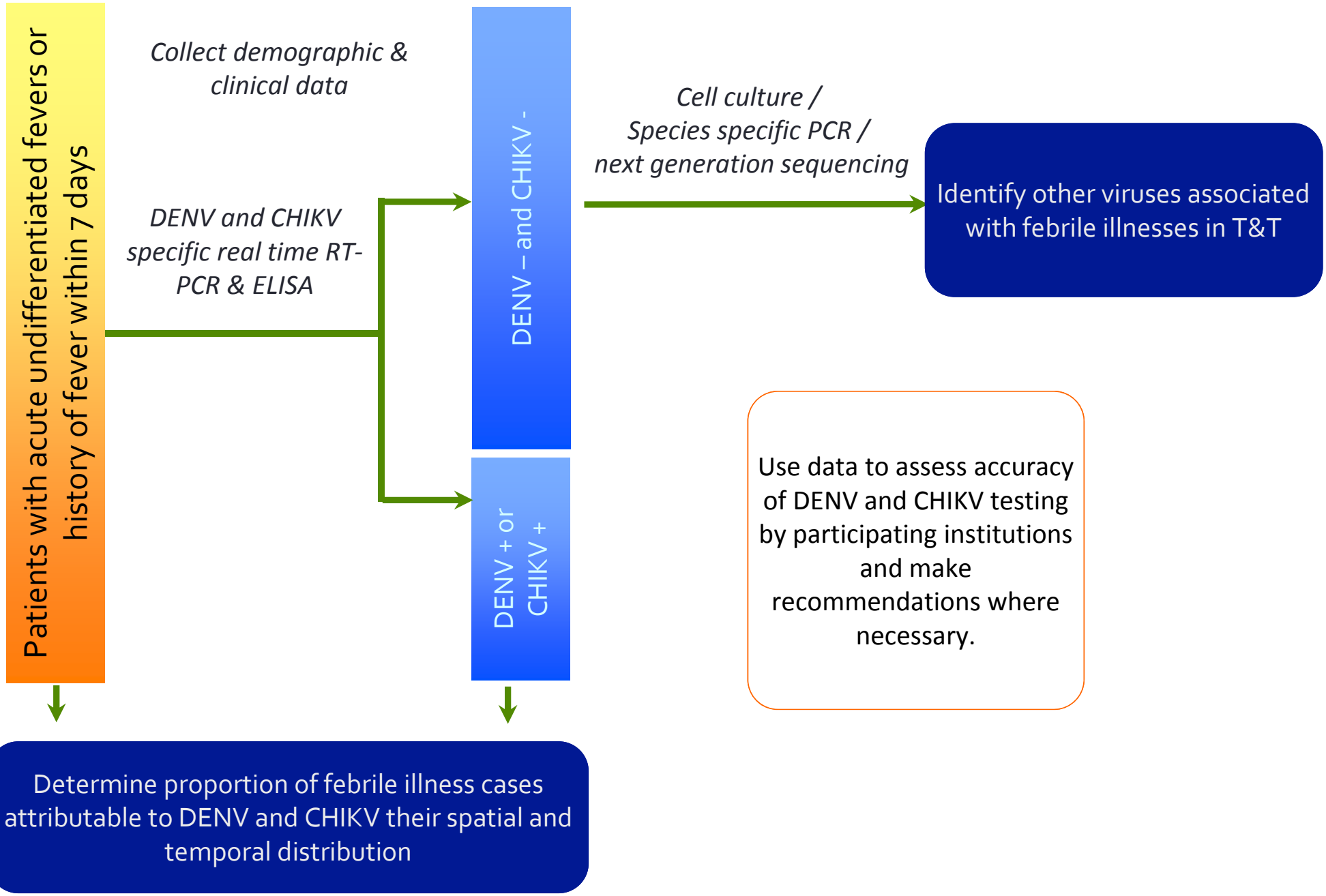
  Cauchemez et al. Euro Surveill. 2014 Jul 17;19(28):20854.

- Consider
  - Immunologically naïve populations
  - Rapid expansion and evolution
  - Adaptive mutations & *Ae. albopictus*

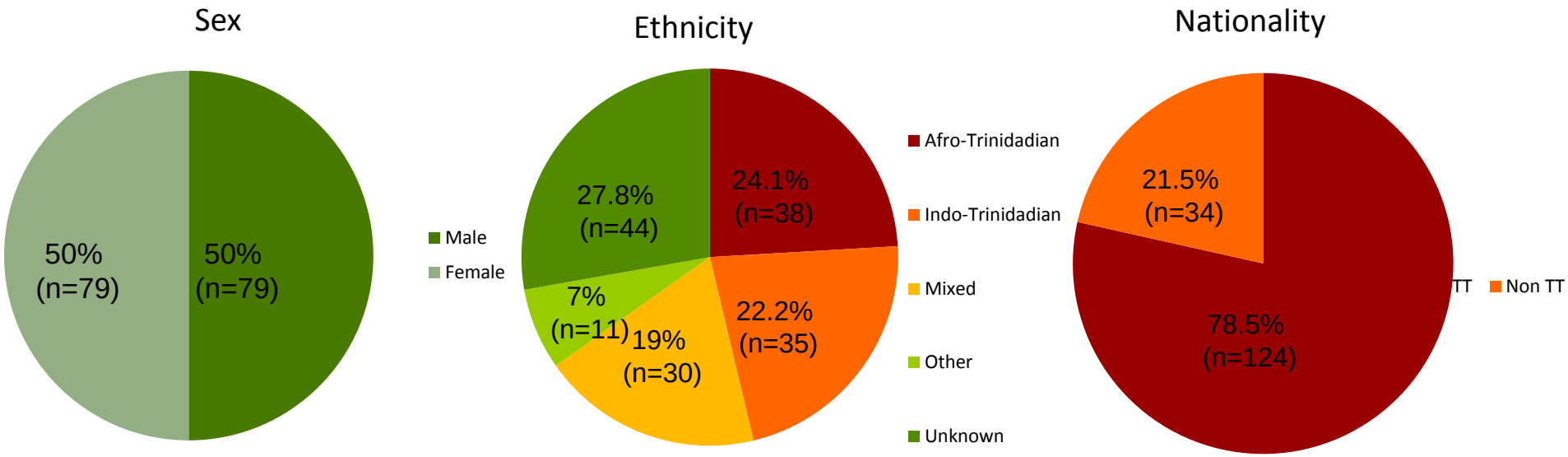
The CHIKV Challenge

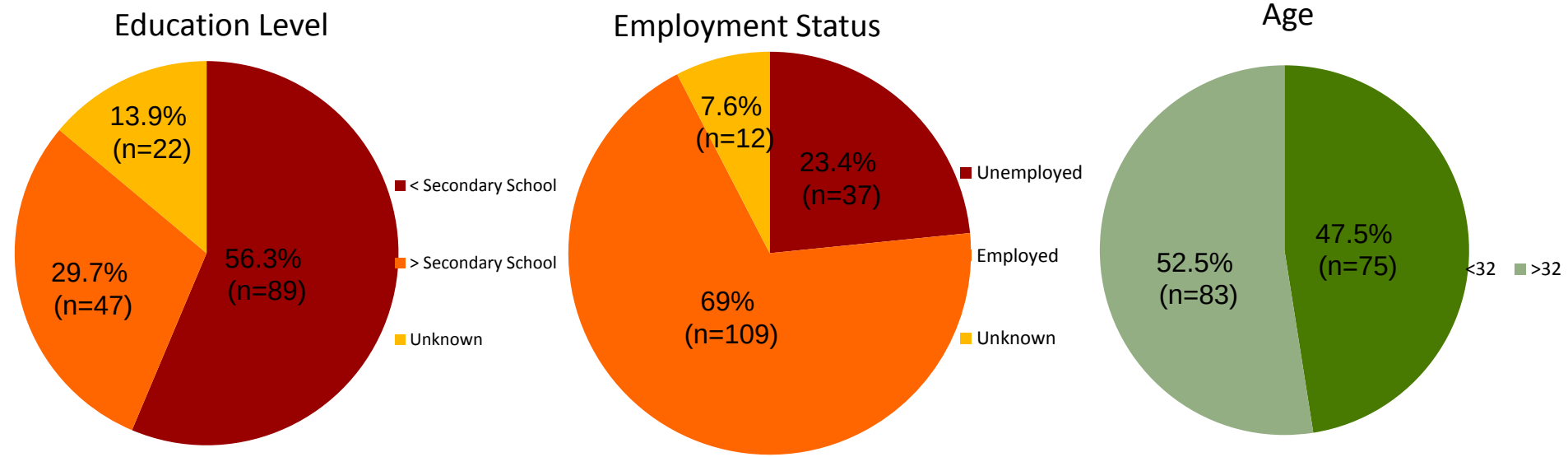
- Debilitating pain, persistent symptoms → economic losses
  - In naïve populations, seroconversion rates = 30 – 60%
  - Decrease in tourism e.g. 60% during 2005-2006 outbreak in La Reunion (Soumahoro et.al. 2011)
  - Jamaica estimated $30 million loss (Jamaica Observer, Oct 20th 2014)
- No vaccine
  - straight forward? (only one serotype; confers life long immunity)
  - Will apparently immune mediated pathology be an issue?
- Rapidly expanding and evolving → increased chance of adaptive mutations
  - *Ae. albopictus* more common in temperate areas, more diverse breeding sites
- Mosquito control difficult

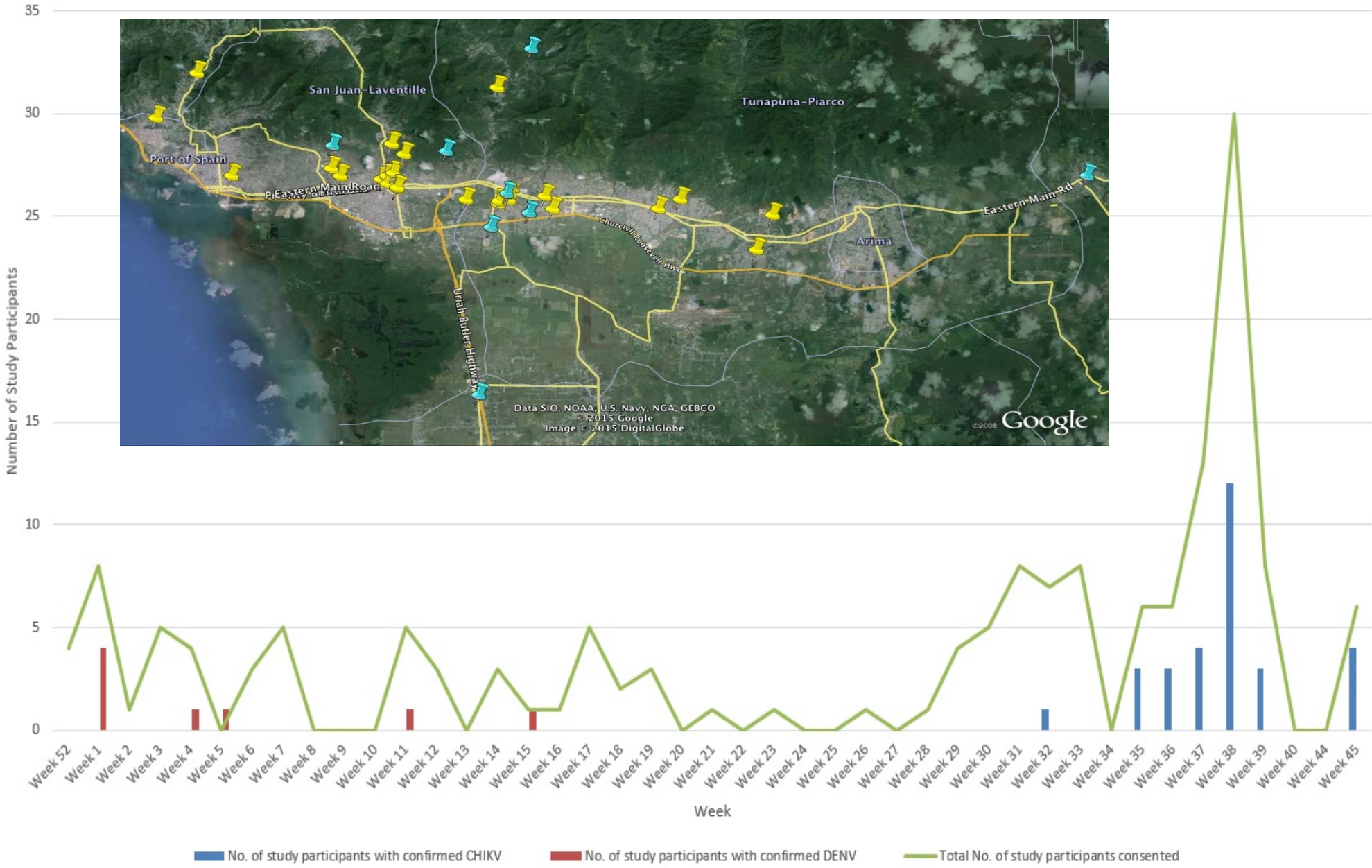
Febrile illness surveillance study



158 individuals presenting at the Adult Priority Care Facility (APCF) of the EWMSC (25 Dec 2013 – 5 Nov 2014)






San Juan-Laventille
Tunapuna-Piarco
Port of Spain
Eastern Main Road
Eastern Main Rd
Arima
Uriah Butler Highway
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2015 Google
Image © 2015 DigitalGlobe
Google
Number of Study Participants
35
30
25
20
15
10
5
0
Week 52
Week 1
Week 2
Week 3
Week 4
Week 5
Week 6
Week 7
Week 8
Week 9
Week 10
Week 11
Week 12
Week 13
Week 14
Week 15
Week 16
Week 17
Week 18
Week 19
Week 20
Week 21
Week 22
Week 23
Week 24
Week 25
Week 26
Week 27
Week 28
Week 29
Week 30
Week 31
Week 32
Week 33
Week 34
Week 35
Week 36
Week 37
Week 38
Week 39
Week 40
Week 44
Week 45
Week
No. of study participants with confirmed CHIKV
No. of study participants with confirmed DENV
Total No. of study participants consented

Results (CHIKV+ vs CHIKV-)

Patients with confirmed CHIKV infection:

- more likely to report joint pain (83%, *p*=0.020)
- less likely to have travelled outside of Trinidad in the 2 weeks prior to interview (17%, *p*=0.044)
- less likely to have reported having laboratory-confirmed dengue previously (7%, *p*=0.026)
- more likely to have presented to the A&E earlier (mean days post onset of fever: 2.39 [CHIKV+] vs. 3.33 [not CHIKV+], *p*=0.021)
- lower mean white blood cell counts (6.52 [CHIKV+] vs. $8.36x10^3$/ml [not CHIKV+], *p*=0.016)

