How can I get the most from my rRNA gene analysis?

- Todd DeSantis

LBL - Ecology Dept.

Developed Greengenes http://greengenes.lbl.gov

Developed the PhyloChip technology

Second Genome, Inc.

Licensed the PhyloChip to provide services outside LBL

Where to read more

- Special Thanks to the Technology Dream Team
  - An improved Greengenes taxonomy for bacteria and archaea with explicit ranks. Daniel McDonald[1], Morgan N. Price[2], Julia Goodrich[1†], Eric P. Nawrocki[3], Todd Z. DeSantis[4], Alexander Probst[4§], Gary L. Andersen[4] Rob Knight[1,5] and Philip Hugenholtz[6*], 2011, ISMEJ, Submitted)
  - The Impact of Classifier Training Sets on Phylogenetic Information From High-Throughput Bacterial 16S rRNA Gene Surveys. (Jeffrey J. Werner[a*], Omry Koren[b*], Philip Hugenholtz[c], Todd Z. DeSantis[d], William A. Walters, J. Gregory Caporaso[e], Largus T. Angenent[a], Rob Knight[e,f] , Ruth E. Ley[b#], 2011, ISMEJ, In Press)
- Applications
  - Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria (Hazen, 2010, Science)
  - Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria (Mendes, 2011, Science)

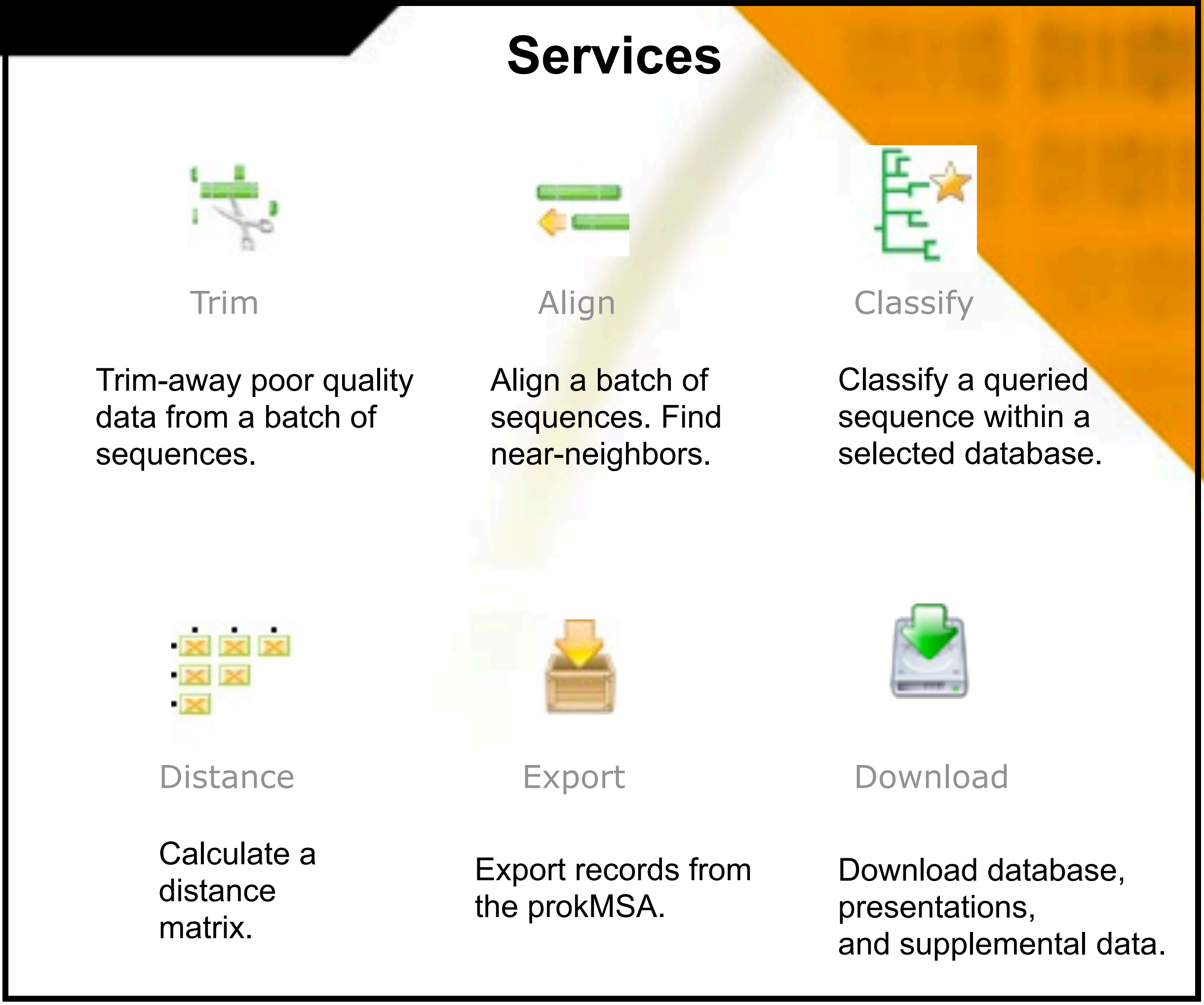
Recent Interesting Use Cases

- http://www.secondgenome.com/2011/03/recent-microbiomics-advances-from-various-fields/

3

http://greengenes.lbl.gov

green genes



Services

Trim

Trim-away poor quality data from a batch of sequences.

Align

Align a batch of sequences. Find near-neighbors.

Classify

Classify a queried sequence within a selected database.

Distance

Calculate a distance matrix.

Export

Export records from the prokMSA.

Download

Download database, presentations,
and supplemental data.

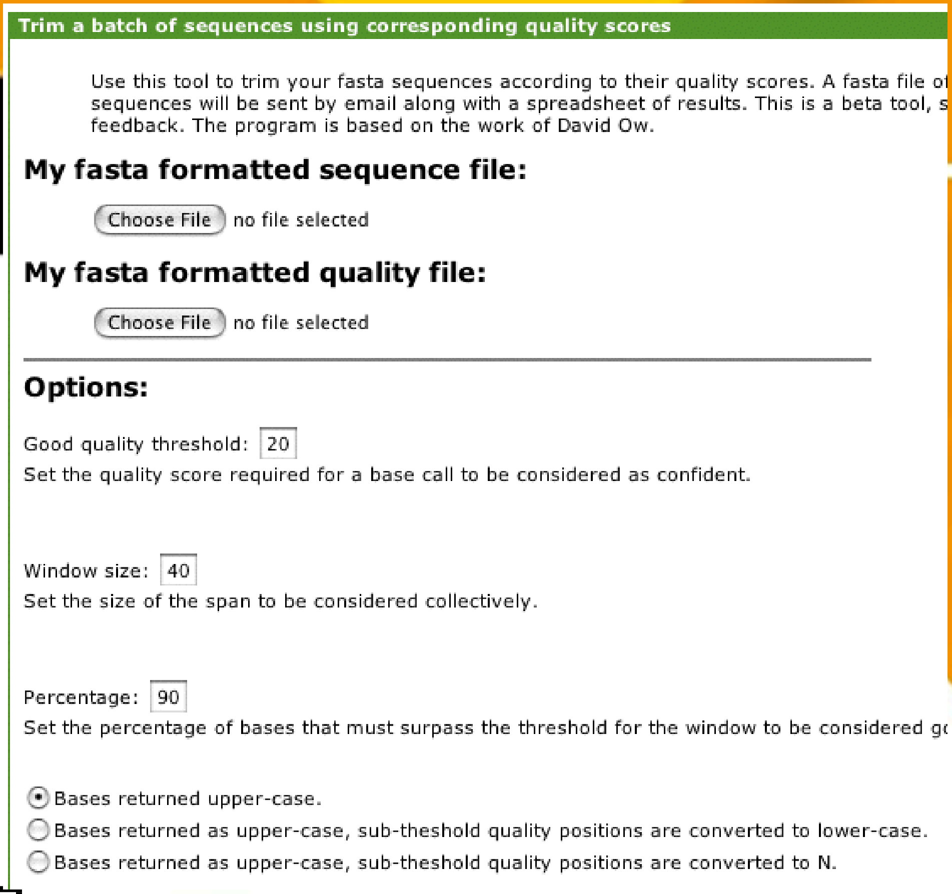
What we'll cover today

- Mini-Background on 16S rRNA gene
- Do quality assessment
- Do alignment
- Do chimera check
- Overview on classification
- Alternate technology - PhyloChip
- Microbiome data visualization

Error probability: scan and trim

```
>actb24
TTTTGGGATTCGCTCCGCCTCGCGGCATCGCAGCCCTTTGTACCGGCCAT
TGTAGCACGTGTGCAGCCCAAGACATAAGGGGCATGATGATTTGACGTCG
TCCCCACCTTCCTCCGAGTTGACCCCGGCAGTCTCCTGTGAGTCCCCGAC
ATTACTCGCTGGCAACACAGAACAAGGGTTGCGCTCGTTGCGGGACTTAA
CCCAACATCTCACGACACGAGCTGACGACAACCATGCACCACCTGTACAC
CGACCACAAGGGGGCTGATATCTCTACCAGTTTCCGGTGTATGTCAAGCC
TTGGTAAGGTTCTTCGCGTTGCGTCGAATTAAGCCACATGCTCCGCTGCT
TGTGCGGGCCCCCGTCAATTCCTTTGAGTTTTAGCCTTGCGGCCGTACTC
CCCAGGCGGGGAACTTAATGCGTTAGCTGCGGCACCGACGACGTGGAATG
TCGCCAACACCTAGTTCCCAACGTTTACGGCGTGGACTACCAGGGTATCT
AATCCTGTTCGCTCCCCACGCTTTCGCTCCTCAGCGTCAGTAATGGCCCA
GAGATCCGCCTTCGCCACCGGTGTTCCTCCTGATATCTGCGCATTTC..
```

```
>actb24
..19 10 11 9 8 8 8 15 9 9 10 9 13 8 10 10 10 16 18 16
16   10 9 11 7 8 8 12 14 25 15 15 6 6 6 6 12 10 14 22
25 21 21 8 10 9 12 11 9 9 17 20 29 29 22 20 11 11 7 7
9 17 13 20 20 31 30 23 23 11 9 9 9 7 7 13 15 25 25 24
21 17 17 17 21 24 24 29 25 25 29 40 32 31 19 19 10 10
9 20 20 25 18 18 25 25 19 19 21 21 23 28 28 29 29 32
22 22 22 32 27 29 25 27 27 22 25 15 15 18 27 27 33 33
33 40 40 47 47 47 32 32 32 32 29 35 40 40 40 40 40 40
31 31 40 32 29 21 21 25 31 26 29 30 30 33 28 31 31 26
26 25 22 22 29 31 28 26 28 27 29 33 25 25 18 27 30 42
37 42 35 35 35 40 40 40 40 42 42 34 34 42 44 47 47 47
47 42 47 42 42 42 42 42 ..
```



Trim a batch of sequences using corresponding quality scores

Use this tool to trim your fasta sequences according to their quality scores. A fasta file o sequences will be sent by email along with a spreadsheet of results. This is a beta tool, s feedback. The program is based on the work of David Ow.

My fasta formatted sequence file:

Choose File no file selected

My fasta formatted quality file:

Choose File no file selected

Options:

Good quality threshold: 20

Set the quality score required for a base call to be considered as confident.

Window size: 40

Set the size of the span to be considered collectively.

Percentage: 90

Set the percentage of bases that must surpass the threshold for the window to be considered go

● Bases returned upper-case.

○ Bases returned as upper-case, sub-theshold quality positions are converted to lower-case.

○ Bases returned as upper-case, sub-theshold quality positions are converted to N.

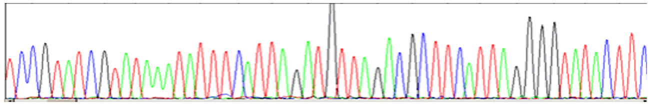
Phred	Error probability		conf.
20	1 in $10^{2.0}$	1/100	99%
15	1 in $10^{1.5}$	1/32	97%
10	1 in $10^{1.0}$	1/10	90%
5	1 in $10^{0.5}$	1/3	66%

qual files should follow sequences throughout pipeline.

