Metagenomics

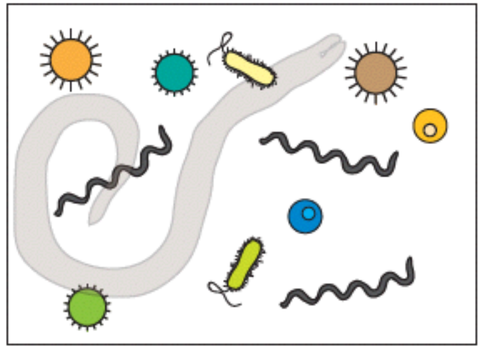
Mark Stenglein

Computational Biology and
Genomics Workshop

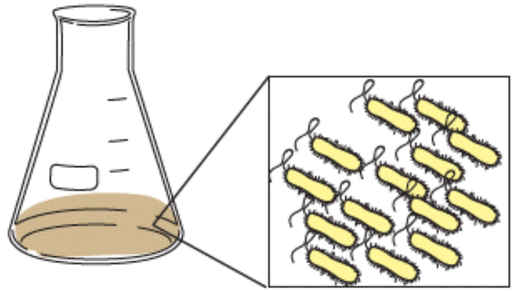
Todos Santos Center
April 9-13, 2018

What is metagenomics?

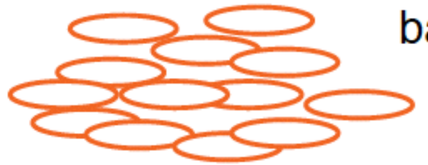
soil community



bacterial isolate

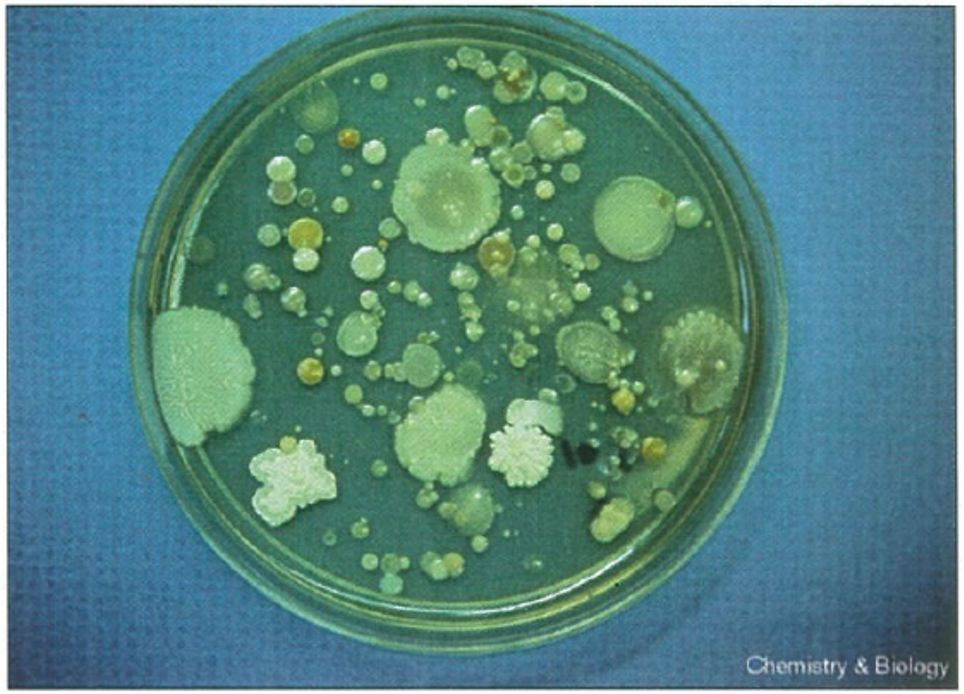


soil ‘metagenome’