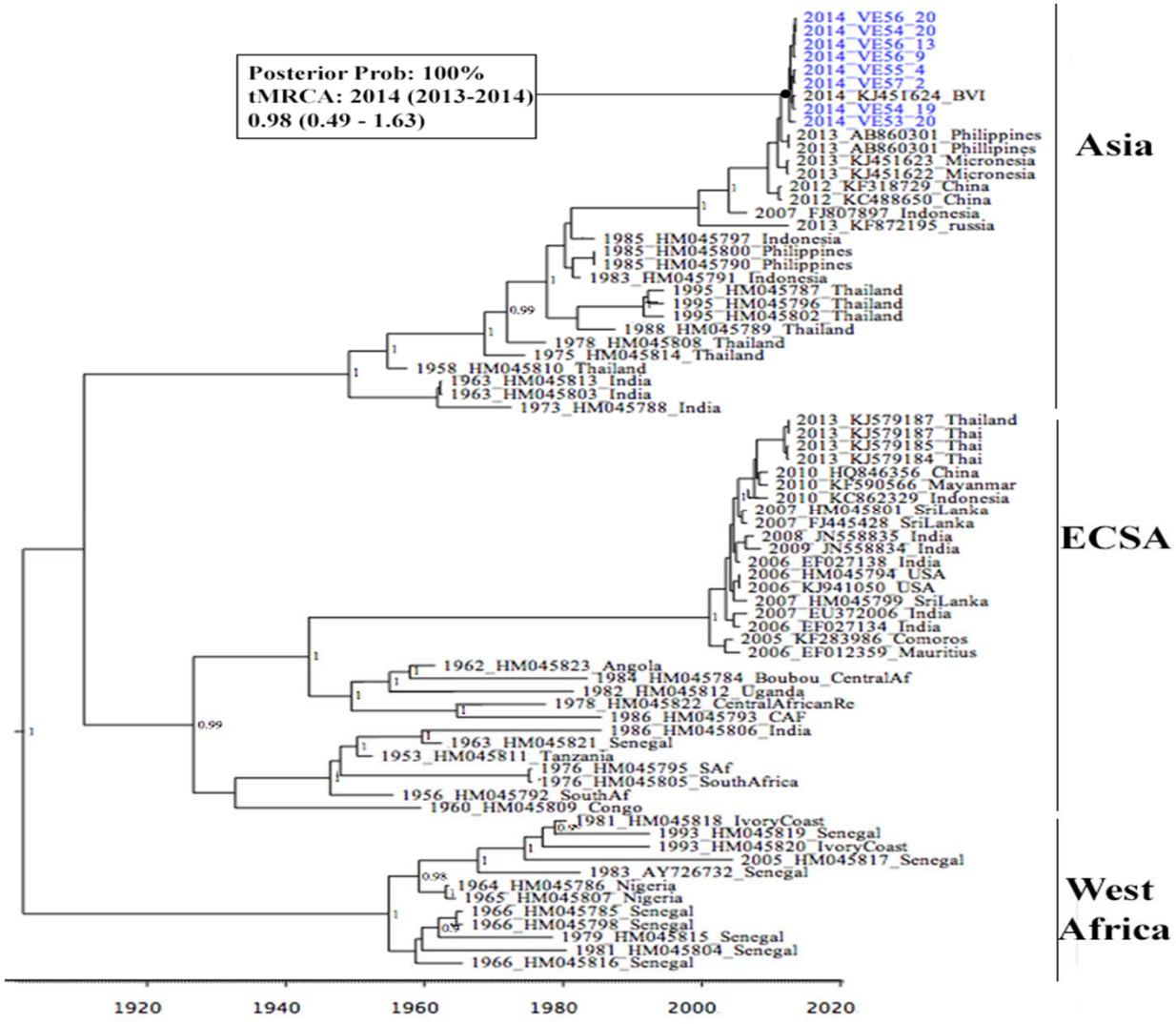
Results (CHIKV+ vs DENV+)

Between the 38 patients CHIKV+(n = 30) and with confirmed DENV (n=8)

- No statistically significant differences ($p < 0.05$) in *symptomatic presentations*
  - when compared to DENV, persons with CHIKV more often reported joint pain (83% vs. 50%, *p*=0.071) and rash (33% vs. 0%, *p*=0.082), and less often reported abdominal pain (7% vs. 38%, *p*=0.053)

- Patients with DENV had significantly lower median white blood cell count (3.50 vs. 6.00 $x10^3$/ml, p=0.028) and platelet counts (1.47 vs. 2.39$x10^5$/ml, p=0.022)

MCC phylogeny of CHIKV whole genomes



Within-Host Variation (CHIKV)

For 3 of the 8 individuals there was sufficient depth of sequencing for reliable SNV analysis.

3 individuals

- Nt position 9377 G→A
  - Residue 604 in E2, Gly → Glu
  - Freq 2.03%, 2.08% and 2.69%

1 individual

- Nt position 1229 T→A, 1.1%
  - Reside 384 nsp1, Leu → Leu
- Nt position 3277 G→T , 1.6%
  - Residue 1067 nsp2, Ser → Ile

1 individual

- Nt position 9039 G →T, 1.18%
  - Residue 491 E2, Glu → Stop codon

Why do CHIKV symptoms persist?

- Molecular mechanism of chronic arthralgia still not well understood
- Nature of joint pain appears to be inflammatory
  - Markers of inflammatory response found in tissues
- Patients are positive for markers of RA (Manimunda *et.al.* 2010)
  - But no RA-related classic symptoms *e.g.* erosion of cartilage/bones
- Theories :
  - Viral persistence in tissue sanctuaries
    - Evidence from mouse studies that CHIKV RNA (of all strains) persists, specifically in joint tissue
    - Adaptive B- and T-cell responses clear virus in muscle tissue but incompletely in joint tissue
    - Prophylactic mAb treatment prevented persistant infection but only tissue-specific effects when administered therapeutically
  - Re-activation of virus
  - Uncontrolled pro-inflammatory cytokine response and/or cross-reactivity with self antigen

Manimunda et.al. (2010); Hawman et.al. (2013)

Deep sequencing of CHIK -/DENV - samples

ID	Contig Length	Virus Protein Hit (VPH)	Length of Virus (nt)	No. of pr Matches	No. of nt Matches
VE41-6	5395	Enterobacteria phage, phiX174	5386	55	1
	489	Cyprinid herpesvirus 3 (Complete genome)	295146	12	0
VE47-4	1375-9447 (2)	Enterobacteria phage, phiX174	5386	110	3
	522	Bovine herpesvirus 1 (Complete genome)	135301	24	2
	462	Suid herpesvirus 1 (Complete genome)	143461	49	3
VE50-2	5419	Enterobacteria phage, phiX174	5386	53	2
	420-610 (3)	Human Immunodeficiency Virus 1 (HIV1) (Complete genome)	9181	48	3
	474-602 (2)	Suid herpesvirus 1 (Complete genome)	143461	99	2
	405	Torque teno virus 3 (Complete genome)	3478	2	1
VE50-8	5411	Enterobacteria phage, phiX174	5386	55	1
	659	Bovine herpesvirus 1 (Complete genome)	135301	74	0
VE51-1	5411	Enterobacteria phage, phiX174	5386	55	1
	564	Cyprinid herpesvirus 3 (Complete genome)	295146	26	0
	487	Suid herpesvirus 1 (Complete genome)	143461	22	4

Acknowledgements

- Orchid Allicock
- Anushka Ramjag
- Nikita Sahadeo
- Collaborators
- UWI-RDI Fund, ISID, Campus Research and Publication Fund

Dr. Rembert Pieper

Proteomics as a Complement to Microbiome and Metagenomic Studies

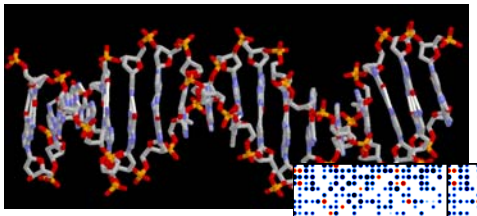
Omics Technologies: Overview

Rembert Pieper, Associate Professor,
J. Craig Venter Institute

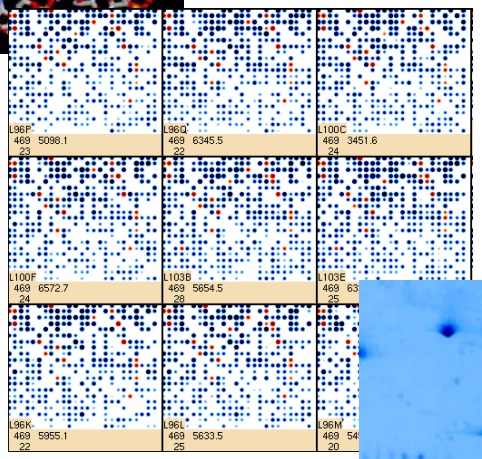
Course, University of West Indies, February 2015

