# Next generation sequencing: applications for bacteria & viruses

Sarah Pendergrass PhD, MS
Center for System Genomics

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- One of the most direct uses of these new technologies
  - Discovery of new microorganisms and viruses
- Previous methods for identification include
  - Viral and pan-microbial oligonucleotide microarray analysis
  - Polymerase-chain Reactions (PCR) assays for known viruses
  - Viral/bacterial culture
  - Depending on method, can be time consuming with poor sensitivity
- NGS can be used when these results are inconclusive

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- Useful for metagenomics
  - Genetic material collected directly from the environment
  - Detection of unknown disease-associated viruses and discovery of novel human viruses
- Analysis of viral genome variability within the host
- Microbiome studies
- Investigation of microbial communities

Center for
Systems Genomics
Pennsylvania State University

# NGS: Applications

- Drug response variability in patients
- Characterization of contagious disease susceptibility across populations
  - Started with GWAS
  - Extended with NGS

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- Changes in
  - Human populations
  - Climate
  - Interfaces between animals, insects, and people
  - Human health status and ability to fight disease

- NGS useful for fast characterization and detection of
  - Existing and emerging zoonotic and arthropod transmitted pathogens
  - Emergence and spread of new illnesses

## Sequencing Platforms

Table 1 | Comparison of next-generation sequencing platforms

| Platform | Library/ template preparation | NGS chemistry | Read length (bases) | Run time (days) | Gb per run | Machine cost (US$) | Pros | Cons | Biological applications | Refs |
|---|---|---|---|---|---|---|---|---|---|---|
| Roche/454's GS FLX Titanium | Frag, MP/ emPCR | PS | 330* | 0.35 | 0.45 | 500,000 | Longer reads improve mapping in repetitive regions; fast run times | High reagent cost; high error rates in homo-polymer repeats | Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics | D. Muzny, pers. comm. |
| Illumina/ Solexa's $GA_{II}$ | Frag, MP/ solid-phase | RTs | 75 or 100 | 4‡, 9§ | 18‡, 35§ | 540,000 | Currently the most widely used platform in the field | Low multiplexing capability of samples | Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics | D. Muzny, pers. comm. |
| | | | | | | | to adapt alternative NGS chemistries | control reagents; shortest NGS read lengths | | |
| Helicos BioSciences HeliScope | Frag, MP/ single molecule | RTs | 32* | 8‡ | 37‡ | 999,000 | Non-bias representation of templates for genome and seq-based applications | High error rates compared with other reversible terminator chemistries | Seq-based methods | 91 |
| Pacific Biosciences (target release: 2010) | Frag only/ single molecule | Real-time | 964* | N/A | N/A | N/A | Has the greatest potential for reads exceeding 1 kb | Highest error rates compared with other NGS chemistries | Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks | S. Turner, pers. comm. |

Bacterial sequencing

454 FLX Roche

Long reads for improved mapping in repetitive regions

*De novo* assembly of sequences without a reference sequence is important!

Metzger, Nature Reviews Genetics 2010

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- Bacterial Basic Steps
  - De novo assembly
    - Mapping sequence reads into contiguous sequences (contigs)
    - Without a reference genome
    - Software approaches to do this (e.x. Velvet)
  - Now compare to a reference genome
    - Order contigs against reference
      - MAUVE
      - ACT: Artemis Comparison Tool

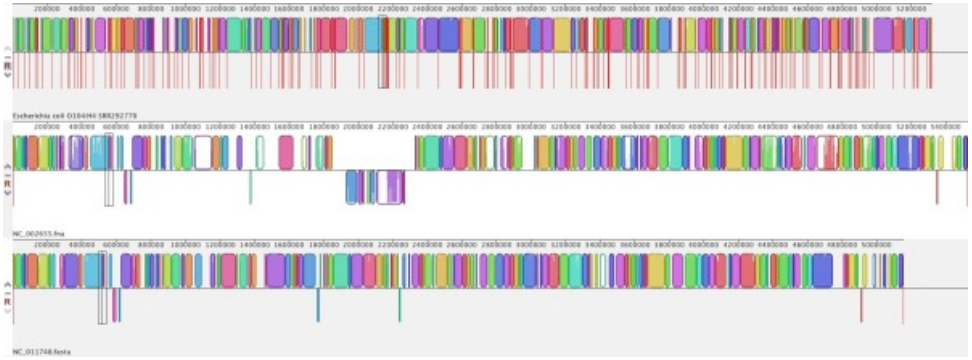
**Beginner's guide to comparative bacterial genome analysis using next-generation sequence data**

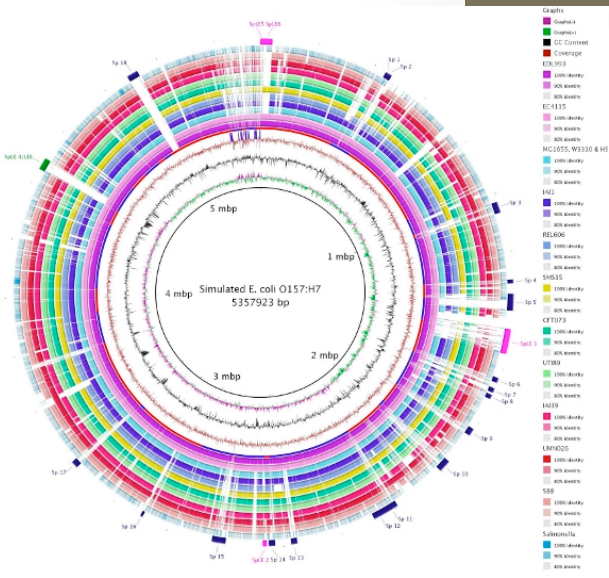
David J Edwards[1,2] and Kathryn E Holt[1]*

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- Bacterial Basic Steps
  - Annotate
    - Gene finding
    - RAST, Prokka, DIYA, RATT
      - RAST can assist with producing a high-quality annotation of the assembly
      - Genes can be viewed and compared to other genomes
  - Comparison to other genomes
    - Visualization of these comparisons
    - BRIG, Mauve, ACT





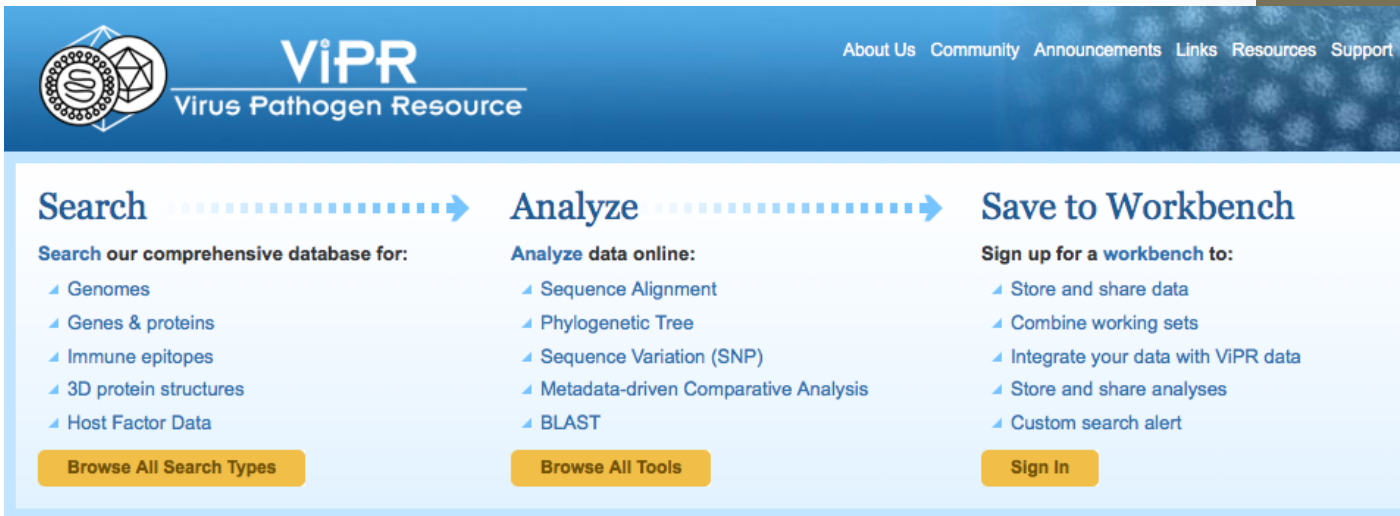

**Beginner's guide to comparative bacterial genome analysis using next-generation sequence data**

David J Edwards[1,2] and Kathryn E Holt[1]*

Center for Systems Genomics
Pennsylvania State University

# NGS: Applications

- Viral Basic Steps
  - Sample library prep
    - DNA or RNA isolated from infected host or purified viruses
  - Intial BLAST and assembly of reads
  - Isolation of potential virus sequences
    - Separate virus to non virus hits
  - Reassemble to longer contigs
  - BLAST to identify virus
  - Filling sequence gaps by PCR and Sanger sequencing




http://www.viprbrc.org/

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Example
- Identification of hemorrhagic fever-associated arenavirus from South Africa (Lujo virus)
  - Arenaviruses: RNA viruses in rodents
  - Most frequently transmitted through exposure to rodent urine