bacterial genome

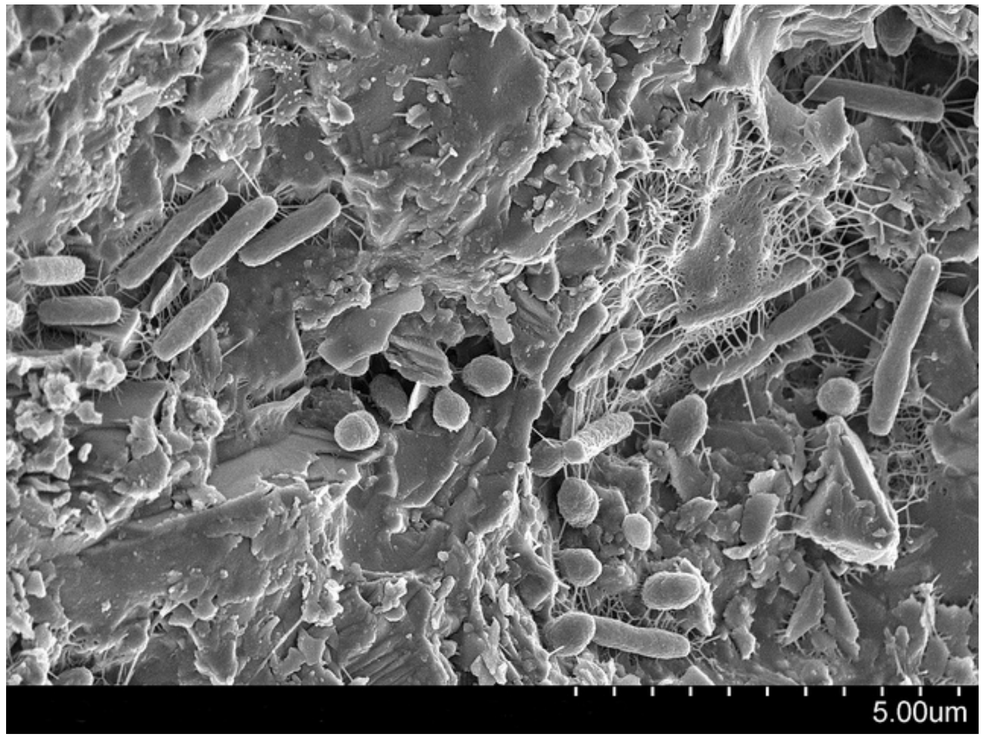
Metagenomics emerged in response to the observation that most micro-organisms can't be cultured



Morphological diversity typical of microorganisms cultured from soil on a broad spectrum medium, tryptic soy agar.

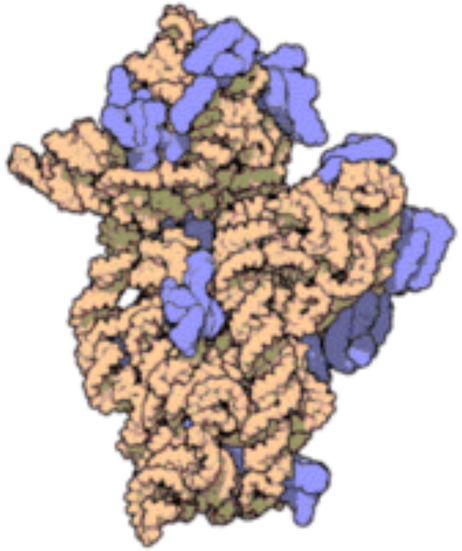
Handelsman et al (1998) Chem & Biol

Estimated: 10^8 bacteria per gram of soil of 6000-8000 different species
Only ~1% culturable (?)

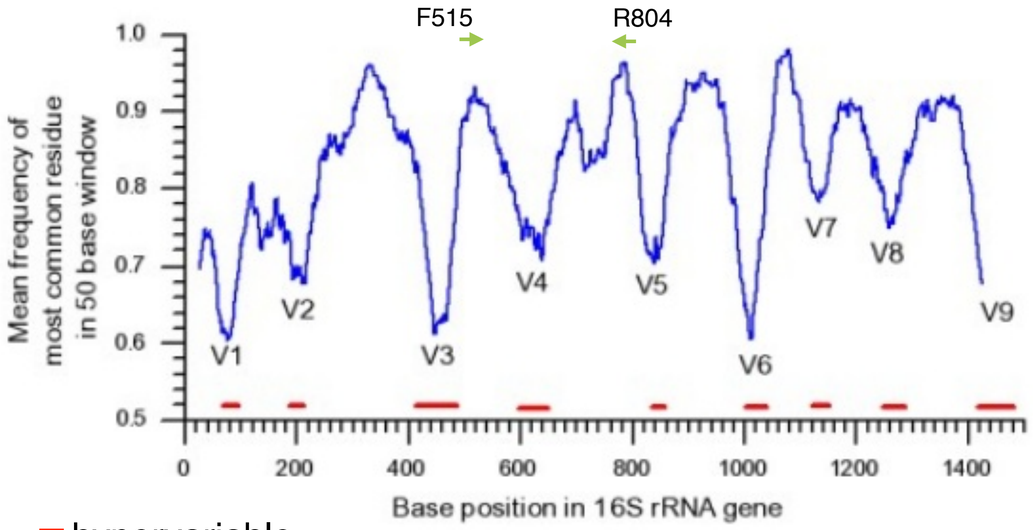


EM: Kim Lewis, Northeastern Univ.