J. Craig Venter™
INSTITUTE

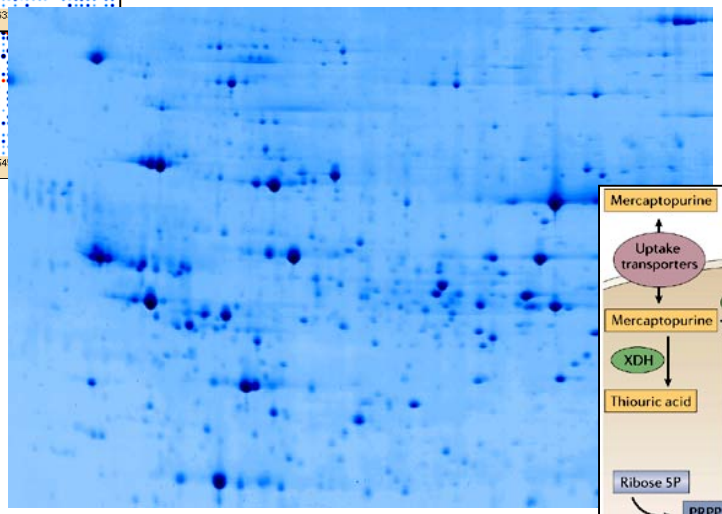
Genomics ……. Proteomics



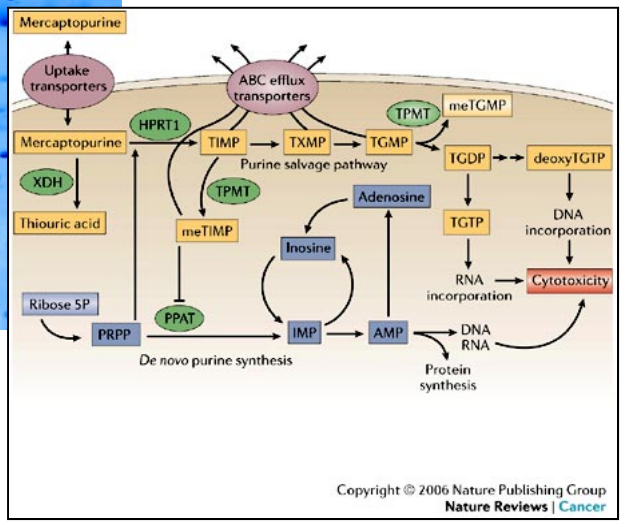
Genome



Transcriptome

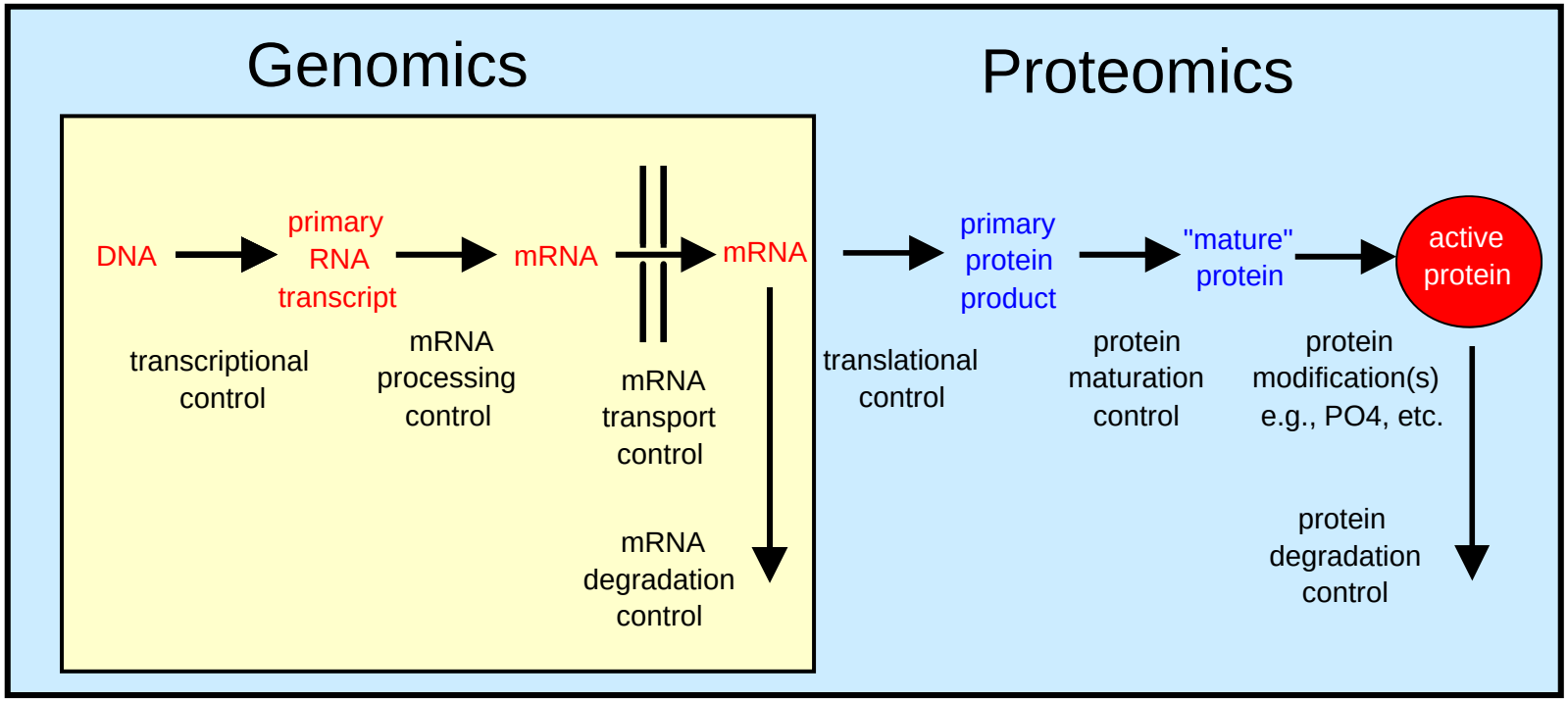


Proteome



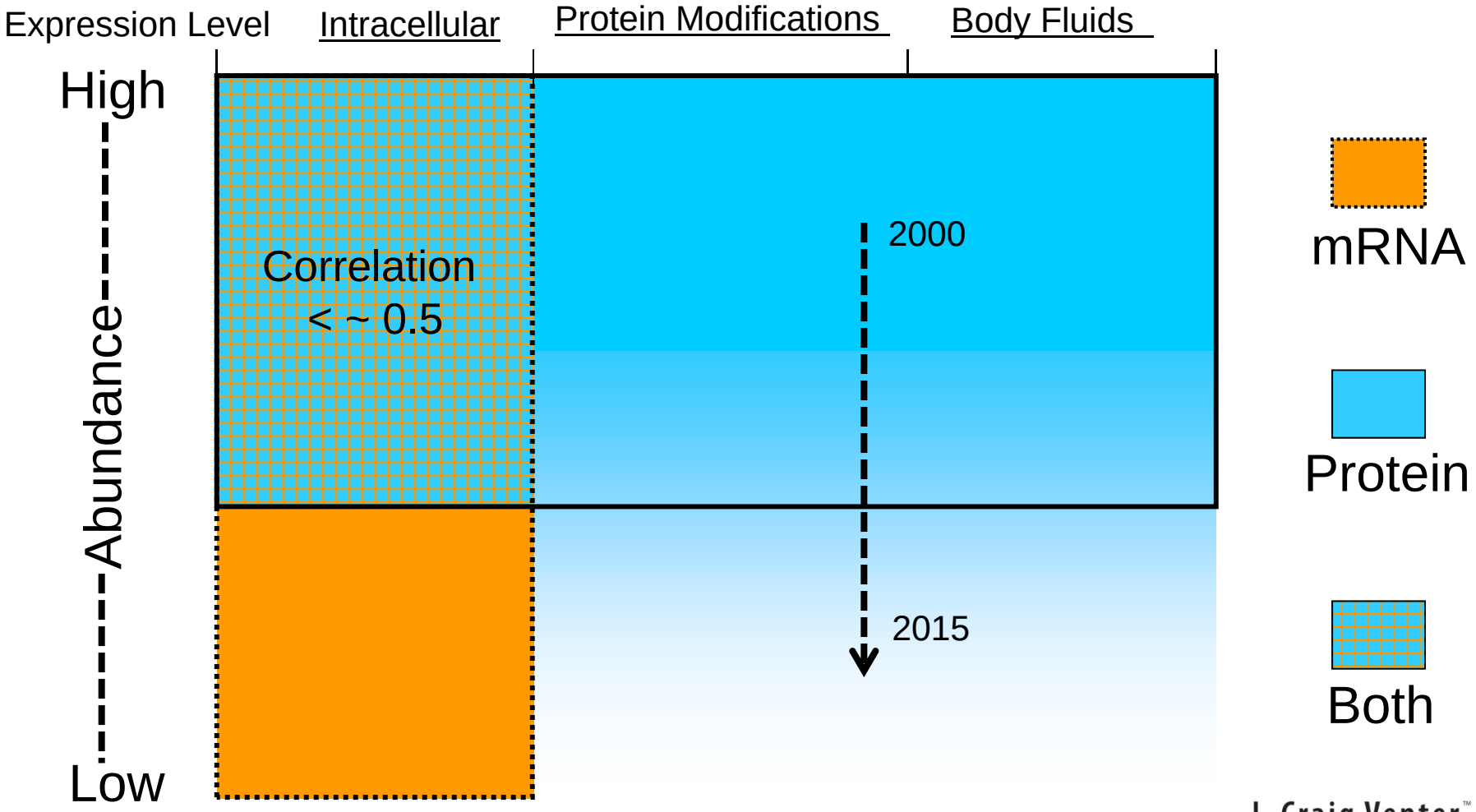
Metabolome

Genomics ……. Proteomics



J. Craig Venter™ INSTITUTE

Synergy and Overlap of Protein and mRNA Measurements



Slide from N. Leigh Anderson

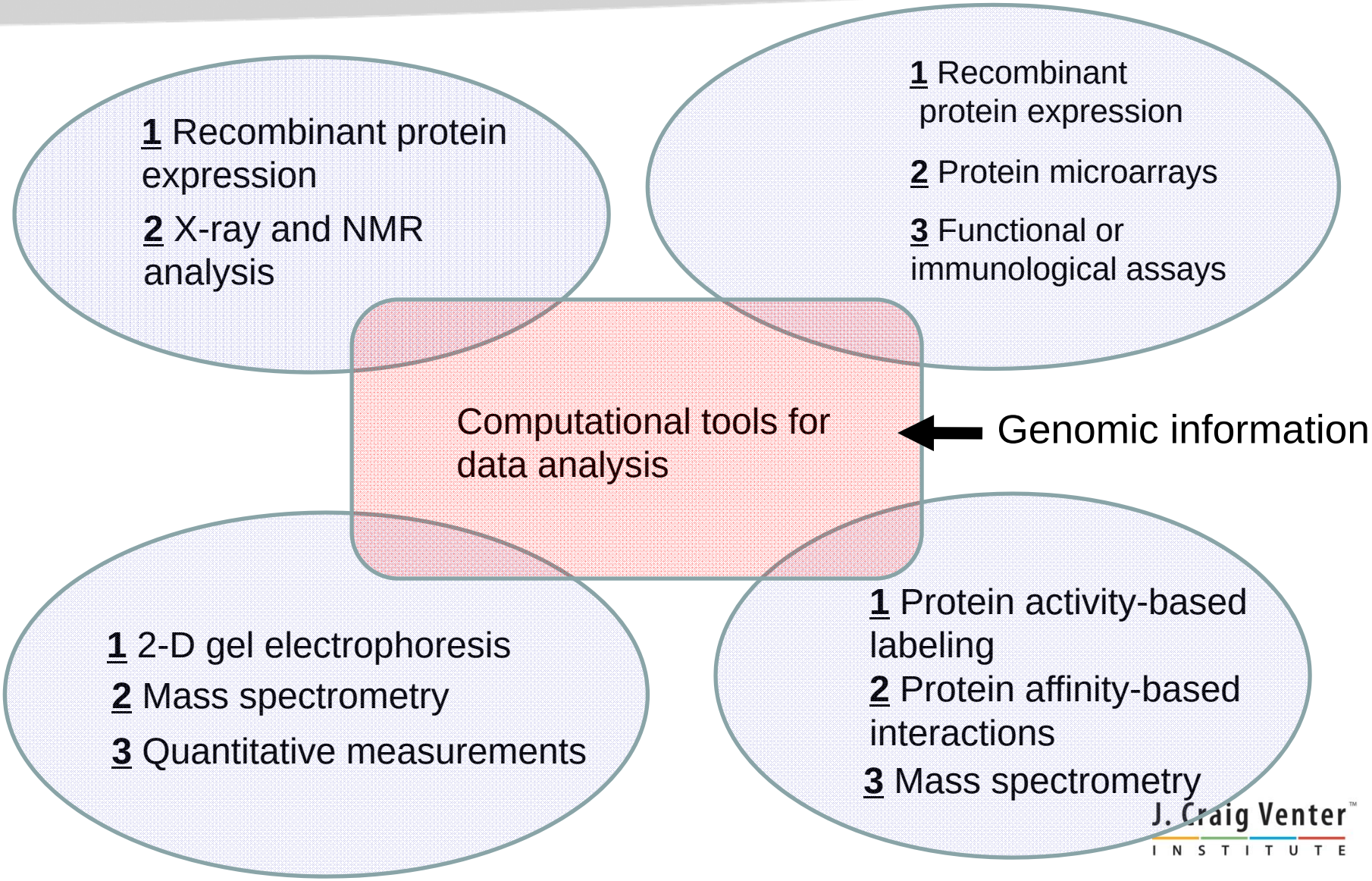
J. Craig Venter INSTITUTE

Proteomics: There Are Many Objectives

- Characterize the proteome of a cell, a subcellular organelle, a protein complex
  - Identification – Quantitation – Functional Analysis – Structural Analysis.
  - …….. with decreasingly parallel applications
- Surrogate biomarker for disease onset or progression, therapeutic treatment or toxicity of a drug
- Metaproteomics in the age of metagenomics
  - Host-pathogen interactions
  - Mutualism in microbial communities
  - Using proteomics to enhance genome annotation

J. Craig Venter™ INSTITUTE

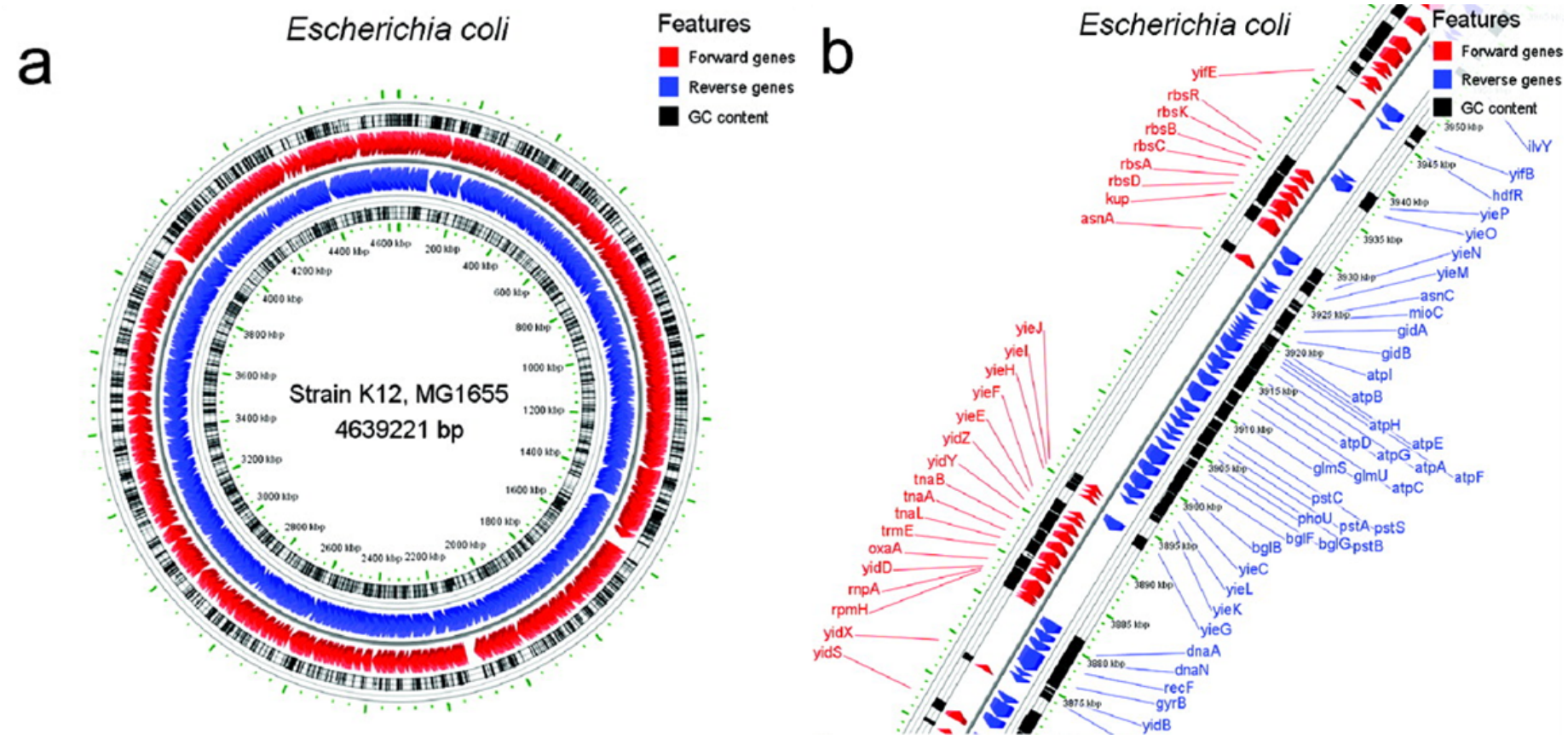
Proteomics - Bioinformatics Interface



J. Craig Venter™ INSTITUTE

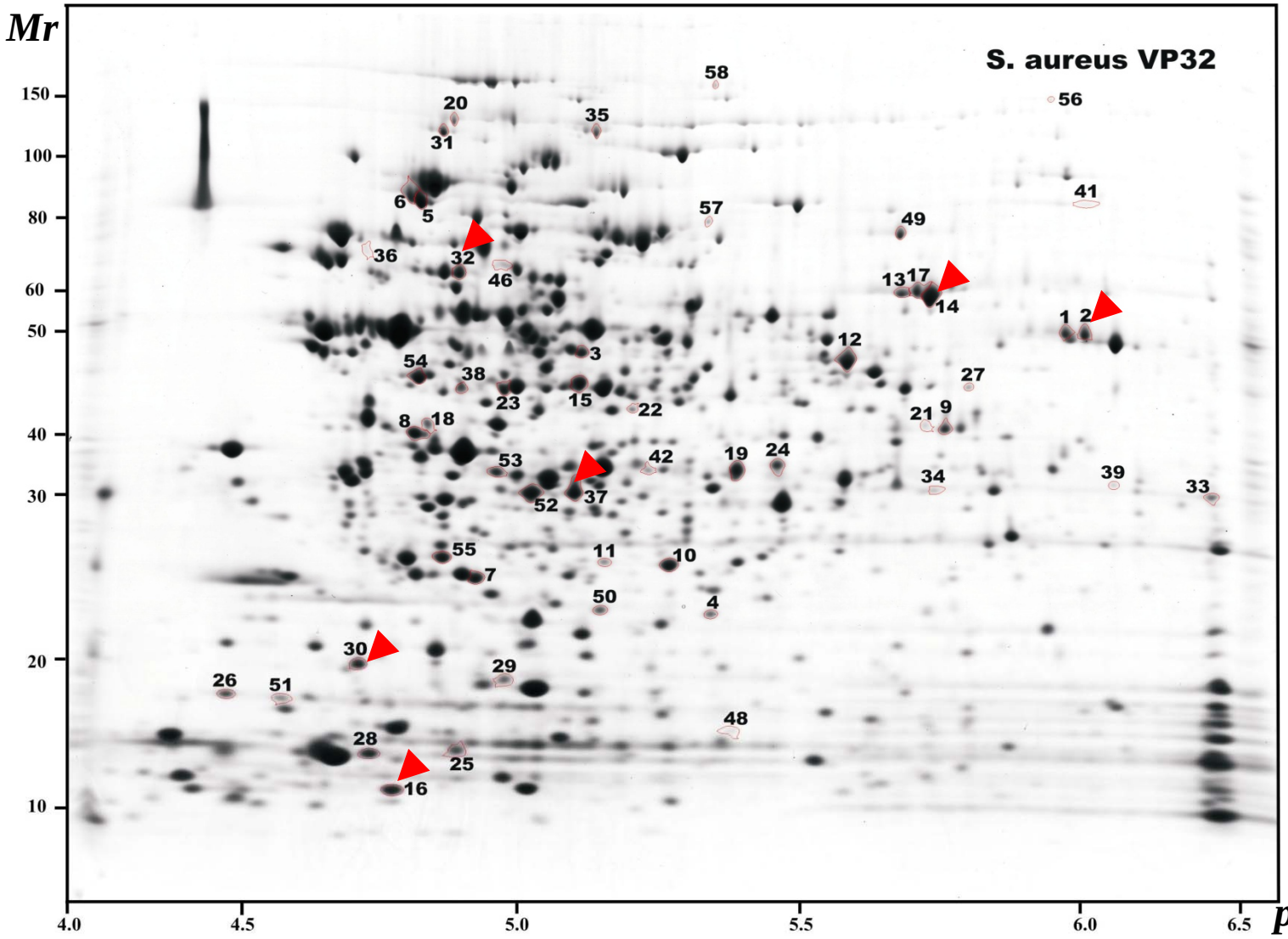
Why the Need for Genome Information in Proteomics ?

