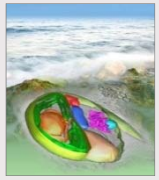
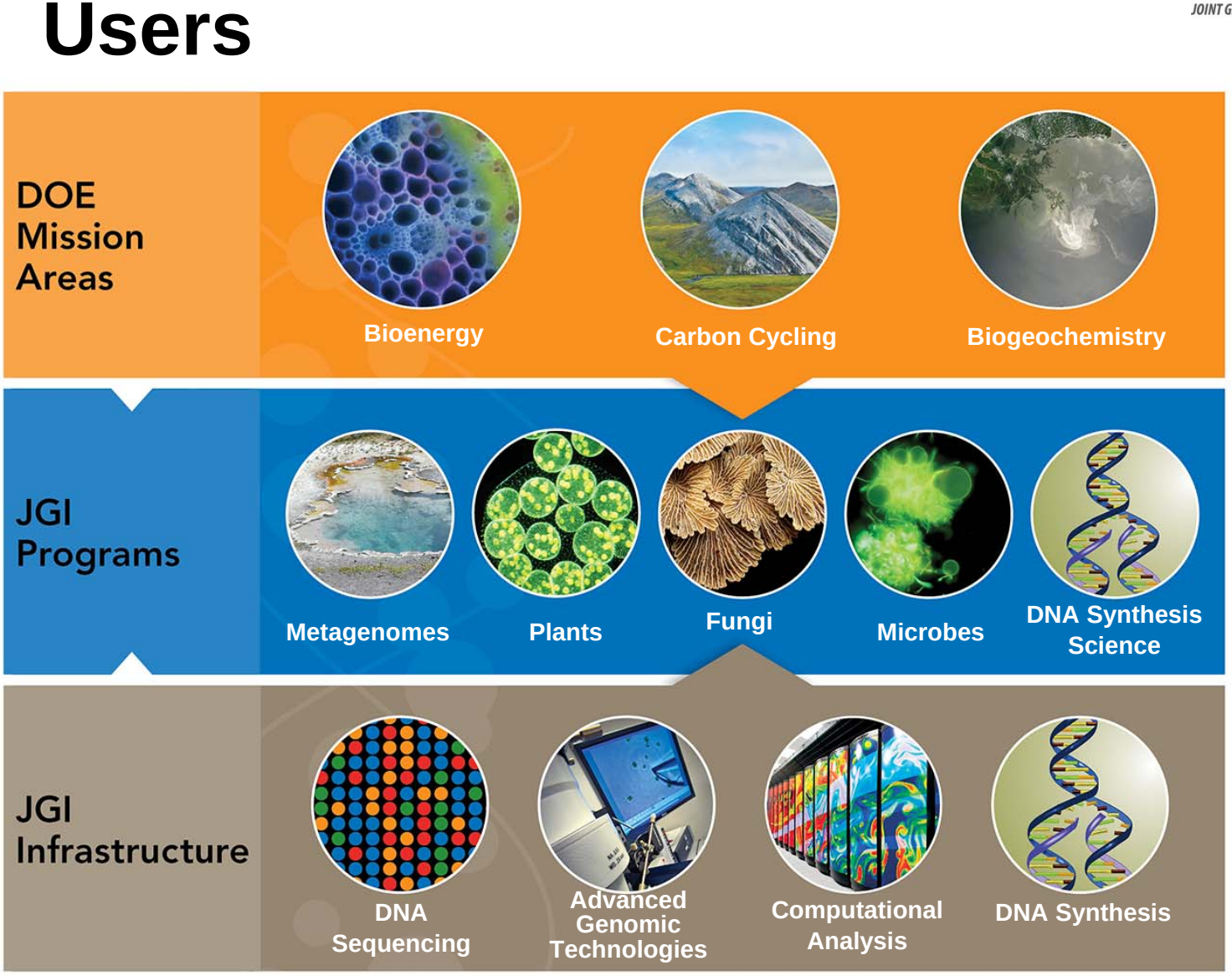
JGI Update: Implementation of the 10 Year Strategic Plan

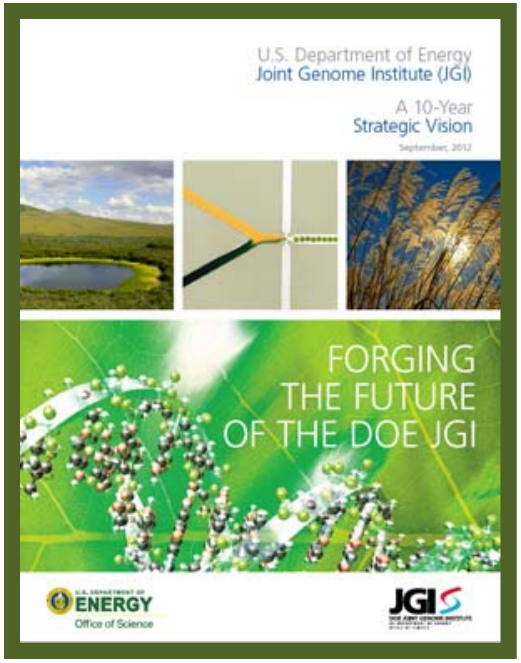


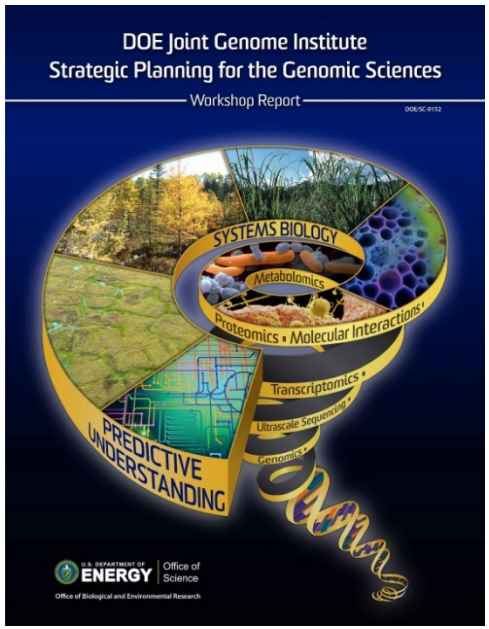



Users


DOE Mission Areas
Bioenergy
Carbon Cycling
Biogeochemistry
JGI Programs
Metagenomes
Plants
Fungi
Microbes
DNA Synthesis Science
JGI Infrastructure
DNA Sequencing
Advanced Genomic Technologies
Computational Analysis
DNA Synthesis

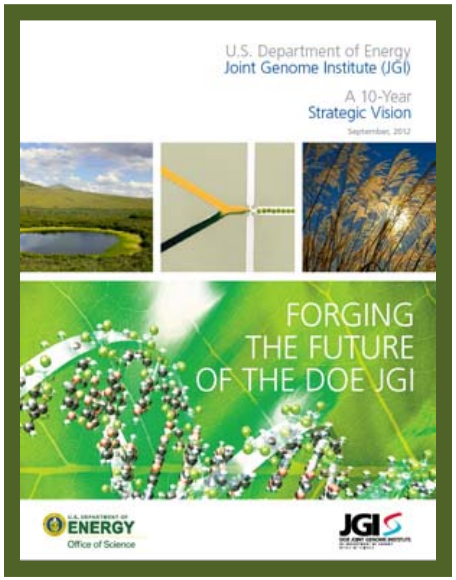
DOE Joint Genome Institute

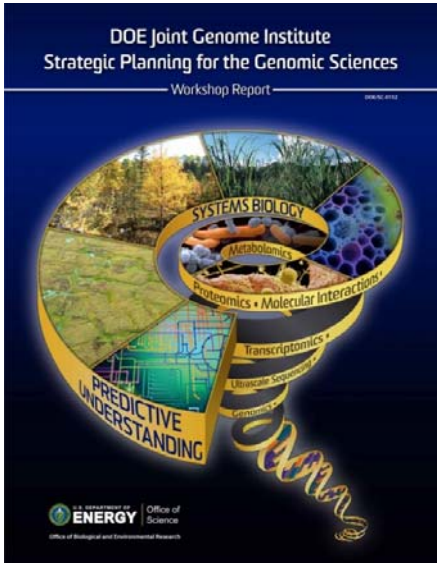




To evolve as a genome SCIENCE user facility to meet the scientific needs of energy and environmental research over the next decade.

High throughput approaches for converting sequence data into functional information



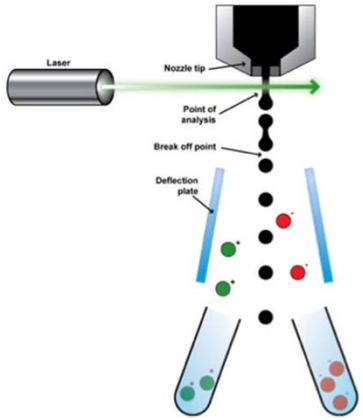


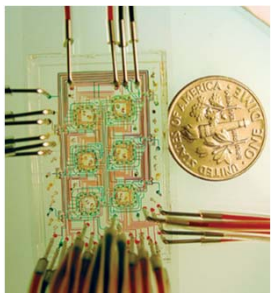
Sequence to Function

JGI will need new capabilities to convert sequence data to functional insights

Emerging Technologies Opportunity Program (ETOP)

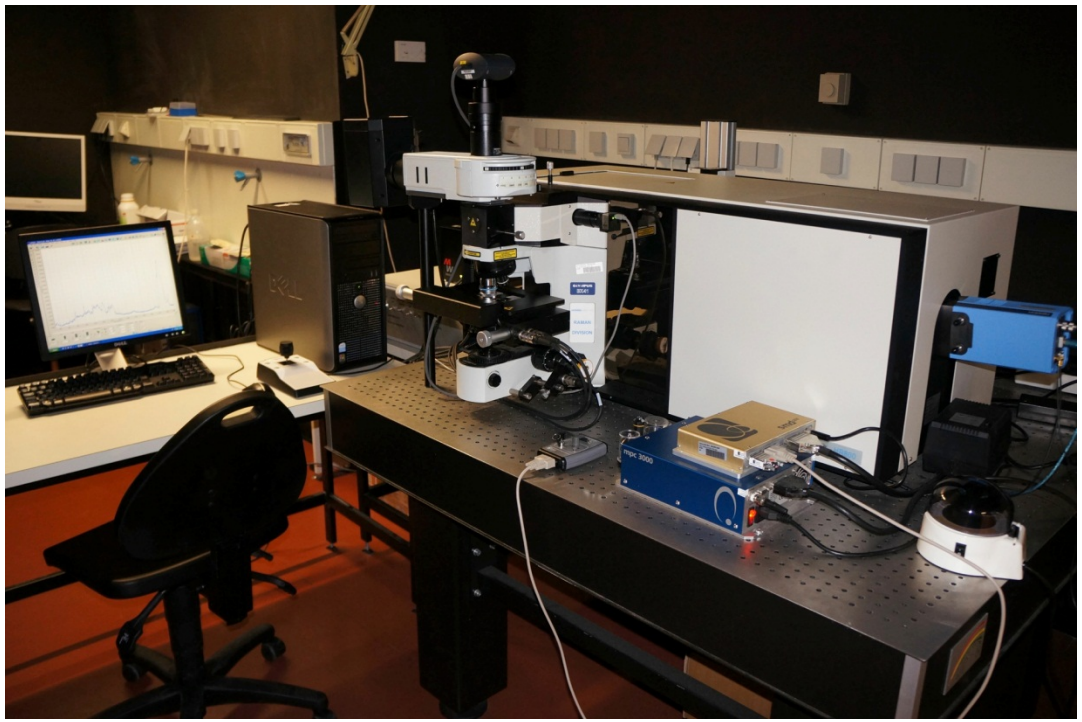
60 applications 6 funded projects





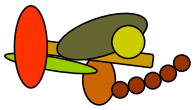
MIT/U.Vienna: “Function-Driven Genomics”

Using Raman-Spectroscopy coupled with microfluidic isolation of targeted single cells