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Identification of hemorrhagic fever-associated arenavirus from South Africa (Lujo virus)
  - 5 cases of undiagnosed hemorrhagic fever, 4 fatal, after air transfer of a critically ill index case from Zambia to South Africa
    - Paramedic and other hospital staff

**Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever–Associated Arenavirus from Southern Africa**

**Thomas Briese[1☯]*, Janusz T. Paweska[2☯], Laura K. McMullan[3], Stephen K. Hutchison[4], Craig Street[1], Gustavo Palacios[1], Marina L. Khristova[5], Jacqueline Weyer[2], Robert Swanepoel[2], Michael Egholm[4], Stuart T. Nichol[3], W. Ian Lipkin[1]***

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Identification of hemorrhagic fever-associated arenavirus from South Africa (Lujo virus)
  - 454 pyrosequencing of RNA extracts and serum and tissues
  - Alignment of the sequences to the GenBank database using the Basic Local Alignment Search Tool (BLAST)
    - BLASTn for nucleotide sequences
    - BLASTx protein databases via nucleotide query

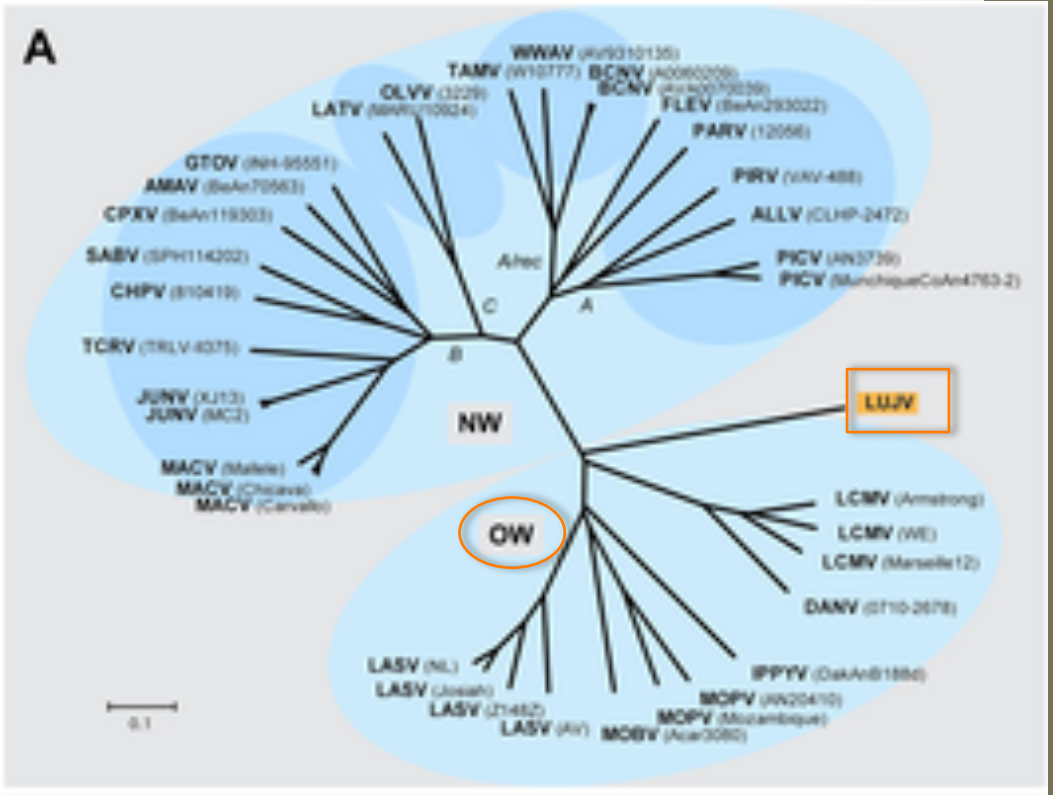
Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Identification of hemorrhagic fever-associated arenavirus from South Africa (Lujo virus)
  - Identification of the disease in 72 hours
  - First hemorrhagic fever–associated arenavirus from the Old World discovered in three decades

# Discovery of New Microorganisms and Viruses

- Lujo virus
  - Phylogenetic relationships between LUJV and reference sequences
  - Characterization showed virus to be a unique, genetically distinct, and highly pathogenic arenavirus



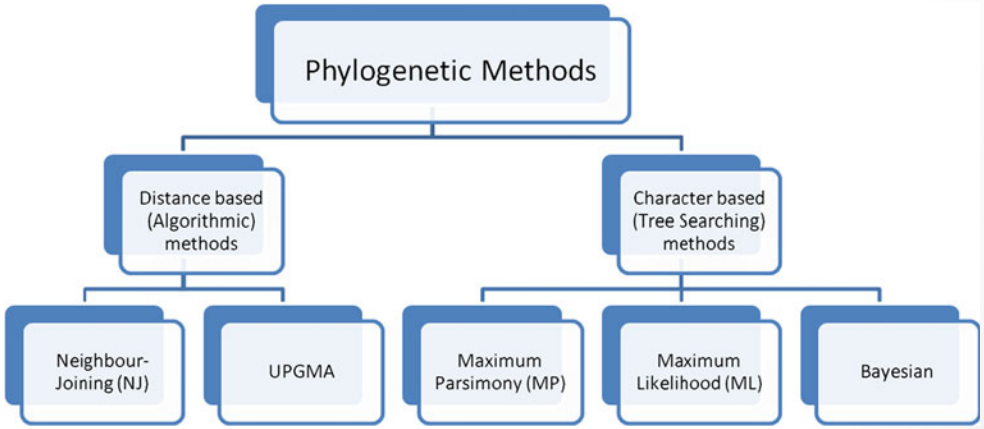

Briese et al. PLOS Pathogens, 2009

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- A note on phylogenetics
  - Alignment of nucleic acid, protein, or reference sequences between organisms
  - Phylogenetic tree—which is simply a graphic representation of the computed results
  - Accumulation of molecular sequence data, arising from concerted large-scale genomic and metagenomic sequencing projects, provides opportunities for uncovering evolutionary relationships

**Fig. 1** Tree-building methods. Schematic overview of the major analytical approaches to Phylogenetic tree building (modified from [1])