Because the sequence of the protein is essential for reagent design and automated identification



J. Craig Venter INSTITUTE

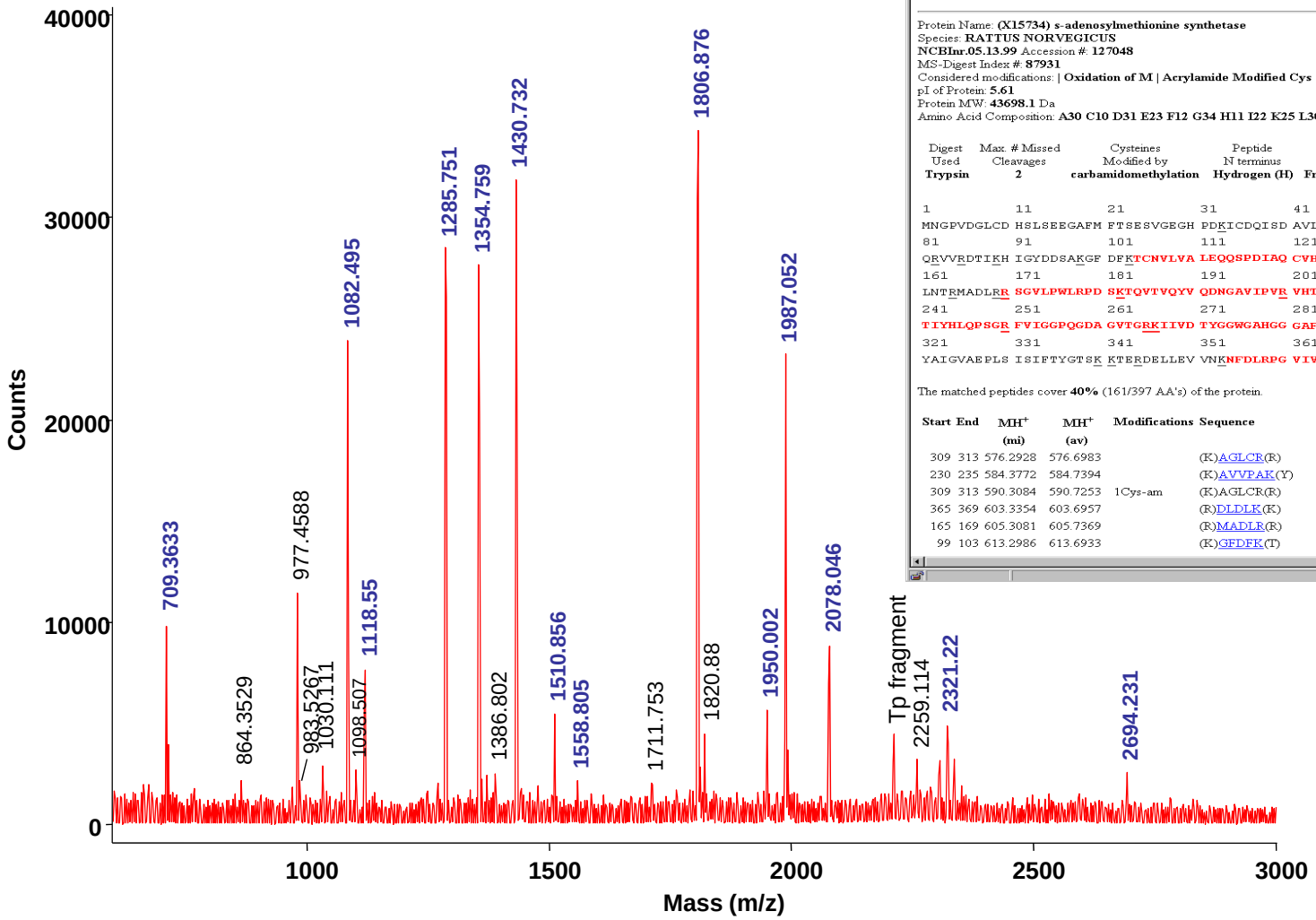
How Analytical Proteomics Started
2-D Gels

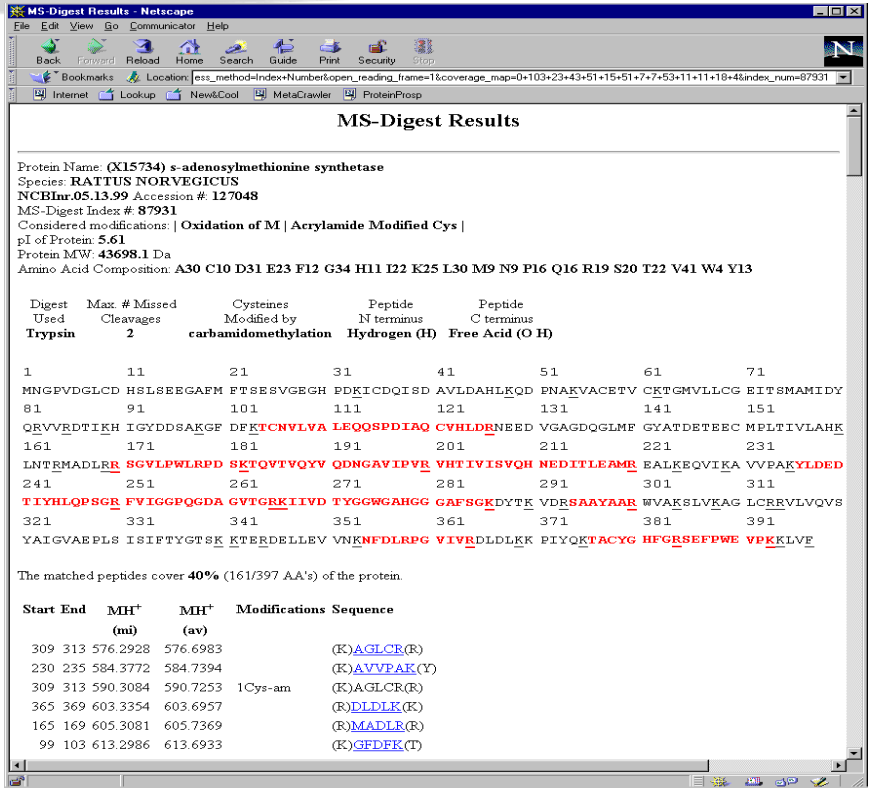


J. Craig Venter™ INSTITUTE

Protein Identification: Mass Spectral Data

MALDI peptide fingerprint of *Rattus norvegicus* S-adenosylmethionine synthetase





MS-Digest Results - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Guide Print Security Stop

Bookmarks Location: ess_method=Index+Number&open_reading_frame=1&coverage_map=0+103+23+43+51+15+51+7+7+53+11+11+18+4&index_num=87931

Internet Lookup New&Cool MetaCrawler ProteinProsp

MS-Digest Results

Protein Name: **(X15734) s-adenosylmethionine synthetase**
Species: **RATTUS NORVEGICUS**
NCBInr.05.13.99 Accession #: **127048**
MS-Digest Index #: **87931**
Considered modifications: | **Oxidation of M** | **Acrylamide Modified Cys** |
pI of Protein: **5.61**
Protein MW: **43698.1** Da
Amino Acid Composition: **A30 C10 D31 E23 F12 G34 H11 I22 K25 L30 M9 N9 P16 Q16 R19 S20 T22 V41 W4 Y13**

Digest Used	Max. # Missed Cleavages	Cysteines Modified by	Peptide N terminus	Peptide C terminus
Trypsin	**2**	**carbamidomethylation**	**Hydrogen (H)**	**Free Acid (O H)**

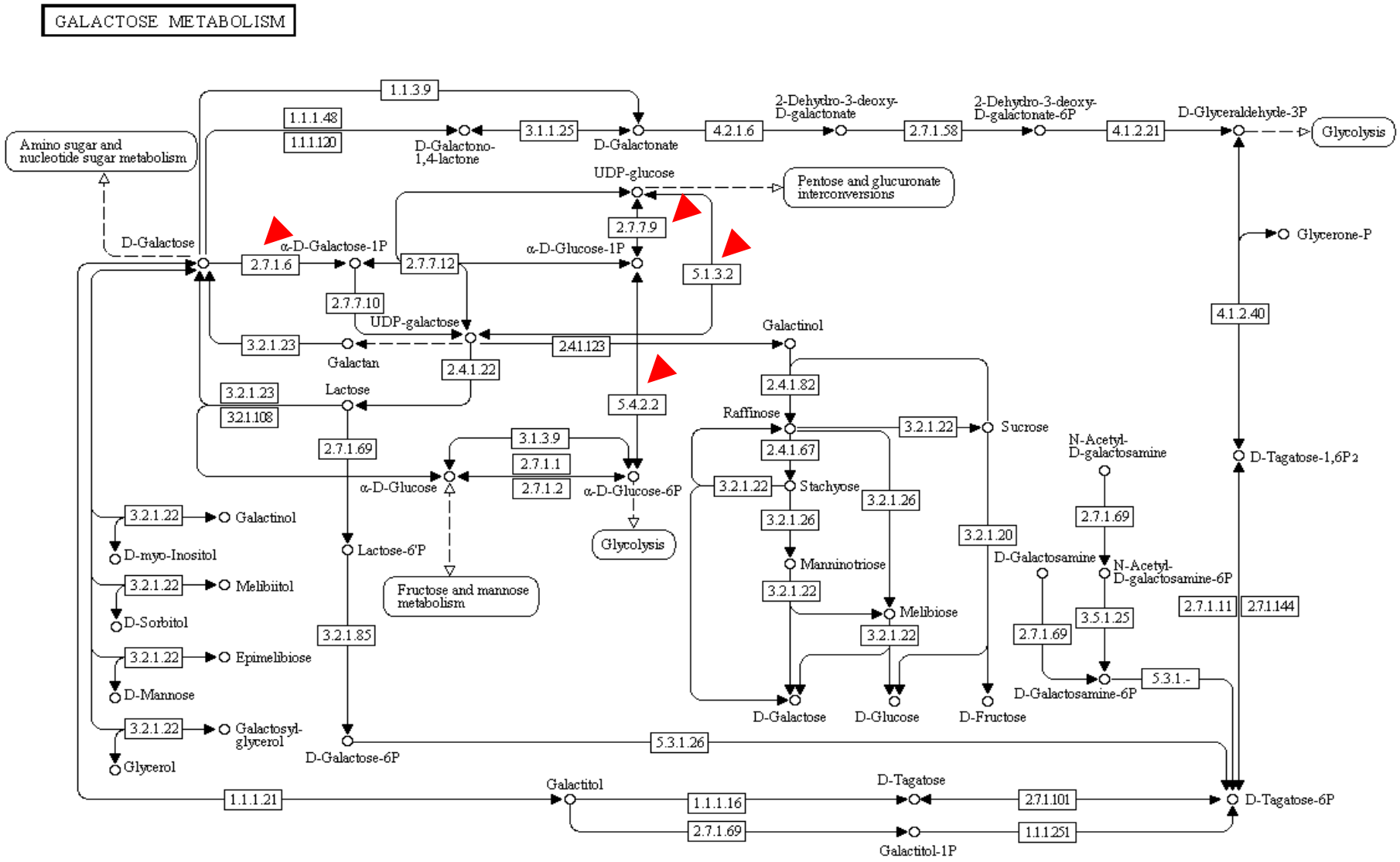
```
1          11         21         31         41         51         61         71
MNGPVDGLCD HSLSEEGAFM FTSESVGEGH PDKICDQISD AVLDAHLKQD PNAKVACETV CKTGMVLLCG EITSMAMIDY
81         91         101        111        121        131        141        151
QRVVRDTIKH IGYDDSAKGF DFKTCNVLVA LEQQSPDIAQ GVHLDRNEED VGAGDQGLMF GYATDETEEC MPLTIVLAHK
161        171        181        191        201        211        221        231
LNTRMADLRR SGVLPWLRPD SKTQVTVQYV QDNGAVIPVR VHTIVISVQH NEDITLEAMR EALKEQVIKA VVPAKYLDED
241        251        261        271        281        291        301        311
TIYHLQPSGR FVIGGPQGDA GVTGRKIIVD TYGGWGAHGG GAFSGKDYTK VDRSAAYAAR WVAKSLVKAG LCRRVLVQVS
321        331        341        351        361        371        381        391
YAIGVAEPLS ISIFTYGTSK KTERDELLEV VNKNFDLRPG VIVRDLDLKK PIYQKTACYG HFGRSEFPWE VPKKLVF
```

The matched peptides cover **40%** (161/397 AA's) of the protein.