PCR using primers targeting conserved regions of the 16S rRNA gene and sequencing enables genotyping of bacteria and archaea without having to culture them



bacterial 30S ribosomal subunit
16S rRNA is in orange
(purple: ribosomal proteins)
image: wikipedia



hypervariable

One of the earliest “metagenomics” paper, based on 16S sequencing

JOURNAL OF BACTERIOLOGY, July 1991, p. 4371–4378
0021-9193/91/144371-08$02.00/0

Vol. 173, No. 14

Analysis of a Marine Picoplankton Community by 16S rRNA Gene Cloning and Sequencing

THOMAS M. SCHMIDT,† EDWARD F. DELONG,‡ AND NORMAN R. PACE*

Department of Biology and Institute for Molecular and Cellular Biology, Indiana University, Bloomington, Indiana 47405

Received 7 January 1991/Accepted 13 May 1991

The phylogenetic diversity of an oligotrophic marine picoplankton community was examined by analyzing the sequences of cloned ribosomal genes. This strategy does not rely on cultivation of the resident microorganisms. Bulk genomic DNA was isolated from picoplankton collected in the north central Pacific Ocean by tangential flow filtration. The mixed-population DNA was fragmented, size fractionated, and cloned into bacteriophage lambda. Thirty-eight clones containing 16S rRNA genes were identified in a screen of 3.2 × 10^4 recombinant phage, and portions of the rRNA gene were amplified by polymerase chain reaction and sequenced. The resulting sequences were used to establish the identities of the picoplankton by comparison with an established data base of rRNA sequences. Fifteen unique eubacterial sequences were obtained, including four from cyanobacteria and eleven from proteobacteria. A single eucaryote related to dinoflagellates was identified; no archaebacterial sequences were detected. The cyanobacterial sequences are all closely related to sequences from cultivated marine *Synechococcus* strains and with cyanobacterial sequences obtained from the Atlantic Ocean (Sargasso Sea). Several sequences were related to common marine isolates of the γ subdivision of proteobacteria. In addition to sequences closely related to those of described bacteria, sequences were obtained from two phylogenetic groups of organisms that are not closely related to any known rRNA sequences from cultivated organisms. Both of these novel phylogenetic clusters are proteobacteria, one group within the α subdivision and the other distinct from known proteobacterial subdivisions. The rRNA sequences of the α-related group are nearly identical to those of some Sargasso Sea picoplankton, suggesting a global distribution of these organisms.

plankton

image: Smithsonian magazine

Schmidt et al actually made a shotgun library from pico plankton gDNA, identified rDNA clones (38/32000 clones) by colony hybridization, and Sanger sequenced them

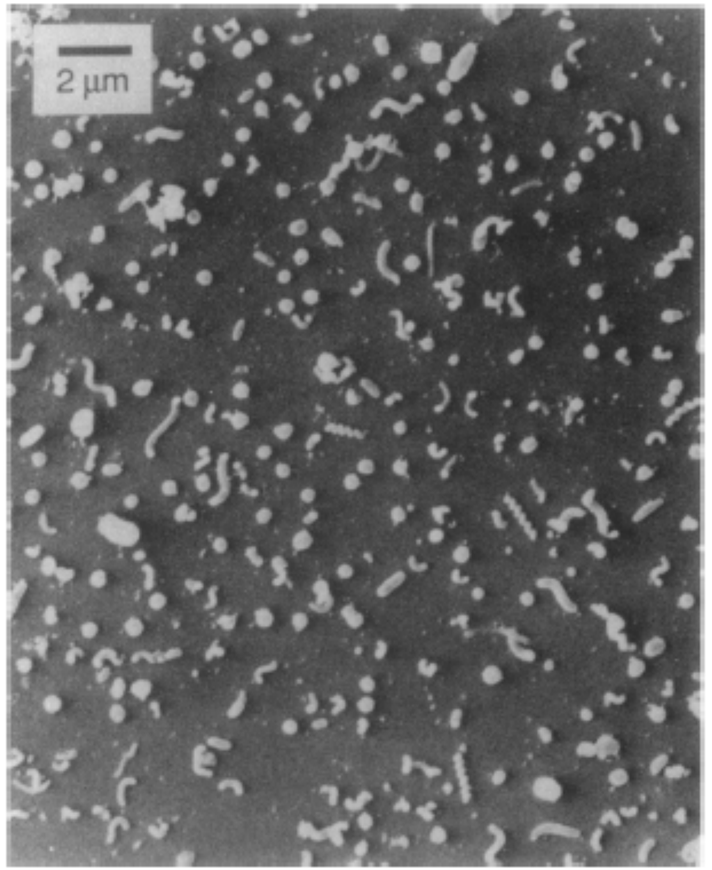


FIG. 2. Scanning electron micrograph of picoplankton represented in the clone library. Picoplankton concentrated from the ALOHA collection site were fixed and prepared for scanning electron microscopy as detailed in Materials and Methods.