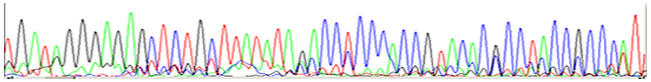
Submit Trim Job

- Do it now, discuss it later …
- Every 3rd person?
  - Goto: http://greengenes.lbl.gov/Download/Tutorial/
  - Get: GG_tut_files.zip
  - Unzip it.
  - View with a text editor:
    - UnAlignSeqsMGM.fasta
    - UnAlignSeqsMGM.qual
  - Goto: http://greengenes.lbl.gov then “Trim”
  - “Begin the Trim”

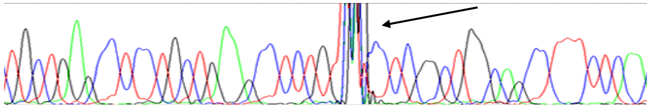
Shall we build our houses on sand?



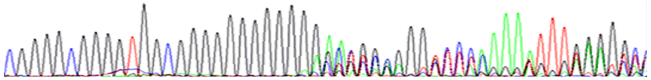
A. High quality data show peaks that are sharp and evenly distributed. Almost no background noise is observed.



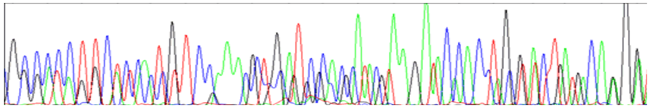
B. Background noise appears as many smaller, irregular peaks under the dominant peaks of interest.



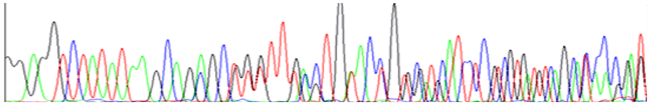
C. Spikes (at arrow) can be caused by air bubbles entering the capillary.



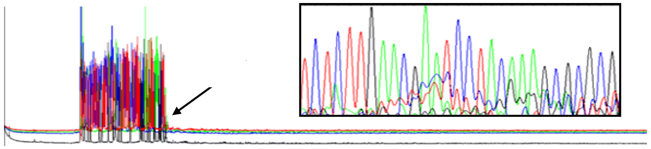
D. Homopolymeric regions allow enzyme "slippage" when the growing strand does not stay paired correctly with the template DNA during polymerization through the low-complexity span. Resulting fragments of varying lengths manifest as double peaks in the chromatogram.



E. Multiple Overlapping Peaks throughout. If derived from direct 16S rRNA gene amplicon sequencing, may indicate multiple genomes due to insufficient colony isolation or endosybiosis. If derived from clone library, may indicate poor transformant colony isolation or plasmid prep contamination.



F. Sporadic Overlapping Peaks. If derived from direct amplicon sequencing, may indicate divergent 16S rRNA genes within the same genome.



G. Abrupt truncations. 16S rRNA amplicons are predisposed to secondary structure formation and may form hairpins restricting the passage of the sequencing polymerase. Inset magnifies chromatogram at region of arrow to display low signal-to-noise ratio.

Chromatogram quality can be occluded by text-only sequence data. In all examples (**A** - **G**), base calls can be produced from chromatograms for most peaks with automated software. In **A,** each peak yields a base-call with low error probability. In contrast, chromatograms **B** - **G** contain peaks that, although are capable of producing automated base-calls, will have vastly varying degrees of error probability. Examples of non-ideal chromatograms in libraries from general sequencing projects shown in **B** - **D** whereas **E** - **G** are examples of additional aberrations encountered in 16S rRNA gene sequencing. Chromatograms produced by Eton Bioscience Inc. (http://www.eatonbio.com) and used with permission.

The Trim Results

- Collect 2 emails
- Save to tutorial folder
- .xls file useful for an overview
  - `UnAlignSeqsMGM_trimmed_XXXXX.xls`
  - Columns
- .fasta file is ready for downstream tools
  - `UnAlignSeqsMGM_trimmed_XXXXX.fasta`
  - Or user can modify

Submit Alignment Job

- Do it now, discuss it later …
  - Goto: http://greengenes.lbl.gov then “Align”
  - Upload UnAlignSeqsMGM_trimmed_XXXXX.fasta
  - Check all boxes.
  - “Submit”

The Ribosome

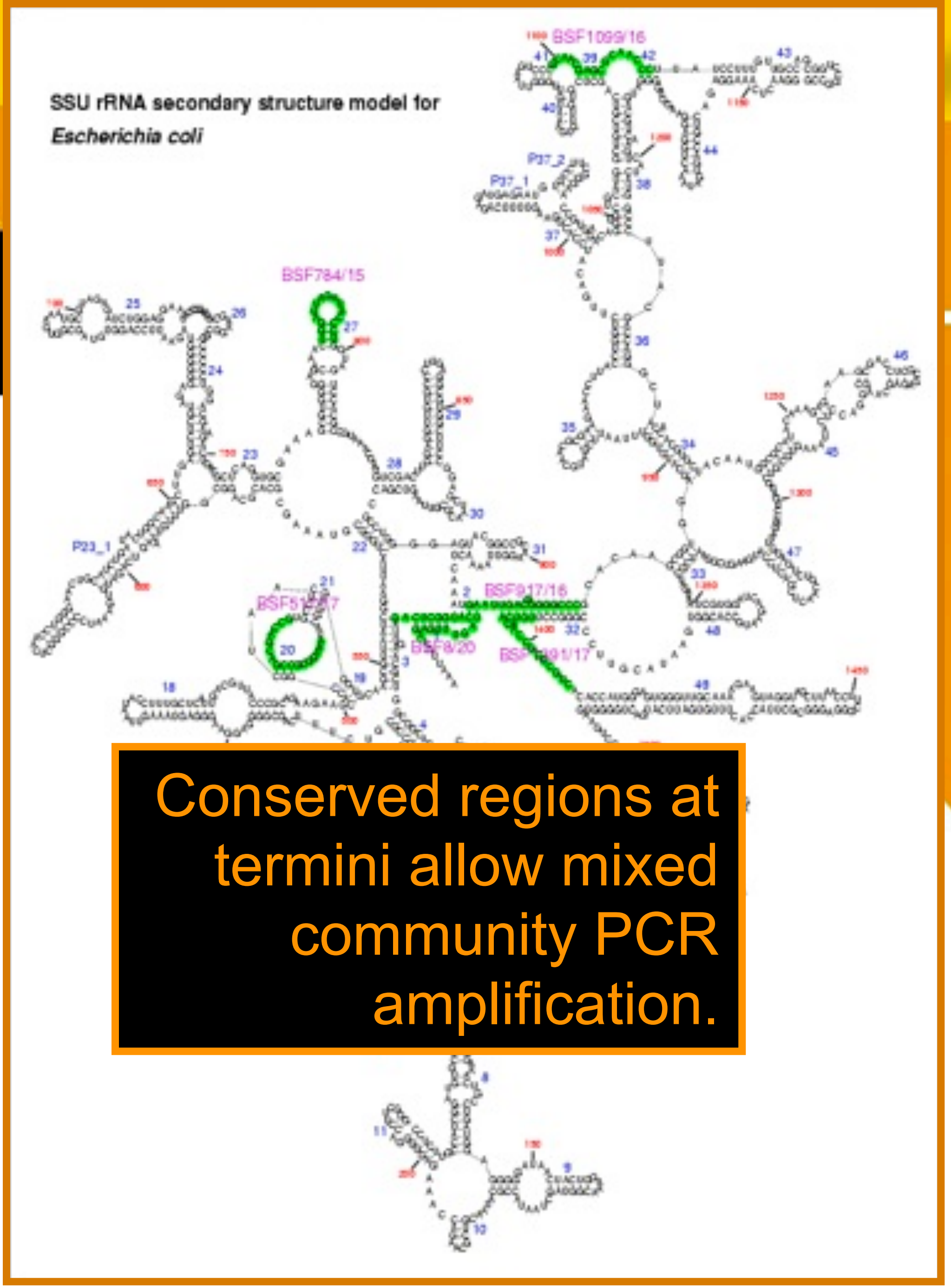
rDNA

rRNA (functional molecule)

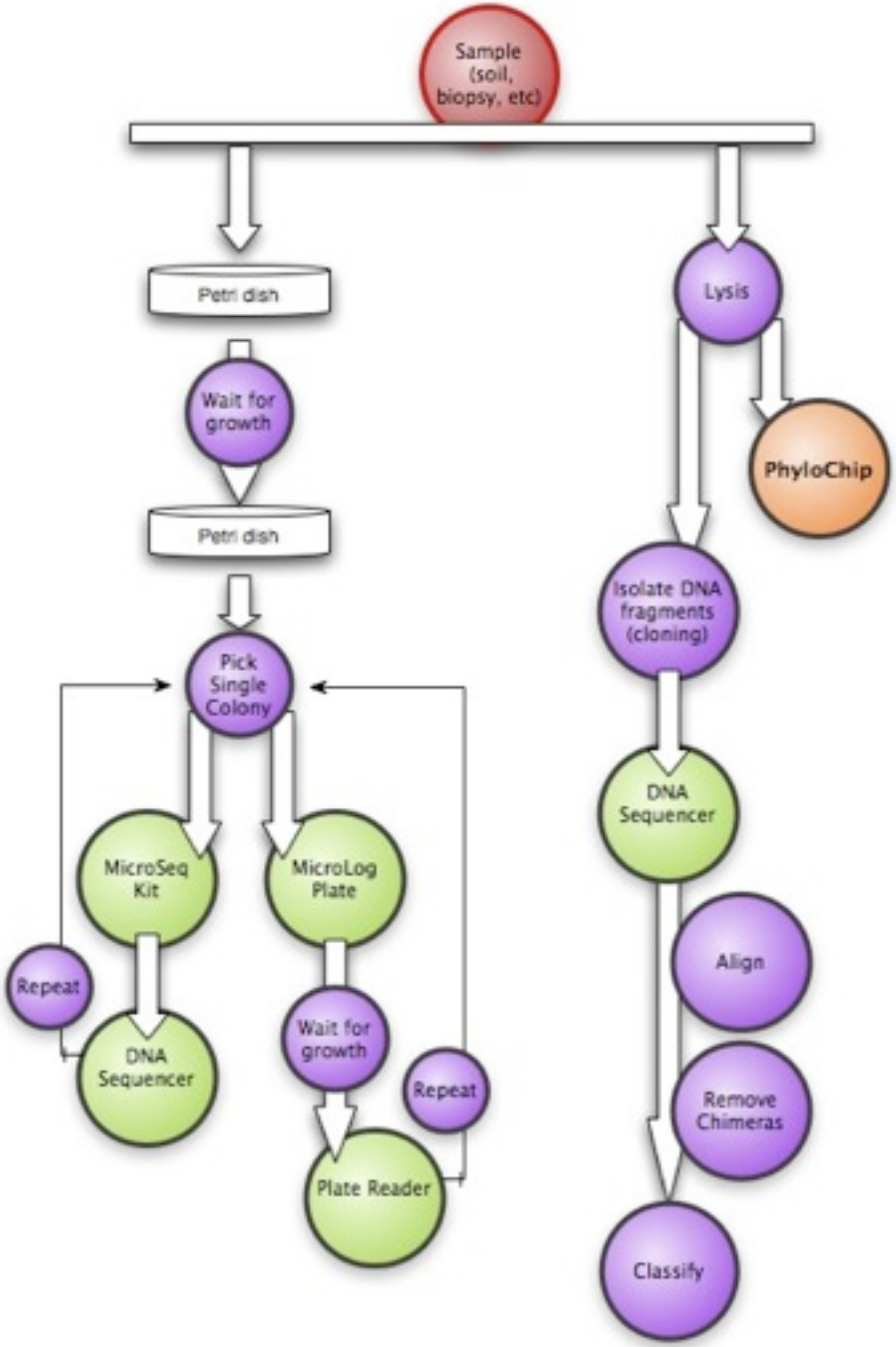
LSU

SSU

16S or 18S



Conserved regions at termini allow mixed community PCR amplification.