**A Beginner's Guide to Phylogenetics**

**Roy D. Sleator**

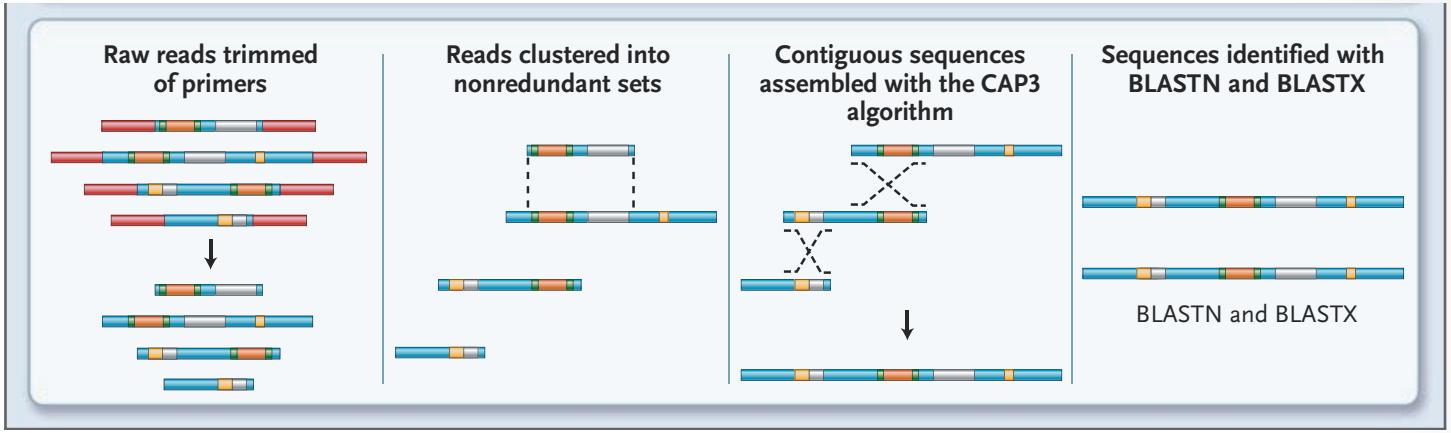
Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Example
- Three patients who died of a febrile illness
  - Transplantation of solid organs from a single donor
    - Three died within 6 weeks of transplantation
    - Donor died of cerebral hemorrhage 10 days after returning to Australia after a trip to rural Yugoslavia
  - Conventional molecular tests had not been informative
    - Bacterial and viral cultures
    - PCR assays for a number of viruses
    - Viral and pan-microbial microarray analysis

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

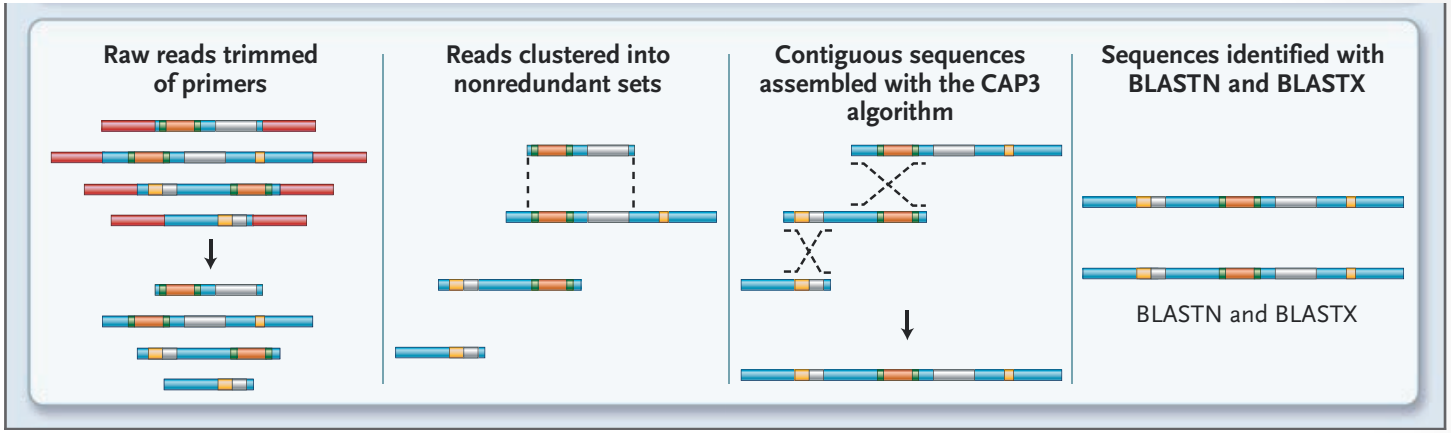



- Three patients who died of a febrile illness
  - RNA purified from brain, cerebrospinal fluid, serum, kidney, and liver from transplant recipients
    - Digestion with DNase enzyme, eliminating human DNA
    - RNA was reverse transcribed and amplified with random primers
    - Sequenced with 454 pyrosequencing technology

Palacios et al. NEJM, 2008

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

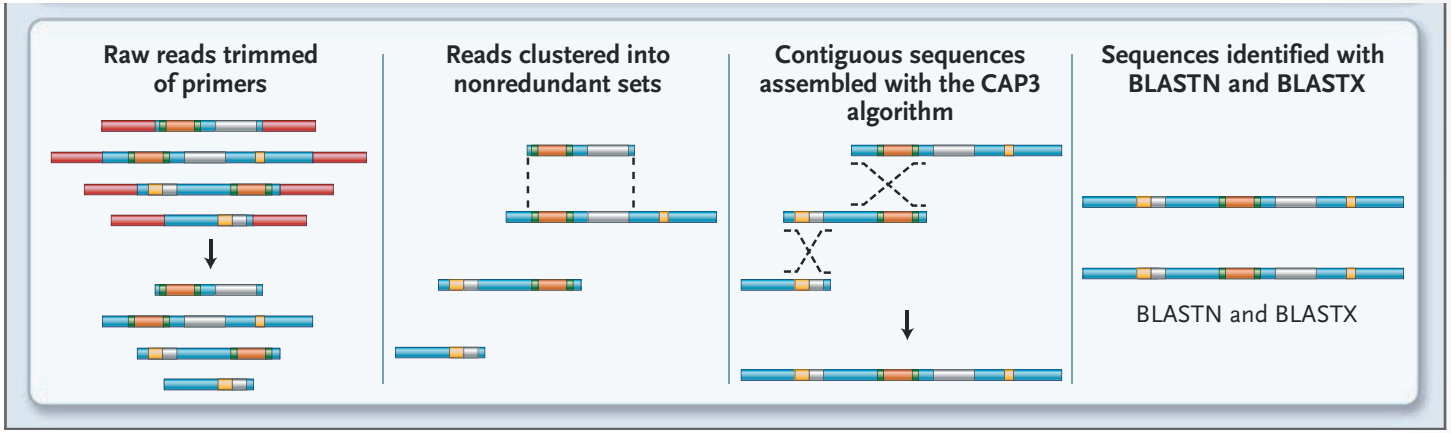



- Three patients who died of a febrile illness
  - Subtraction of vertebrate and highly repetitive sequences
  - Clustered non-redundant sequences (Cd-hit) and assembled into contiguous sequences

Palacios et al. NEJM, 2008

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses



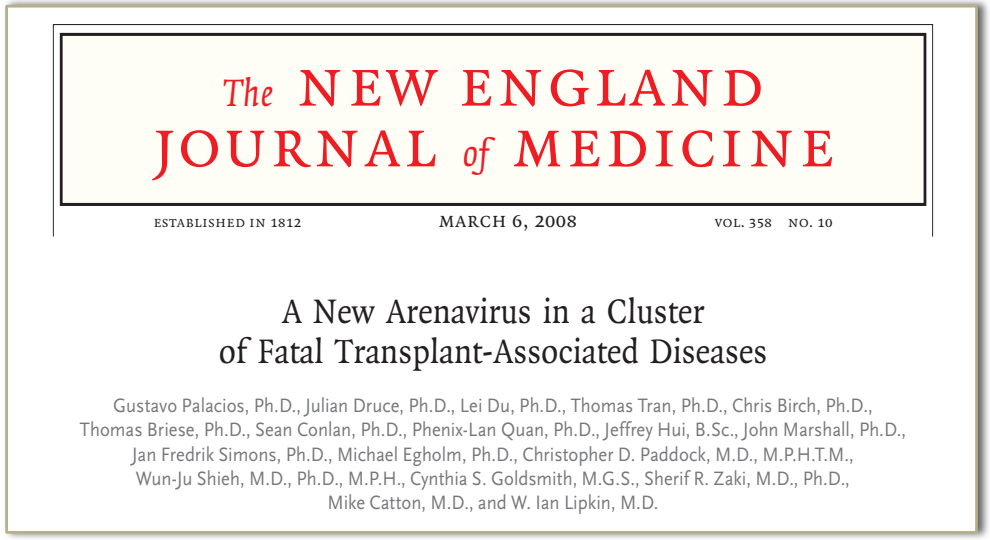

- Three patients who died of a febrile illness
  - Sequences assembled and compared with motifs represented in databases of microbes
  - Sequences consistent with an Old World arenavirus
  - A new arenavirus related to lymphocytic choriomeningitis viruses

