

Rocky 2008, December

High Throughput Metagenome Analysis

Rob Edwards

Department of Computer Science,
San Diego State University
Mathematics and Computer Sciences Division,
Argonne National Laboratory

Free workshops on NMPDR, RAST, mg-RAST, SEED
Contact Leslie McNeil (lkmcneil@ncsa.uiuc.edu)
or visit <http://www.nmpdr.org/>

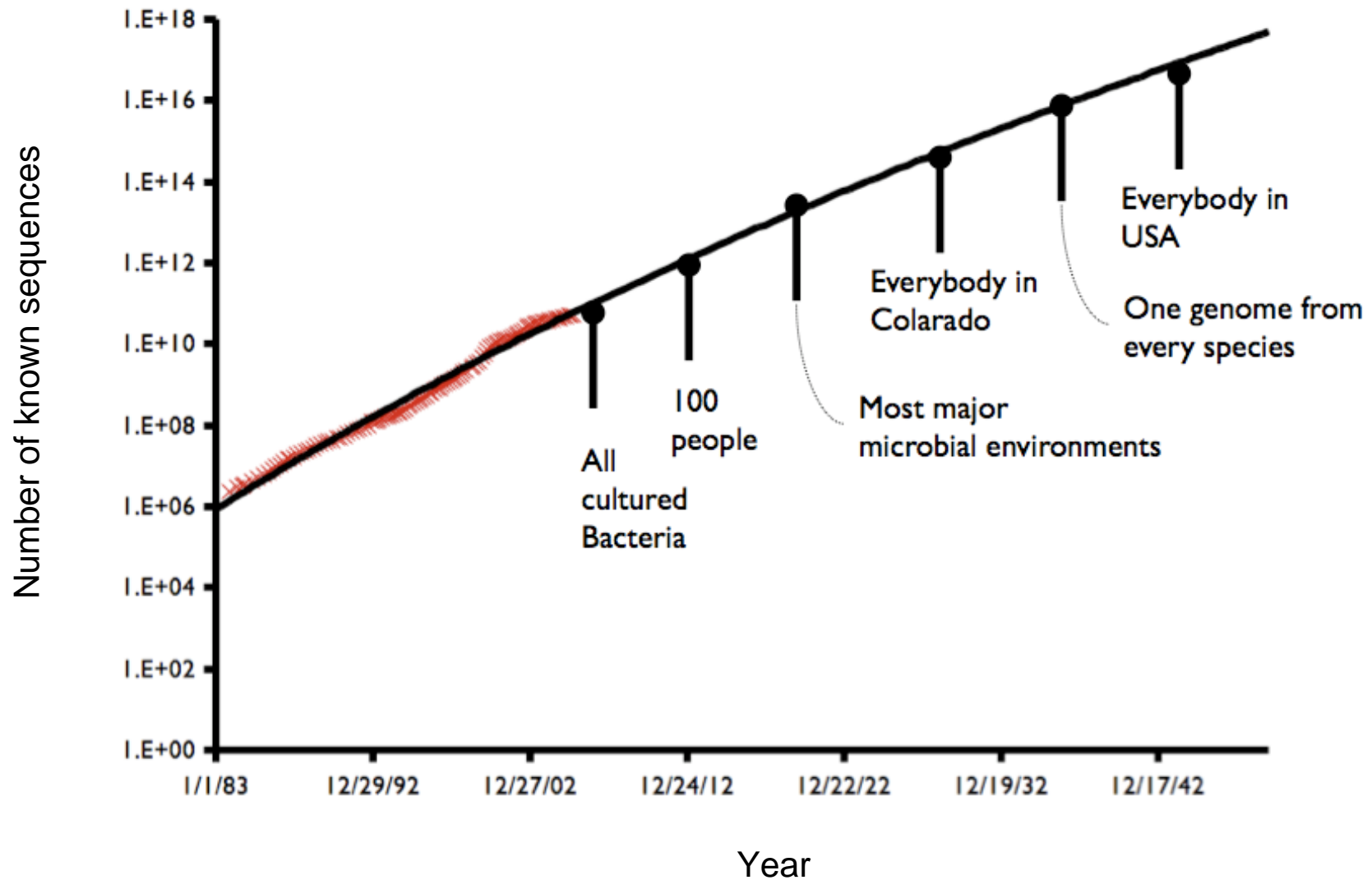


www.nmpdr.org

www.theseed.org



How much sequence?