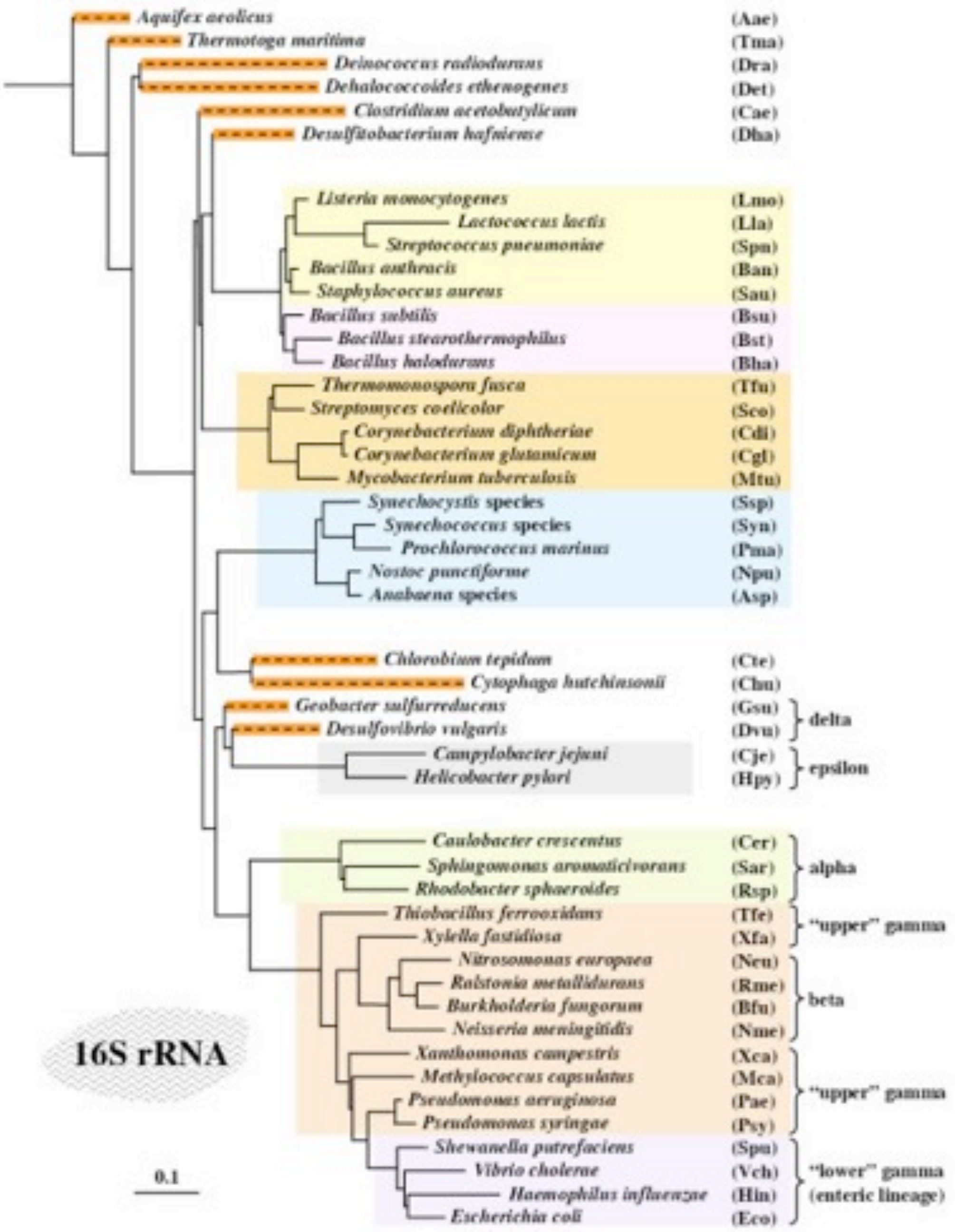


Lab Workflows

- Culturable?
- Extractable?
- Able to ligate?
- Able to yield viable transformants?
- Clean sequencing reaction?

Classify

- 16S rRNA aids in taxonomic placement of sequences
- Compare your sequence to others.


Aquifex aeolicus (Aae)
Thermotoga maritima (Tma)
Deinococcus radiodurans (Dra)
Dehalococcoides ethenogenes (Det)
Clostridium acetobutylicum (Cae)
Desulfitobacterium hafniense (Dha)
Listeria monocytogenes (Lmo)
Lactococcus lactis (Lla)
Streptococcus pneumoniae (Spn)
Bacillus anthracis (Ban)
Staphylococcus aureus (Sau)
Bacillus subtilis (Bsu)
Bacillus stearothermophilus (Bst)
Bacillus halodurans (Bha)
Thermomonospora fusca (Tfu)
Streptomyces coelicolor (Sco)
Corynebacterium diphtheriae (Cdi)
Corynebacterium glutamicum (Cgl)
Mycobacterium tuberculosis (Mtu)
Synechocystis species (Ssp)
Synechococcus species (Syn)
Prochlorococcus marinus (Pma)
Nostoc punctiforme (Npu)
Anabaena species (Asp)
Chlorobium tepidum (Cte)
Cytophaga hutchinsonii (Chu)
Geobacter sulfurreducens (Gsu)
Desulfovibrio vulgaris (Dvu)
delta
Campylobacter jejuni (Cje)
Helicobacter pylori (Hpy)
epsilon
Caulobacter crescentus (Ccr)
Sphingomonas aromaticivorans (Sar)
Rhodobacter sphaeroides (Rsp)
alpha
Thiobacillus ferrooxidans (Tfe)
Xylella fastidiosa (Xfa)
"upper" gamma
Nitrosomonas europaea (Neu)
Ralstonia metallidurans (Rme)
Burkholderia fungorum (Bfu)
Neisseria meningitidis (Nme)
beta
Xanthomonas campestris (Xca)
Methylococcus capsulatus (Mca)
Pseudomonas aeruginosa (Pae)
Pseudomonas syringae (Psy)
"upper" gamma
Shewanella putrefaciens (Spu)
Vibrio cholerae (Vch)
Haemophilus influenzae (Hin)
Escherichia coli (Eco)
"lower" gamma
(enteric lineage)
16S rRNA
0.1

Xie et al. BMC Biology 2004 2:15

Greengenes maintains a high-quality Core Set

- Start by collecting a large set:
  - All matching GOLD (genomesonline.org) organisms

  AND

  - All meeting specific criteria
    - ≥ 1300 nt.
    - ≤ 2 homopolymers
    - ≤ 0.3% ambiguity
    - ≤ 30 “small gap intrusions” - bases not supported by the 2° structure.
    - Non-chimeric

Greengenes maintains high-quality Core Sets

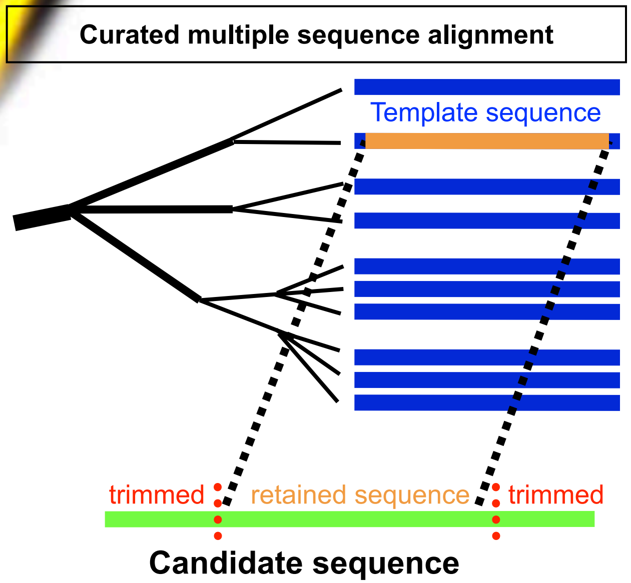
- Then de-replicate at >95% Uclust identity.
- 36,550 genes represents the known 16S diversity.

- Other sub-sets
  - All named isolates
  - All HMP Genome Strains
  - Knight, Caporaso: QIIME-ready reference sets.
  - Older Core Sets

- http://greengenes.lbl.gov/Download/Sequence_Data/Fasta_data_files/

Download a Core Set to screen your contigs for 16S content...

NAST align against the core set.



Curated multiple sequence alignment

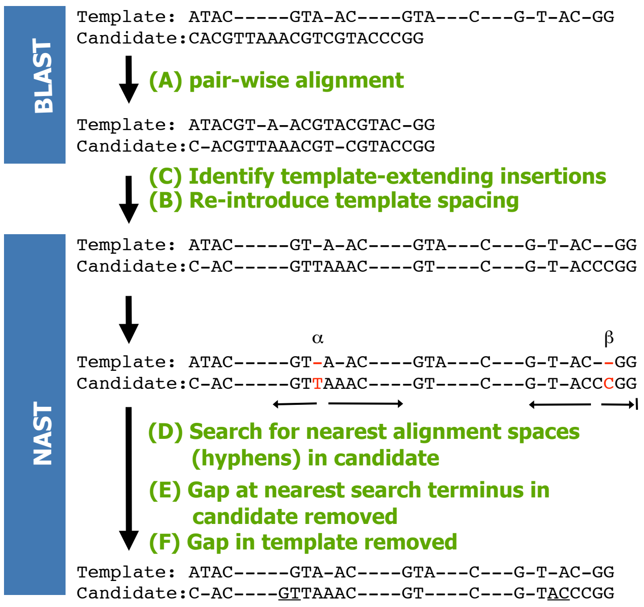
Template sequence

trimmed

retained sequence

trimmed

Candidate sequence



```
Template: ATAC-----GTA-AC----GTA---C---G-T-AC-GG
Candidate:CACGTTAAACGTCGTACCCGG
```

(A) pair-wise alignment

```
Template: ATACGT-A-ACGTACGTAC-GG
Candidate:C-ACGTTAAACGT-CGTACCGG
```

(C) Identify template-extending insertions

(B) Re-introduce template spacing

```
Template: ATAC-----GT-A-AC----GTA---C---G-T-AC--GG
Candidate:C-AC-----GTTAAAC----GT----C---G-T-ACCCGG
```

```
Template: ATAC-----GT-A-AC----GTA---C---G-T-AC--GG
Candidate:C-AC-----GTTAAAC----GT----C---G-T-ACCCGG
```

(D) Search for nearest alignment spaces (hyphens) in candidate

(E) Gap at nearest search terminus in candidate removed

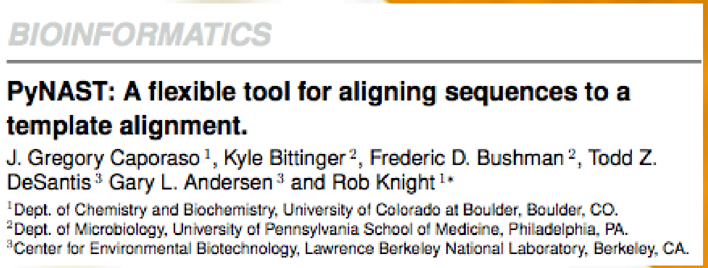
(F) Gap in template removed

```
Template: ATAC-----GTA-AC----GTA---C---G-T-AC-GG
Candidate:C-AC----GTTAAAC----GT----C---G-TACCCGG
```

Example of NAST (Nearest Alignment Space Termination) compression of a BLAST pair-wise alignment using a 38 character aligned template. Template and candidate is extended to 40 characters after BLAST gap insertion (A) and retention of original template spacing (B). Nucleotide insertions in the candidate relative to the template which force additional characters to be added in the template are identified at positions α and β (C). A bi-directional search for the nearest alignment space (hyphen) relative to the insertion terminates at the positions indicated by the black arrows (D). The leftward search from the α position was shorter in distance compared to the rightward, thus the space left of 'GT' was removed. The search from the β position encountered the alignment edge on the right, thus the position to the left of 'AC' was removed (E). Lastly, the two template-extending spaces are deleted from the template (F). Notice that sequence data is not added to nor overwritten in the candidate. The NAST removal of two characters from both sequences allowed local misalignments (underlined) while preserving the 38 character format of the global multiple sequence alignment.