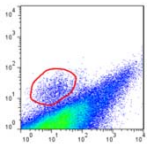


Single Cell Genomics

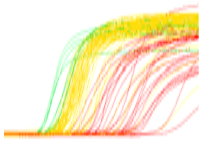
Sample



Single cell isolation

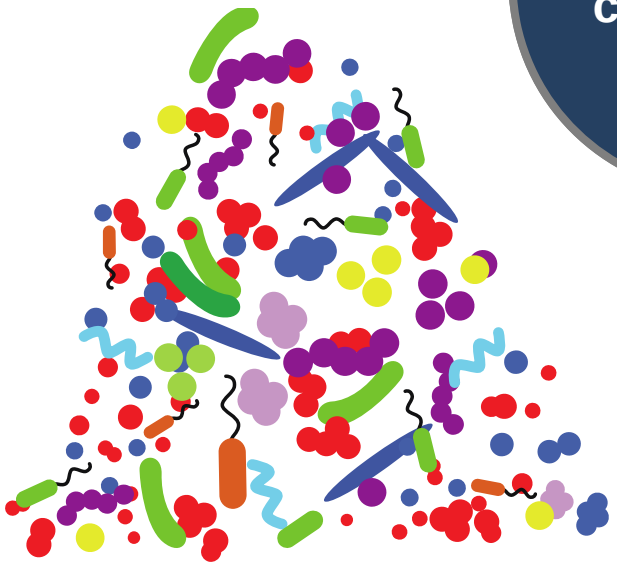


Whole genome amplification

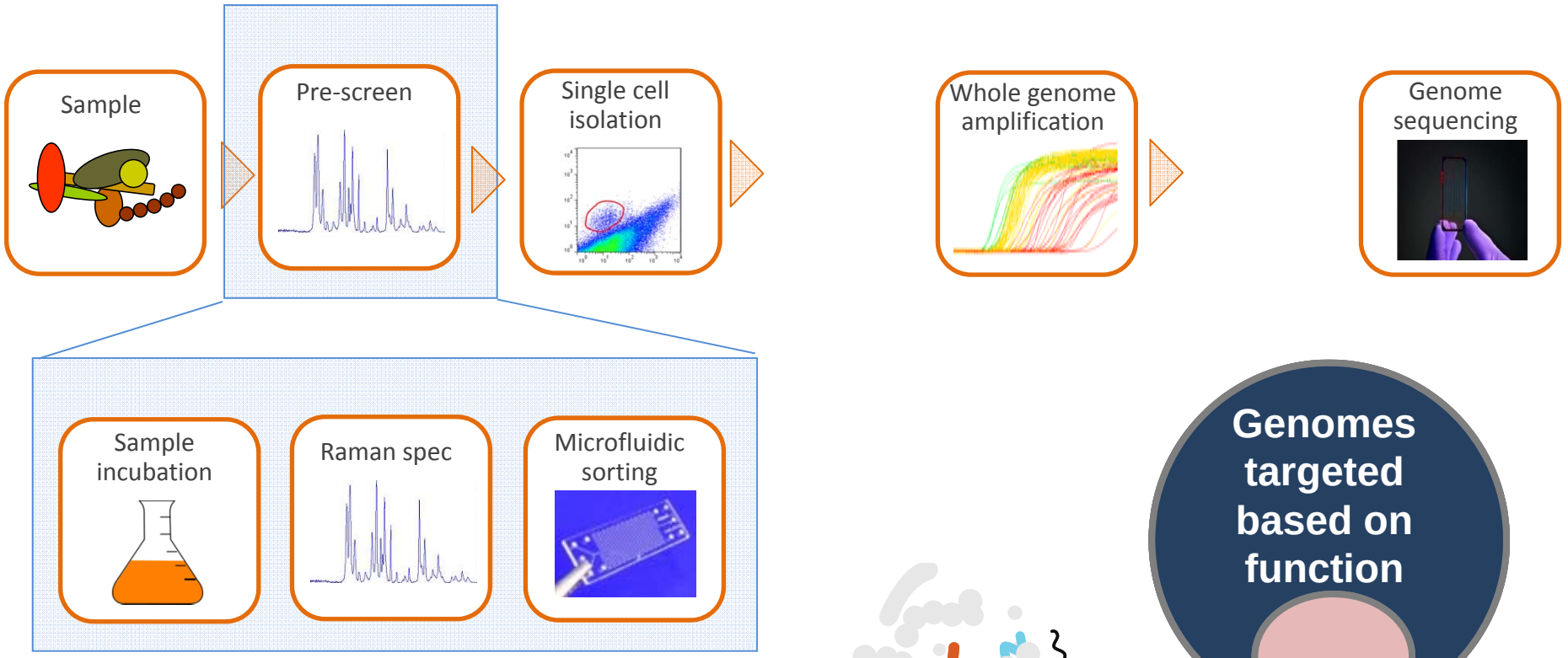


Genome sequencing

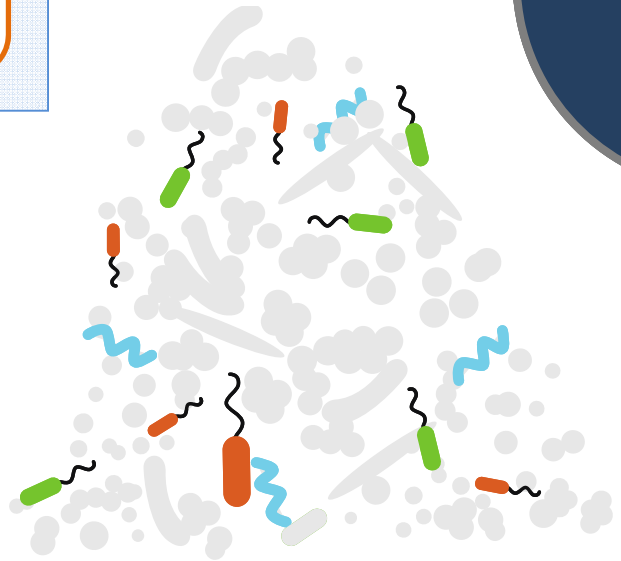
Individual genomes from total community



Function-Driven Genomics



Genomes targeted based on function



U.S. DEPARTMENT OF ENERGY

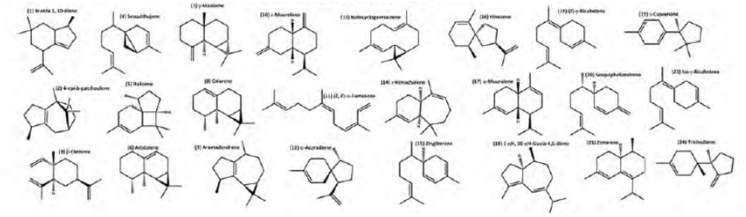
Sequence to Function

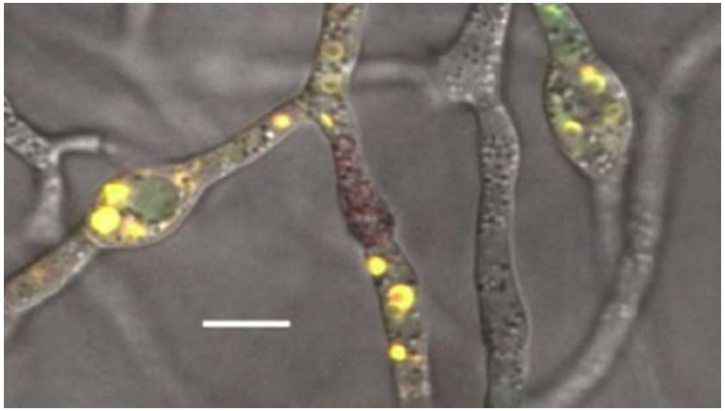
JGI-EMSL Joint User Program
Synergistically Link JGI and EMSL Capabilities

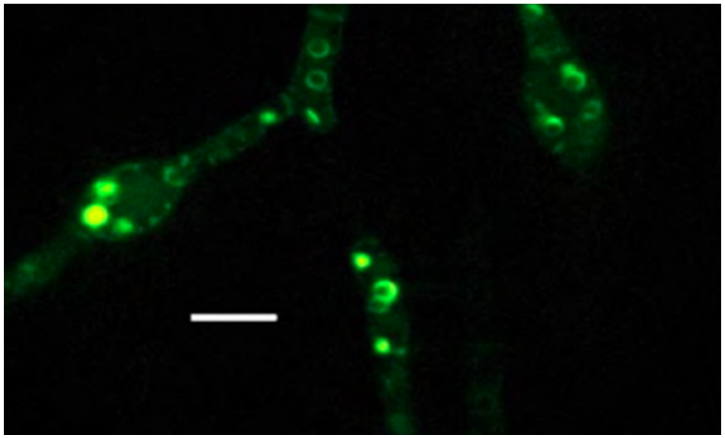
- ***30 applications 8 proposals approved***

UM/USDA

Terpinoid Biosynthesis in the Fungus *Fusarium*



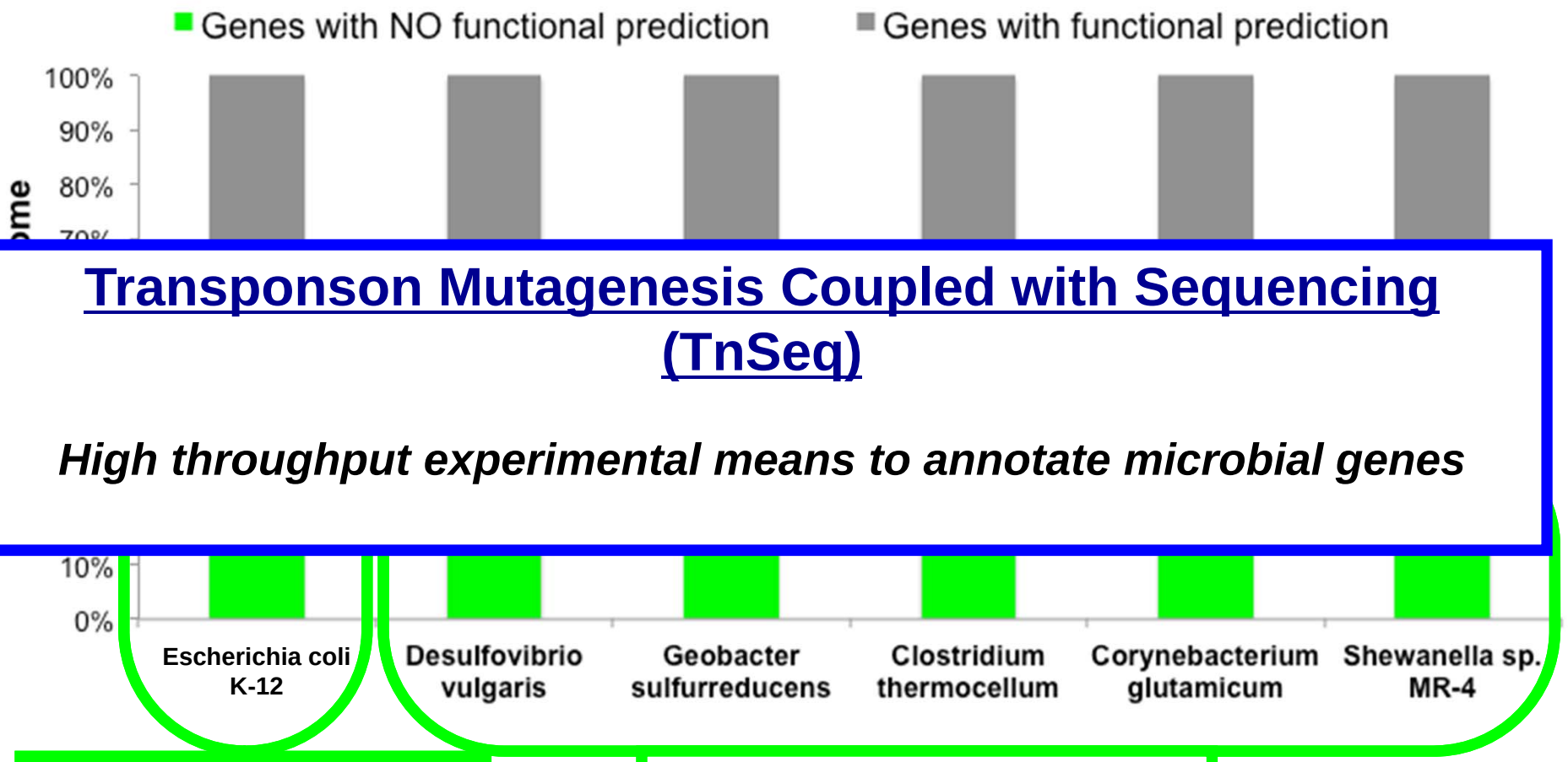




- ***Fusarium graminearum***
  - ***makes many terpinoids***
  - ***synthesis in organelles***
- ***EMSL***
  - ***organelle sorting***
  - ***proteomics & metabolomics***
- ***JGI***
  - ***genomics***
  - ***transcriptomics***

Sequence to Function

Annotation of Microbial Genomes



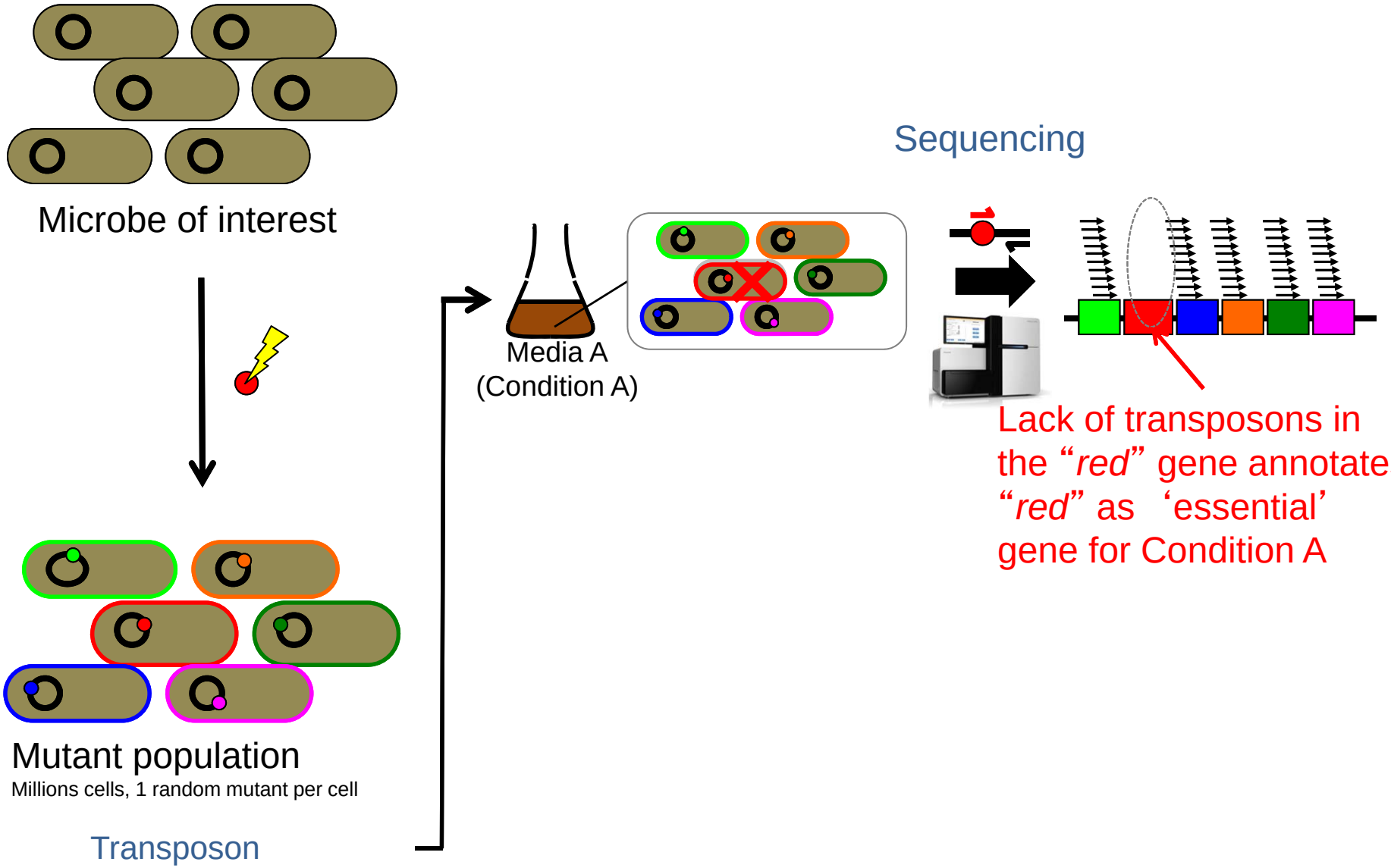
Transponson Mutagenesis Coupled with Sequencing (TnSeq)