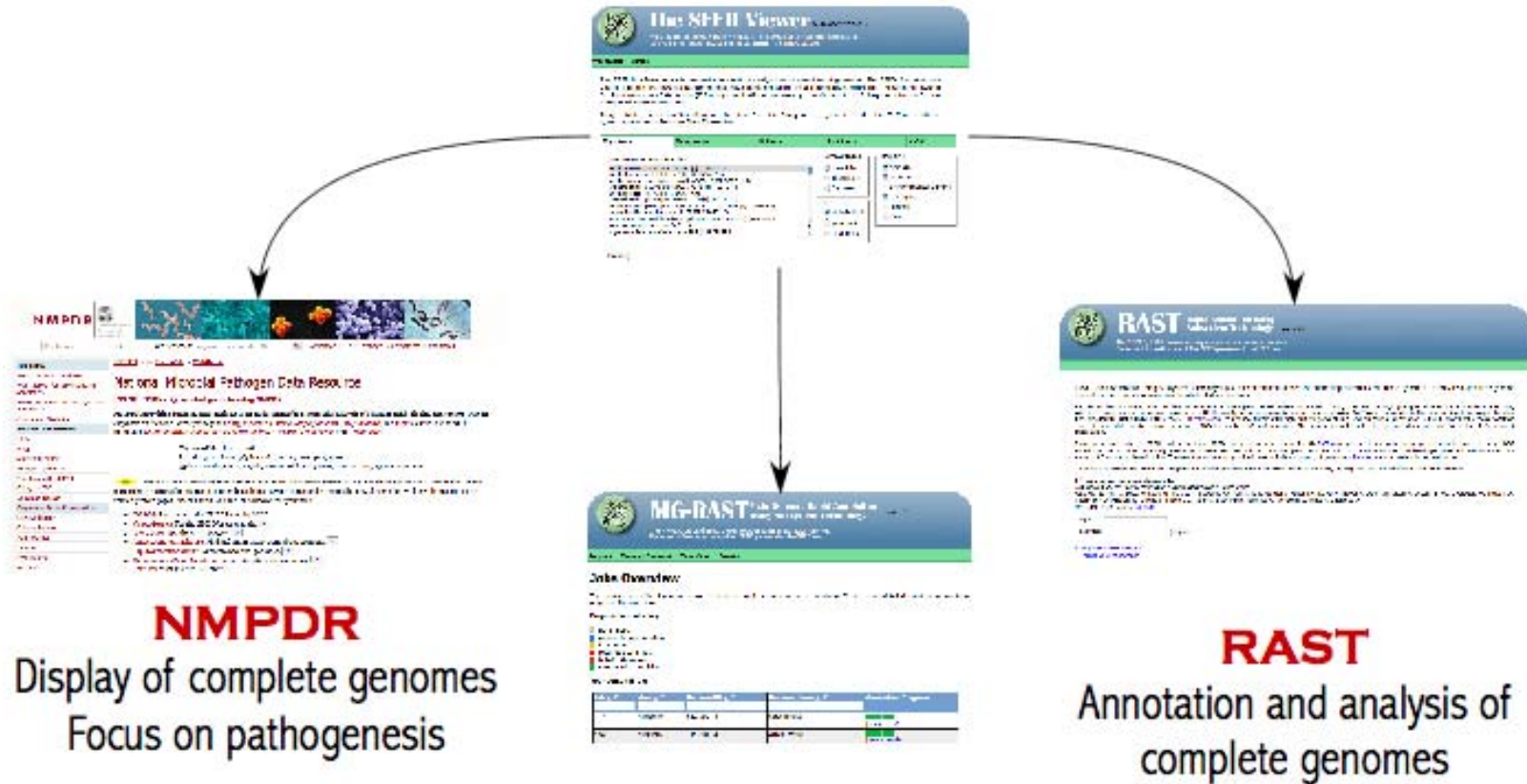
www.nmpdr.org

www.theseed.org



The SEED Family

THE SEED Environmental, Viral, Bacterial, Archaeal, and Eukaryal Genome Interpretation



NMPDR
Display of complete genomes
Focus on pathogenesis

MG-RAST
Annotation and analysis of
metagenomes

RAST
Annotation and analysis of
complete genomes



www.nmpdr.org

www.theseed.org



Metagenomics

(Just sequence it)