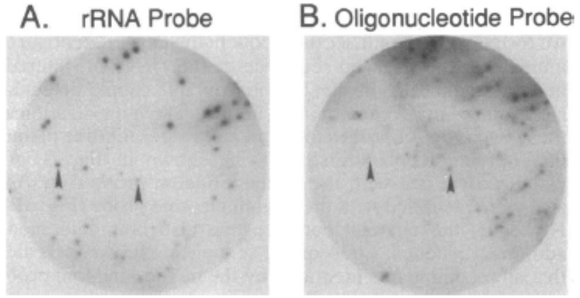


FIG. 3. Identification of rDNA-containing clones. Filter lifts of the recombinant bacteriophage library were probed as detailed in Materials and Methods with a mixed-kingdom rRNA probe consisting of 16S rRNAs from *O. linum* (eubacterium), *S. solfataricus* (archaebacterium), and *S. cerevisiae* (eucaryote) or an oligonucleotide probe complementary to a universally conserved 16S rRNA sequence.

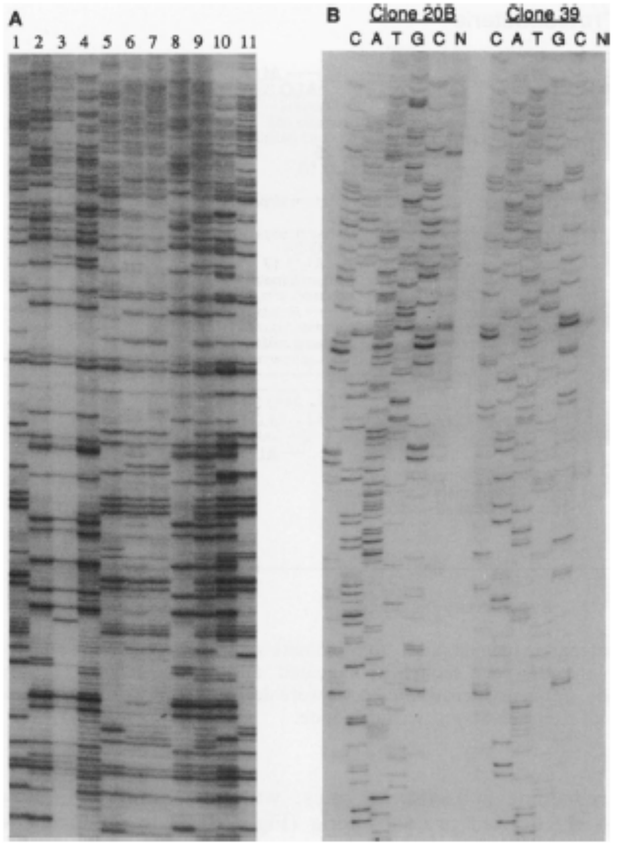
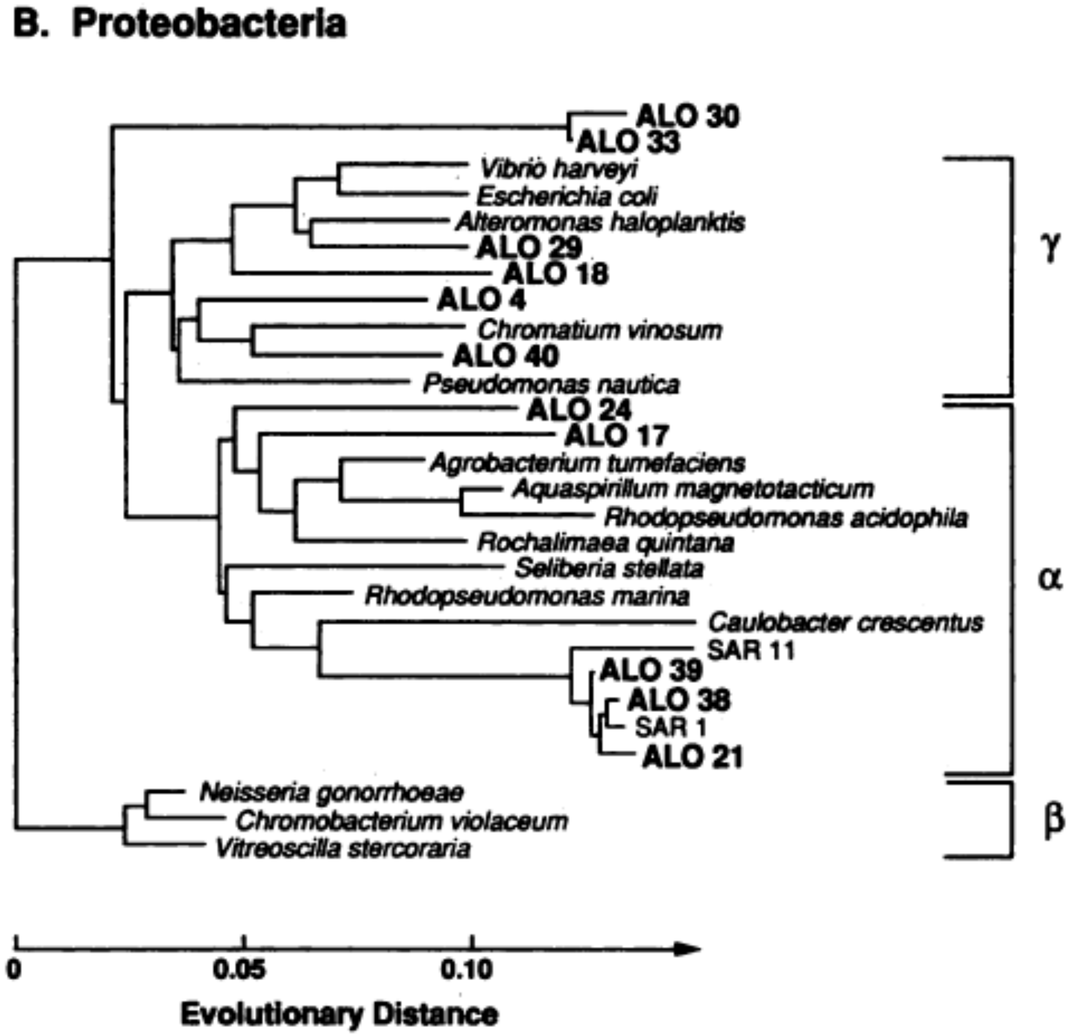
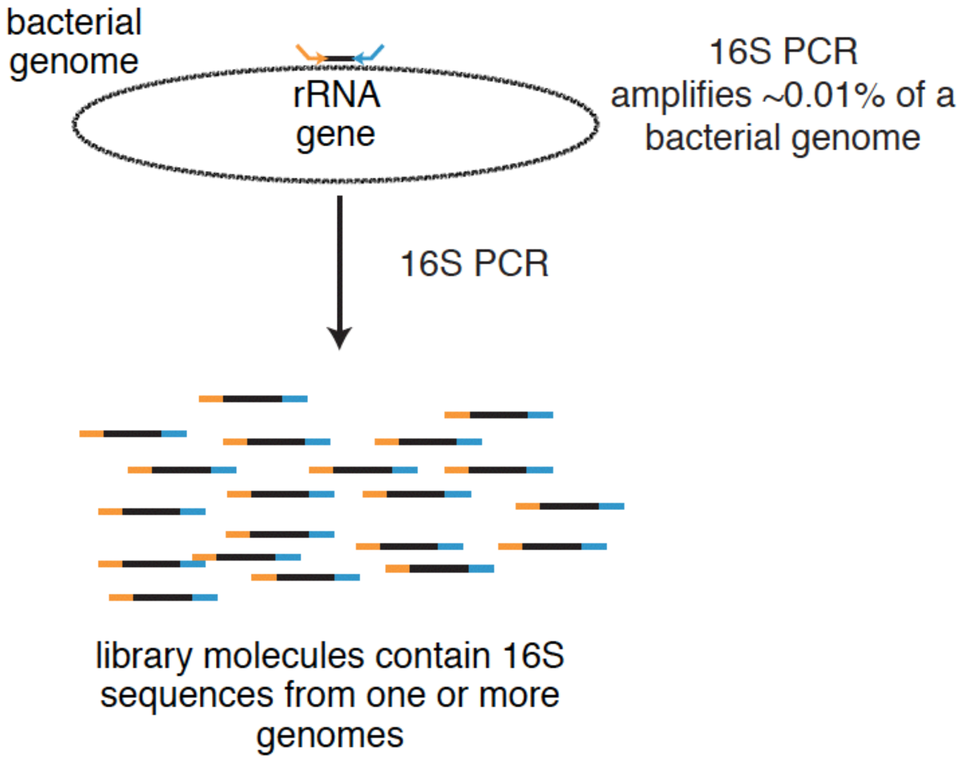