High throughput experimental means to annotate microbial genes

15% *E. coli* genes have no known function

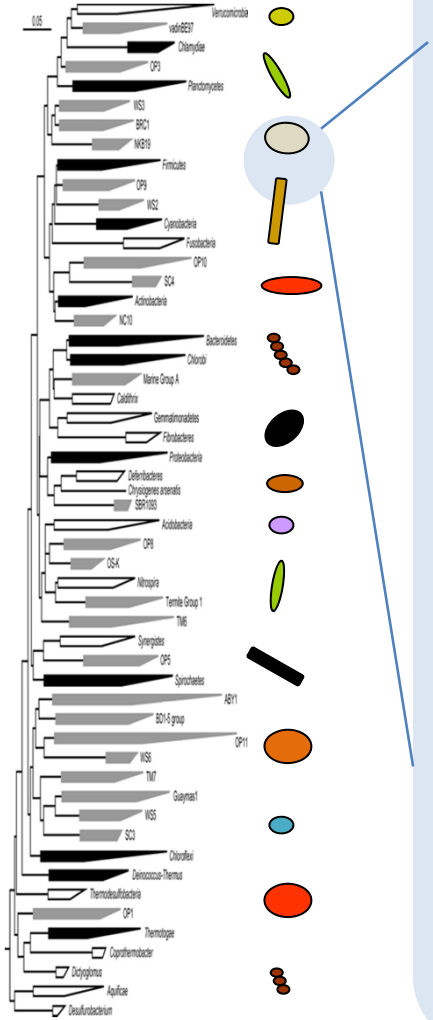
25-40% genes in JGI genomes have no known function

Transposon Mediated Mutagenesis – Sequencing (TnSeq)

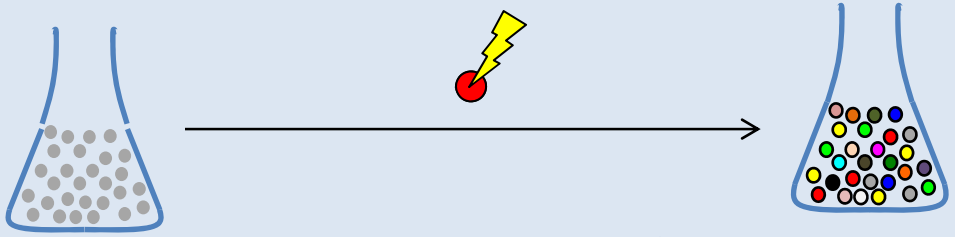


Functional Encyclopedia of Bacteria and Archaea (FEBA)

50 phylogenetically diverse bacteria

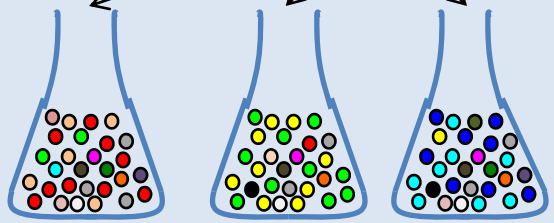


For each bacteria:



1. Culture

2. Transposon (Tn) mutant population

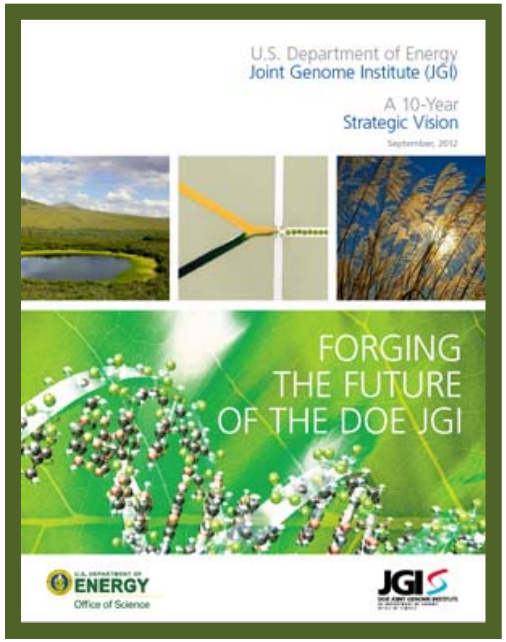


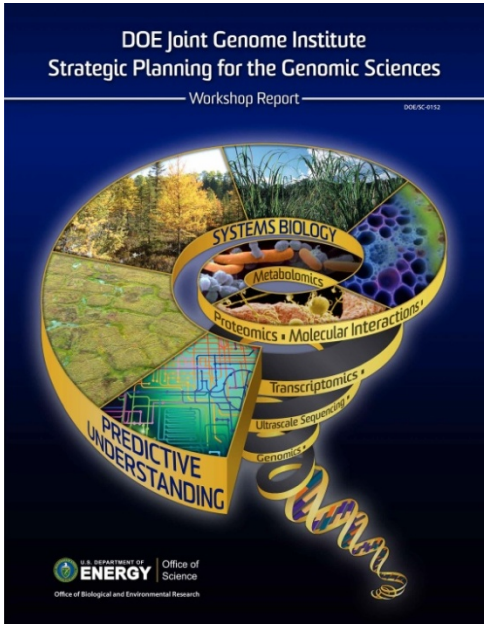
Carbon sources,
Nitrogen sources,
Small molecule stresses,
pH, temp …etc

3. Selection under 50 conditions

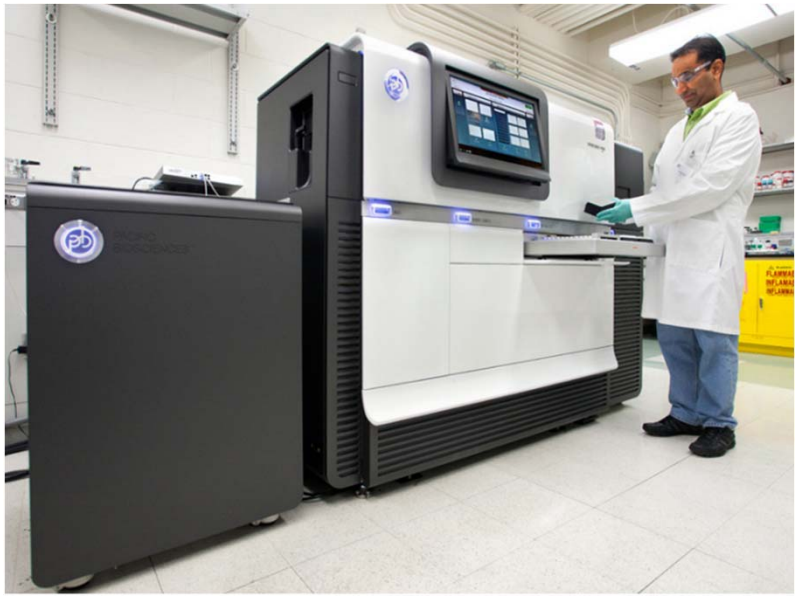
Large scale annotation of microbial genes

State of the Art DNA Sequencing





State of the Art in DNA Sequencing

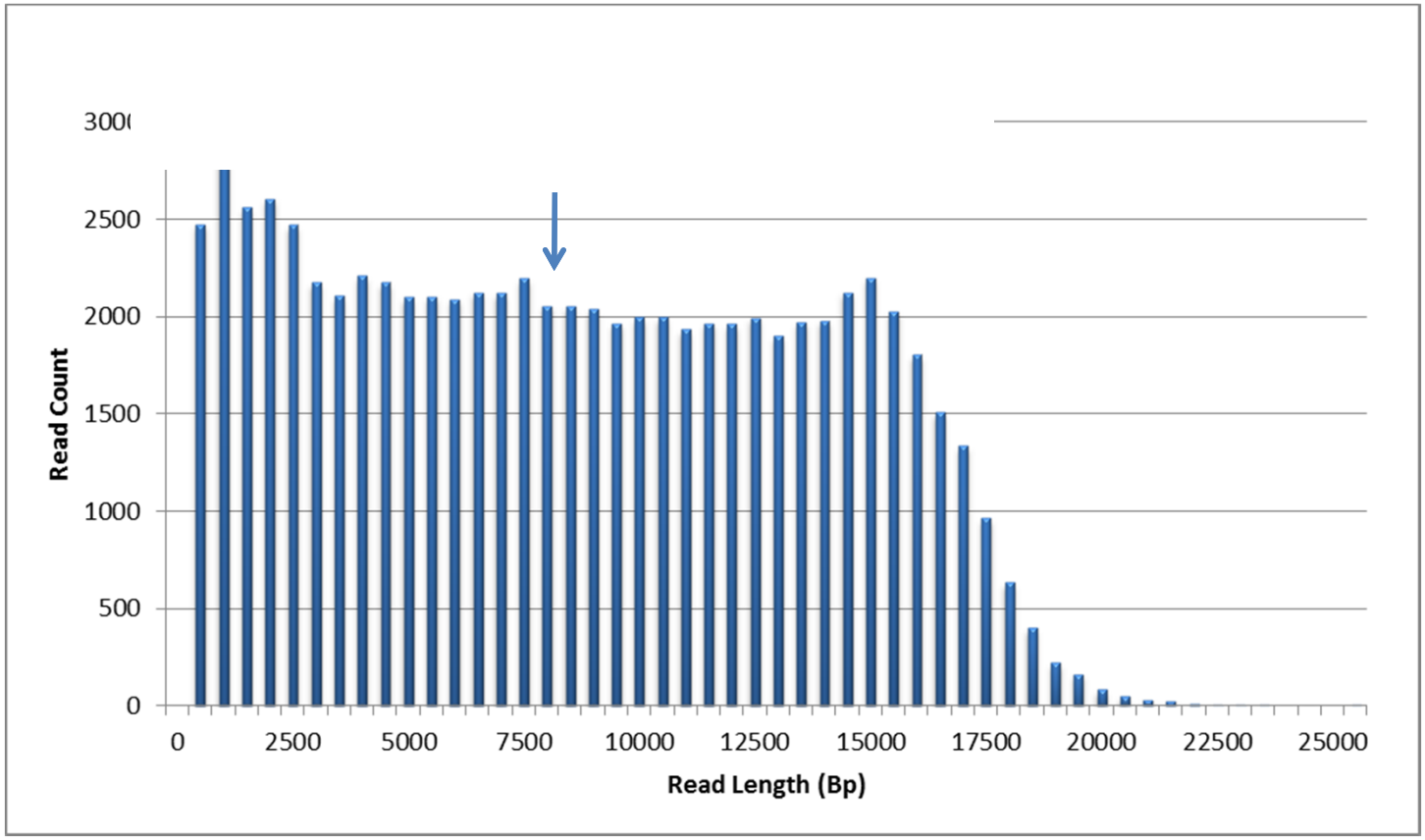


Pacific Biosystems DNA Sequencer

-Extremely Long Read Lengths

-Detects Methylated Bases

PacBio Read Lengths



PacBio Metrics	
Base Count (Mb)	615
Mean Read Length (Bp)	8,346
Max Read Length (Bp)	25,129
Average Quality	0.84

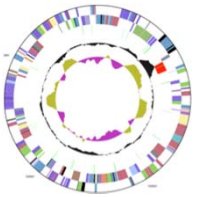
(Illumina Read Length 200 bp)

Finished Microbial Genomes

Historic timeline of JGI sequencing of bacteria and archaea:

2002-2006	2006-2008	2008-2011	2011	2011-2013
Sanger	Sanger/454	454/Illumina	Illumina	Illumina
$50k	$35k	$10k	$1.5-3k	$5k
49 contigs*	22 contigs	44 contigs	69 contigs	6 contigs

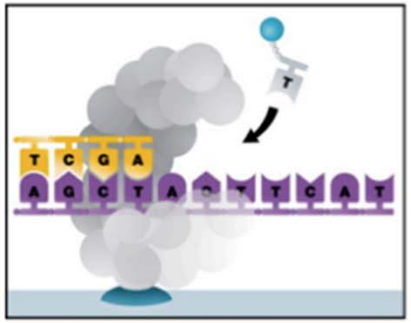
Manual Finishing ($35k/genome) | **No Finishing**

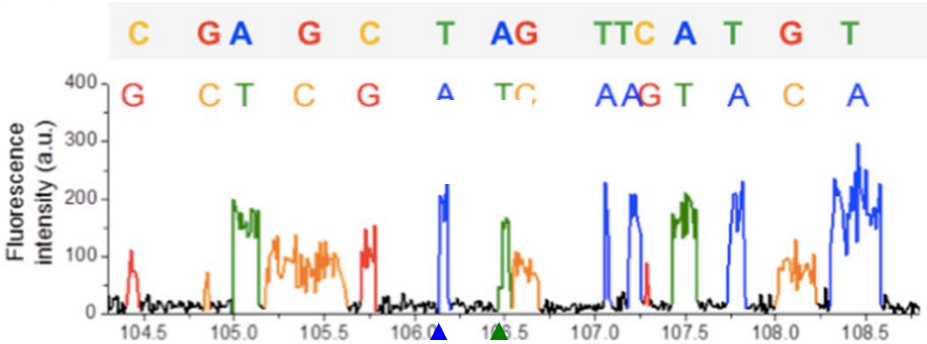


$45-85k

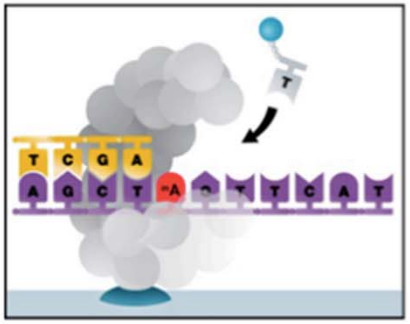
Direct detection of DNA methylation via PacBio sequencing