200 liters water
5-500 g fresh fecal matter
50 g soil



Concentrate and purify bacteria,
viruses, etc



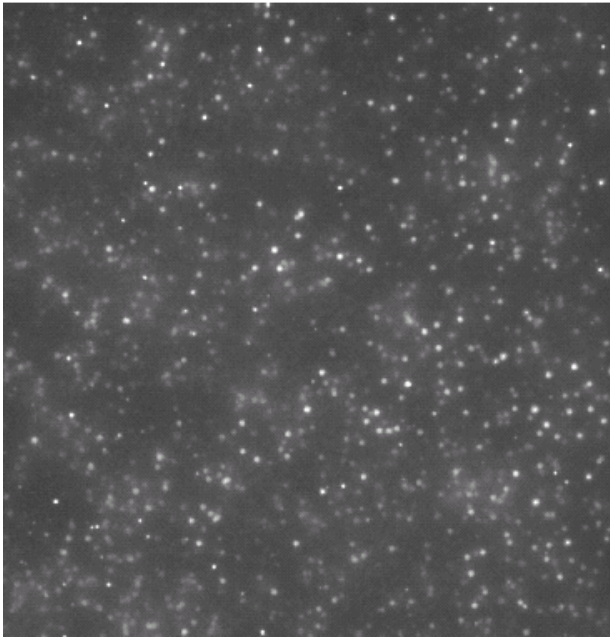
Extract nucleic acids



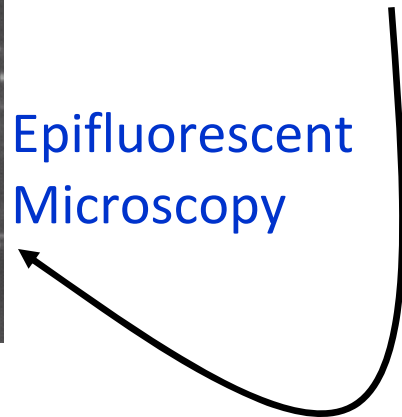
Sequence



Publish papers



Epifluorescent
Microscopy



The Metagenomics RAST Server



MG-RAST Meta Genome Rapid Annotation using Subsystem Technology

Welcome to the Metagenomics SEED Viewer.
For more information about The SEED please visit theSEED.org.

»Navigate »Help

MG-RAST is a fully-automated service for annotating metagenome samples.

It provides:

- **annotation** of sequence fragments,
- their **phylogenetic classification**,
- **metabolic reconstructions** and
- **comparison tools**

The service is built as a modified version of the RAST server which was originally designed to support high-quality annotation of complete or draft microbial genomes.



If you use our service, please cite:

The Metagenomics RAST server - A public resource for the automatic phylogenetic and functional analysis of metagenomes F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, R. Stevens, A. Wilke, J. Wilkening and R. A. Edwards submitted



» [Manage your uploaded data](#) » [Re](#)

You have access to t

Number of publicly available m

start typing to narrow selection

Public: 5-Way (CG) Acid Mine Drainag
Public: 640F6 (4440355.3) from proje
Public: 710F6 (4440387.3) from proje
Public: 80F6 (4440356.3) from projec
Public: ALVINELLA (4441102.3) from p
Public: ArcticVir2002 (4440306.3) fro
Public: Australian Phosphorus Removi
Public: BBCVir96to04 (4440305.3) fro

[View Metagenome](#)

User uploads DNA sequences

Pre-processing:
Remove duplicates, QC/QA
Renumber

Similarities

SEED proteins

Boutique nucleotides

Post-processing

Export

Display and analysis

Metagenomics Tools

- Download all data
- Annotation Summaries
- Annotation via Subsystems
- Connections to KEGG maps
- Recruitment Plots
- Phylogenomic Reconstruction
- Metagenome Comparisons



www.nmpdr.org

www.theseed.org



How much so far

Public metagenomes:

- 192 metagenomes
- 28,935,057 sequences
- 6,880,430,153 bp

Total metagenomes:

- 1,414 metagenomes
- 126,113,077 sequence
- 30,166,714,783 bp

Compute time:

~19 hours of compute per input megabyte
573,154 hours = 23,881 days = 65 years



www.nmpdr.org

www.theseed.org