Start	End	MH+ (mi)	MH+ (av)	Modifications	Sequence
309	313	576.2928	576.6983		(K)AGLCR(R)
230	235	584.3772	584.7394		(K)AVVPAK(Y)
309	313	590.3084	590.7253	1Cys-am	(K)AGLCR(R)
365	369	603.3354	603.6957		(R)DLDLK(K)
165	169	605.3081	605.7369		(R)MADLR(R)
99	103	613.2986	613.6933		(K)GFDFK(T)

J. Craig Venter™ INSTITUTE

From Proteome Data to Biochemical Pathway Analyses



Tomorrow's Tutorial

- Use Proteome Discoverer
  - Mass spectral data
  - Peptide identification
  - Protein assignment
  - Probability-based scoring
- Use MeV
  - Upload proteomic data
  - Clustering datasets
  - Draw conclusions from quantitative proteomic changes
  - Use databases to assess functional significance

J. Craig Venter™
INSTITUTE

Mass Spectrometry for Proteomics

Rembert Pieper, Associate Professor,
J. Craig Venter Institute

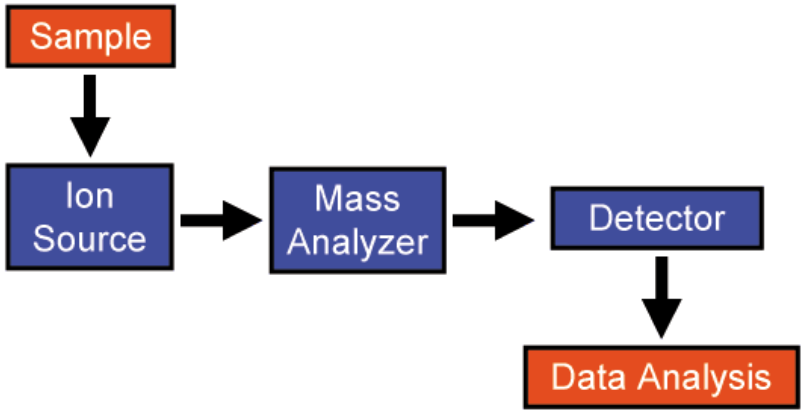
Course, University of West Indies, February 2015

J. Craig Venter™
INSTITUTE

Mass Spectrometry

Mass spectrometry (MS) is an analytical technique for the determination of the elemental composition of a sample or a molecule. It is used to elucidate the chemical structures of molecules, such as peptides and other chemical compounds.

1. Ionization
2. Mass Analysis
3. Detection



J. Craig Venter™ INSTITUTE

Mass Spectrometry Technologies

1. Ionization

- Matrix-Assisted Laser Desorption Ionization (MALDI)
- Electrospray Ionization (ESI)
- Atmospheric Pressure Chemical Ionization (APCI)
  - Nobel Prizes for J. Fenn, K. Tanaka, M. Karas & F. Hillenkamp, 2003

2. Mass Analysis

- Time-Of-Flight (TOF)
- Quadrupole (Q)
- Quadrupole Ion Trap
- Linear Ion Trap Quadrupole (LTQ)
- Fourier Transfrom Ion Cyclotron Resonance (FTMS)
- Orbitrap

3. Detection

- Electron Multipliers
- Ion-to-Photon Detectors
- Pairs of Metal Surface Inductive Detectors

J. Craig Venter™ INSTITUTE

The Physics Laws of Mass Analyzers

Mass analyzers separate the ions according to their mass-to-charge ratio. The following two laws govern the dynamics of charged particles in electric and magnetic fields in vacuum:

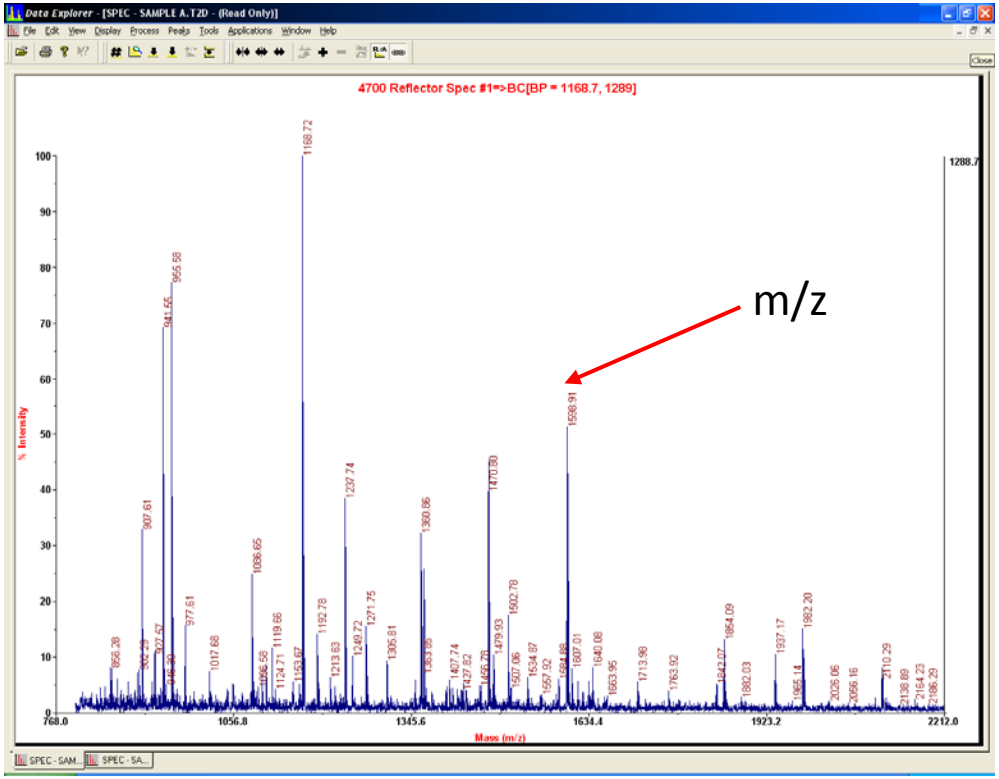
$$\mathbf{F} = Q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$ (Lorentz force law);

$$\mathbf{F} = m\mathbf{a}$$ (Newton's second law of motion in non-relativistic case, i.e. valid only at ion velocity much lower than the speed of light).

Here **F** is the force applied to the ion, *m* is the mass of the ion, **a** is the acceleration, *Q* is the ion charge, **E** is the electric field, and **v** x **B** is the vector cross product of the ion velocity and the magnetic field

Equating the above expressions for the force applied to the ion yields:

$$(m/Q)\mathbf{a} = \mathbf{E} + \mathbf{v} \times \mathbf{B}.$$

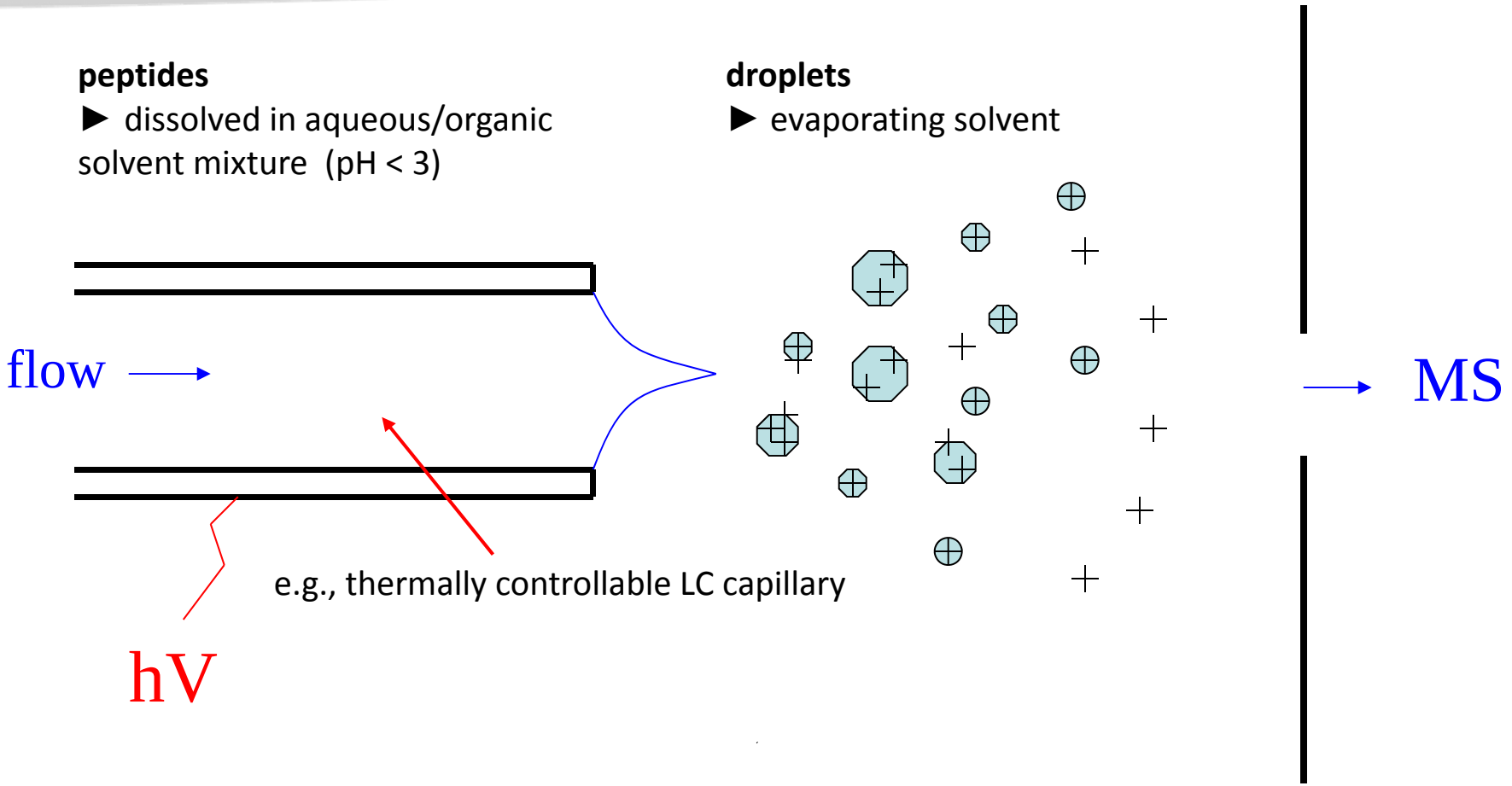


- *m/Q = m/z*

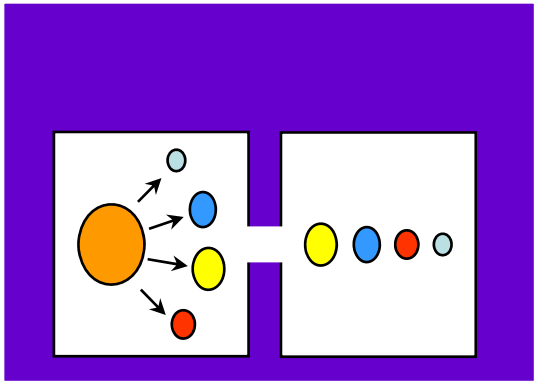
  m/z ratios are represented by the spectral peaks detected after ion passage through the mass analyzer