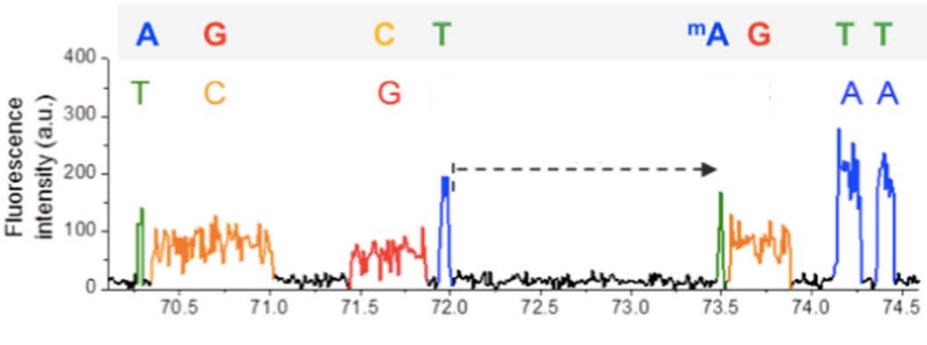
Unmodified DNA template





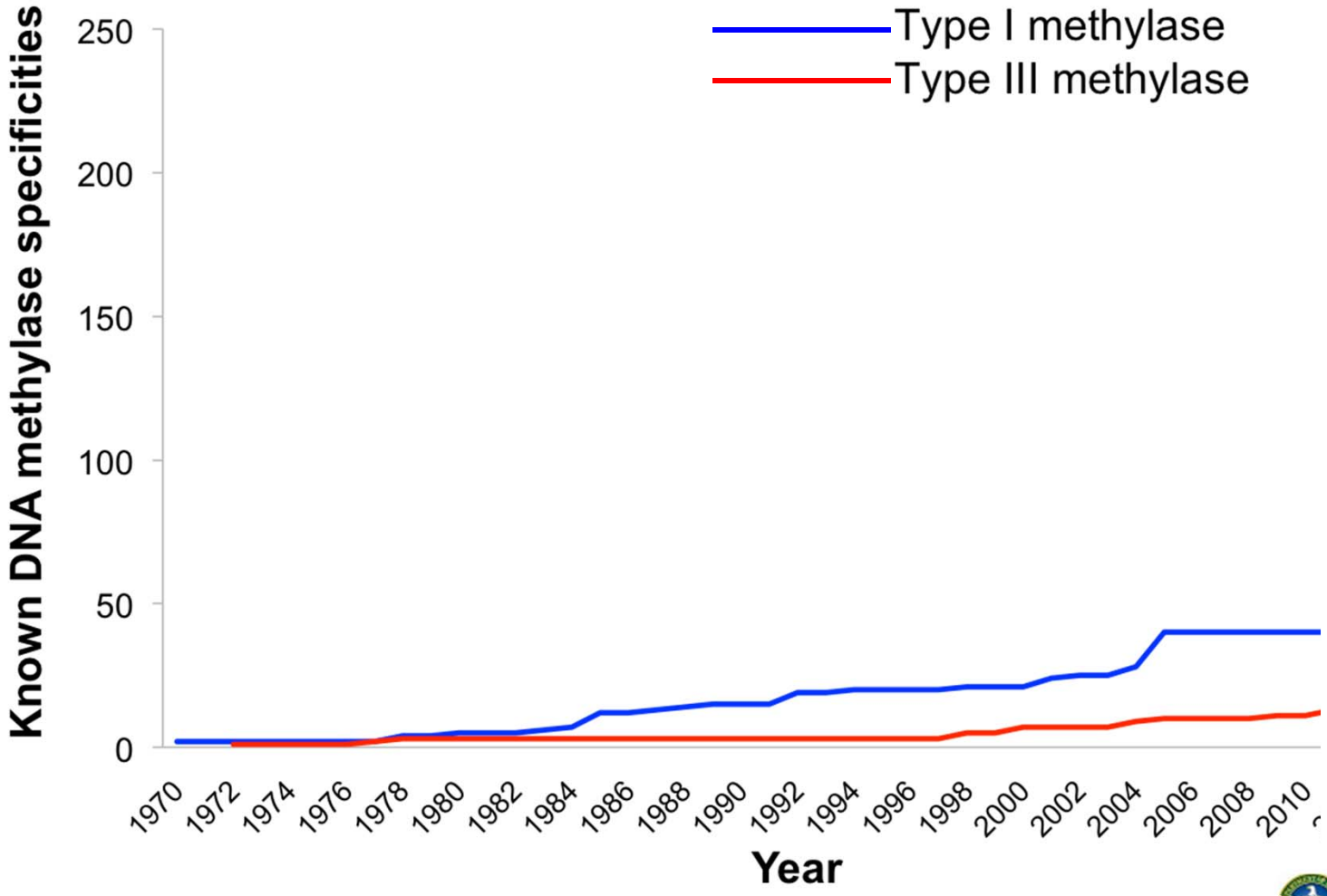
Modified DNA template



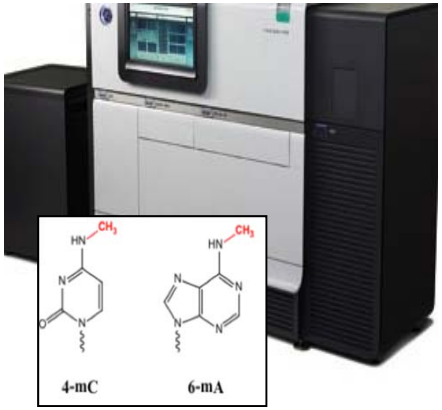


Delay in base incorporation opposite methyl adenine.

New insights into the specificities of DNA methylases



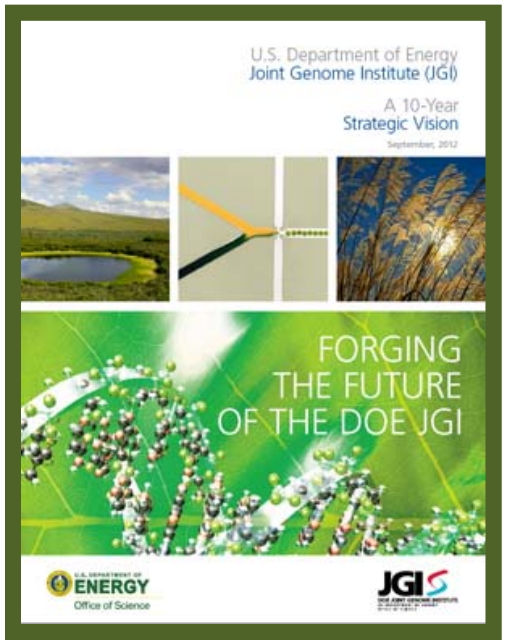
The Epigenomic Landscape of Bacteria

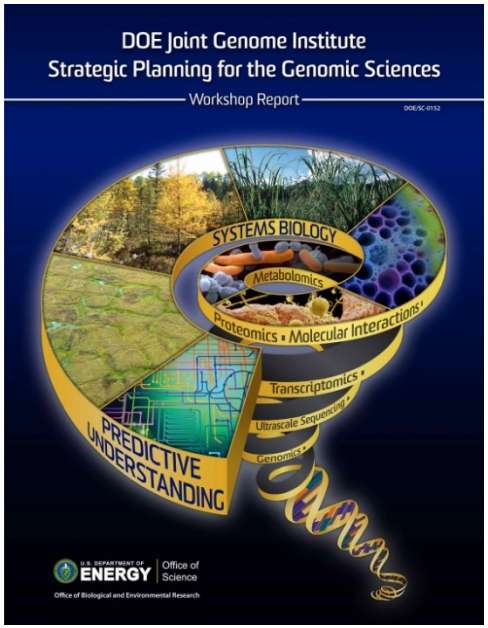


>200 bacterial DNA base modification datasets on the PacBio platform (>90% of genomes with modifications)

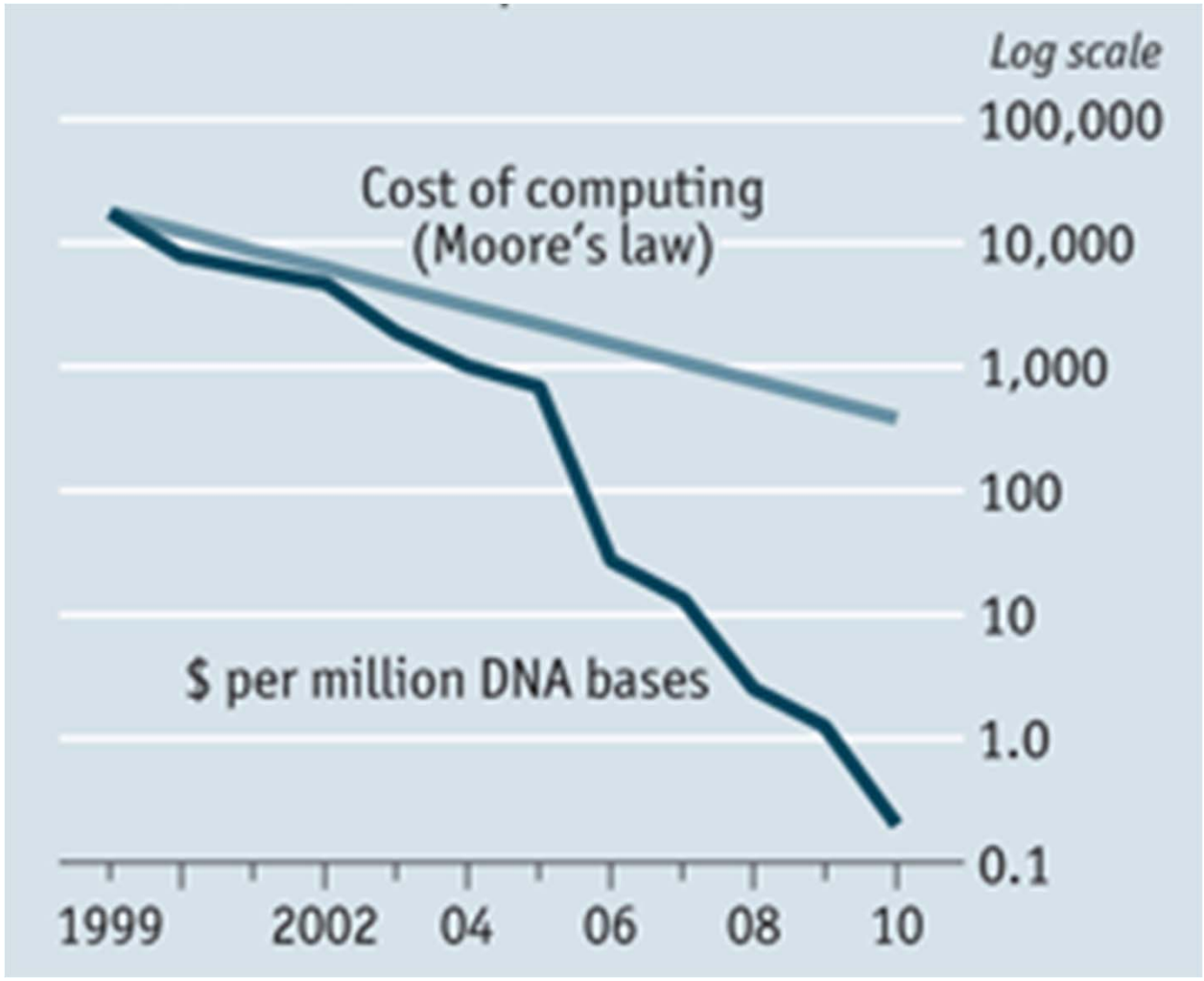
JGI offering capability to users to explore the functions of DNA base modification
(*CSP call*)

High-performance computing infrastructure for genomic big data challenge





HPC Infrastructure for Genomic Big Data Challenge



Major Effort Underway Between the JGI & National Energy Research Supercomputing Center (NERSC)

To tackle big genomic data computational challenges

Goal: **Move Appropriate JGI workflows to HPC**

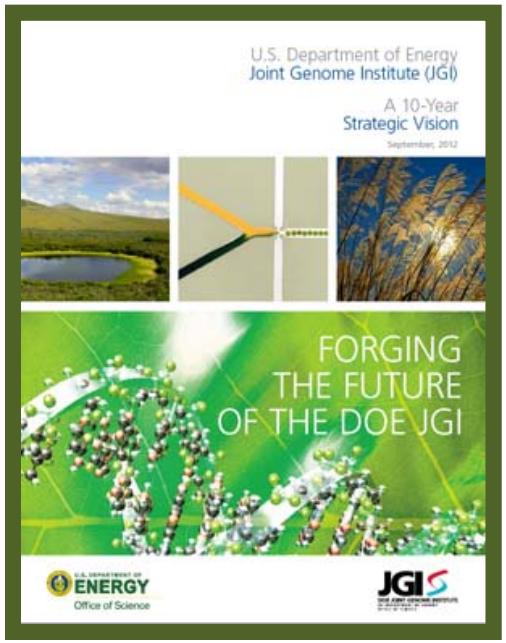
.