The NAST Alignment Results

- .xls file useful for an overview
  - Walk through columns
  - NAST vrs NASTnot
  - Purpose of nn and nni
    - Clearcut, RAX-ml, ITOL, etc., tree ….



BIOINFORMATICS

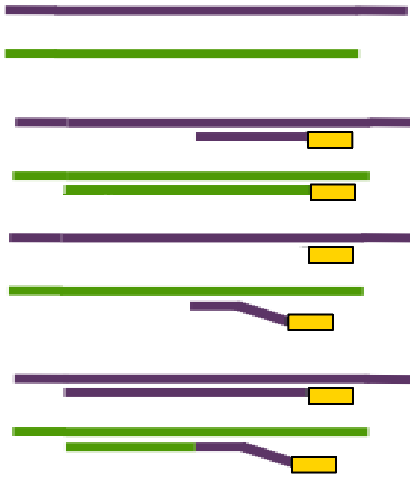
PyNAST: A flexible tool for aligning sequences to a template alignment.

J. Gregory Caporaso [1], Kyle Bittinger [2], Frederic D. Bushman [2], Todd Z. DeSantis [3] Gary L. Andersen [3] and Rob Knight [1*]

[1] Dept. of Chemistry and Biochemistry, University of Colorado at Boulder, Boulder, CO.
[2] Dept. of Microbiology, University of Pennsylvania School of Medicine, Philadelphia, PA.
[3] Center for Environmental Biotechnology, Lawrence Berkeley National Laboratory, Berkeley, CA.

- Other NAST implementations
  - PyNAST - Knight
  - NASTier - Hass
  - NAST MOTHUR - Schloss

How chimera’s form



PCR templates from distinct phyla

Primer extension with premature termination

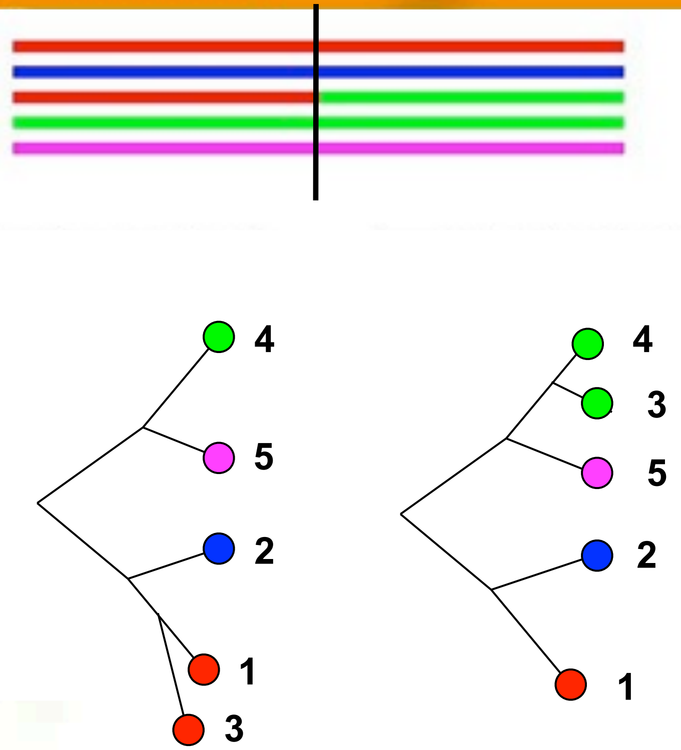
Fragment re-annealing to DNA strand of dissimilar template

Fragment polymerization forming a chimera

An example of a chimeric artifact generated during PCR amplification of a mixed population using broad-specificity 16S rRNA primers Partial amplicons may form hybrids with dissimilar templates because conserved regions exist at positions medial to the PCR primer targets. The partial amplicon can be extended using the dissimilar 16S gene as a template.

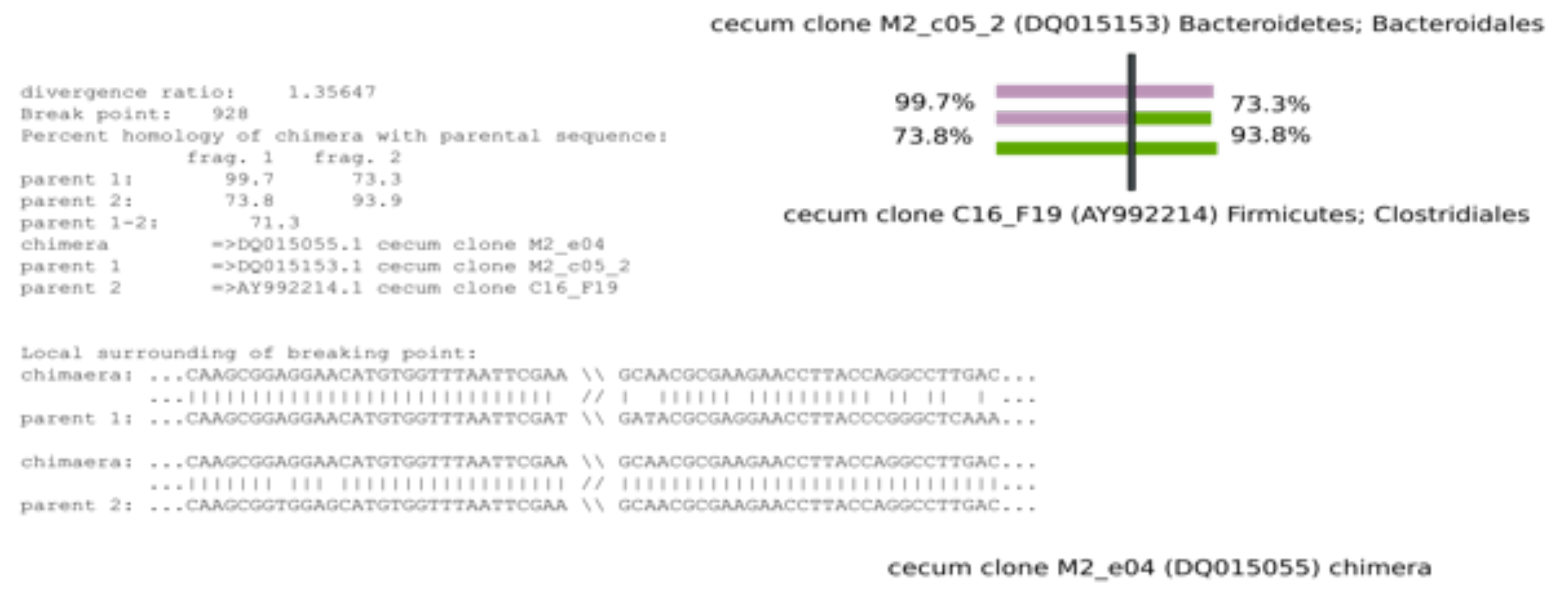
Find them with the partial-partial tree building approach

- Build MSA
- Divide MSA at all possible break points
- Construct 2 distance matrices for each break test.
- Compare consistency of distances.



(Wang and Wang, 1997; Hugenholtz , 2003)

Setting thresholds to flag chimeras.



```
divergence ratio:     1.35647
Break point:    928
Percent homology of chimera with parental sequence:
                    frag. 1    frag. 2
parent 1:            99.7       73.3
parent 2:            73.8       93.9
parent 1-2:             71.3
chimera           =>DQ015055.1 cecum clone M2_e04
parent 1          =>DQ015153.1 cecum clone M2_c05_2
parent 2          =>AY992214.1 cecum clone C16_F19


Local surrounding of breaking point:
chimaera: ...CAAGCGGAGGAACATGTGGTTTAATTCGAA \\ GCAACGCGAAGAACCTTACCAGGCCTTGAC...
          ...|||||||||||||||||||||||||||||  // |  |||||| |||||||||| |||    | ...
parent 1: ...CAAGCGGAGGAACATGTGGTTTAATTCGAT \\ GATACGCGAGGAACCTTACCCGGGCTCAAA...

chimaera: ...CAAGCGGAGGAACATGTGGTTTAATTCGAA \\ GCAACGCGAAGAACCTTACCAGGCCTTGAC...
          ...||||||| ||| |||||||||||||||||| // ||||||||||||||||||||||||||||||...
parent 2: ...CAAGCGGTGGAGCATGTGGTTTAATTCGAA \\ GCAACGCGAAGAACCTTACCAGGCCTTGAC...
```

cecum clone M2_c05_2 (DQ015153) Bacteroidetes; Bacteroidales
99.7%
73.8%
73.3%
93.8%
cecum clone C16_F19 (AY992214) Firmicutes; Clostridiales
cecum clone M2_e04 (DQ015055) chimera

Figure 4. Chimeras can be detected within large data sets using modifications to the existing software, Bellerophon (Huber et al, 2004). In this example a chimera (clone M2_eO4) was found in a 16S rRNA gene clone library prepared from cecal samples. The modified Bellerophon is able to search for parents over intra-library (putative parent M2_c05_2) and inter-library (putative parent C16_F19) sequences. **The *divergence ratio* of 1.36 indicates the parents are 36% more divergent from each other than the chimeric fragments are from their respective parents.** Further investigation placed the parents in distinct phyla.

Submit Chimera Check

- Goto: http://greengenes.lbl.gov
  - Then “More Tools”
  - Then “Chimera Check with Bellerophon”
- “Submit”

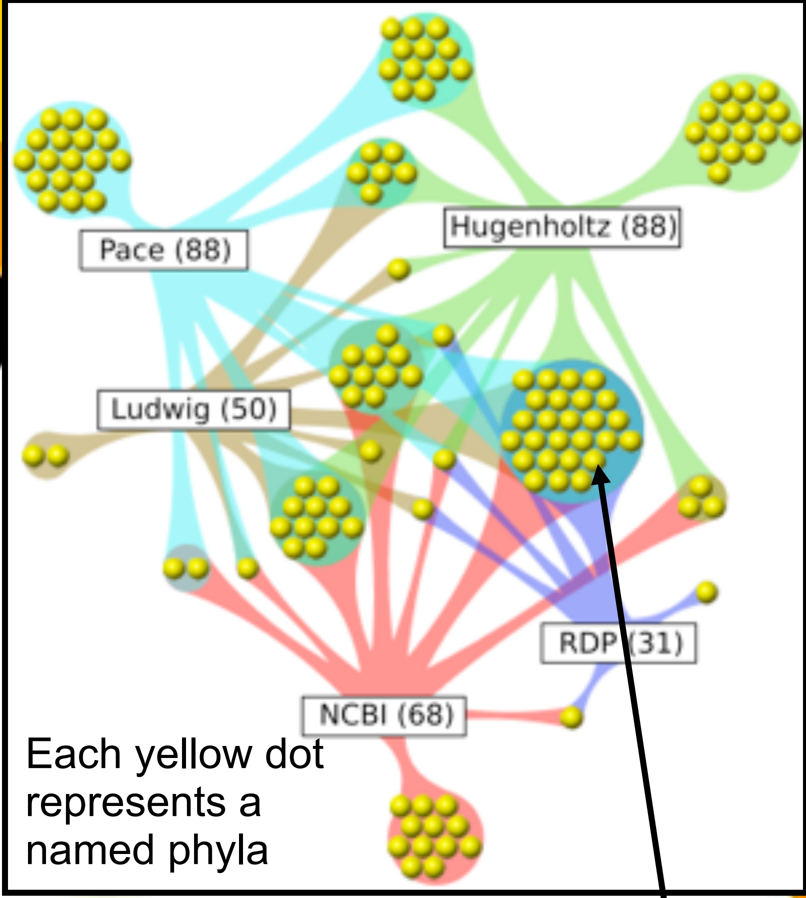
The Bellerophon Results

- .xls file useful for an overview
  - Walk through columns
- Bclean, Bambig, Bchimera