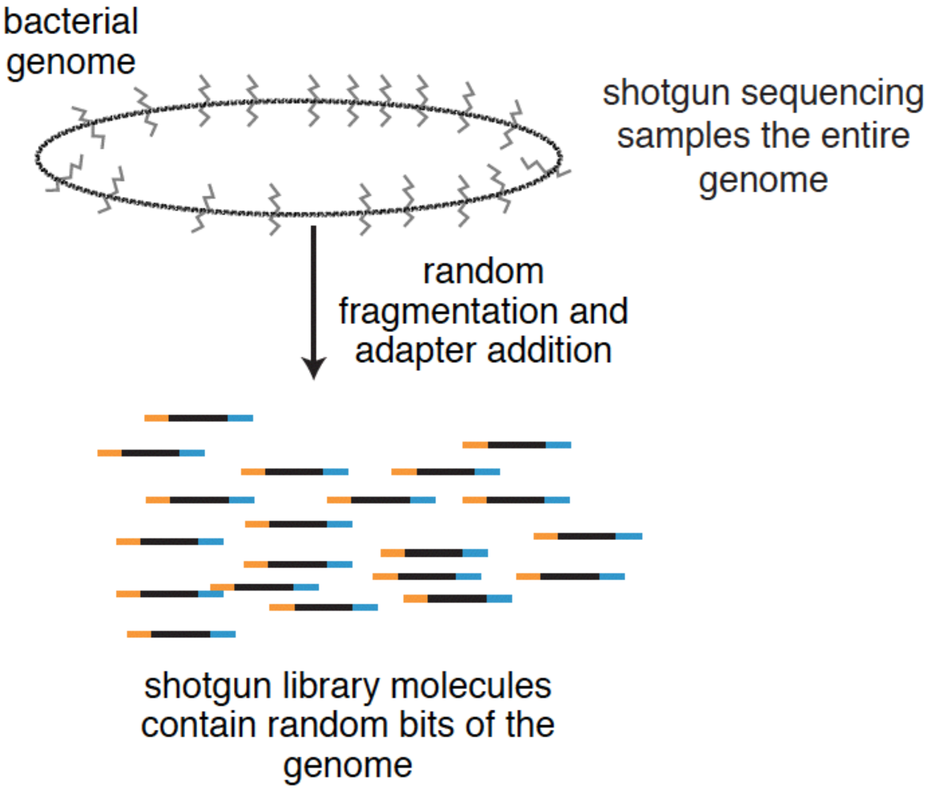
FIG. 5. Sorting and sequence analysis of rDNA-containing clones. (A) Single-nucleotide sequence pattern for 11 picoplankton clones, produced as detailed in Materials and Methods by using a PCR-amplified template, primer 519R, and a single dideoxynucleotide (ddA). (B) Sequence determination by dideoxynucleotide chain termination, using a PCR-generated, single-stranded template from two of the picoplankton clones and primer 519R.

This unbiased survey revealed that the picoplankton contained many previously unknown bacterial species



16S sequencing vs. shotgun metagenomics





- Only bacteria and archaea surveyed
- Deeper sampling of bacterial diversity per $
- Relatively easy to make libraries and interpret results
- Appropriate if all you care about is microbial diversity / ecology

- All organisms studied*
- Decreased sampling depth per $
- Enables analysis of other genomic features of organisms, e.g. antimicrobial resistant genes
- Analysis is significantly more difficult

Warning! Some will object if you refer to 16S-based studies as “metagenomics”

THE TREE OF LIFE

Blog of Jonathan Eisen, Professor at UC Davis.

Wednesday, August 22, 2012

Referring to 16S surveys as "metagenomics" is misleading and annoying #badomics #OmicMimicry

Aargh. I am a big fan if of ribosomal RNA based surveys of microbial diversity. Been doing them for 20+ years and still continue to - even though I have moved on to more genomic/metagenomic based studies. But it drives me crazy to see rRNA surveys now being called "metagenomics".

Here are some examples of cases where rRNA surveys are referred to as metagenomics:

Deep 16S rRNA metagenomics and quantitative PCR analyses of the premature infant fecal

http://phylogenomics.blogspot.com/2012/08/referring-to-16s-surveys-as.html

Viral metagenomics

JOURNAL OF VIROLOGY, July 2010, p. 6955–6965 Vol. 84, No. 14
0022-538X/10/$12.00 doi:10.1128/JVI.00501-10

Bat Guano Virome: Predominance of Dietary Viruses from Insects and Plants plus Novel Mammalian Viruses▽

Linlin Li,[1,2] Joseph G. Victoria,[1,2] Chunlin Wang,[3] Morris Jones,[4] Gary M. Fellers,[5] Thomas H. Kunz,[6] and Eric Delwart[1,2]*