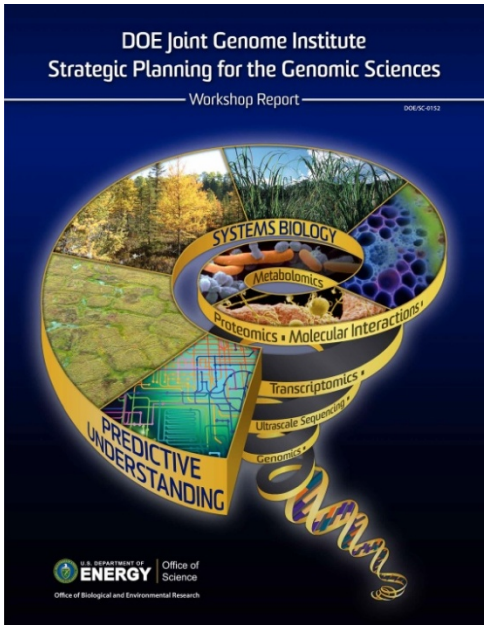
1: All against all homology searches

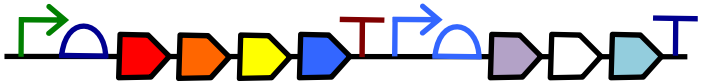
2: Development of a HPC metagenome assembly pipeline

3: Optimization of data quality control pipelines for HPC

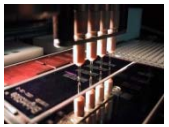
Expand capacity to synthesize ("write") DNA to study genes, pathways and manipulate genomes

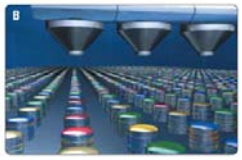






Pathway design

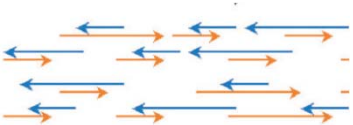




Sequence databases

Design Software

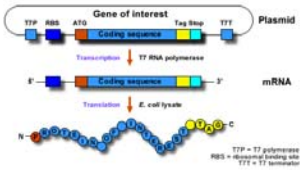
Oligonucleotides

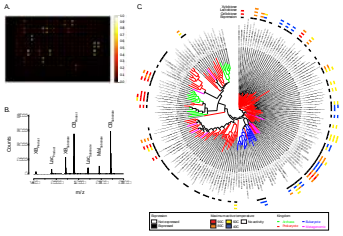


Liquid handling systems

Assembly

Cloning and picking





Sequence validation

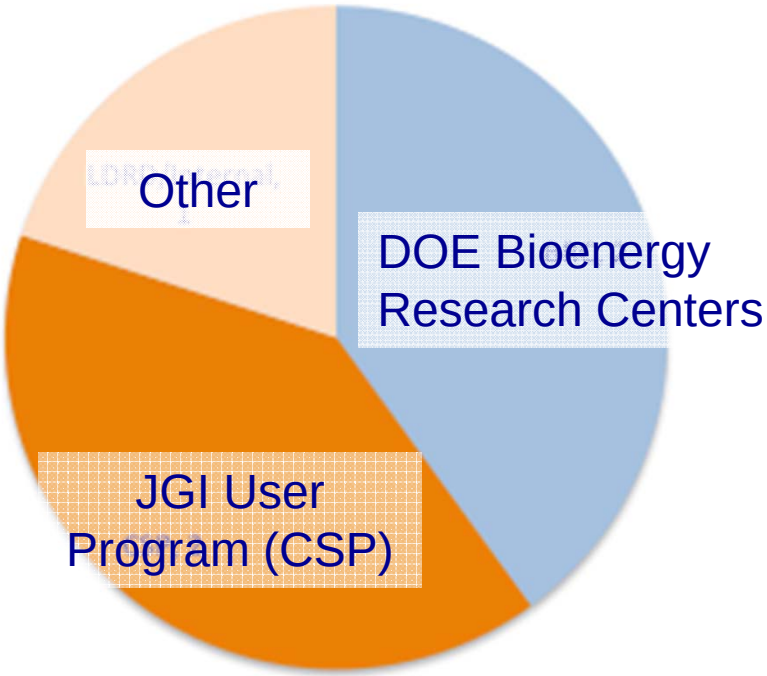
Gene expression

Functional studies

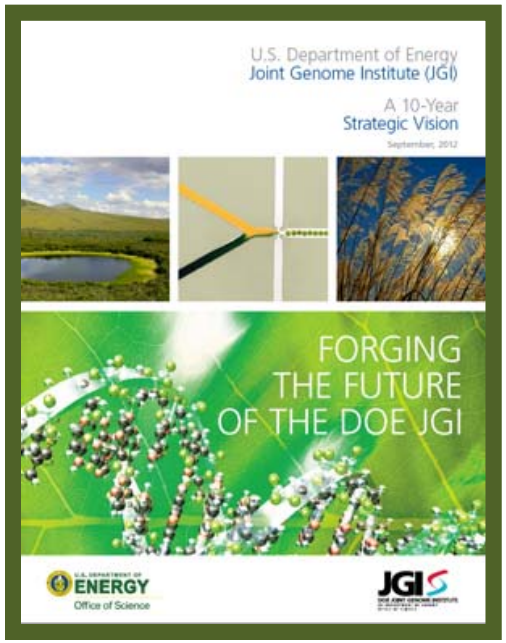
Synthesis Projects

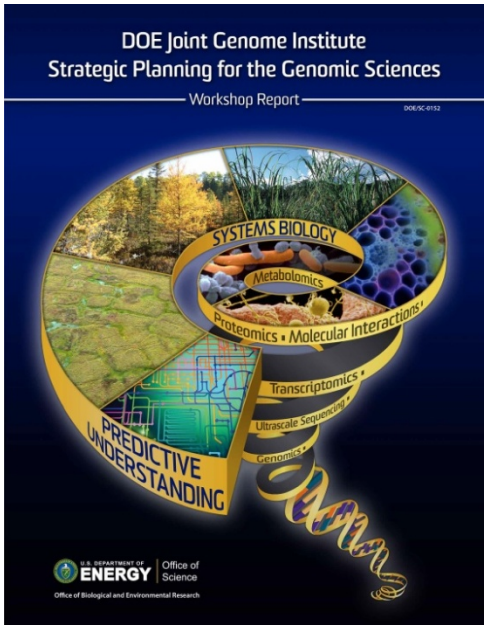
FY14

30 User Projects 4.0 Mb



Encourage and Support the Organization and Building of User Communities Around Specific Topics

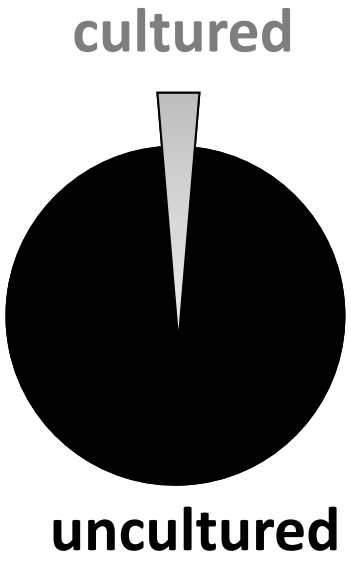


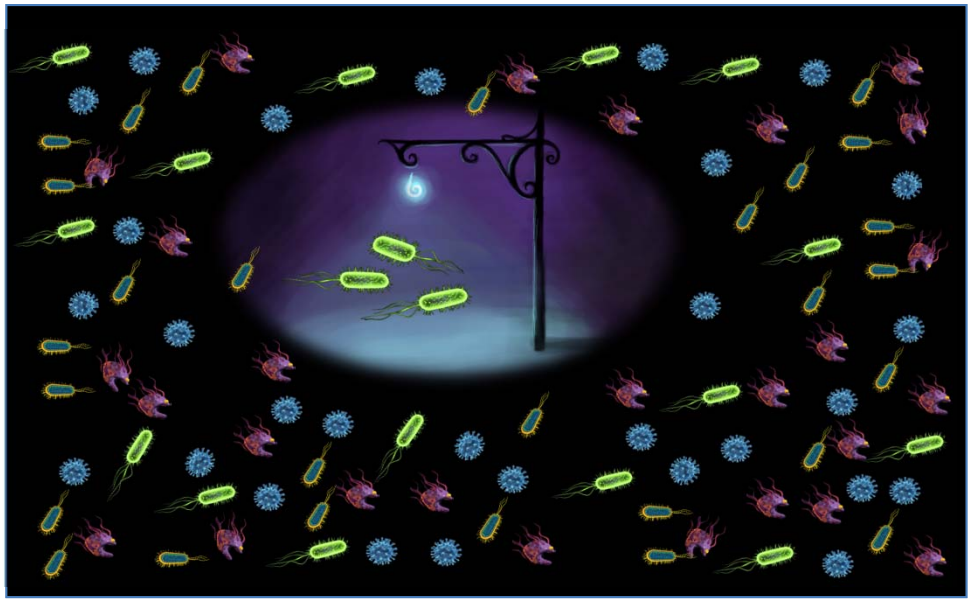


Organize/Build Communities

Organize/Build Communities

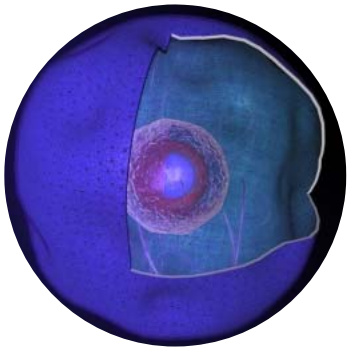
Microbial Dark Matter Project (MDM)