Taxonomy in Flux

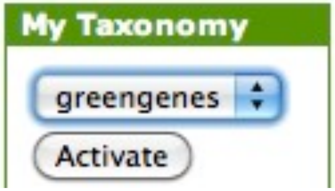
- Incongruent taxonomic nomenclature even at phylum level.
- Making multiple taxonomic classifications available through Greengenes will aid in standardizing classification, particularly for environmental lineages.
- Greengenes integrates each, allowing a balanced approach to nomenclature of newly discovered organisms.
- Example Search:
  - NCBI: CP000866
  - or “Nitrosopumilus maritimus”

http://greengenes.lbl.gov/cgi-bin/show_one_record_v2.pl?prokMSA_id=247303

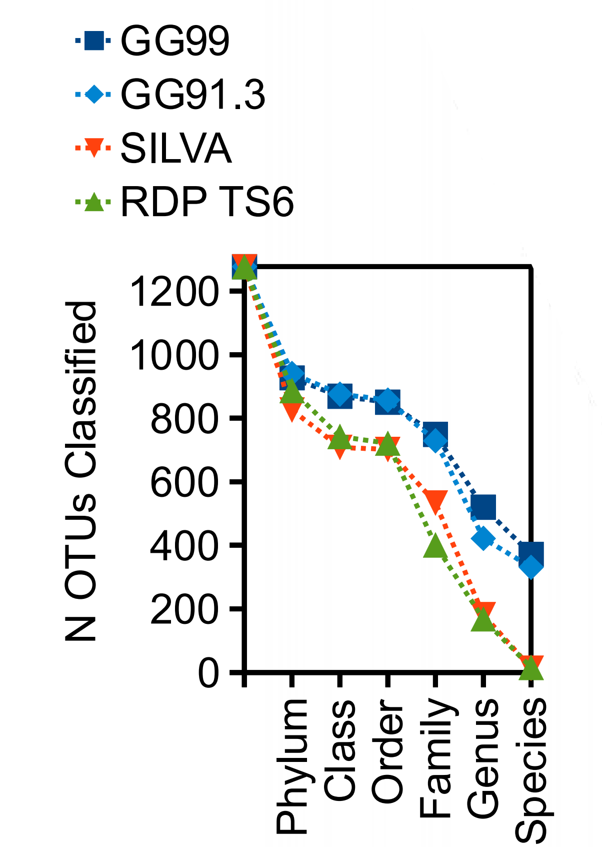


Each yellow dot represents a named phyla

Only a fraction of the phyla are recognized by all five major curators.



Why use the greengenes taxonomy?



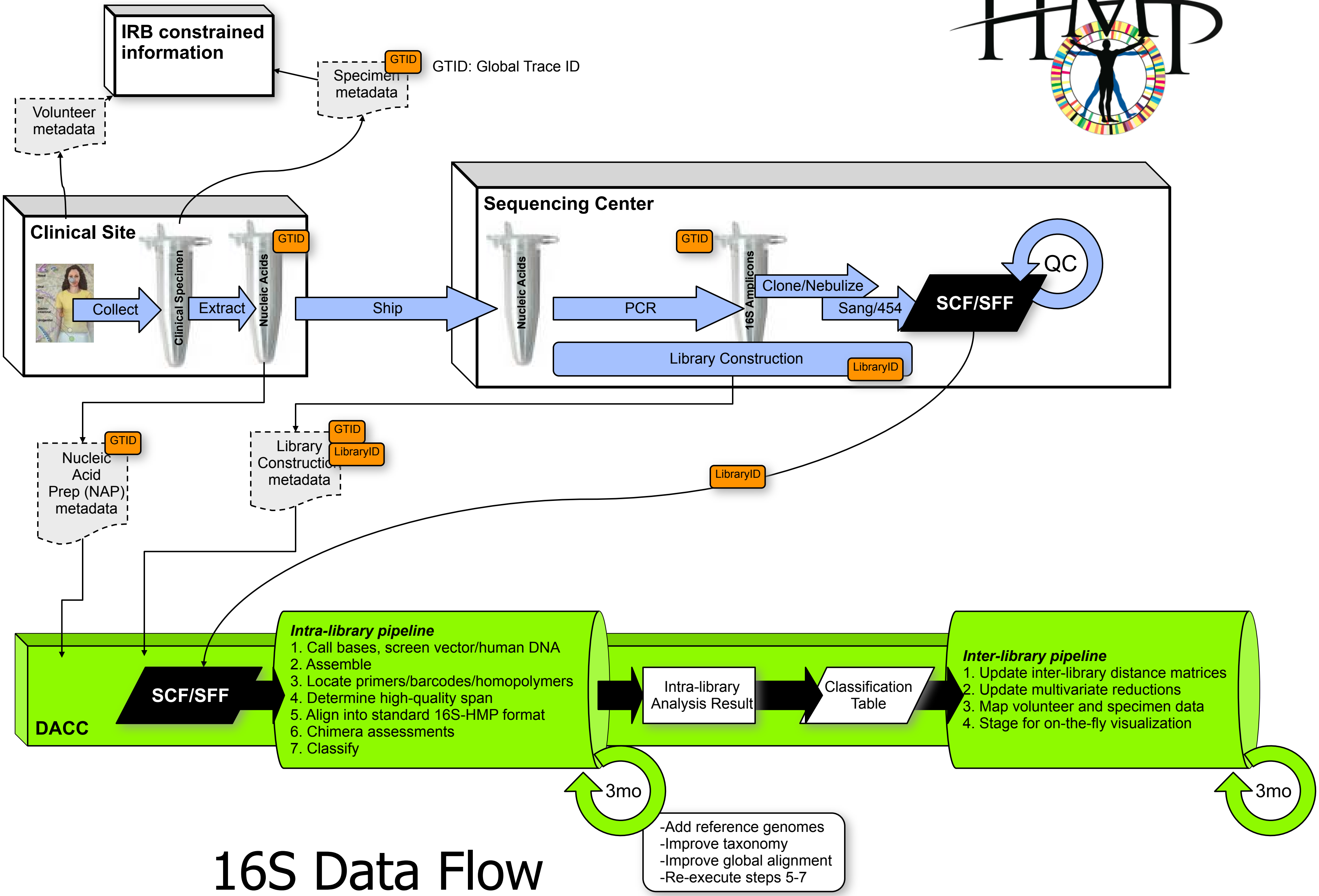
- A high-quality reference tree with nomenclature is maintained
- More of your experimental reads are classified at a high resolution

24

The Classification Results

- classify_xxxxx.xls file
  - Walk through columns

16S Data Flow



Intra-library pipeline

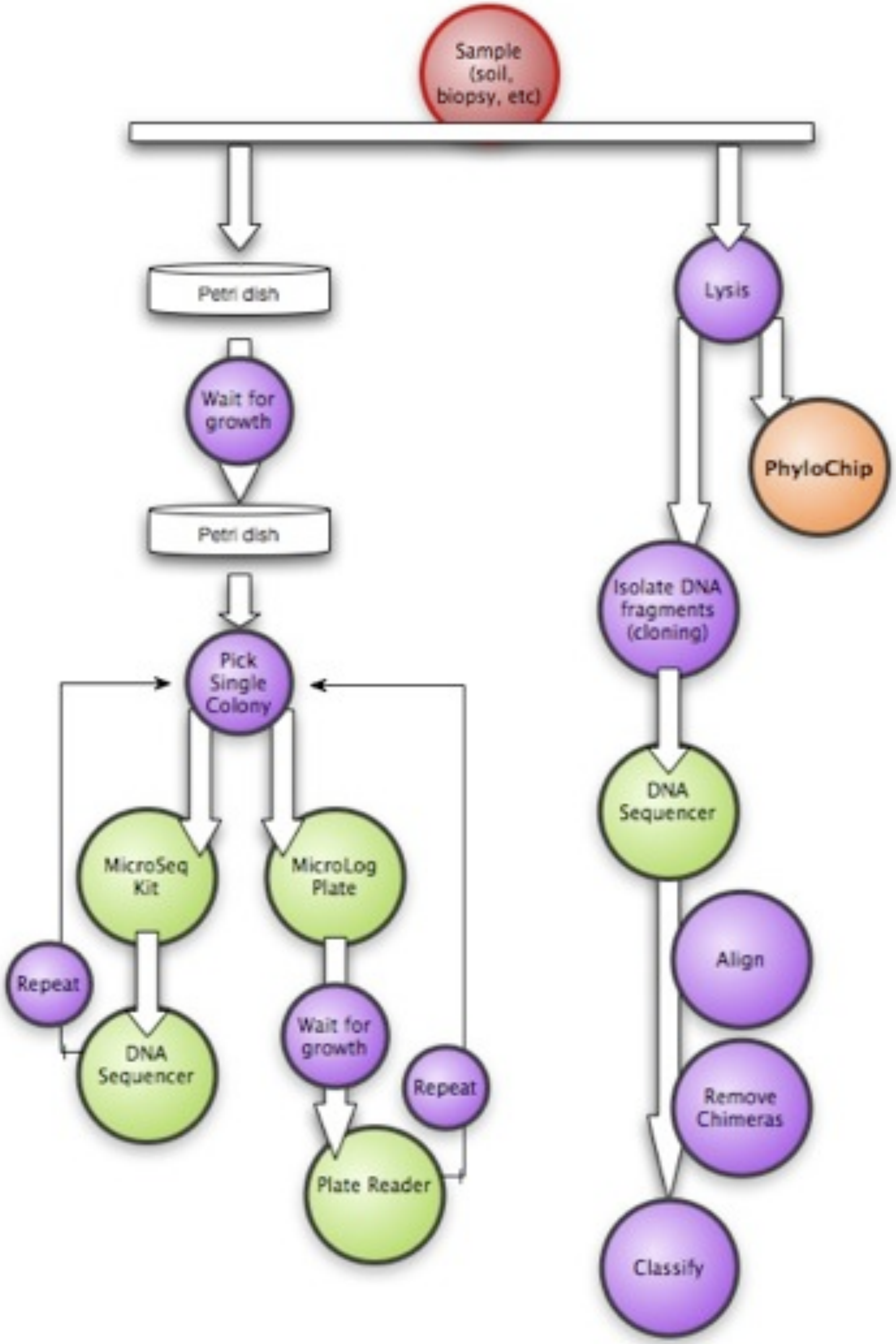
1. Call bases, screen vector/human DNA
2. Assemble
3. Locate primers/barcodes/homopolymers
4. Determine high-quality span
5. Align into standard 16S-HMP format
6. Chimera assessments
7. Classify

Inter-library pipeline

1. Update inter-library distance matrices
2. Update multivariate reductions
3. Map volunteer and specimen data
4. Stage for on-the-fly visualization

-Add reference genomes
-Improve taxonomy
-Improve global alignment
-Re-execute steps 5-7

Merging Sequencing and PhyloChip™ Assay Results



PhyloChip

- Do biological replicates
- Pre-screen samples before doing metagenomics
- Rapid
- Comprehensive
- “1 trillion”