Palacios et al. NEJM, 2008

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Arenaviruses
  - Infection with lymphocytic choriomeningitis virus (LCMV) is usually mild
  - Human to human transmission of LCMV has been known for recipients of solid-organ transplants with fatal consequences



The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MARCH 6, 2008 VOL. 358 NO. 10

A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases

Gustavo Palacios, Ph.D., Julian Druce, Ph.D., Lei Du, Ph.D., Thomas Tran, Ph.D., Chris Birch, Ph.D., Thomas Briese, Ph.D., Sean Conlan, Ph.D., Phenix-Lan Quan, Ph.D., Jeffrey Hui, B.Sc., John Marshall, Ph.D., Jan Fredrik Simons, Ph.D., Michael Egholm, Ph.D., Christopher D. Paddock, M.D., M.P.H.T.M., Wun-Ju Shieh, M.D., Ph.D., M.P.H., Cynthia S. Goldsmith, M.G.S., Sherif R. Zaki, M.D., Ph.D., Mike Catton, M.D., and W. Ian Lipkin, M.D.

Palacios et al. NEJM, 2008

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

**Novel, Divergent Simian Hemorrhagic Fever Viruses in a Wild Ugandan Red Colobus Monkey Discovered Using Direct Pyrosequencing**

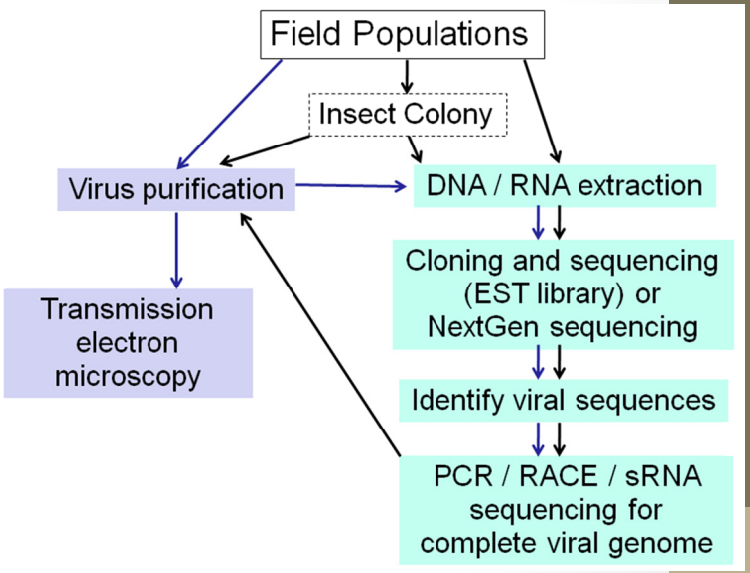
**Michael Lauck[1], David Hyeroba[2,3], Alex Tumukunde[3], Geoffrey Weny[3], Simon M. Lank[4], Colin A. Chapman[2,5,6], David H. O'Connor[1,4], Thomas C. Friedrich[4,7], Tony L. Goldberg[3,4,6,7]***

- Kibale EcoHealth Project
  - Tropical forest changes and how they alter health-related outcomes for people, domestic animals, and wildlife
  - Interface between environmental change, animals, and humans
- Kibale National Park in western Uganda
  - Zoonotic disease emergence
    - Fast rate of population growth
    - High human disease burden (e.g. malaria, AIDS)
    - Exceptional diversity of animal disease reservoirs.
- The project focuses on wild non-human primates
  - Known to exchange pathogens with local people and domestic animals.

http://svmweb.vetmed.wisc.edu/KibaleEcoHealth/research.html

# Discovery of New Microorganisms and Viruses

- Insects are often affected with multiple viruses
- Traditionally, viruses isolated from insects displaying abnormal phenotypes due to infection
- NGS is allowing for the identification of new viruses from insects
- Viral sequences can now be derived from whole insects or specific tissues




Liu et al., 2011

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Arthropod-borne viruses explored in Aedes aegypti mosquitos
  - Experimentally infected with dengue virus
  - Pooled with noninfected mosquitos
  - RNA purified from mosquito pools subjected to 454 sequencing
  - Led to correct identification of infected mosquito pools

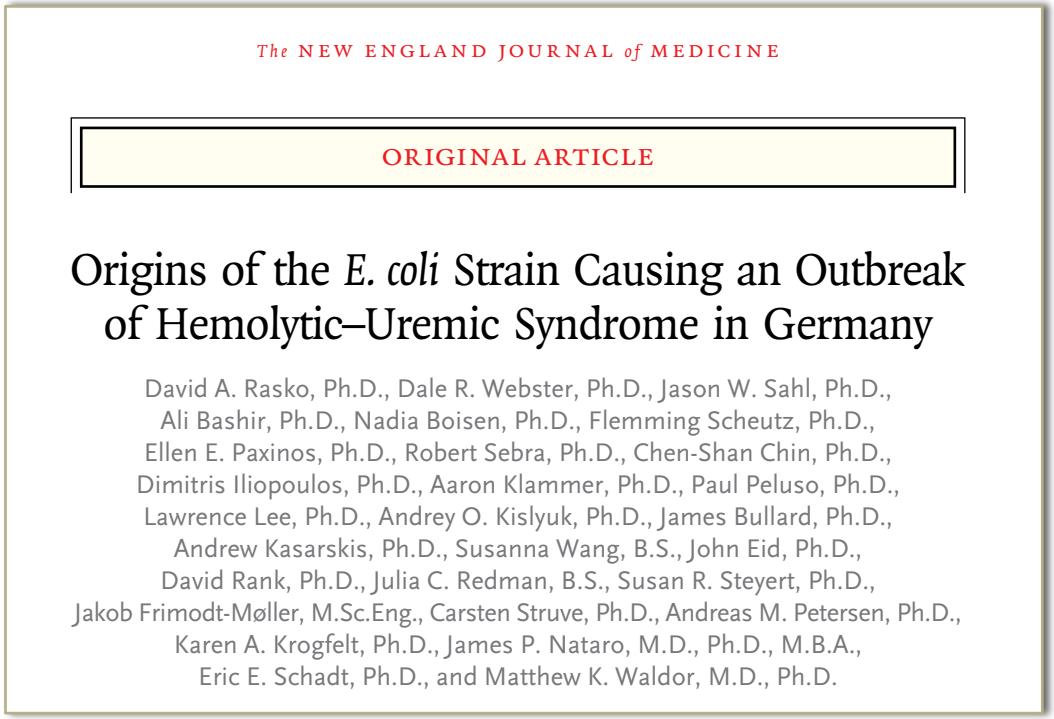
**Arbovirus Detection in Insect Vectors by Rapid, High-Throughput Pyrosequencing**

**Kimberly A. Bishop-Lilly[1]*, Michael J. Turell[2], Kristin M. Willner[1], Amy Butani[1], Nichole M. E. Nolan[1], Shannon M. Lentz[1], Arya Akmal[1], Al Mateczun[1], Trupti N. Brahmbhatt[1], Shanmuga Sozhamannan[1], Chris A. Whitehouse[2], Timothy D. Read[1¤]**

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- "Next-next generation" sequencing of microorganisms
  - *E. Coli* outbreak in Germany in May 2011
    - Series of deaths as a result



The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

## Origins of the *E. coli* Strain Causing an Outbreak of Hemolytic–Uremic Syndrome in Germany

David A. Rasko, Ph.D., Dale R. Webster, Ph.D., Jason W. Sahl, Ph.D., Ali Bashir, Ph.D., Nadia Boisen, Ph.D., Flemming Scheutz, Ph.D., Ellen E. Paxinos, Ph.D., Robert Sebra, Ph.D., Chen-Shan Chin, Ph.D., Dimitris Iliopoulos, Ph.D., Aaron Klammer, Ph.D., Paul Peluso, Ph.D., Lawrence Lee, Ph.D., Andrey O. Kislyuk, Ph.D., James Bullard, Ph.D., Andrew Kasarskis, Ph.D., Susanna Wang, B.S., John Eid, Ph.D., David Rank, Ph.D., Julia C. Redman, B.S., Susan R. Steyert, Ph.D., Jakob Frimodt-Møller, M.Sc.Eng., Carsten Struve, Ph.D., Andreas M. Petersen, Ph.D., Karen A. Krogfelt, Ph.D., James P. Nataro, M.D., Ph.D., M.B.A., Eric E. Schadt, Ph.D., and Matthew K. Waldor, M.D., Ph.D.