J. Craig Venter INSTITUTE

Electrospray Ionization



MS and Tandem Mass Spectrometry

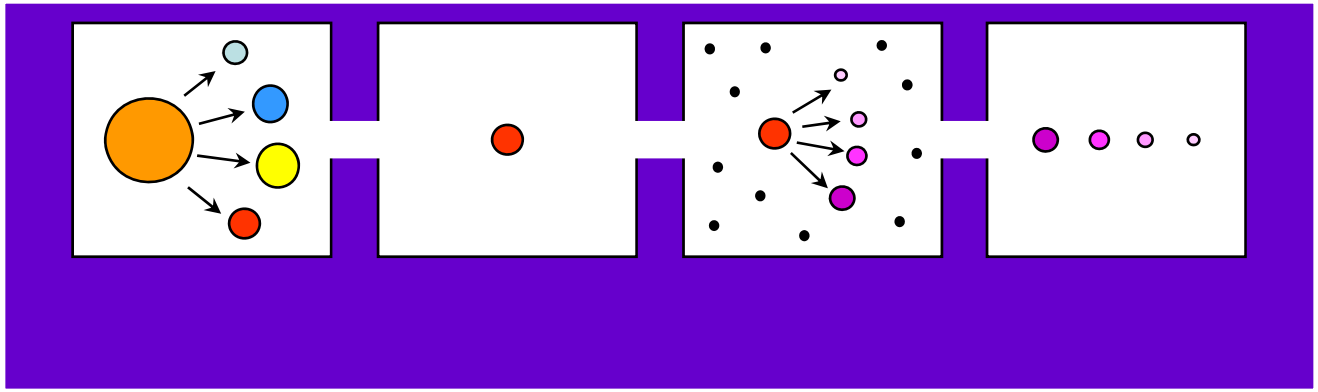


MS

Ionization Mass Analysis

Tandem MS

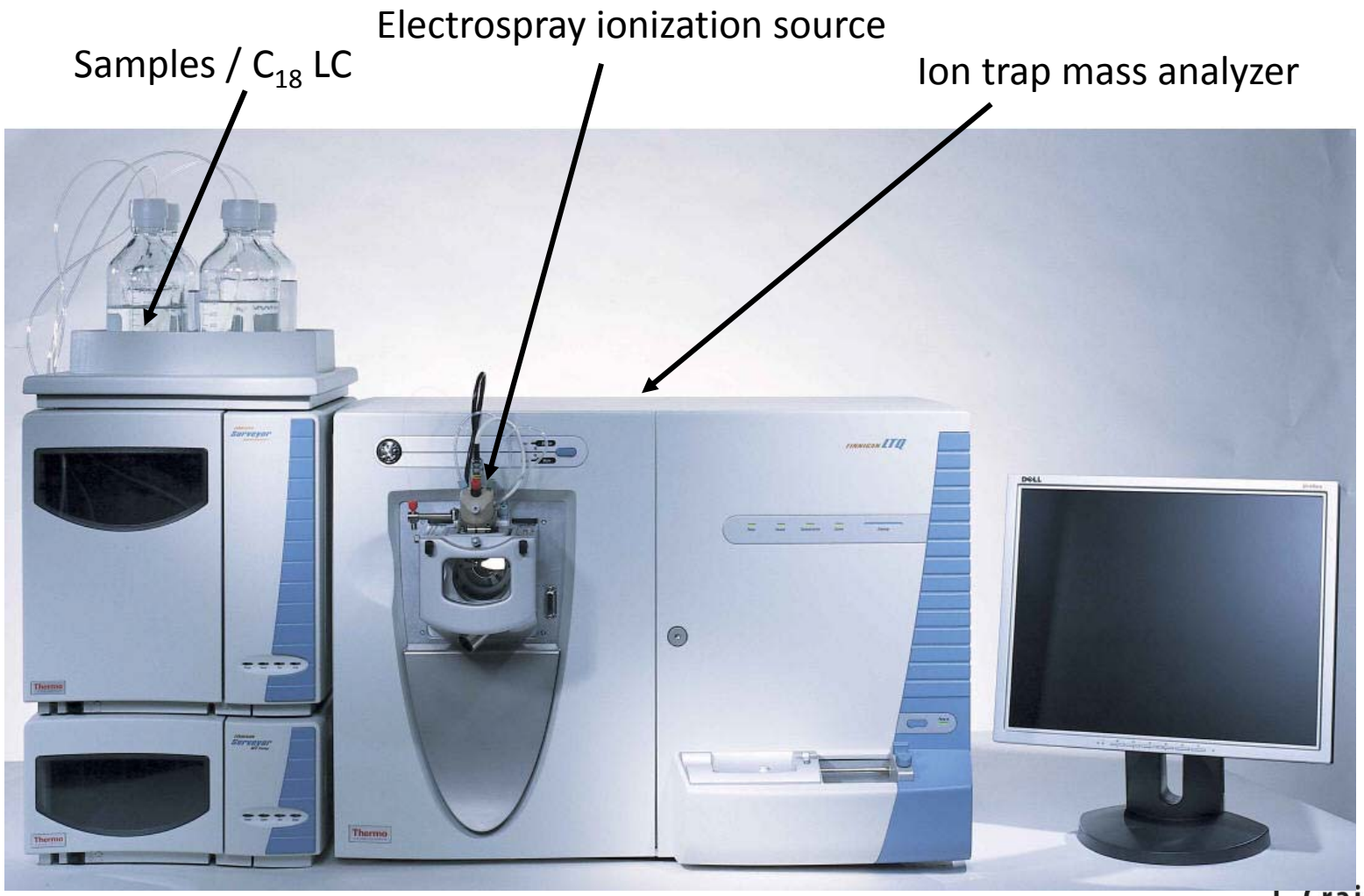
MS MS^2



Ionization Separation Activation Mass Analysis

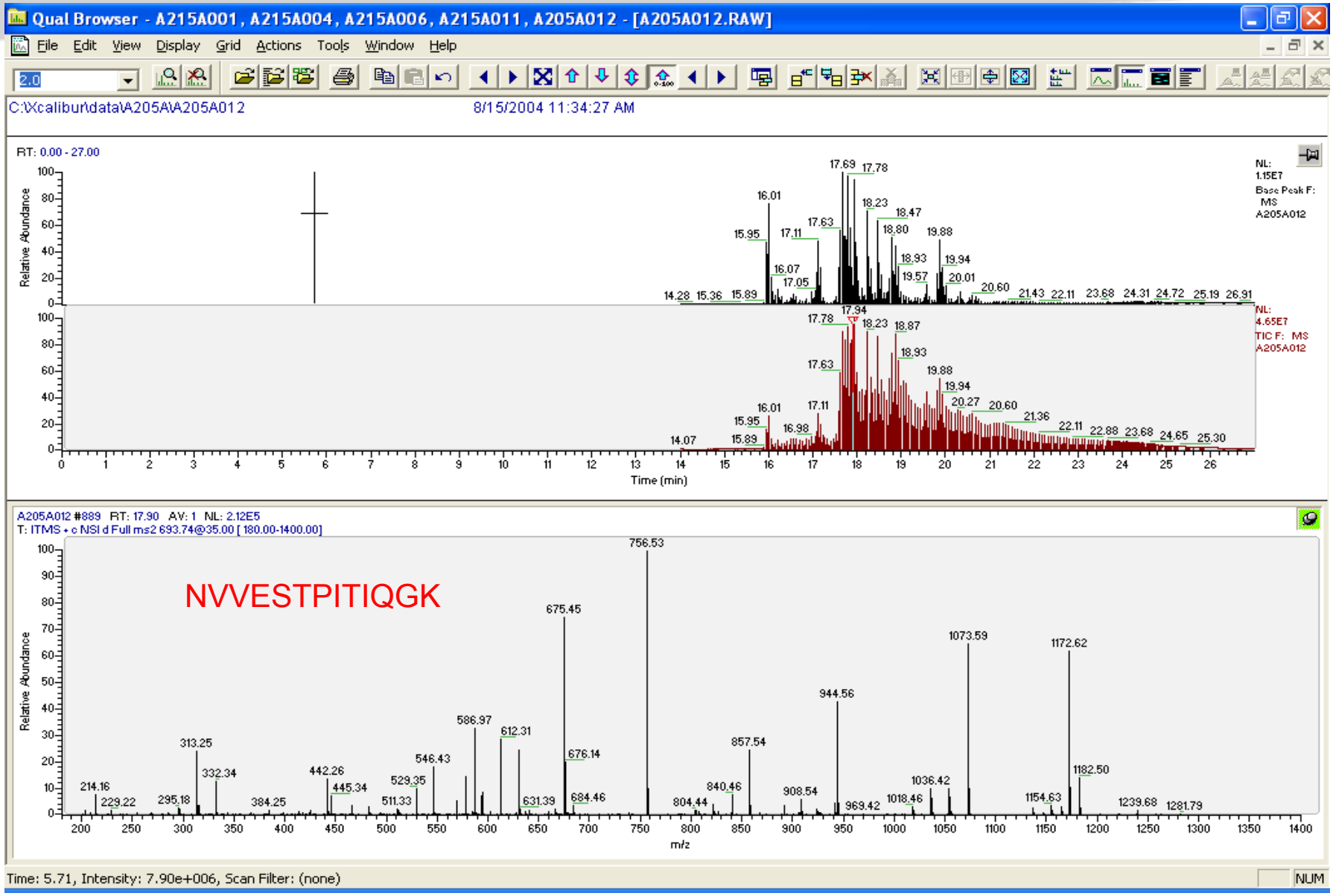
J. Craig Venter™ INSTITUTE

A quadrupole linear ion trap instrument

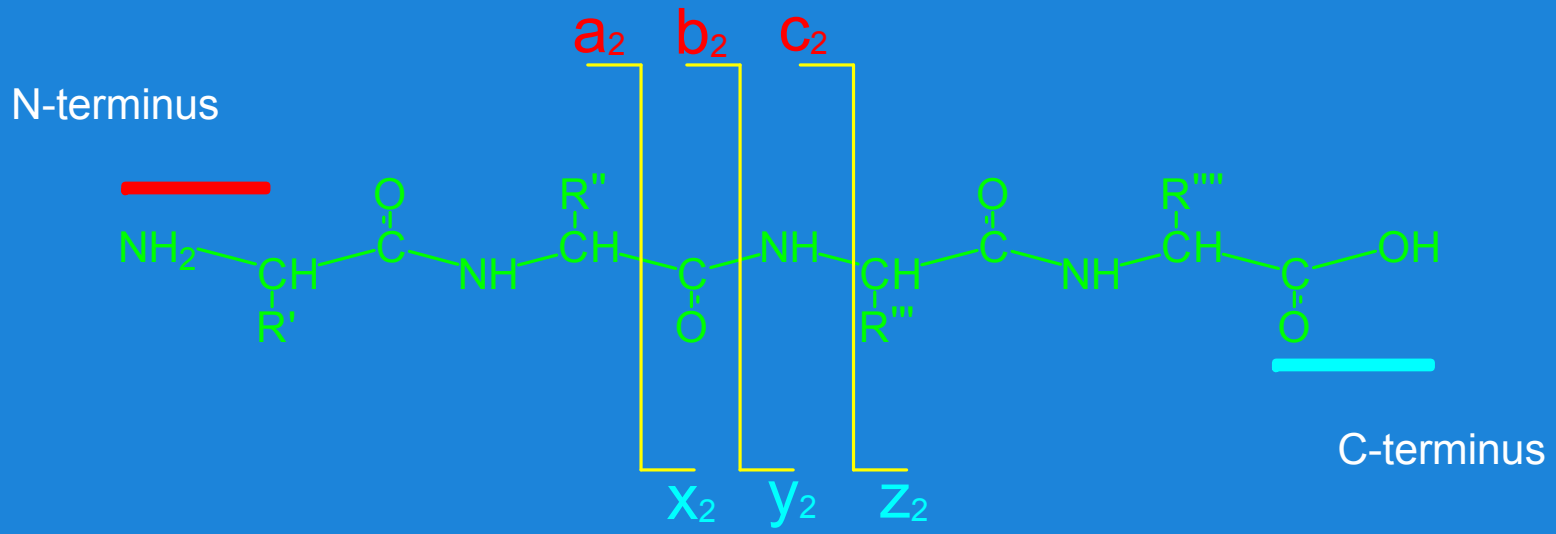


J. Craig Venter INSTITUTE

LC-nESI-MS/MS - Software

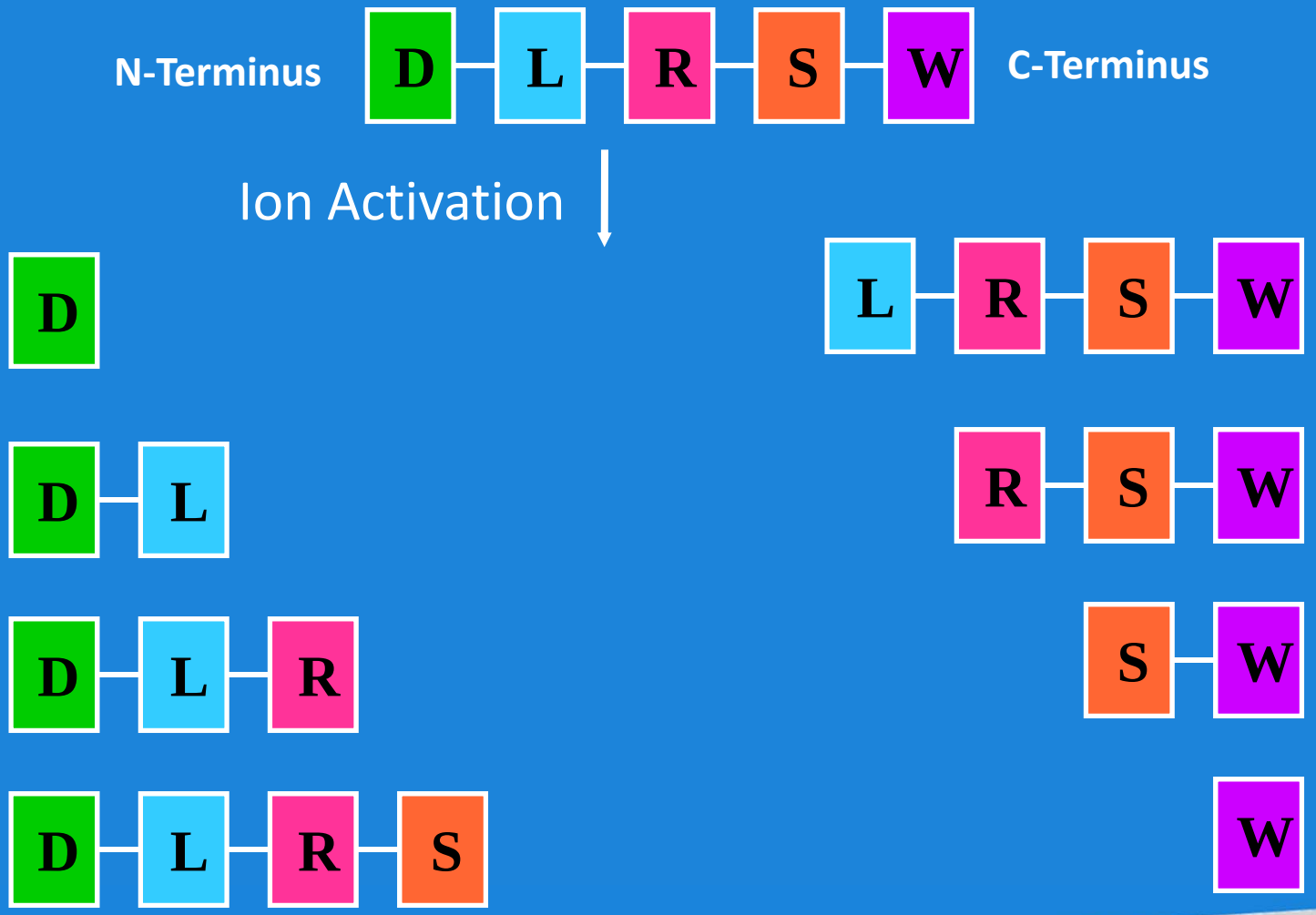


Peptide Fragmentation in the Collision Cell



J. Craig Venter™ INSTITUTE

Peptide Fragmentation in the Collision Cell



a, b, c-ions

x, y, z-ions

J. Craig Venter INSTITUTE

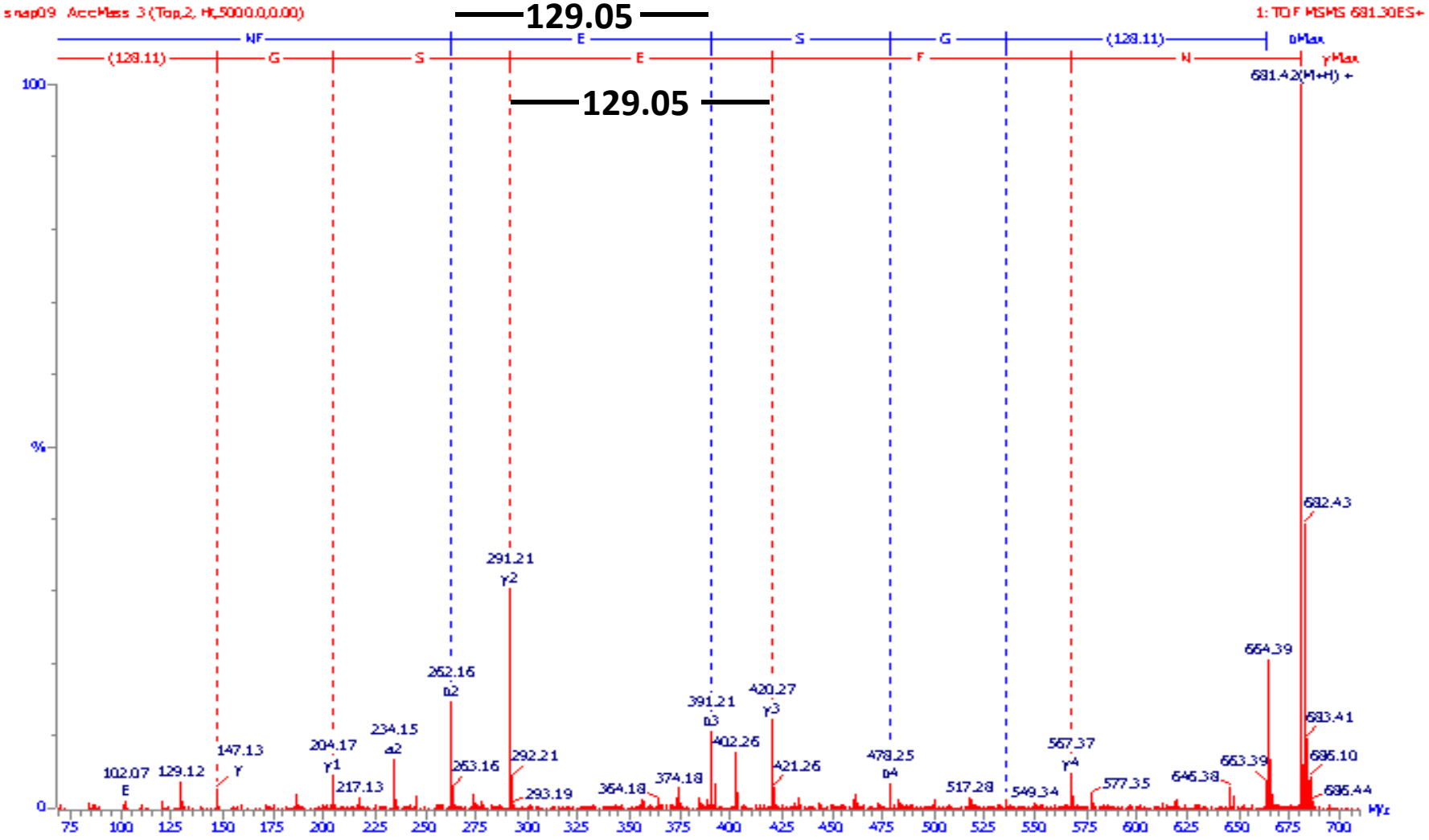
Table of Amino Acid Residue Masses

Symbol	Structure	Mass (Da)
Ala A	-NH.CH.(CH_3).CO-	71.0
Arg R	-NH.CH.[(CH_2)$_3$.NH.C(NH).NH_2].CO-	156.1
Asn N	-NH.CH.(CH_2CONH_2).CO-	114.0
Asp D	-NH.CH.(CH_2COOH).CO-	115.0
Cys C	-NH.CH.(CH_2SH).CO-	103.0
Gln Q	-NH.CH.($CH_2CH_2CONH_2$).CO-	**128.1**
Glu E	-NH.CH.(CH_2CH_2COOH).CO-	129.0
Gly G	-NH.CH_2.CO-	57.0
His H	-NH.CH.($CH_2C_3H_3N_2$).CO-	137.1
Ile I	-NH.CH.[CH.(CH_3)CH_2.CH_3].CO-	**113.1**
Leu	-NH.CH.[$CH_2CH(CH_3)_2$].CO-	**113.1**
Lys K	-NH.CH.[(CH_2)$_4NH_2$].CO-	**128.1**
Met M	-NH.CH.[(CH_2)$_2$.SCH_3].CO-	131.0
Phe F	-NH.CH.(CH_2Ph).CO-	147.1
Pro P	-NH.(CH_2)$_3$.CH.CO-	97.1
Ser S	-NH.CH.(CH_2OH).CO-	87.0
Thr T	-NH.CH.[CH(OH)CH_3].CO-	101.0
Trp W	-NH.CH.[CH_2.C_8H_6N].CO-	186.1
Tyr Y	-NH.CH.[(CH_2).C_6H_4.OH].CO-	163.1
Val V	-NH.CH.[$CH(CH_3)_2$].CO-	99.1