Blood Systems Research Institute, San Francisco, California[1]; Department of Laboratory Medicine, University of California, San Francisco, California[2]; Stanford Genome Technology Center, Stanford, California[3]; Clinical Investigation Facility, David Grant USAF Medical Center, Travis Air Force Base, California[4]; U.S. Geological Survey, Western Ecological Research Center, Point Reyes, California[5]; and Center for Ecology and Conservation Biology, Department of Biology, Boston University, Boston, Massachusetts[6]

Received 5 March 2010/Accepted 30 April 2010

Bats are hosts to a variety of viruses capable of zoonotic transmissions. Because of increased contact between bats, humans, and other animal species, the possibility exists for further cross-species transmissions and ensuing disease outbreaks. We describe here full and partial viral genomes identified using metagenomics in the guano of bats from California and Texas. A total of 34% and 58% of 390,000 sequence

Tadarida brasiliensis

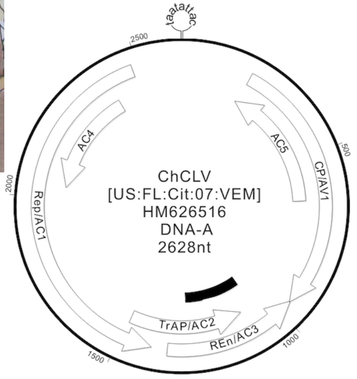
image: Wikipedia/NPS

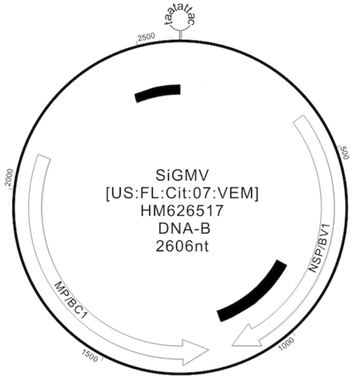
They found:

- Known bat viruses
- Putative new bat viruses
- Viruses likely infecting the plants and insects that the bats ate

‘Vector-enabled metagenomics’

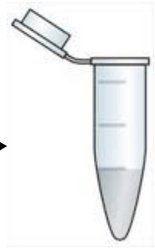
Mya Breitbart lab, Univ of South Florida