Prior to the DNA Era - 2 Domains

life

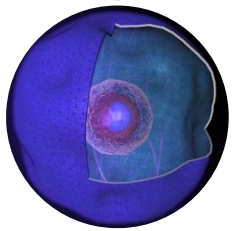


Eukarya

Bacteria

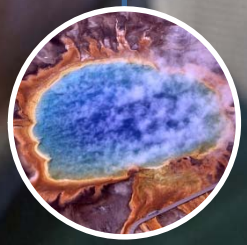
Carl Woese
1928-2012

3 Domains



Eukarya

Bacteria



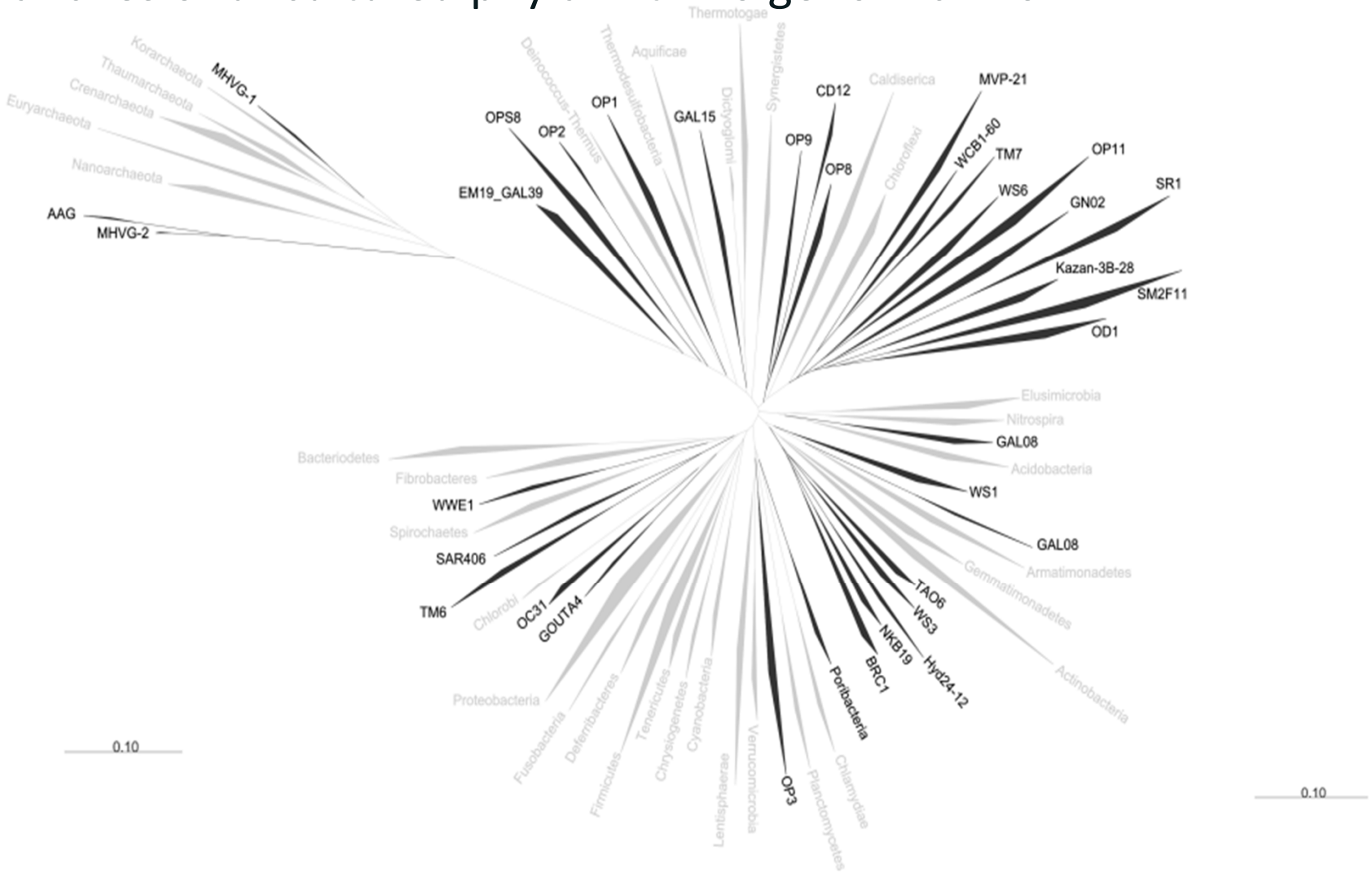
Archaea

life

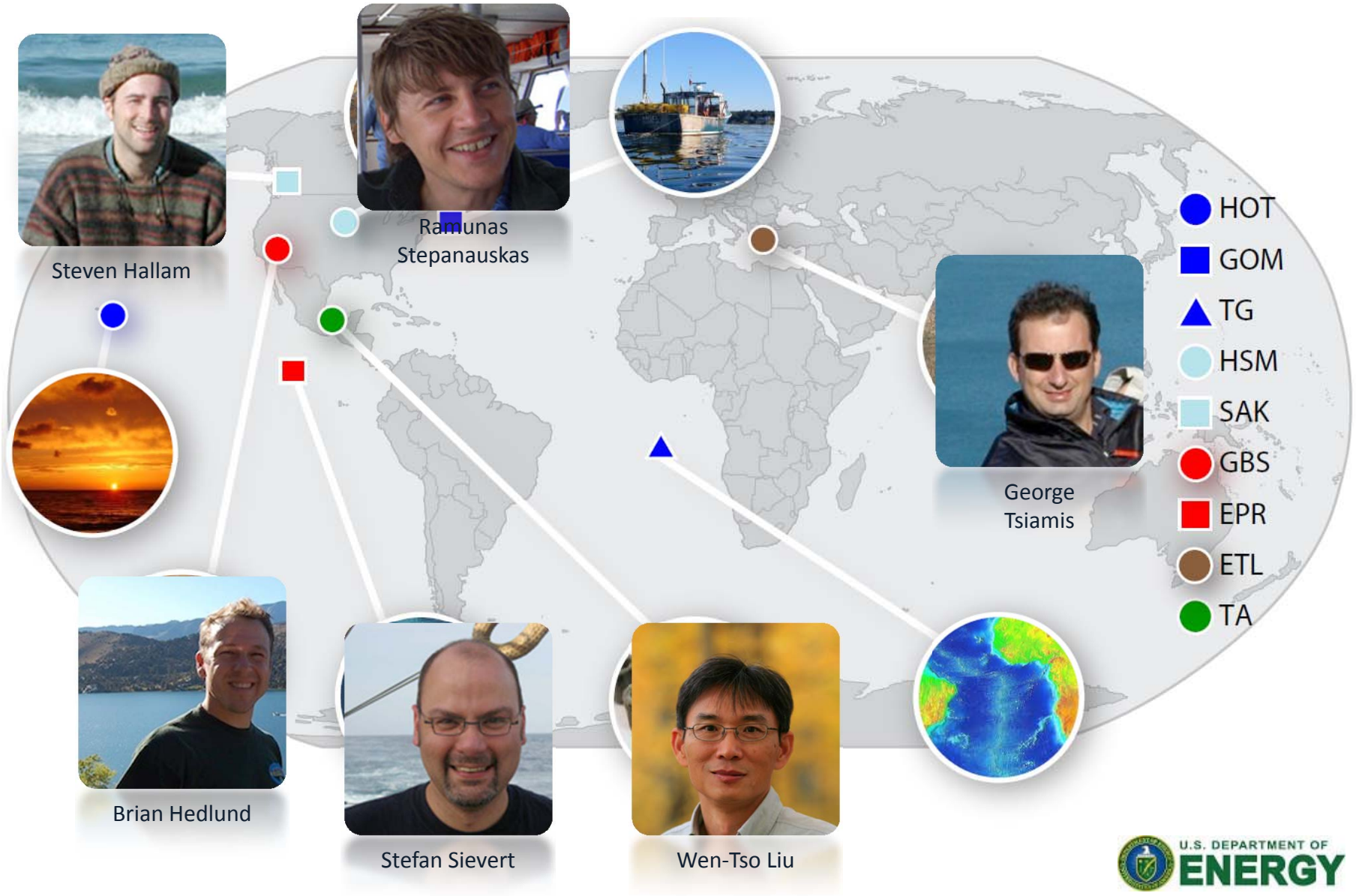
Carl Woese
1928-2012

16S rRNA Tree of all Known Phyla

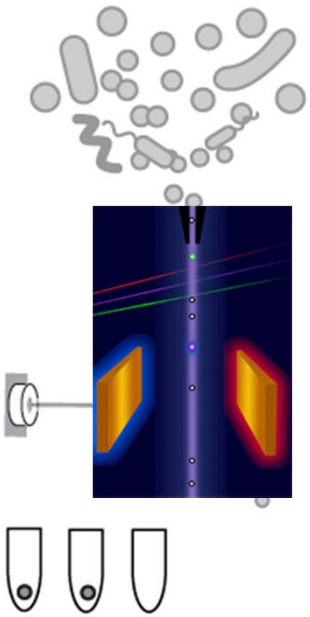
Branches of uncultured phyla with no genomic info



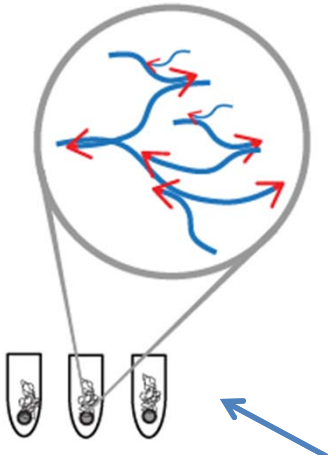
Samples from 9 sites were selected from user projects



Targeted our Sequencing to “Dark Matter” Genomes

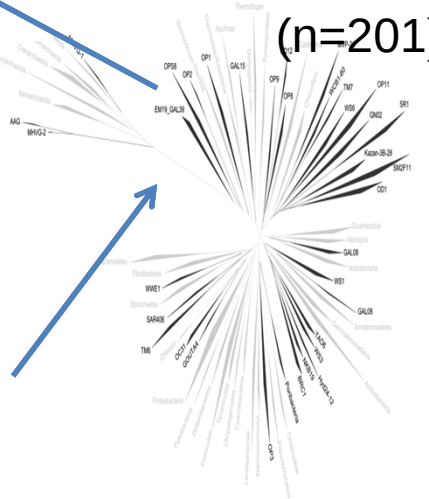


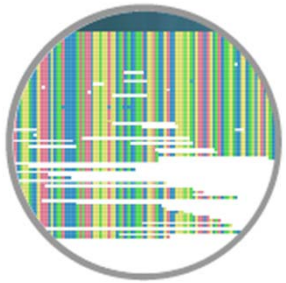
isolation of single cells **(n~10,000)**



whole genome amplification **(n=3,300)**

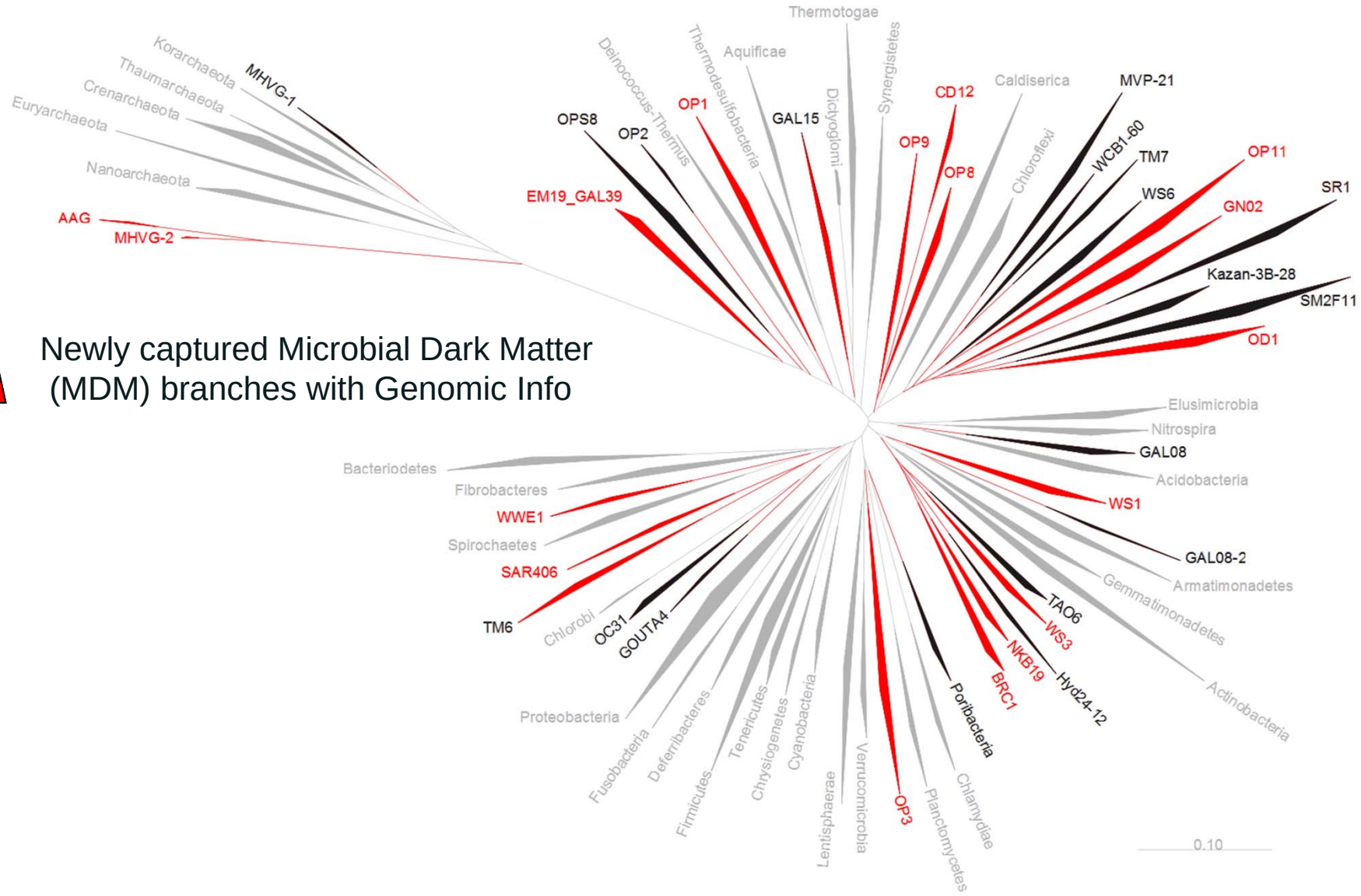
16S identification





genome sequence and assembly (n=201)

Shedding light on the tree of life

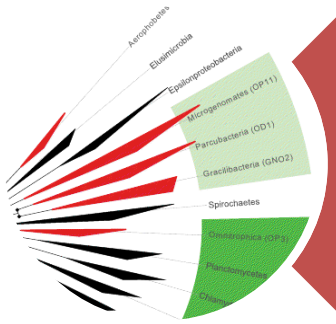


Newly captured Microbial Dark Matter (MDM) branches with Genomic Info

16S rRNA tree of microbial phyla

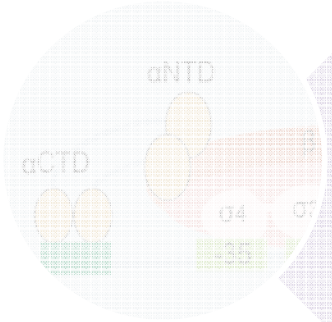
Insights from “Dark Matter Genomes”

Nature 2013 “Insights into phylogeny and coding potential of microbial dark matter”



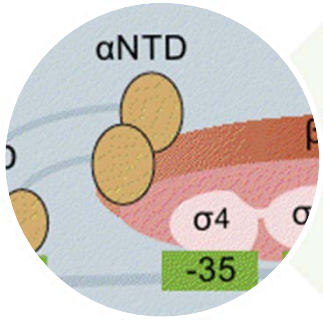
Reshaping the Tree of Life

Assigning of metagenomic data to an organism

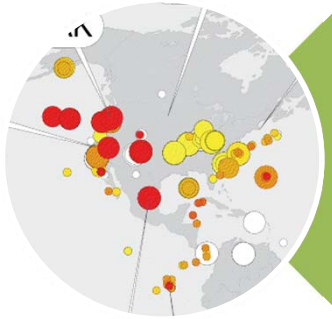


Discoveries

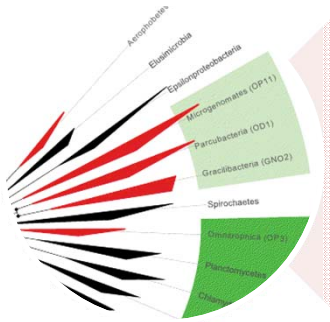
Insights from “Dark Matter Genomes”



Functional novelty

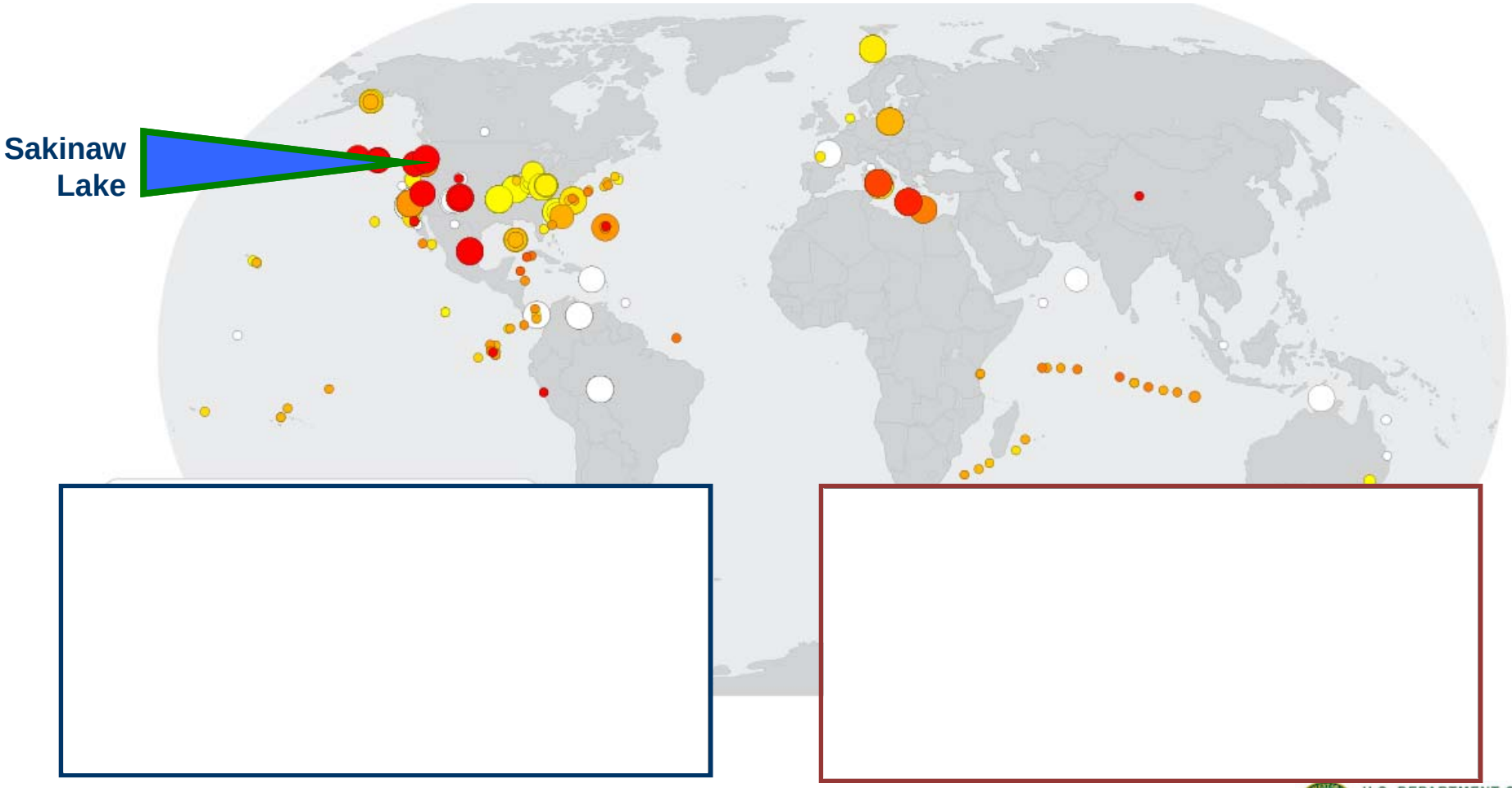


Assigning of metagenomic data to an organism

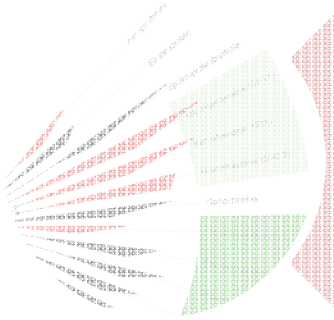


Reshaping the tree of life

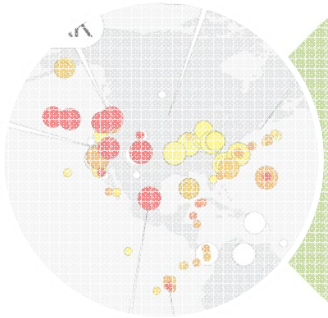
Assigning Metagenomic Reads to an Organism with the Microbial Dark Matter Genomes