J. Craig Venter™
INSTITUTE

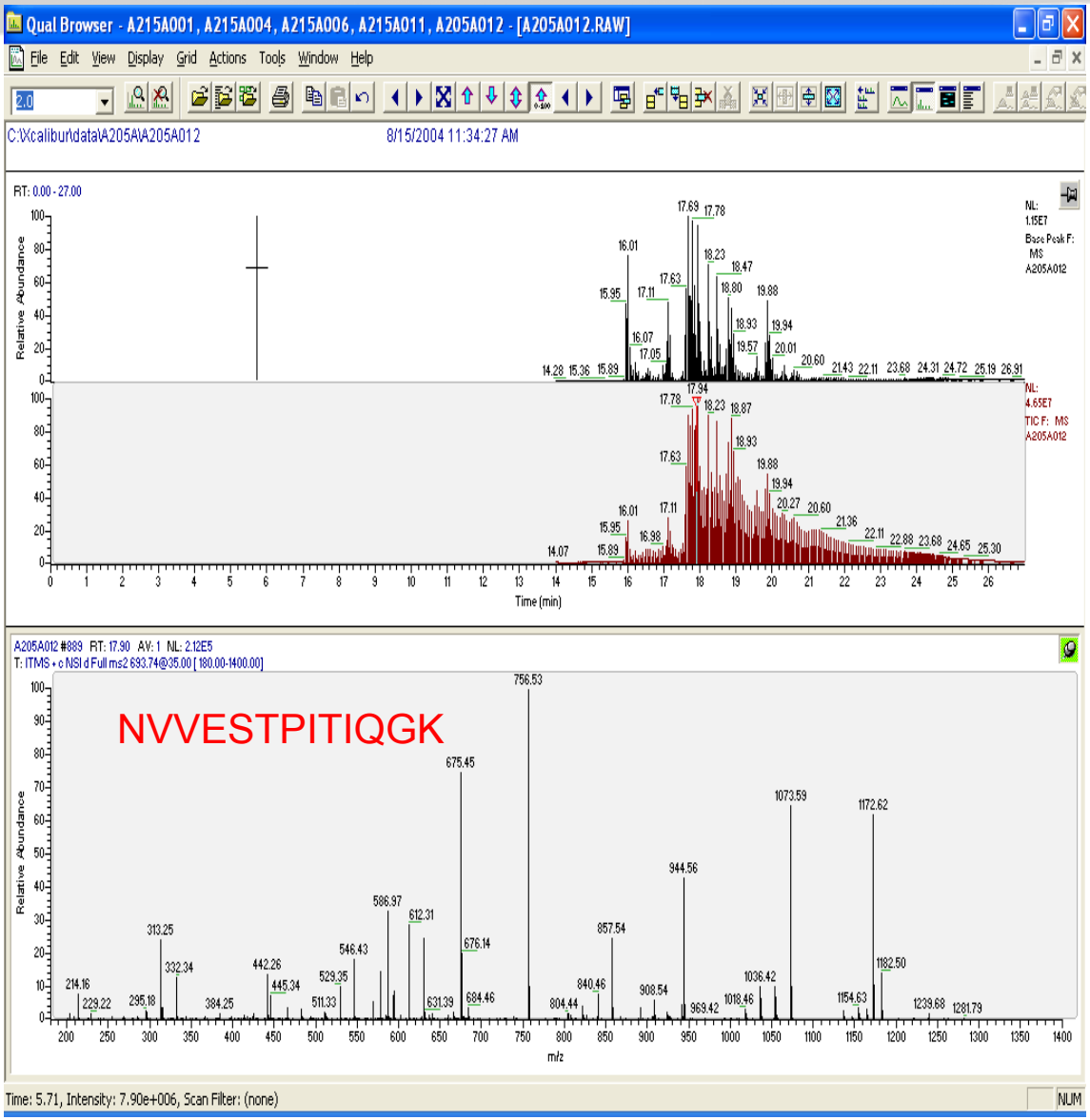
Peptide fragmentation generating b, a and y-ions

NFESGK: amino acid residue masses – 114, 147, 129, 87, 57, 128



Peptide sequencing by tandem mass spectrometry - an MS-MS daughter or product ion spectrum.

Data-Dependent MS^2 Acquisition



Tryptic peptide digest

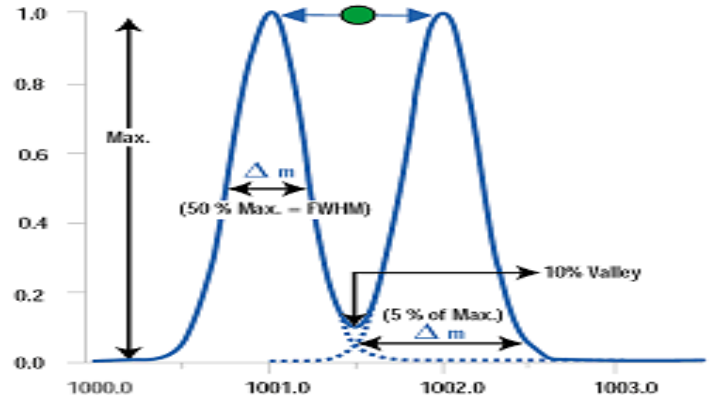
- LC separation of peptides (30 min)
- MS peak chromatogram: total ion counts
- Select 5 most abundant ions for MS^2
- Apply CID
- Acquire MS^2 spectra
- Repeat cycle of MS + 5 MS^2 hundreds of times
- Submit tandem MS data to database searches

J. Craig Venter™ INSTITUTE

Parameters for Mass Spectrometry Analysis

Resolution = R = $\frac{m}{\Delta m}$

i.e. R = $\frac{1000}{0.5}$ = 2000



Mass accuracy: difference between theoretical and experimental mass

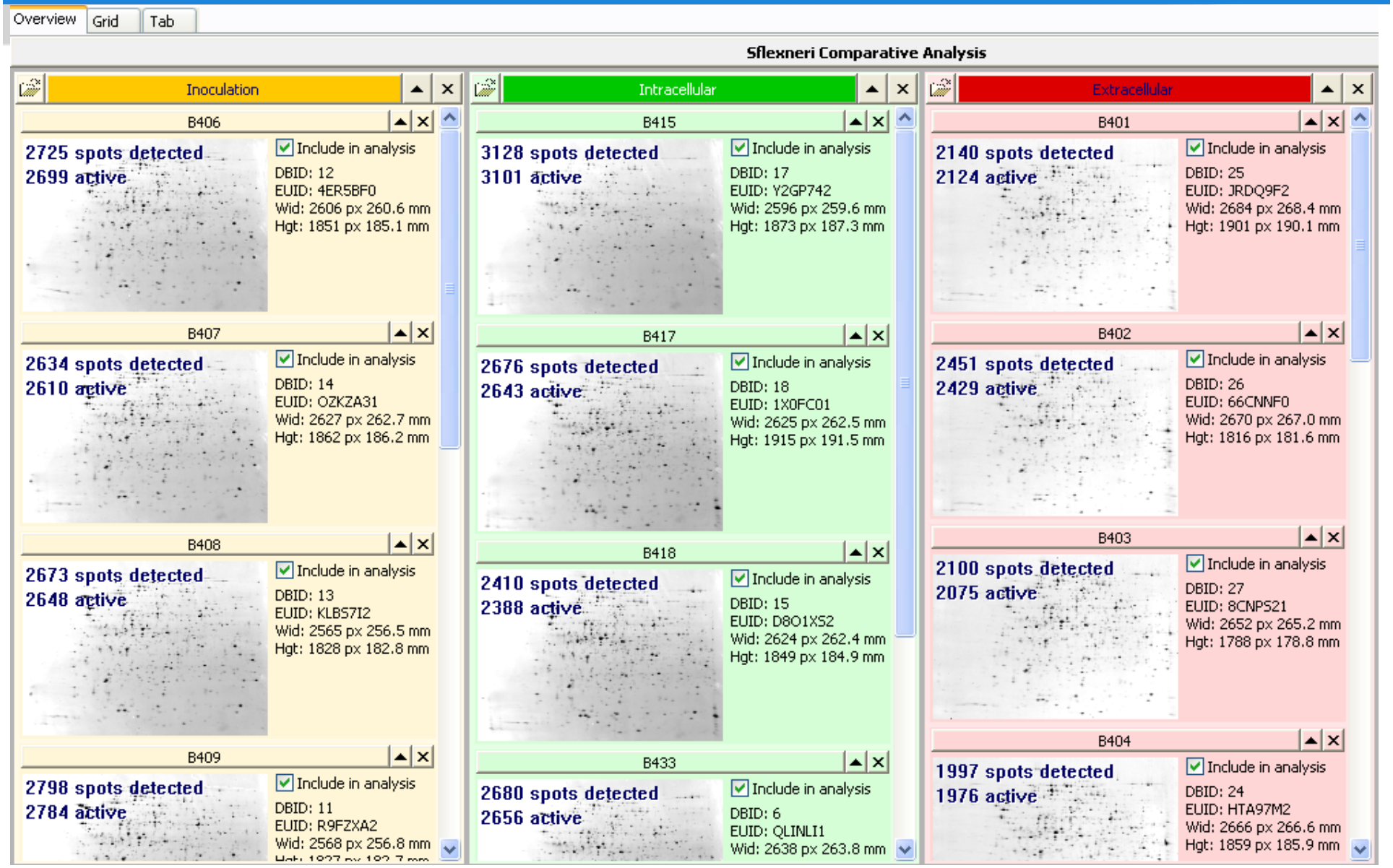
$$ppm = 10^6 * (m_{real} - m_{measured}) / m_{measured}$$

i.e.: theoretical mass: 1000, measured mass: 999.9 error: 100 ppm

Sensitivity = signal/noise = S/N

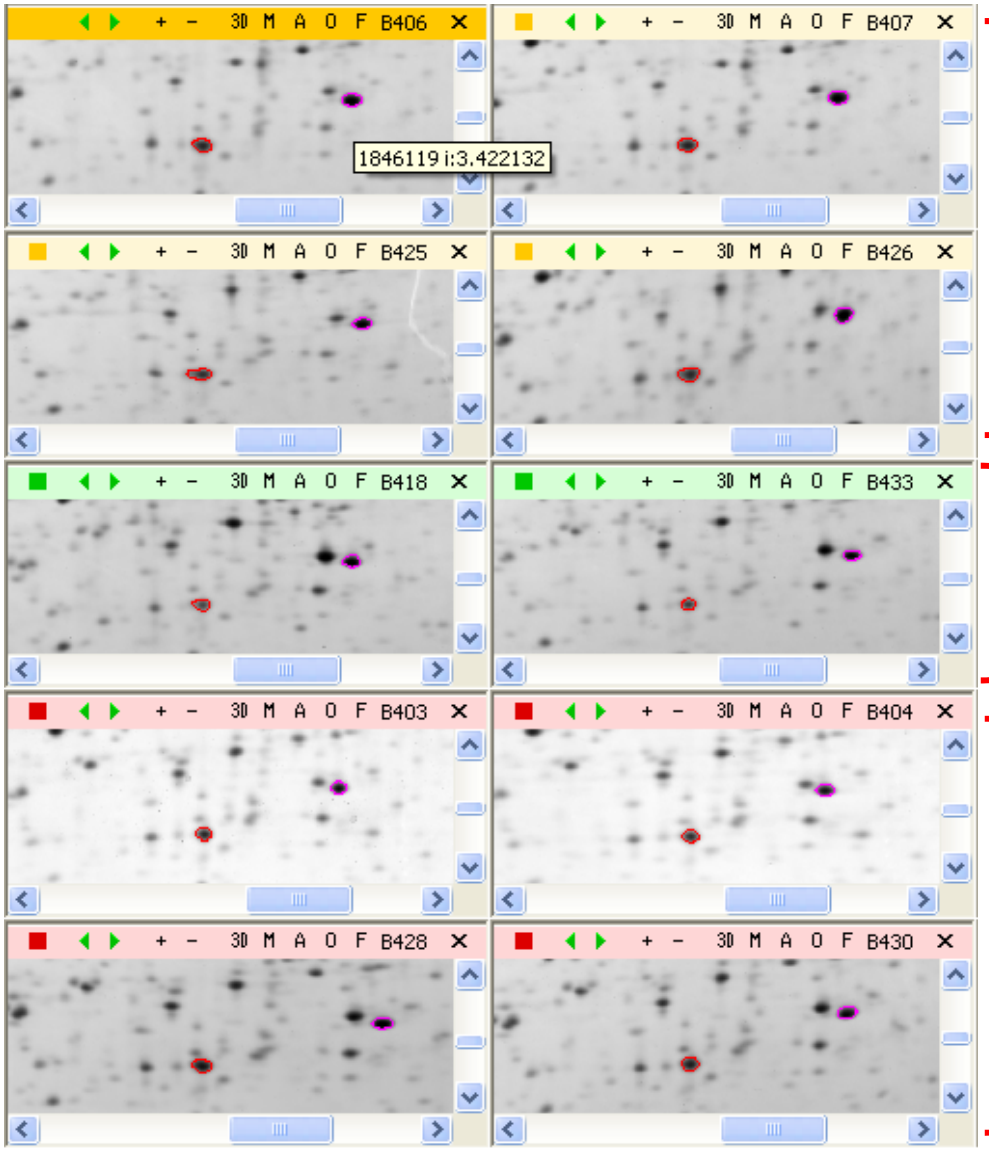
J. Craig Venter™ INSTITUTE

Differential Gel Display



Post-Spot Match Analysis of Data

- matched spots
- unmatched spots
- spot quantity averages



J. Craig Venter INSTITUTE

Tryptic Peptides: Most Common Analytes in Proteomic Research

{MATRIX SCIENCE} Mascot Search Results

Protein View

Match to: **gi|82778621** Score: **445**
GTP-binding protein chain elongation factor EF-G [Shigella dysenteriae Sd197]
Found in search of C:\Program Files\Matrix Science\Mascot Daemon\mgf\1470 Sd(