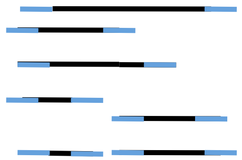


Pathogen discovery using metagenomics sequencing

case and control tissues

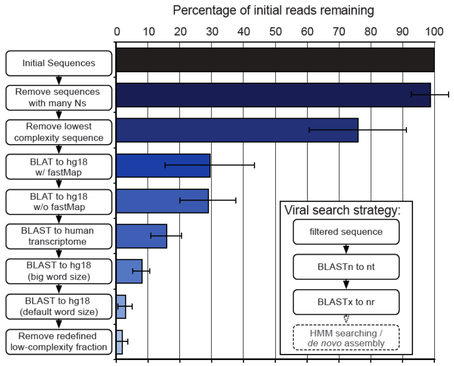


Nucleic acid



Library prep / barcode

Illumina sequencing



Computational Analysis

Follow-up

Sequencing can only give you sequences

A rabbit facility in TN experienced an outbreak of fatal gastroenteritis

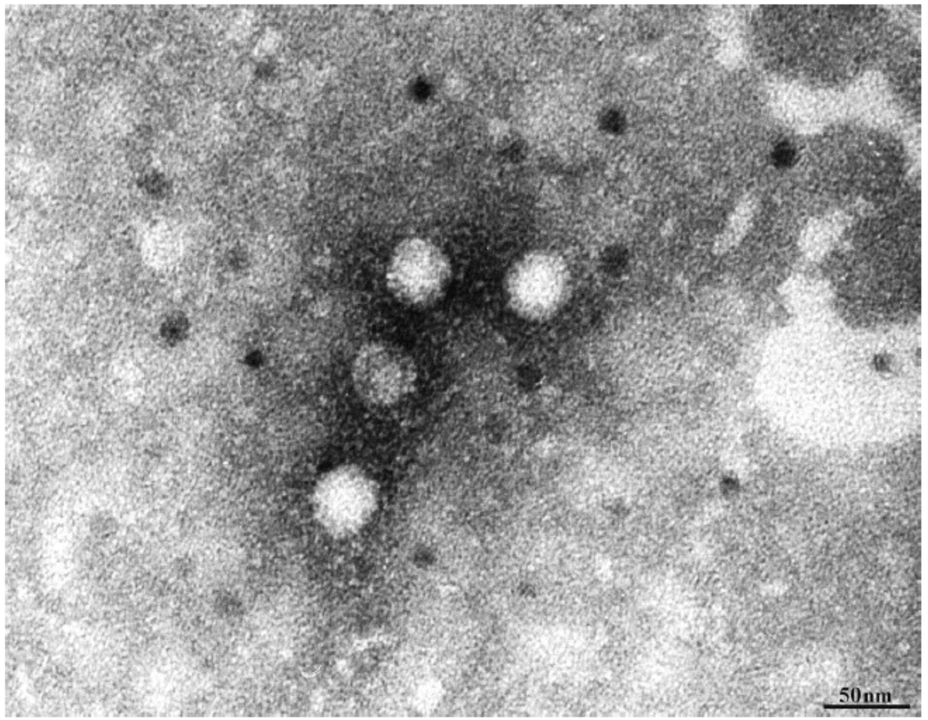


Figure 1 Electron micrograph of virus like particles in the stool of one animal (Table 1). Scale bar indicates 50 nm.

image: wikipedia

An astrovirus is the likely cause of gastroenteritis in these rabbits

Stenglein *et al. Virology Journal* 2012, **9**:216
http://www.virologyj.com/content/9/1/216

VIROLOGY JOURNAL

RESEARCH Open Access

Complete genome sequence of an astrovirus identified in a domestic rabbit (*Oryctolagus cuniculus*) with gastroenteritis

Also lots of plant sequences, presumably from the rabbit's diet

Astroviruses cause diarrhea in a variety of animals, including humans.

Astrovirus particles

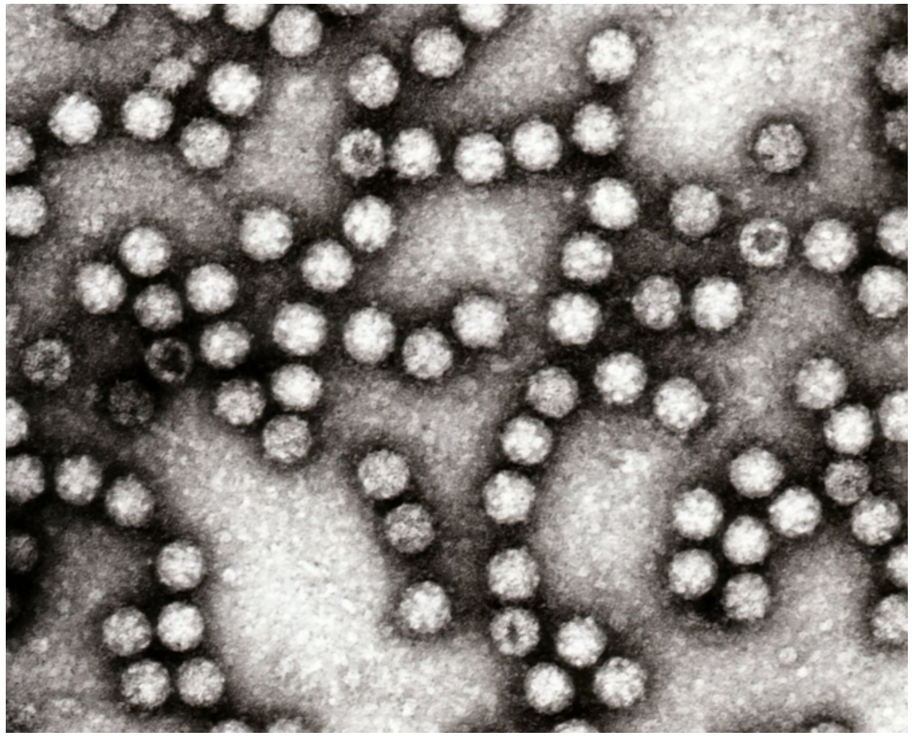


image: wikipedia

JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 1993, p. 955–962
0095-1137/93/040955-08$02.00/0

Vol. 31, No. 4

Characterization and Seroepidemiology of a Type 5 Astrovirus Associated with an Outbreak of Gastroenteritis in Marin County, California

KAREN MIDTHUN,[1]†* HARRY B. GREENBERG,[1]‡ JOHN B. KURTZ,[2] G. WILLIAM GARY,[3] FENG-YING C. LIN,[4] AND ALBERT Z. KAPIKIAN[1]

RESULTS

Volunteer study. Nineteen adult volunteers were orally administered a filtrate prepared from a 0.1% suspension of stool from one of the ill individuals in the original Marin County outbreak. None of 17 volunteers who received a 1-ml inoculum became ill. Because of this, the amount of inoculum was increased to 20 ml. Of two volunteers who received the larger inoculum, one developed a gastrointestinal illness characterized by nausea, vomiting, diarrhea, and malaise.