# Discovery of New Microorganisms and Viruses

- Isolated and sequenced E. Coli from stool samples
- Real time DNA sequencing to determine the complete genome sequence of the strain, as well as seven other E. Coli strains from elsewhere – 5 hours per isolate
- Compared the sequences with other previously sequenced E. Coli isolates

**Table 1. Isolates of Enteroaggregative *Escherichia coli* Sequenced or Analyzed in This Study.***

| | Isolate† | Serotype | Location of Isolate | Source of Isolate |
|---|---|---|---|---|
| Representatives of best studied strains of E.coli | O42 | O44:H18 | Peru | Stool sample from child with diarrhea |
| | 17-2 | O3:H2 | Chile | Stool sample from child with diarrhea |
| | JM221 | O92:H33 | Mexico | Stool sample from adult with diarrhea |
| | C1010-00 | Orough:H- | Denmark | Stool sample from child with diarrhea |
| | 55989 | O104:H4 | Central African Republic | Stool sample from adult with diarrhea |
| Six enteroaggregative E.Coli clinical isolates | C35-10‡ | O104:H4 | Africa | Stool sample from child with diarrhea |
| | C682-09‡ | O104:H4 | Africa | Stool sample from child with diarrhea |
| | C734-09‡ | O104:H4 | Africa | Stool sample from child without diarrhea |
| | C754-09‡ | O104:H4 | Africa | Stool sample from child without diarrhea |
| | C760-09‡ | O104:H4 | Africa | Stool sample from child without diarrhea |
| | C777-09‡ | O104:H4 | Africa | Stool sample from child with diarrhea |
| Germany Samples | **TY2482** | O104:H4 | Germany | Stool sample from adult with diarrhea |
| | **LB226692** | O104:H4 | Germany | Diarrhea in adult |
| | **H112180280** | O104:H4 | Germany | Diarrhea in adult |
| | **C227-11** | O104:H4 | Denmark | Diarrhea in adult |

Rasko et al., NEJM 2011

Center for Systems Genomics
Pennsylvania State University

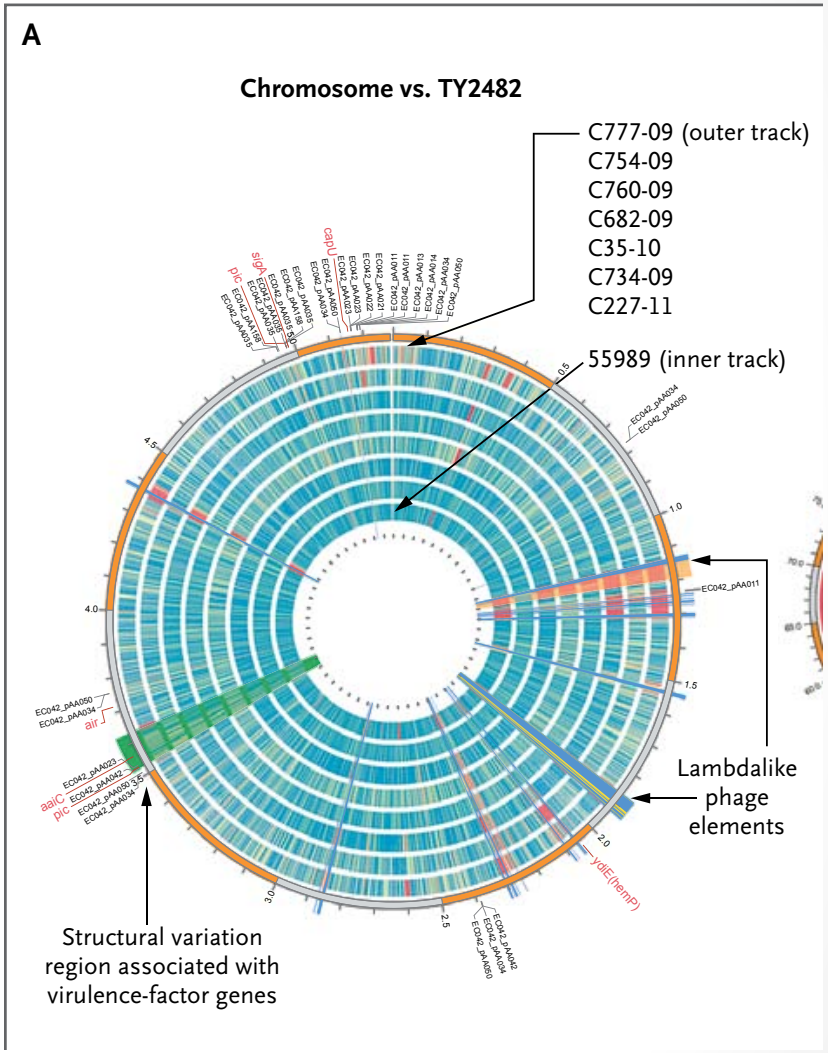
# Discovery of New Microorganisms and Viruses

- Isolated and sequenced E. Coli from stool samples
- Diarrhea associated with the hemolytic-uremic syndrome is usually caused by a certain kind of *E. Coli*
- This strain
  - Did not have the pathogenicity island of enterohemorrhagic *E. Coli,* which facilitates colonization of the large intestine
  - Similar to an *E. Coli* isolate from a patient in the Central African Republic, that did not produce Shiga-toxin

Rasko et al., NEJM 2011

# Discovery of New Microorganisms and Viruses

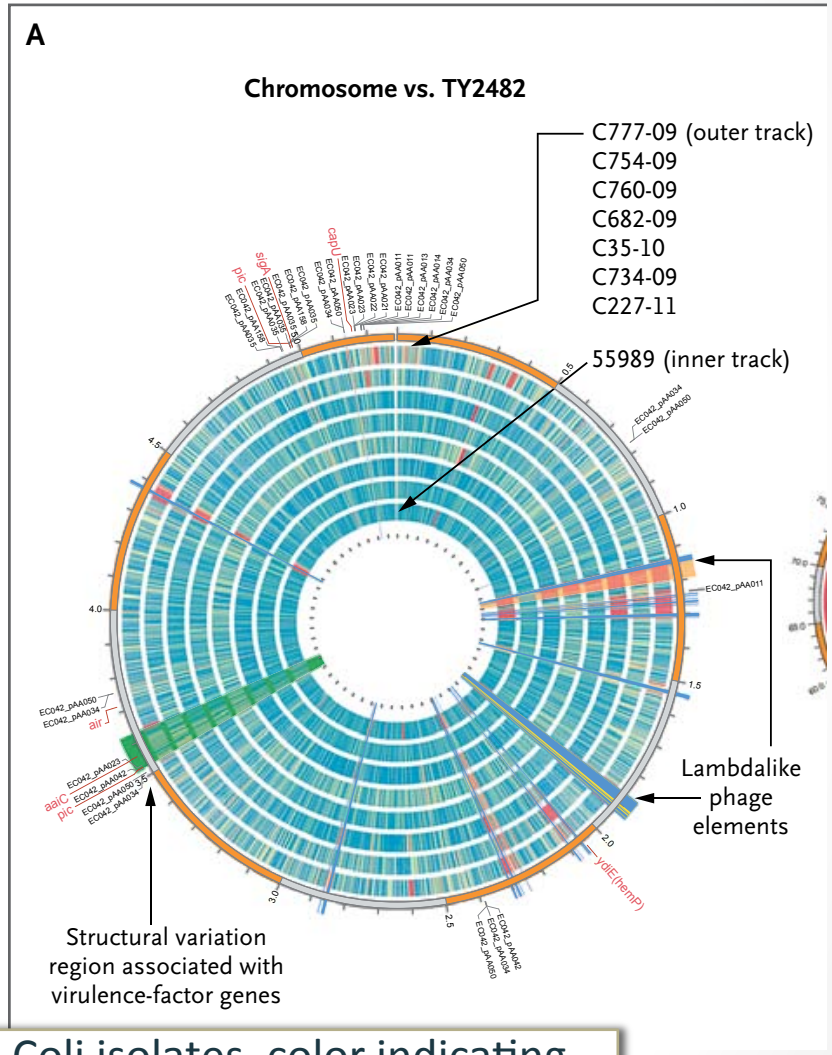
- Eight circular bands represent tracks of the isolates sequence mapped against the reference genome
- Annotations indicate positions of known virulence factor genes