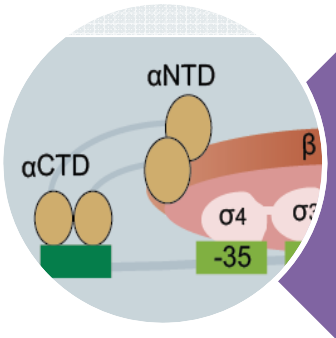
Insights from “Dark Matter Genomes”



Reshaping the Tree of Life

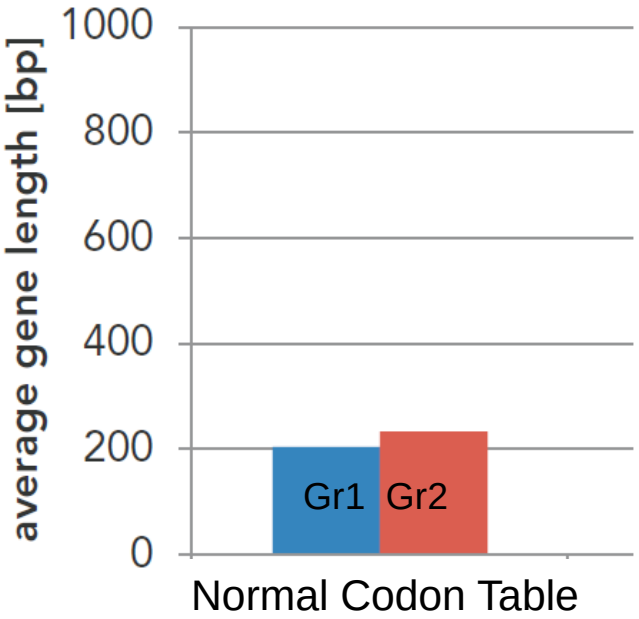


Assigning of metagenomic data to an organism



Discoveries

Two Gracillibacteria had Remarkably Small Genes



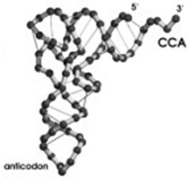
The Canonical Genetic Code

tRNAs

64 Codons (61 Amino Acids & 3 Stop Codons)

UUU	F	UCU	S	UAU	Y	UGU	C
UUC	F	UCC	S	UAC	Y	UGC	C
UUA	L	UCA	S	UAA	✸	UGA	✸
UUG	L	UCG	S	UAG	✸	UGG	W
CUU	L	CCU	P	CAU	H	CGU	R
CUC	L	CCC	P	CAC	H	CGC	R
CUA	L	CCA	P	CAA	Q	CGA	R
CUG	L	CCG	P	CAG	Q	CGG	R
AUU	I	ACU	T	AAU	N	AGU	S
AUC	I	ACC	T	AAC	N	AGC	S
AUA	I	ACA	T	AAA	K	AGA	R
AUG	M	ACG	T	AAG	K	AGG	R
GUU	V	GCU	A	GAU	D	GGU	G
GUC	V	GCC	A	GAC	D	GGC	G
GUA	V	GCA	A	GAA	E	GGA	G
GUG	V	GCG	A	GAG	E	GGG	G

tRNA (UGA)



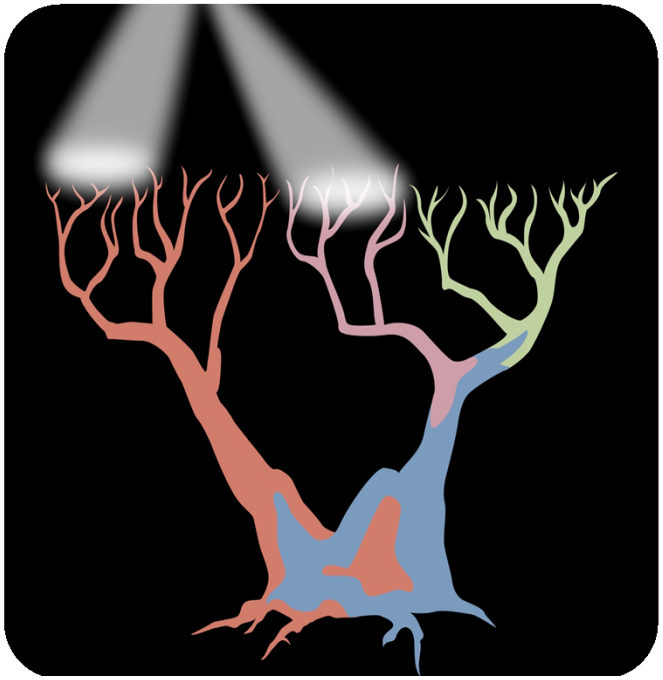
Glycine

Releasing Factors 2

Microbial Dark Matter Project

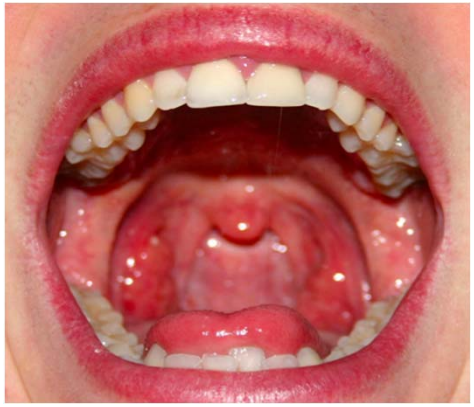
Expands the genomic representation of the tree of life and a systematic step forward to an unbiased view of microbial evolution on our planet

Microbial Dark Matter Project



SUGGESTS THAT THE CANONICAL GENETIC CODE MAY NOT BE ALL THAT IS OUT THERE IN THE WILD

Genetic Codes Operating in the Wild



Analysis of metagenomic and phage data sets for codon reassignment

~1100 individual sampling sites

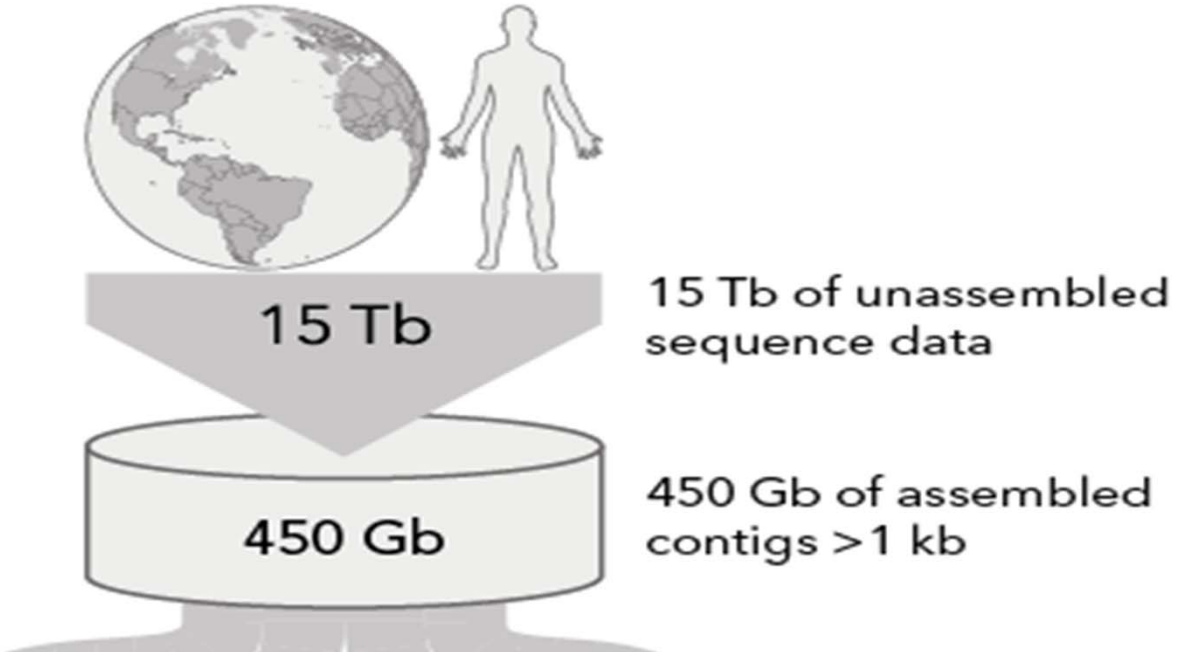
>15 Tb of metagenomic data and plus phage data

Search for Recoding Events

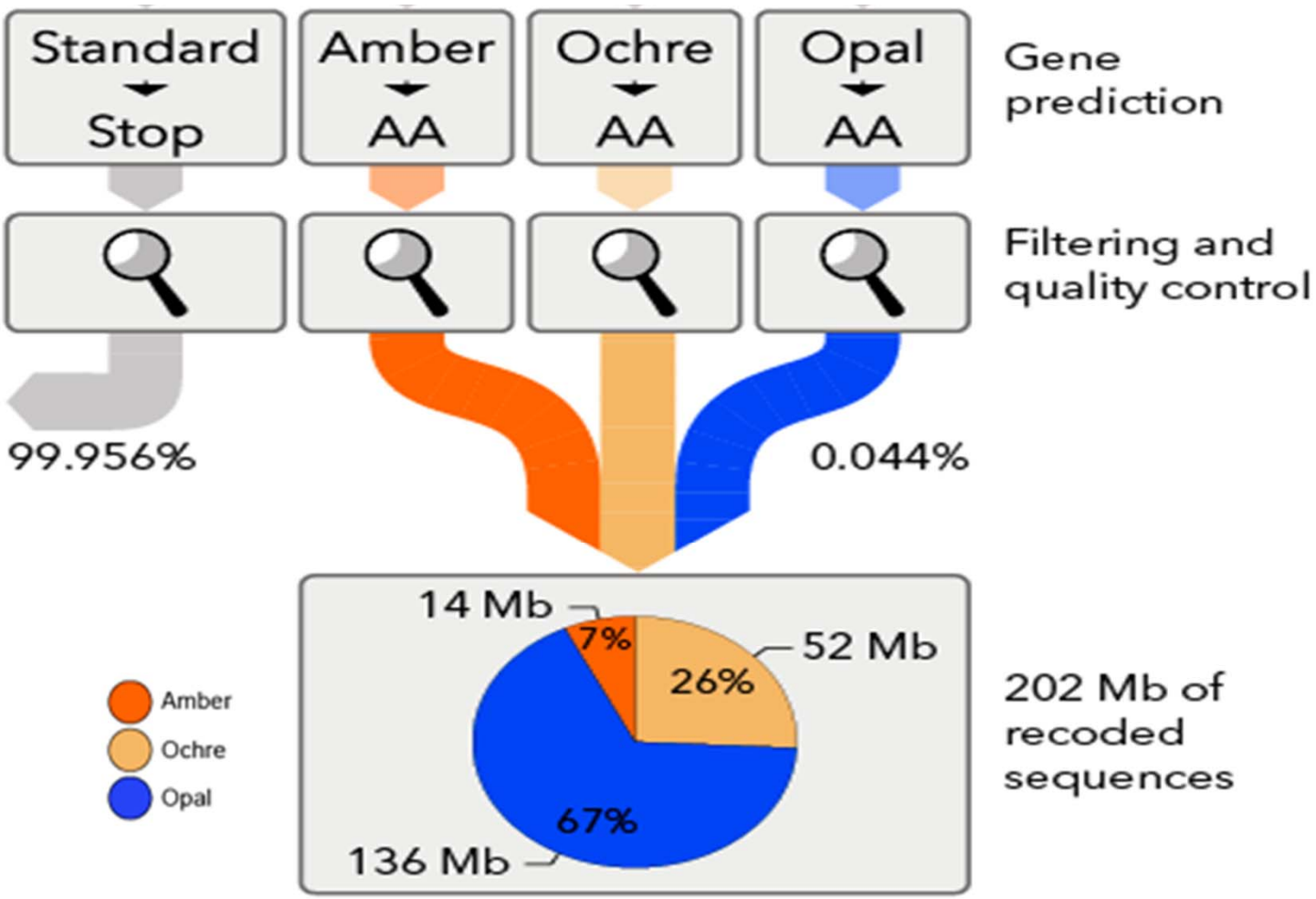
Looked specifically for the most radical and easy to detect conversion of a stop to an amino acid codon