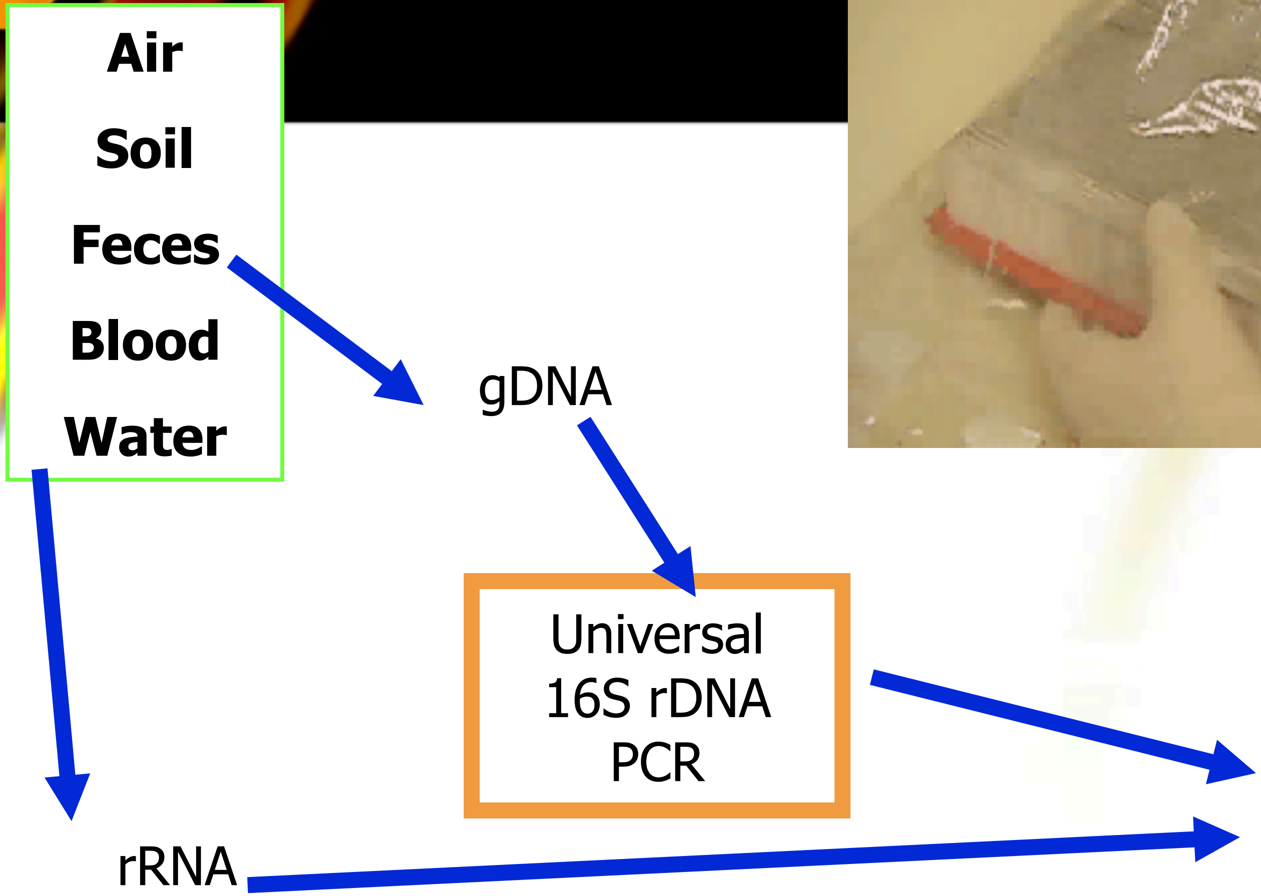
Sampling Effort

- typical 16S rRNA gene PCR yields 500 to 1,000 ng in a 20 uL volume
  - 1500bp @ 660 g/mole/bp
  - 5E-14 moles / uL
  - 3E+10 molecules /uL
  - 6E+11 total sequences
- How many should we observe?
  - 600? 60,000 (1 out of every 10 million)?
- Hybridize them all on a PhyloChip ...
  - Dominant populations do not occlude minority populations

Project Overview

- Goal
  - Create a single microarray capable of detecting and categorizing the bacteria and archaea in a complex sample.
- Approach
  - GeneChip targeted at 16S rDNA sequence variations to distinguish taxa.

General Protocol



Air

Soil

Feces

Blood

Water

gDNA

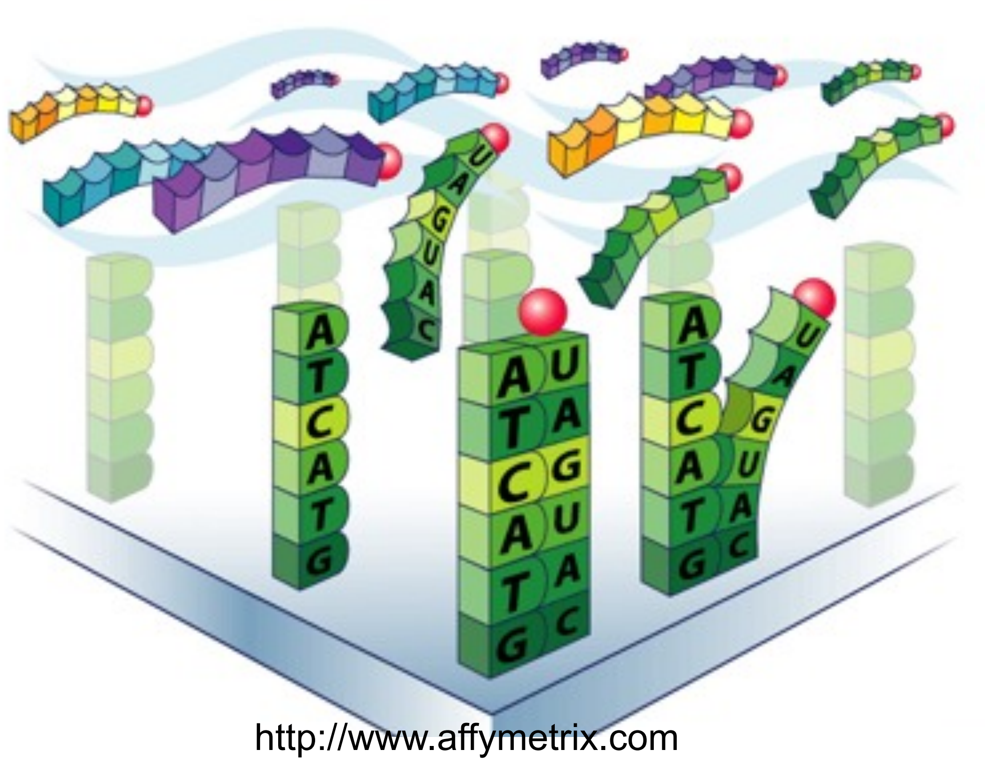
Universal 16S rDNA PCR

rRNA

Contains probes adhered to glass surface in grid pattern.

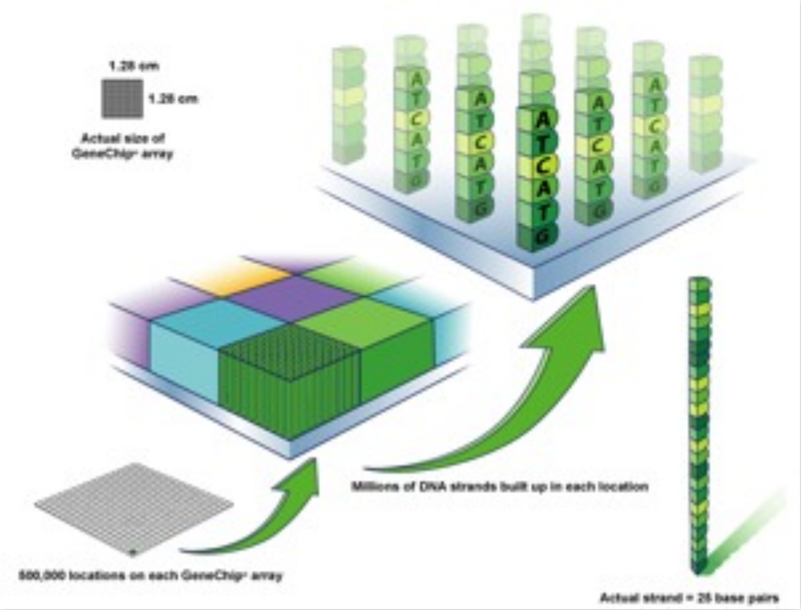
Hybridization

- Notice all NA is labeled (florescence)
- Non-binding NA is washed away
- If surface "glows", then target was captured by probe.