Rasko et al., NEJM 2011

# Discovery of New Microorganisms and Viruses

- Genetic exchange allowed for the emergence of the highly virulent Shiga-toxin producing E. Coli strain that caused the German outbreak
- An enteroaggregative *E. coli* strain acquired a Shiga-toxin–encoding phage




Eight Bands, Representing the different E.Coli isolates, color indicating coverage of the reference strain

Rasko et al., NEJM 2011

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Clinical syndromes suspected to be of viral etiology, and isolation of the causative agent
  - Discovery of a new polyomavirus associated with most cases of Merkel cell carcinoma
    - Rare and aggressive skin cancer
      - Typically affects elderly and immunocompromised individuals
      - Suggestive of an infectious origin

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Discovery of a new polyomavirus associated with most cases of Merkel cell carcinoma (MCC)
  - RNA purified from MCC samples and analyzed with 454 pyrosequencing
  - Digital transcriptome subtraction of all human sequences
  - Identified sequence used as a starting point for whole genome sequencing of this polyomavirus

**Clonal Integration of a Polyomavirus in Human Merkel Cell Carcinoma**

Huichen Feng, Masahiro Shuda, Yuan Chang,* Patrick S. Moore*

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Further novel human polyomavirus identification
  - Isolation of circular viral DNA from skin swabs
  - Further exploration of Merkel cell polyomavirus and other polyomaviruses that inhabit healthy human skin
  - Discovery of two previously unknown polyomavirus species
    - Human polyomavirus-6 and 7
  - Polyomavirus DNA shed from the skin in the form of assembled virons

**Merkel Cell Polyomavirus and Two Previously Unknown Polyomaviruses Are Chronically Shed from Human Skin**

Rachel M. Schowalter,[1] Diana V. Pastrana,[1] Katherine A. Pumphrey,[1] Adam L. Moyer,[1] and Christopher B. Buck[1,*]

Center for Systems Genomics
Pennsylvania State University

# Discovery of New Microorganisms and Viruses

- Characterization of Ebola
  - 99 Ebola virus genomes collected from 78 Ebola patients
  - 300 genetic changes making these distinct from previous outbreaks

HARVARD gazette

SCIENCE & HEALTH > HEALTH & MEDICINE

## Ebola genomes sequenced

Speedy analysis reveals mutations, insights into outbreak, along with clues to origin, spread

August 28, 2014 |

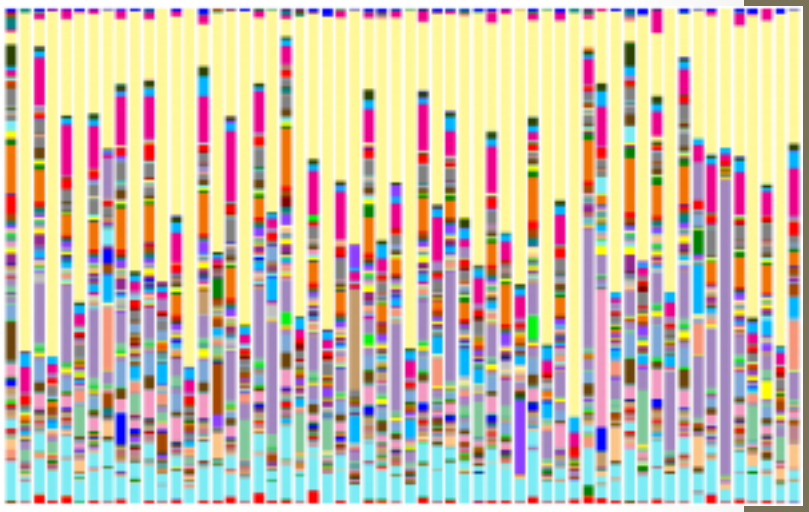
Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Metagenomics
  - Studies of the collective set of genomes in mixed microbial communities
  - Material obtained directly from the environment, such as clinical samples
  - Evaluation of the diversity of an actual sample
  - Metagenomes of microbial communities include human body niches
    - Microbiome – the entire population of microbes, bacteria, fungi, and viruses that colonize the human body

# Investigation of microbial communities

- Home microbiome
  - Each vertical bar is one shoe or cell phone
  - Each color represents a different species of bacteria
  - The larger the colored block, the more prevalent the species.



http://homemicrobiome.com/microbial-biodiversity-of-cell-phones-and-shoes/

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

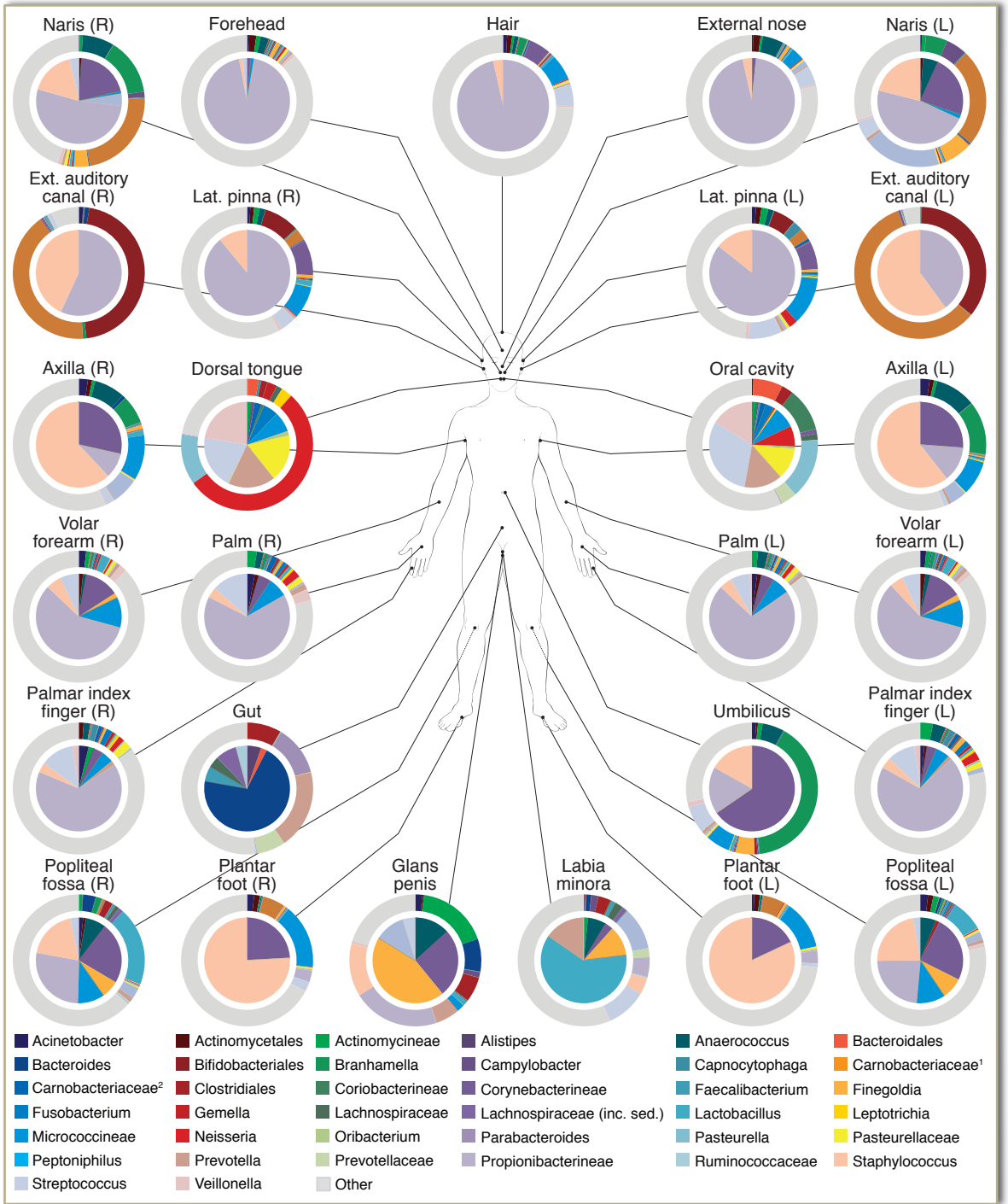
- Human Microbiome
  - Study surveying bacteria from up to 27 sites in healthy adults on four occasions
  - Interpersonal variability was high, but individuals exhibited minimal temporal variability