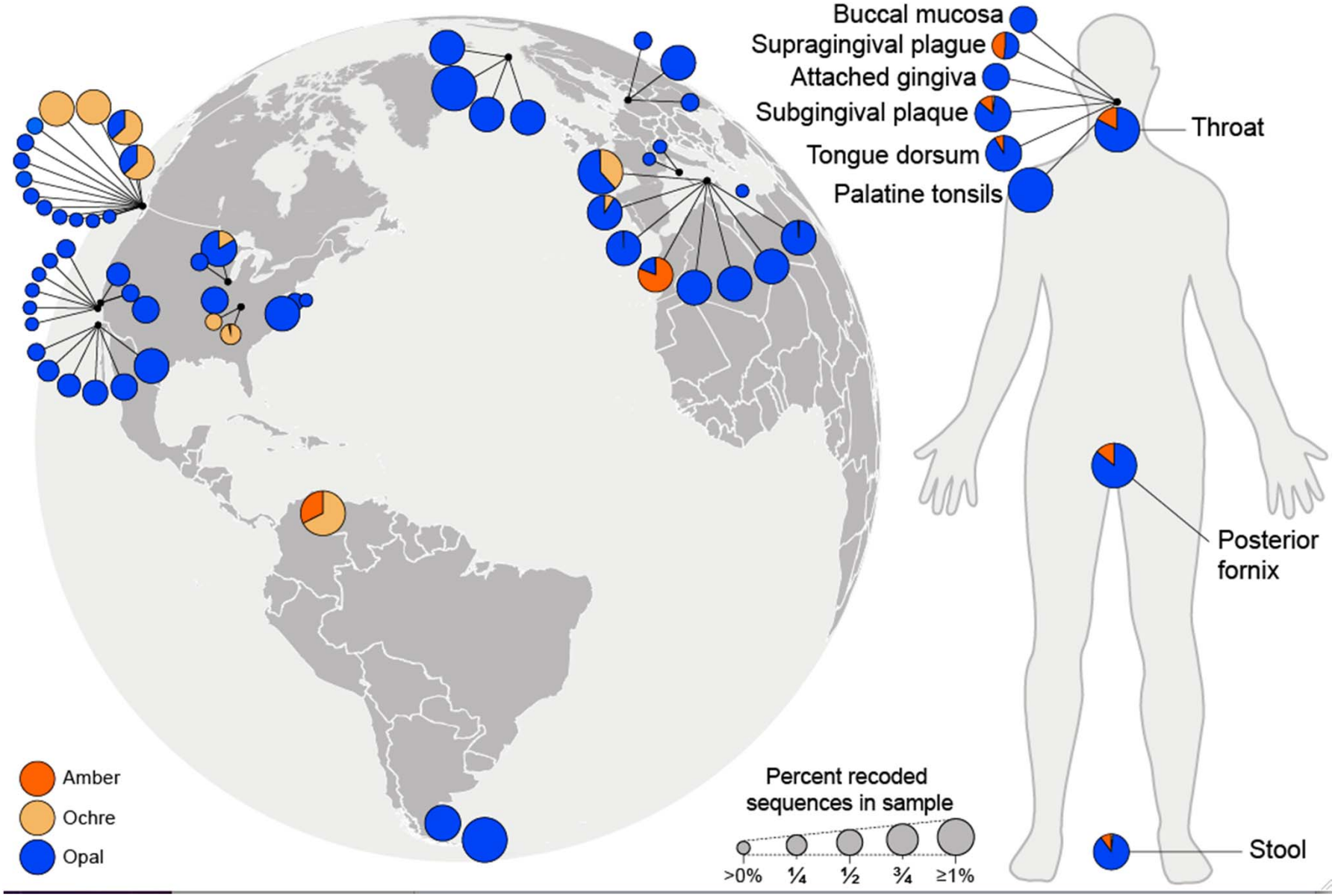
	Opal, Ochre, Amber
Stop codons	UGA, UAA, UAG

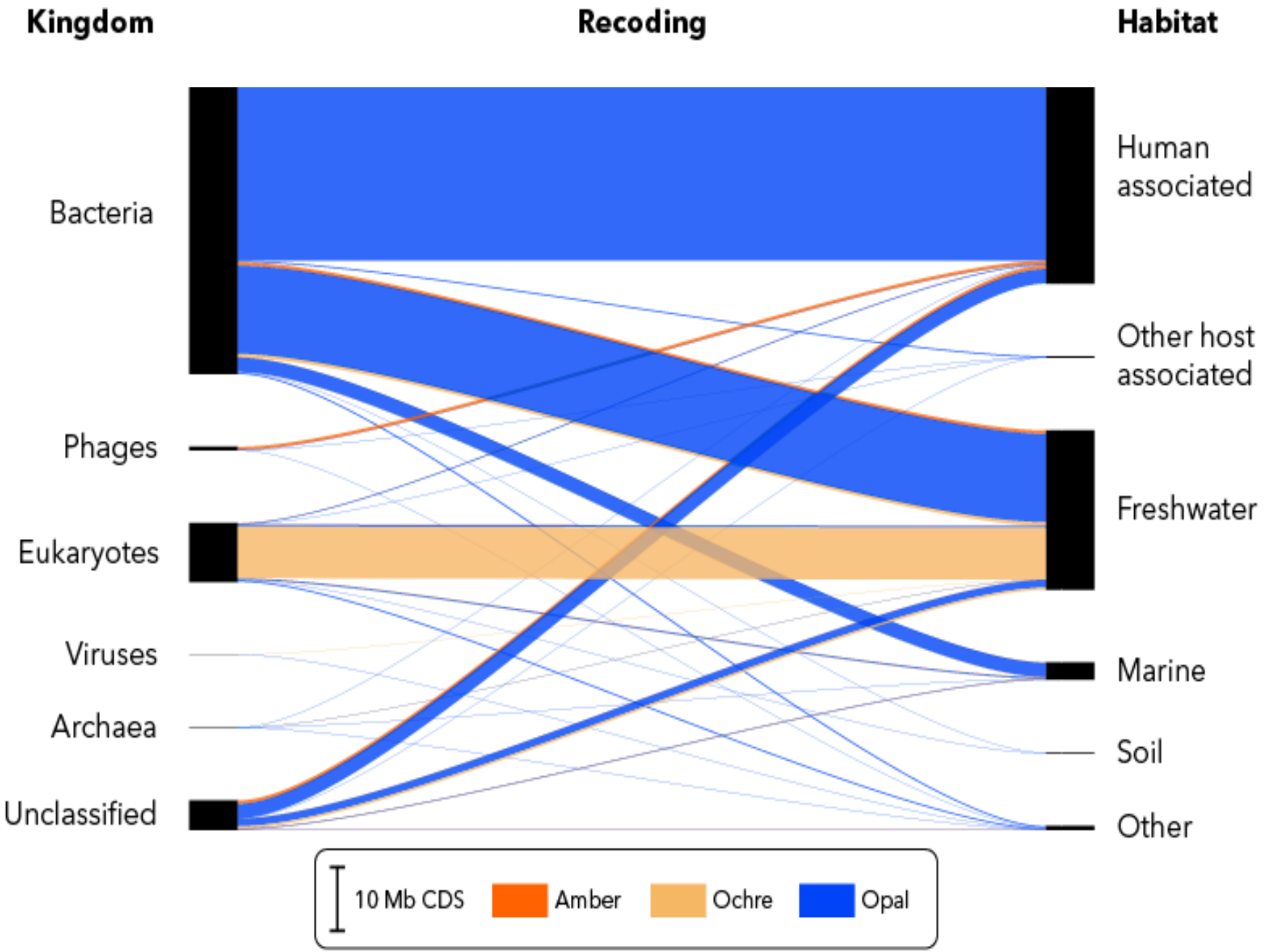
Search for Recoding Events



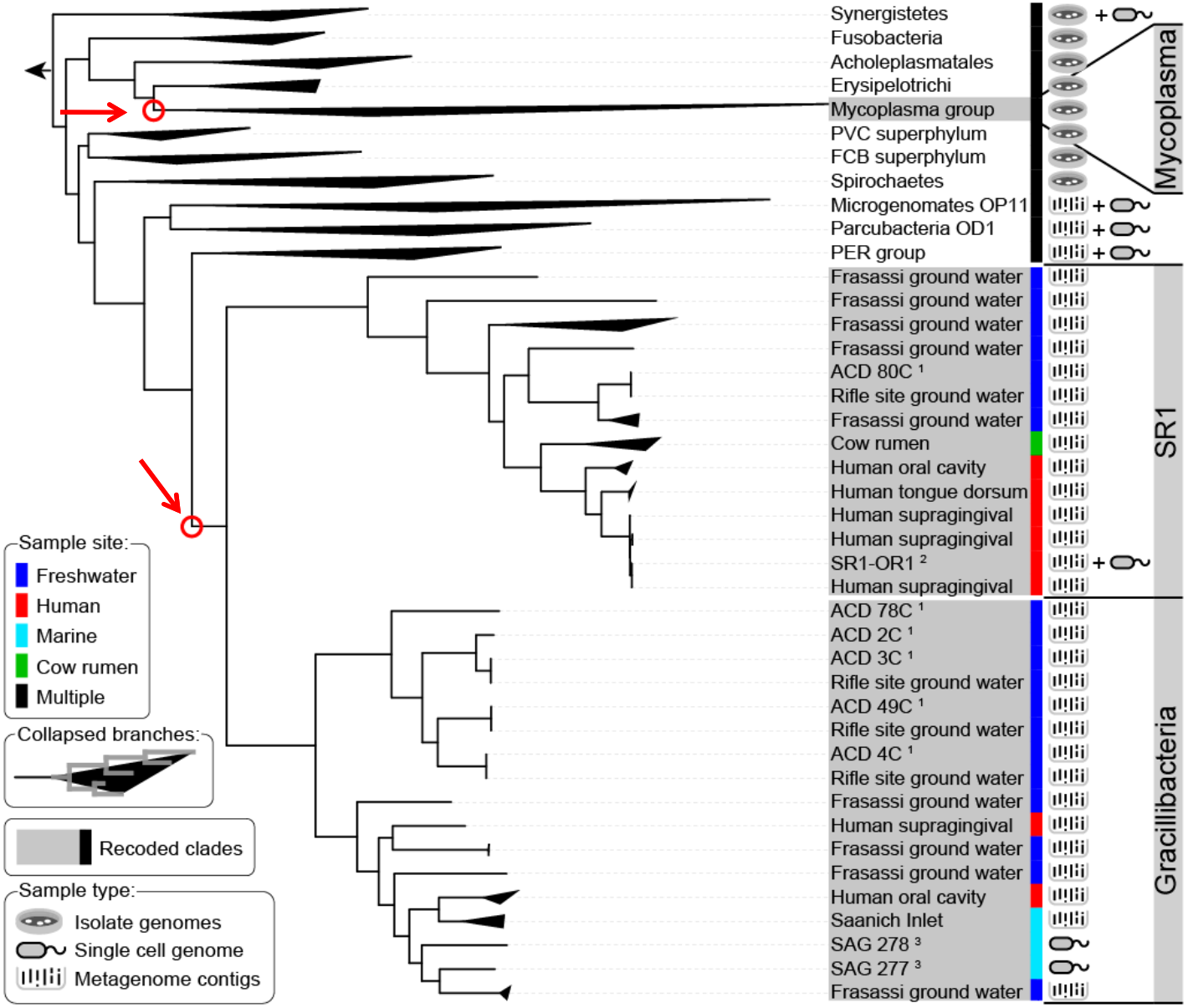
Search for Recoding Events




Buccal mucosa
Supragingival plague
Attached gingiva
Subgingival plaque
Tongue dorsum
Palatine tonsils
Throat
Posterior fornix
Stool
Amber
Ochre
Opal
Percent recoded sequences in sample
>0%
¼
½
¾
≥1%


Kingdom
Recoding
Habitat
Bacteria
Phages
Eukaryotes
Viruses
Archaea
Unclassified
Human associated
Other host associated
Freshwater
Marine
Soil
Other
10 Mb CDS
Amber
Ochre
Opal

Phylogenetic tree of Prokaryotic Recoding



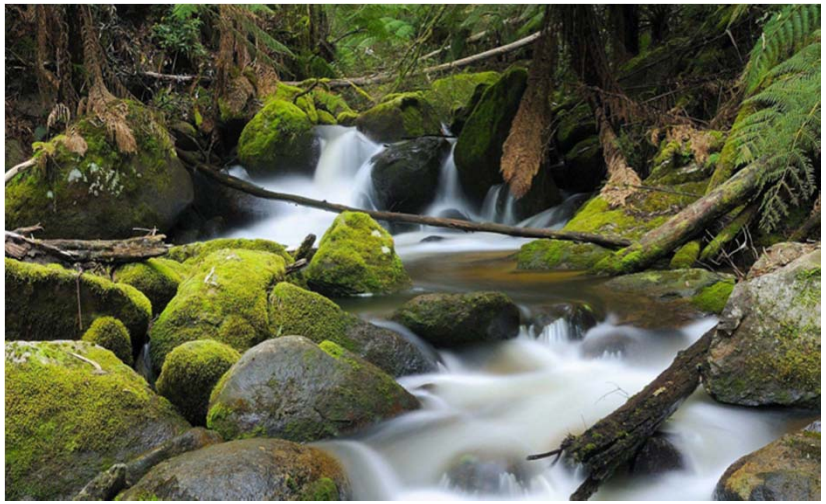
Tryptophan

Glycine

Summary

- **Our assumptions about the code largely based on studies of cultured microbes.**

- **Metagenomics offers a broad unbiased window into codon usage in nature**



With the implementation of the strategic plan initiatives the JGI is evolving as a genome SCIENCE user facility to meet the scientific needs of energy and environmental research going forward

Emerging Technologies Opportunity Program (ETOP)

KBase PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

EMSL

NERSC

ENIGMA

U.S. DEPARTMENT OF ENERGY

JGI
JOINT GENOME INSTITUTE

U.S. DEPARTMENT OF
ENERGY