Nominal mass (M_r): **77670**; Calculated pI value: **5.24**
NCBI BLAST search of gi|82778621 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Shigella dysenteriae Sd197

Fixed modifications: Methylthio (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Sequence Coverage: **21%**

Matched peptides shown in **Bold Red**

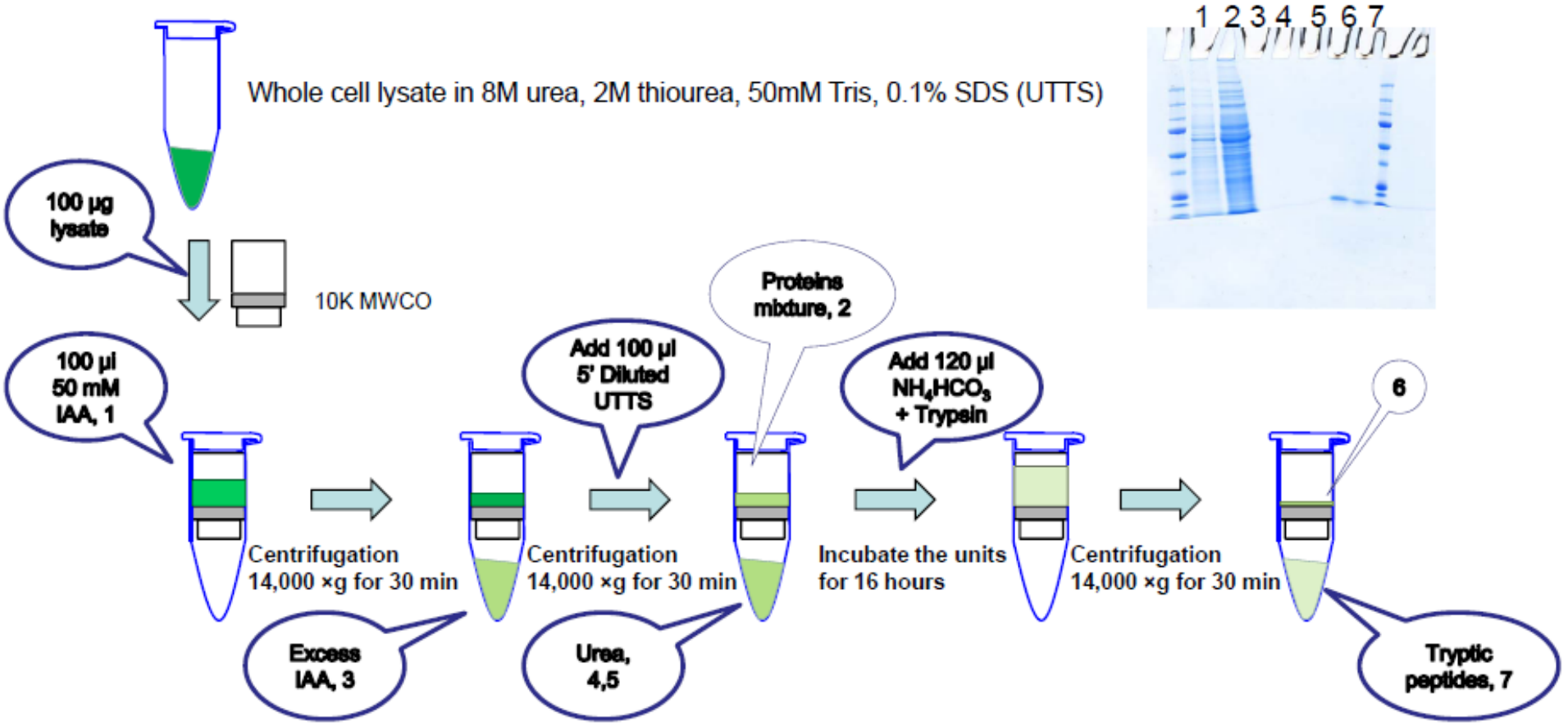
```
    1 MARTTPIARY RNIGISAHID AGKTTTTERI LFYTGVNHKI GEVHDGAATM
   51 DWMEQEQERG ITITSAATTA FWSGMAKQYE PHRINIIDTP GHVDFTIEVE
  101 RSMRVLDGAV MVYCAVGGVQ PQSETVWRQA NKYKVPRIAF VNKMDRMGAN
  151 FLKVVNQIKT RLGANPVPLQ LAIGAEEHFT GVVDLVKMKA INWNDADQGV
  201 TFEYEDIPAD MVELANEWHQ NLIESAAEAS EELMEKYLGG EELTEAEIKG
  251 ALRQRVLNNE IILVTCGSAF KNKGVQAMLD AVIDYLPSPV DVPAINGILD
  301 DGKDTPAERH ASDDEPFSAL AFKIATDPFV GNLTFFRVYS GVVNSGDTVL
  351 NSVKAARERF GRIVQMHANK REEIKEVRAG DIAAAIGLKD VTTGDTLCDP
  401 DAPIILERME FPEPVISIAV EPKTKADQEK MGLALGRLAK EDPSFRVWTD
  451 EESNQTIIAG MGELHLDIIV DRMKREFNVE ANVGKPQVAY RETIRQKVTD
  501 VEGKHAKQSG GRGQYGHVVI DMYPLEPGSN PKGYEFINDI KGGVIPGEYI
  551 PAVDKGIQEQ LKAGPLAGYP VVDMGIRLHF GSYHDVDSSE LAFKLAASIA
  601 FKEGFKKAKP VLLEPIMKVE VETPEENTGD VIGDLSRRRG MLKGQESEVT
  651 GVKIHAEVPL SEMFGYATQL RSLTKGRASY TMEFLKYDEA PSNVAQAVIE
```

Tryptic peptides identified for EF-G

1. R / YRNIGISAHIDK
2. R / ILFYTGVNHK
3. R / MGANFLKVVNQKITR
4. R / HASDDEPFALAFK
5. K / IATDPFVGNLTFER
6. K / EVRAGDIAAAIGLK
7. R / MEFPEPVISIVEPK
8. K / LAASIAFK
9. K / EGFKK
10. K / AKPVLLEPIMK
11. K / GQESEVTGVK
12. K / IHAEVPLSEMFGYATQLR

J. Craig Venter™ INSTITUTE

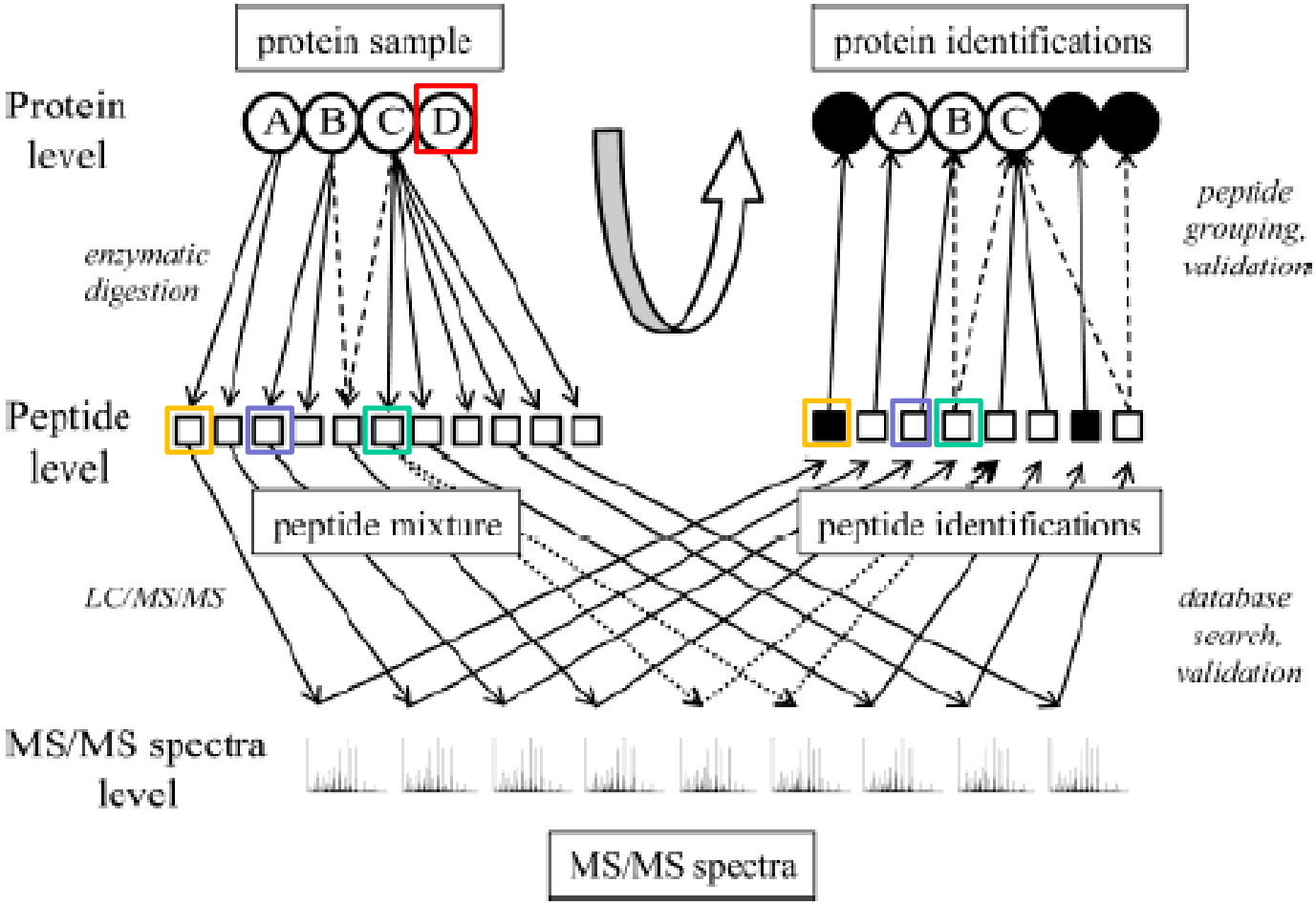
FASP Sample Processing



Mann et al. Nat Methods. 2009, 6, 359

J. Craig Venter™ INSTITUTE

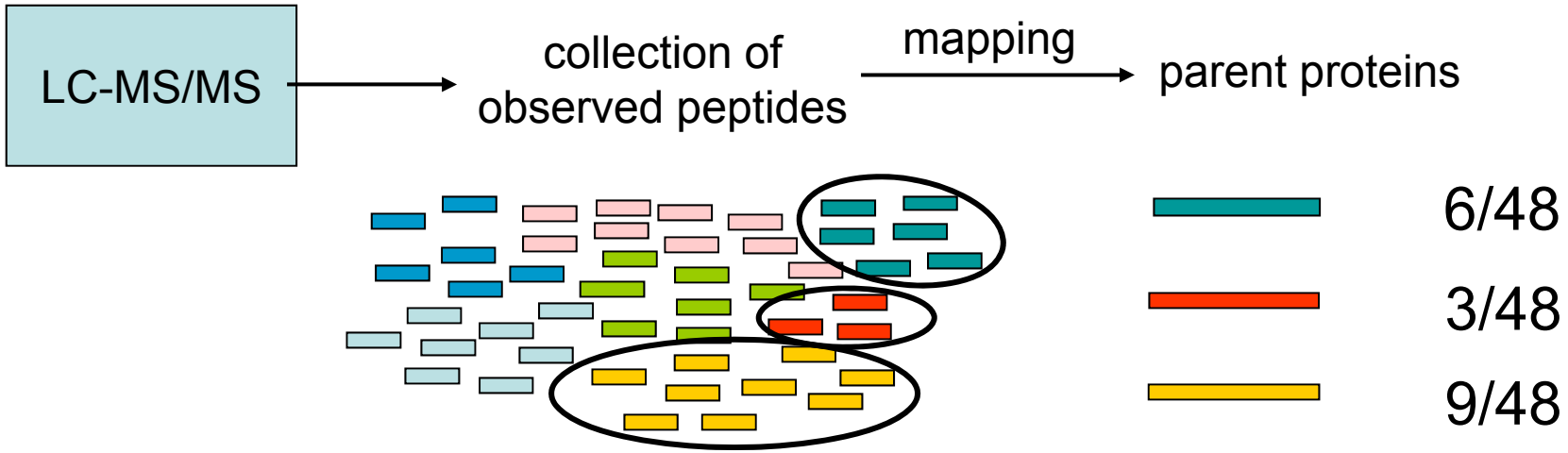
LC-MS/MS Principles



Nesvizhskii et al., Anal. Chemistry, 2003

J. Craig Venter INSTITUTE

Peptide Spectral Counting



- peptides map to parent proteins
- spectral counting attempts to infer protein abundance from the number of peptides observed for each protein
- APEX quantitation method corrects for variable MS peptide detection