## Bacterial Community Variation in Human Body Habitats Across Space and Time

Elizabeth K. Costello,[1] Christian L. Lauber,[2] Micah Hamady,[3] Noah Fierer,[2,4] Jeffrey I. Gordon,[5] Rob Knight[1,6]*

Elucidating the biogeography of bacterial communities on the human body is critical for establishing healthy baselines from which to detect differences associated with diseases. To obtain an integrated view of the spatial and temporal distribution of the human microbiota, we surveyed bacteria from up to 27 sites in seven to nine healthy adults on four occasions. We found that community composition was determined primarily by body habitat. Within habitats, interpersonal variability was high, whereas individuals exhibited minimal temporal variability. Several skin locations harbored more diverse communities than the gut and mouth, and skin locations differed in their community assembly patterns. These results indicate that our microbiota, although personalized, varies systematically across body habitats and time; such trends may ultimately reveal how microbiome changes cause or prevent disease.

Center for
Systems Genomics
Pennsylvania State University




Costello et al., Science 2009

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Metagenomes of microbial communities that occupy human body niches
  - Estimated to have a gene content 100-fold greater than the human genome
  - Collections of genes encoding a wide array of biochemical and physiological functions
    - May be relevant in healthy and disease conditions

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Metagenomes of microbial communities that occupy human body niches
  - Deep sequencing of clinical samples allows identification and characterization
    - novel pathogens
    - host response to infection

**Metagenomics and development of the gut microbiota in infants**

Y. Vallès[1], M. J. Gosalbes[1,2], L. E. de Vries[3,4], J. J. Abellán[1,2] *and* M. P. Francino[1,5]

Center for
Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- The human virome
  - Description of viral communities, including bacteriophages in the human body
    - Their relationship with health and disease
  - Characterization of fecal viromes and relations with bacterial metagenomes
  - Characterization of the virome in the skin of healthy individuals

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- The human virome
  - Bioinformatic annotation strategies may help in the identification and quantitative description of pathogenic viruses

**Viral metagenome analysis to guide human pathogen monitoring in environmental samples**

K. Bibby, E. Viau and J. Peccia

Department of Chemical and Environmental Engineering, Yale University, New Haven, CT, USA

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Metagenomic analyses in the honey bee
  - One of the first uses of NGS for virus discovery in insects
  - Colony collapse disorder (CCD)
  - Determined Israeli acute paralysis virus (IAPV) was a significant indicator of CCD
- Microbiome of the honey bee
  - NGS has been used to identify four novel viruses in honey bees

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Viral metagenomics may enhance the chances of identifying viruses in clinical samples
  - Monitoring the emergence of mutations that confer virulence or resistance to antiviral drugs in influenza
  - Detection of influenza virus from nasopharyngeal aspirates of influenza patients
  - Norovirus from fecal specimens of patients infected with norovirus
    - Nakamura et al. PLoS One 2009
  - Identification of viral quasi-species

**A Metagenomic Analysis of Pandemic Influenza A (2009 H1N1) Infection in Patients from North America**

Alexander L. Greninger[1], Eunice C. Chen[2,3], Taylor Sittler[2,3], Alex Scheinerman[2,3], Nareg Roubinian[4], Guixia Yu[2,3], Edward Kim[5,6], Dylan R. Pillai[7,8], Cyril Guyard[7,8], Tony Mazzulli[7,9], Pavel Isa[10], Carlos F. Arias[10], John Hackett Jr.[11], Gerald Schochetman[11], Steve Miller[2,3], Patrick Tang[5,6], Charles Y. Chiu[2,3,4]*

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Viral metagenomics
  - Metagenomic data originate from heterogeneous microbial communities
  - Vast majority of viruses isolated from environmental samples are novel and consequently most of their genes do not have homologous sequences in the public databases, making functional annotation even more difficult

**TheViral MetaGenome Annotation Pipeline(VMGAP):an automated tool for the functional annotation of viral Metagenomic shotgun sequencing data**

Hernan A. Lorenzi[1], Jeff Hoover[1], Jason Inman[1], Todd Safford[1], Sean Murphy[1], Leonid Kagan[1], Shannon J. Williamson[2*]

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Study of fecal viromes
  - Sequencing of virus-like particles and bacteria isolated from fecal samples
  - Collected from healthy monozygotic twins and their mothers
  - Three time points over a one year period

**Viruses in the faecal microbiota of monozygotic twins and their mothers**

Alejandro Reyes[1], Matthew Haynes[2], Nicole Hanson[2], Florent E. Angly[2,3], Andrew C. Heath[4], Forest Rohwer[2] & Jeffrey I. Gordon[1]

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Study of fecal viromes
  - Co-twins and mothers shared a significantly greater degree of similarity in their bacterial faecal communities than unrelateds
  - Viromes were unique to individuals regardless of relatedness
  - Interpersonal variation in viromes was low, with > 95% of virotypes maintained over the period surveyed

**Viruses in the faecal microbiota of monozygotic twins and their mothers**

Alejandro Reyes[1], Matthew Haynes[2], Nicole Hanson[2], Florent E. Angly[2,3], Andrew C. Heath[4], Forest Rohwer[2] & Jeffrey I. Gordon[1]

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Quasispecies
  - Viral populations rapidly adapt to changing environments by selection of variants with better fitness
  - Important viral properties may not be best explained by a consensus sequence
    - Requires knowledge about microvariants present in viral populations
    - Knowledge relevant to
      - Viral evolution
      - Spread
      - Evasion of immune response
      - Anti-viral drug resistance
      - Vaccine development and manufacture

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Quasispecies
  - HIV quasispecies
  - HIV is highly heterogeneous
    - High turnover rates
    - High viral load
    - Replication mediated by error prone reverse transcriptase enzyme
    - Recombination shuttling mutations between viral genomes, leading to antigenic shifts and alterations in virulence

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Quasispecies
  - HIV quasispecies
  - 454 pyrosequencing methods analzying the variable regions of heavy and light chains of neutralizing antibodies against HIV in the blood obtained from HIV-1 infected individuals
  - Understanding how broadly neutralizing antibodies develop