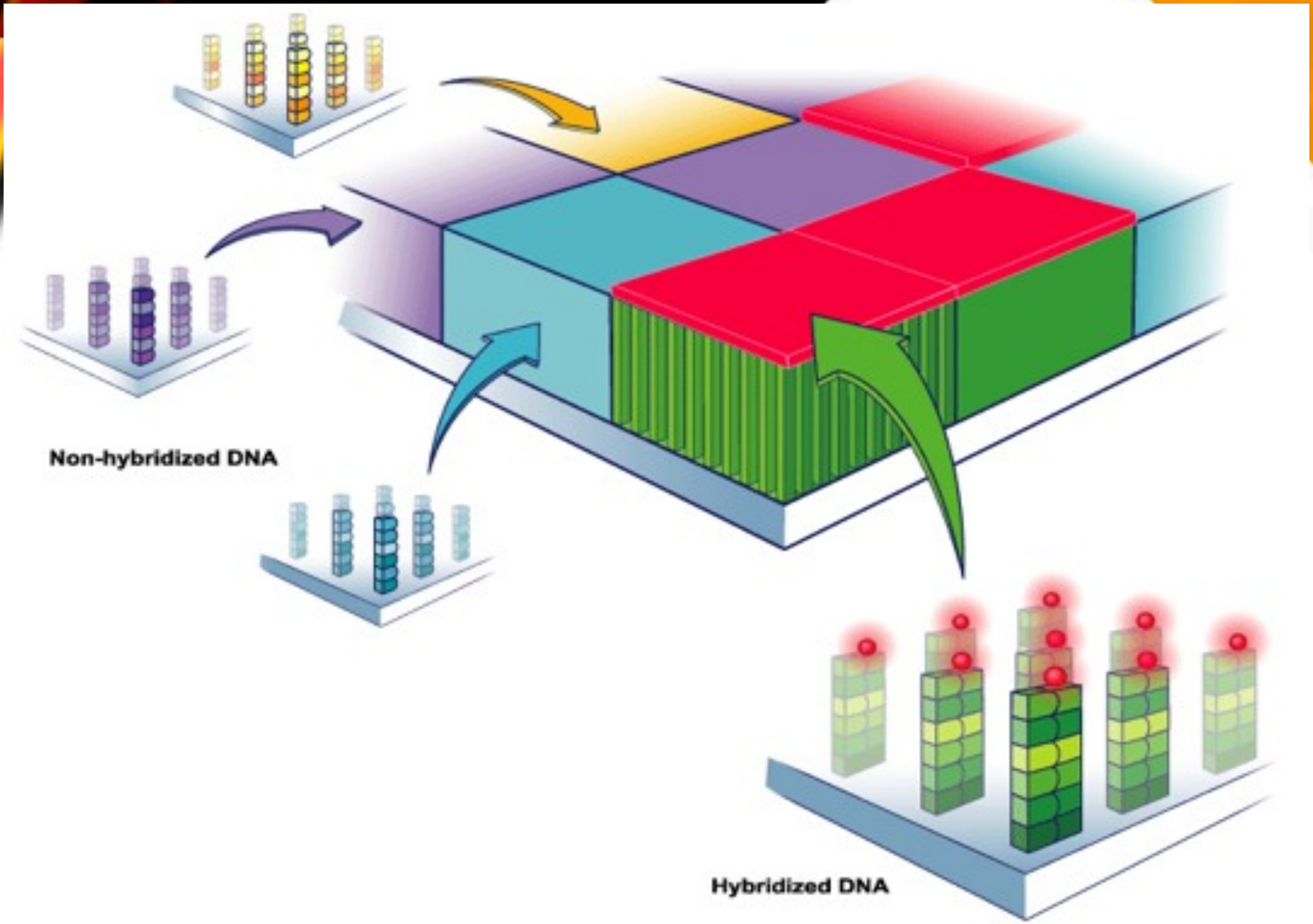


http://www.affymetrix.com

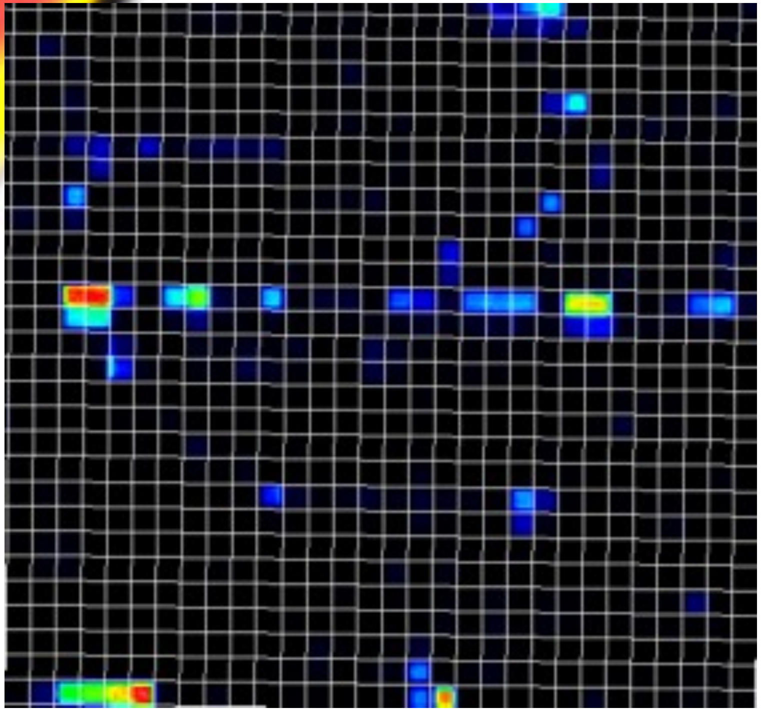
Millions of copies per feature

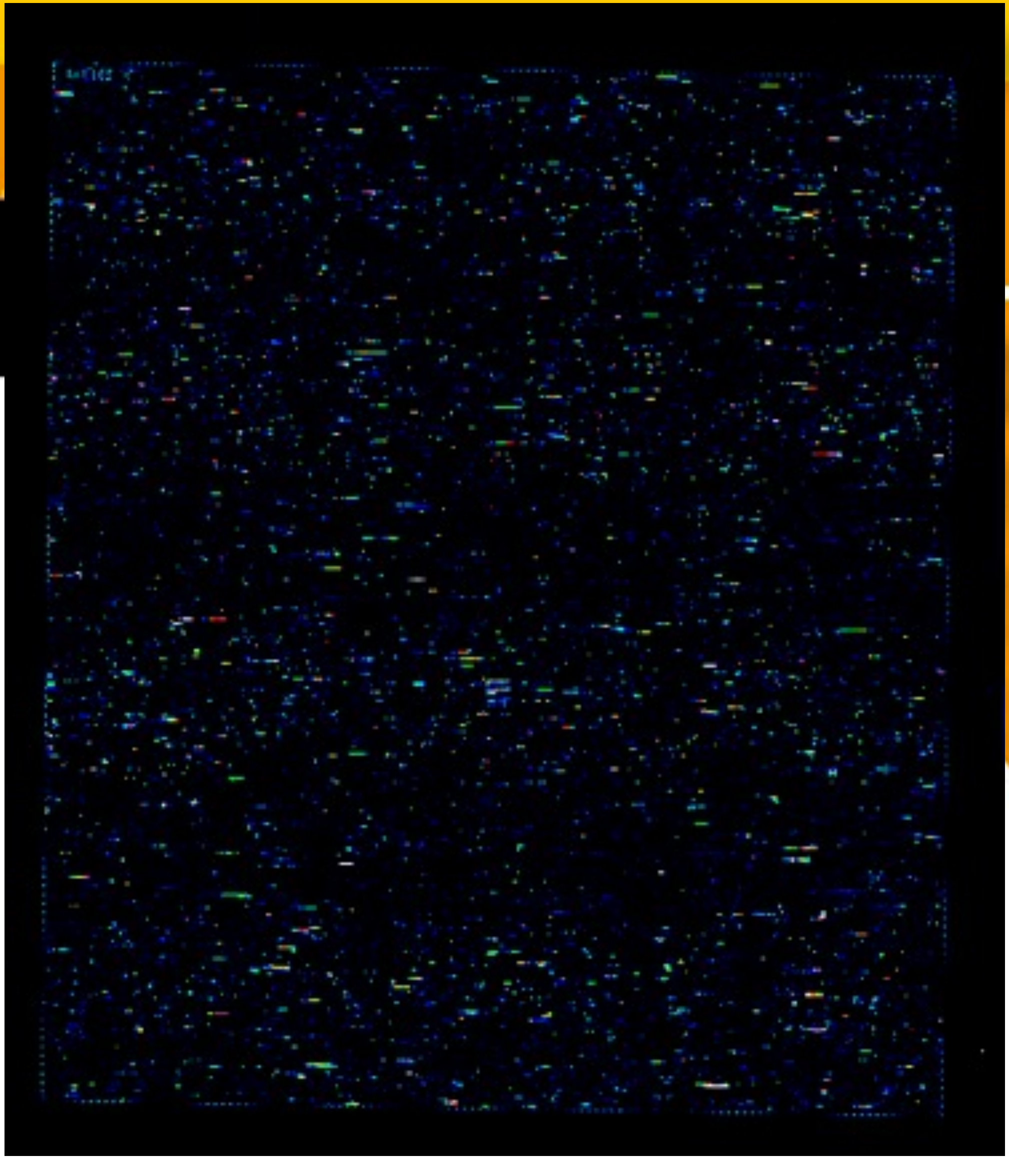

1.28 cm
1.28 cm
Actual size of GeneChip® array
A T C A T G
Millions of DNA strands built up in each location
500,000 locations on each GeneChip® array
Actual strand = 25 base pairs

Coordinates of fluorescence determines test results.

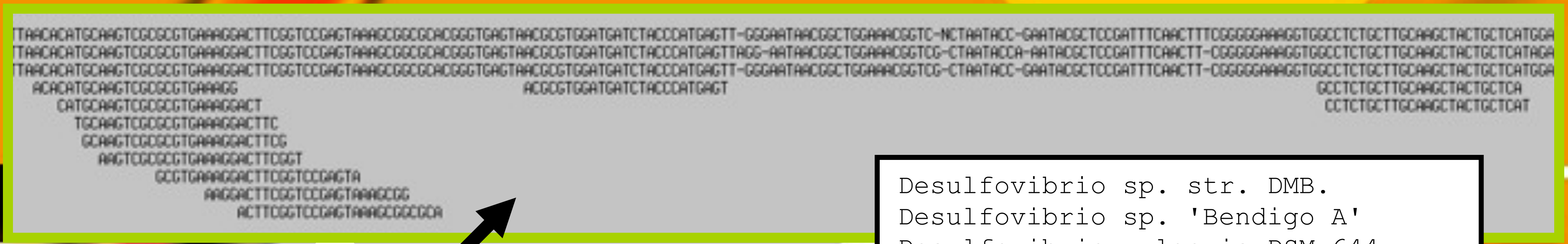


DOE165





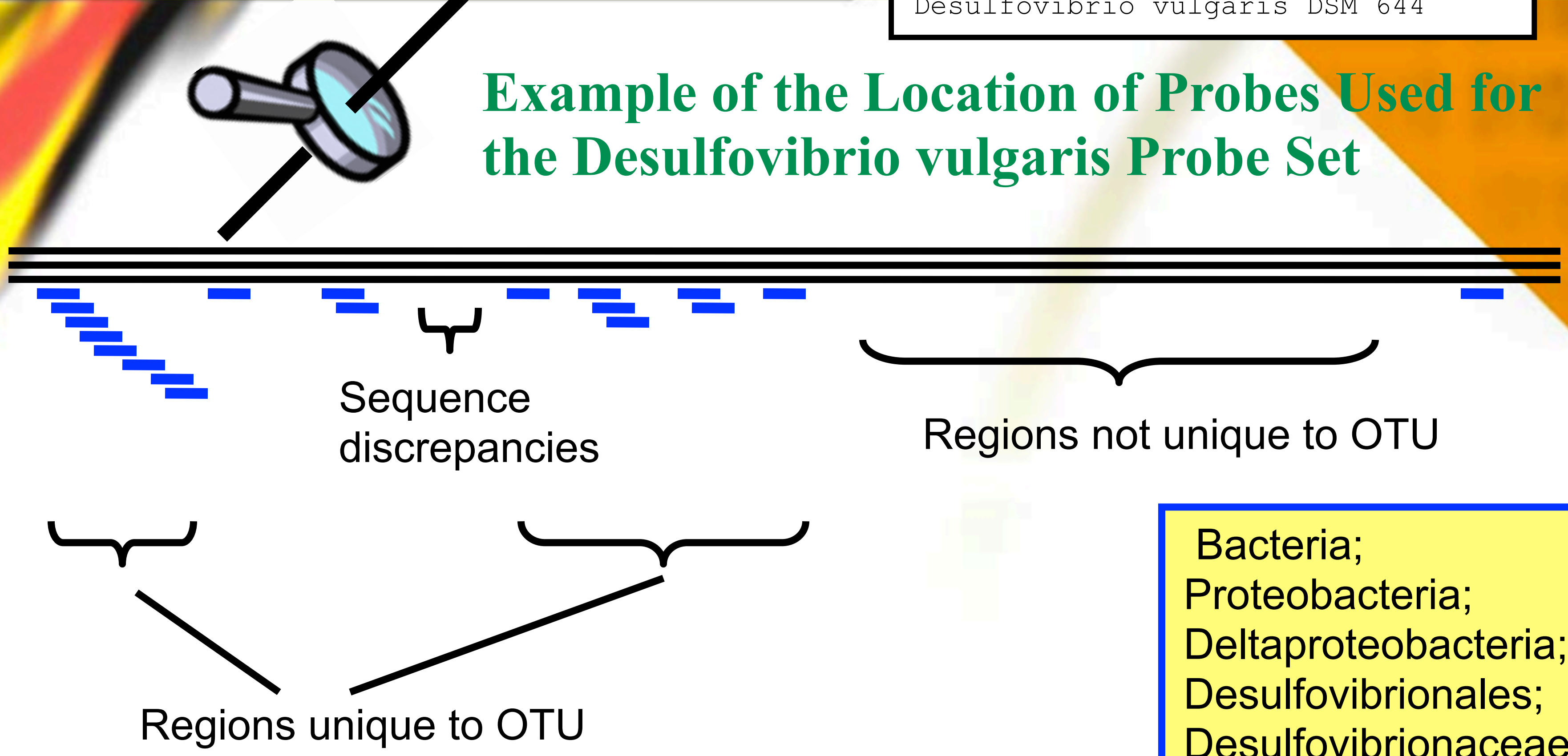
Probe Design



```
TTAACACATGCAAGTCGCGCGTGAAAGGACTTCGGTCCGAGTAAAGCGGCGCACGGGTGAGTAACGCGTGGATGATCTACCCATGAGTT-GGGAATAACGGCTGGAAACGGTC-NCTAATACC-GAATACGCTCCGATTTCAACTTTCGGGGGAAAGGTGGCCTCTGCTTGCAAGCTACTGCTCATGGA
TTAACACATGCAAGTCGCGCGTGAAAGGACTTCGGTCCGAGTAAAGCGGCGCACGGGTGAGTAACGCGTGGATGATCTACCCATGAGTTAGG-AATAACGGCTGGAAACGGTCG-CTAATACCA-AATACGCTCCGATTTCAACTT-CGGGGGAAAGGTGGCCTCTGCTTGCAAGCTACTGCTCATAGA
TTAACACATGCAAGTCGCGCGTGAAAGGACTTCGGTCCGAGTAAAGCGGCGCACGGGTGAGTAACGCGTGGATGATCTACCCATGAGTT-GGGAATAACGGCTGGAAACGGTCG-CTAATACC-GAATACGCTCCGATTTCAACTT-CGGGGGAAAGGTGGCCTCTGCTTGCAAGCTACTGCTCATGGA
   ACACATGCAAGTCGCGCGTGAAAGG                                   ACGCGTGGATGATCTACCCATGAGT                                                                        GCCTCTGCTTGCAAGCTACTGCTCA
      CATGCAAGTCGCGCGTGAAAGGACT                                                                                                                                  CCTCTGCTTGCAAGCTACTGCTCAT
        TGCAAGTCGCGCGTGAAAGGACTTC
         GCAAGTCGCGCGTGAAAGGACTTCG
           AAGTCGCGCGTGAAAGGACTTCGGT
                  GCGTGAAAGGACTTCGGTCCGAGTA
                        AAGGACTTCGGTCCGAGTAAAGCGG
                            ACTTCGGTCCGAGTAAAGCGGCGCA
```

```
Desulfovibrio sp. str. DMB.
Desulfovibrio sp. 'Bendigo A'
Desulfovibrio vulgaris DSM 644
```

Example of the Location of Probes Used for the Desulfovibrio vulgaris Probe Set



Sequence discrepancies

Regions not unique to OTU

Regions unique to OTU

Bacteria; Proteobacteria; Deltaproteobacteria; Desulfovibrionales; Desulfovibrionaceae; sf_1; otu_10051

Is Cyanobacteria OTU 5157 Present?