**Focused Evolution of HIV-1 Neutralizing Antibodies Revealed by Structures and Deep Sequencing**

Xueling Wu,[1*] Tongqing Zhou,[1*] Jiang Zhu,[1*] Baoshan Zhang,[1] Ivelin Georgiev,[1] Charlene Wang,[1] Xuejun Chen,[1] Nancy S. Longo,[1] Mark Louder,[1] Krisha McKee,[1] Sijy O'Dell,[1] Stephen Perfetto,[1] Stephen D. Schmidt,[1] Wei Shi,[1] Lan Wu,[1] Yongping Yang,[1] Zhi-Yong Yang,[1] Zhongjia Yang,[1] Zhenhai Zhang,[1,2] Mattia Bonsignori,[3] John A. Crump,[4,5] Saidi H. Kapiga,[6] Noel E. Sam,[5,6] Barton F. Haynes,[3] Melissa Simek,[7] Dennis R. Burton,[8,9] Wayne C. Koff,[7] Nicole A. Doria-Rose,[10] Mark Connors,[10] NISC Comparative Sequencing Program,[11] James C. Mullikin,[11] Gary J. Nabel,[1] Mario Roederer,[1] Lawrence Shapiro,[1,2] Peter D. Kwong,[1]† John R. Mascola[1]†

Center for Systems Genomics
Pennsylvania State University

# Investigation of microbial communities

- Quasispecies
  - Other quasispecies
    - Total RNAs extracted from a patient who died of viral pneumonia due to pandemic 2009 influenza A virus
    - Human rhinovirus genome in patient infected with the same HRV strain for two years

Center for Systems Genomics
Pennsylvania State University

# Drug resistance mutations in patients

- Drug resistance mutations in patients
  - Antiviral drug resistance
  - Deep sequencing to detect low abundance drug resistance
    - HIV variants
    - Hepatitis C virus
      - HCV minor variants
  - High sensitivity for minor variants

## Extensive Genome-Wide Variability of Human Cytomegalovirus in Congenitally Infected Infants

**Nicholas Renzette[1], Bornali Bhattacharjee[1], Jeffrey D. Jensen[2], Laura Gibson[3], Timothy F. Kowalik[1,4]***

**1** Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, **2** Program in Bioinformatics & Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, **3** Departments of Pediatrics and Medicine, Divisions of Infectious Diseases and Immunology, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America, **4** Immunology and Virology Program, University of Massachusetts Medical School, Worcester, Massachusetts, United States of America

Center for
Systems Genomics
Pennsylvania State University

# Drug resistance mutations in patients

- Drug resistance mutations in patients
  - Clinical setting, careful consideration necessary of technical errors in data collection and analysis to discriminate between error and true variants

Center for Systems Genomics
Pennsylvania State University

# Drug resistance mutations in patients

## Characterization of Quasispecies of Pandemic 2009 Influenza A Virus (A/H1N1/2009) by *De Novo* Sequencing Using a Next-Generation DNA Sequencer

**Makoto Kuroda[1]*, Harutaka Katano[2], Noriko Nakajima[2], Minoru Tobiume[2], Akira Ainai[2], Tsuyoshi Sekizuka[1], Hideki Hasegawa[3], Masato Tashiro[3], Yuko Sasaki[4], Yoshichika Arakawa[4], Satoru Hata[5], Masahide Watanabe[6], Tetsutaro Sata[2]**

## Deep Sequencing Reveals Highly Complex Dynamics of Human Cytomegalovirus Genotypes in Transplant Patients over Time$^{\nabla}$

Irene Görzer,[1] Christian Guelly,[2] Slave Trajanoski,[2] and Elisabeth Puchhammer-Stöckl[1]*

## Use of Massive Parallel Pyrosequencing for Near Full-Length Characterization of a Unique HIV Type 1 BF Recombinant Associated with a Fatal Primary Infection

Alessandro Bruselles, Gabriella Rozera, Barbara Bartolini, Mattia Prosperi, Franca Del Nonno, Pasquale Narciso, Maria R. Capobianchi, and Isabella Abbate

Center for Systems Genomics
Pennsylvania State University

# Disease Susceptibility Across Populations

- DNA arrays and next-generation sequencing
  - International HapMap Project
  - 1000 Genomes project
  - Better understanding of the entire genome variation across populations
  - Allow us to evaluate the contribution of host genetic diversity to differences in susceptibility
    - Rare
    - Common infectious disease

Center for Systems Genomics
Pennsylvania State University

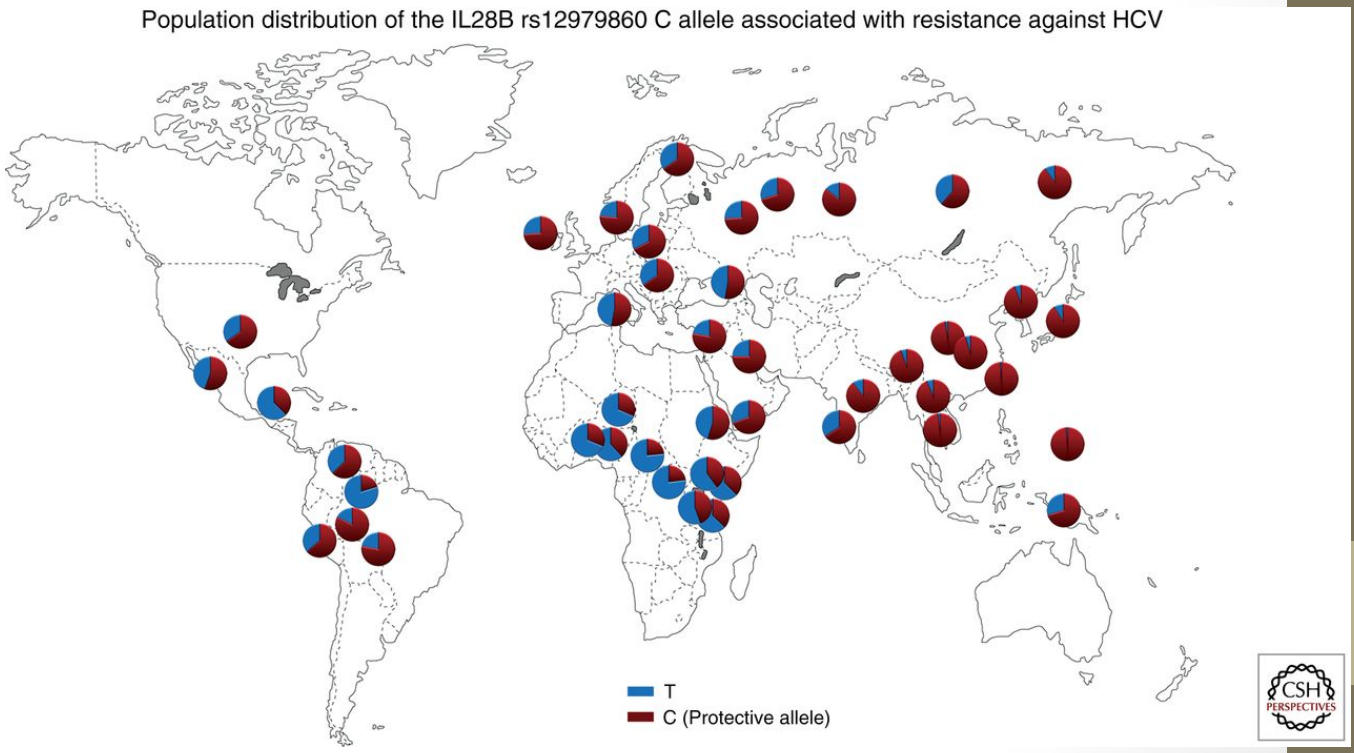
# Disease Susceptibility Across Populations

- DNA arrays and next-generation sequencing
  - Allow us to evaluate the contribution of host genetic diversity to differences in susceptibility
    - Infectious disease: powerful agents impacting evolution of human genetic diversity

# Disease Susceptibility Across Populations

- DNA arrays and next-generation sequencing
  - GWA of common viral, bacterial, and parasitic infections have shown the genetic basis of susceptibility

The "protective" C variant (red) has been associated with both a better response to hepatitis C drug treatment and enhanced spontaneous HCV viral clearance




Ge et al. 2009; Thomas et al. 2009

Center for
Systems Genomics
Pennsylvania State University

# NGS: Applications

- Questions?