Clone library says “NO”

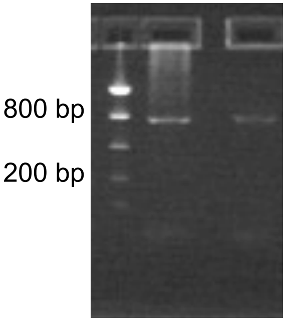
Chip says “YES”

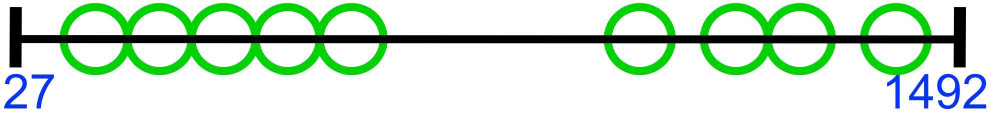
Confirmation

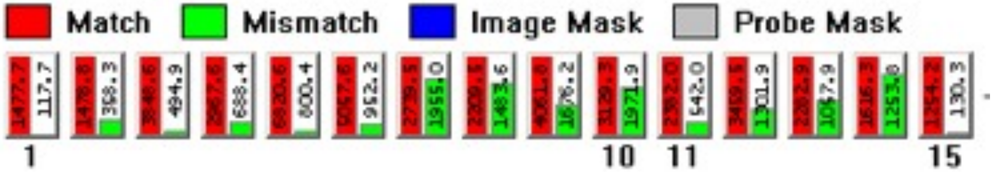
02210104000000.5157_

```
2.21.1.4.5157 OTU 9 seqs
prokMSA_id:3279 Leptolyngbya boryanum PCC 73110. NONE
prokMSA_id:3280 Leptolyngbya foveolarum str. Komarek 1964/112
prokMSA_id:3281 "Plectonema boryanum" UTEX 485
prokMSA_id:3282 "Oscillatoria" sp. str. M-117
prokMSA_id:3283 Phormidium sp. str. M-99
prokMSA_id:39175 Phormidium tenue
prokMSA_id:41034 Phormidium tenue AF337652
prokMSA_id:43289 Phormidium molle
prokMSA_id:45010 Phormidium pachydematicum
```

PCR with OTU Specific Primers says “YES”







Sequencing and BLAST of PCR says “YES”

BLAST of sequenced PCR shows 99% Identity:
Cyanobacteria
    Subgroup: Leptolyngbya
        prokMSA_id:3280 Leptolyngbya foveolarum str. Komarek 1964/112
        prokMSA_id:3281 "Plectonema boryanum" UTEX 485

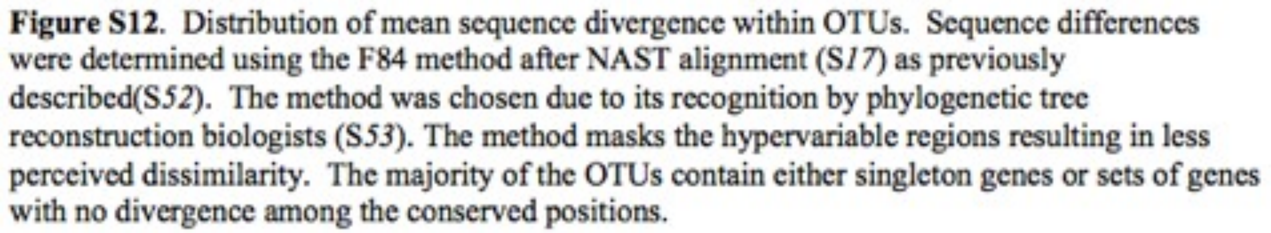


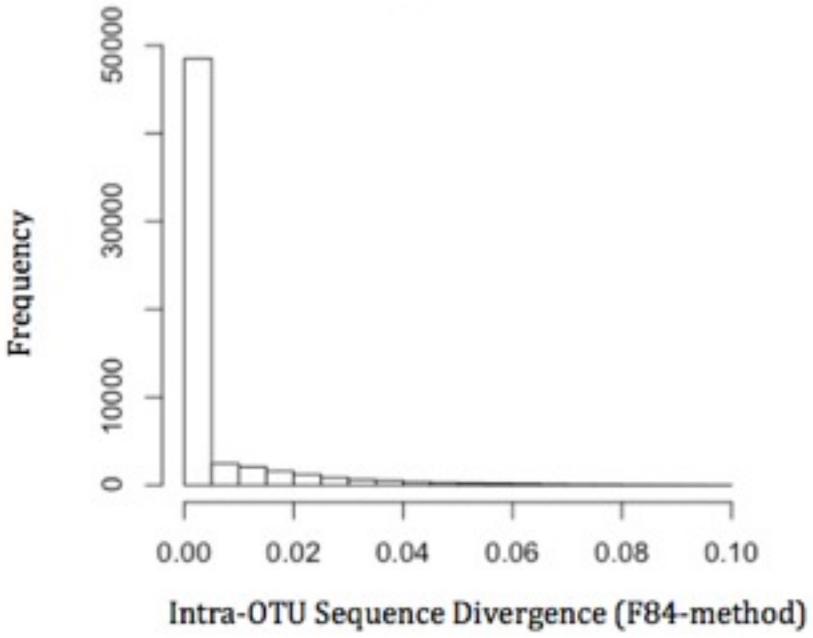
Figure S12. Distribution of mean sequence divergence within OTUs. Sequence differences were determined using the F84 method after NAST alignment (S*17*) as previously described(S*52*). The method was chosen due to its recognition by phylogenetic tree reconstruction biologists (S*53*). The method masks the hypervariable regions resulting in less perceived dissimilarity. The majority of the OTUs contain either singleton genes or sets of genes with no divergence among the conserved positions.

approximate density of 10,000 molecules per µm2

“midi 100 format” hybridization cartridges

1,016,064 probe features, arranged as a grid of 1,008 rows and columns.

Probes complementary to lower confidence 16S sequences were included to enable broadening the phylogenetic scope of analysis, when those sequences are validated with unambiguous entries into public repositories.



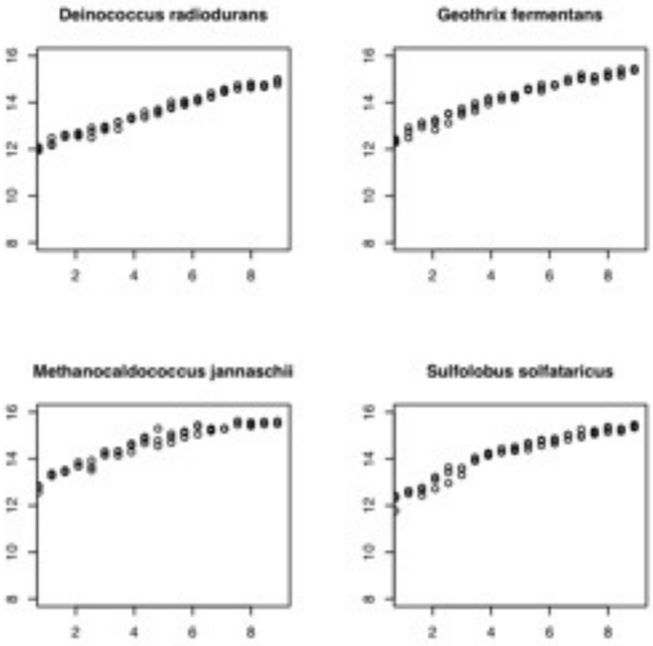
Taxonomic overlay

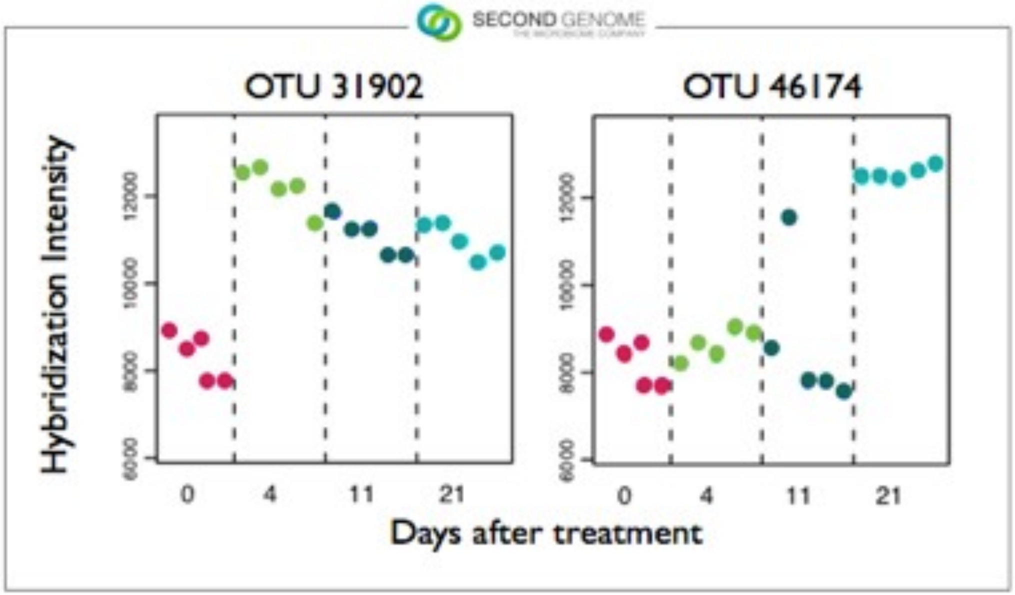
The OTUs represented **2 domains, 147 phyla, 1,123 classes**, and **1, 219 orders** demarcated within the archaea and bacteria. Each OTU was assigned to one of **1,464** families according to the placement of its member organisms in the taxonomic outline as maintained by Philip Hugenholtz (S*23*). The OTUs comprising each family were clustered into sub-families by transitive (single linkage) sequence identity of 72% common heptamers. Altogether, **10,993** sub-families were found.

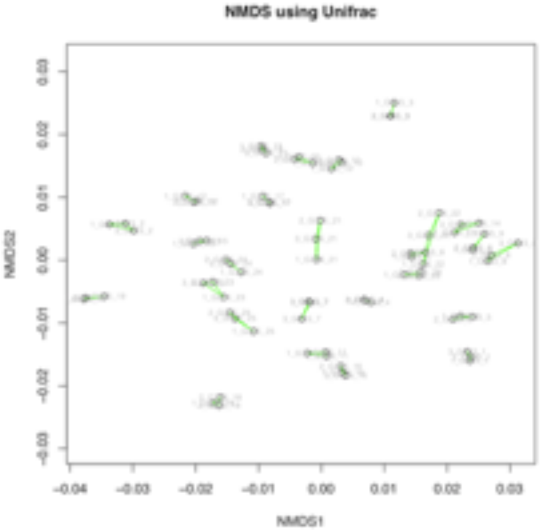
The average number of probe pairs assigned to each OTU was 37 (s.d. 9.6).

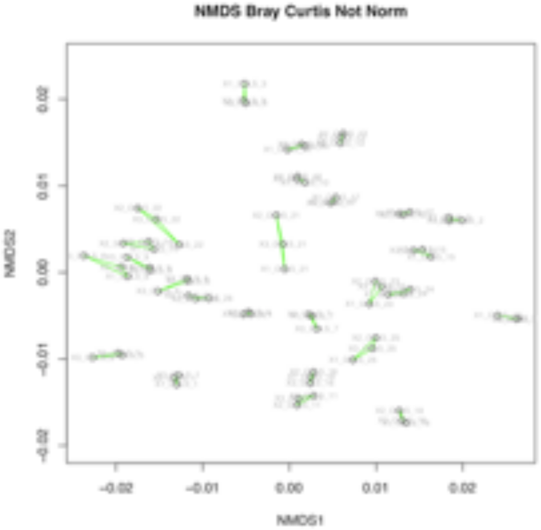
59,959 OTUs

Quantitative Tracking

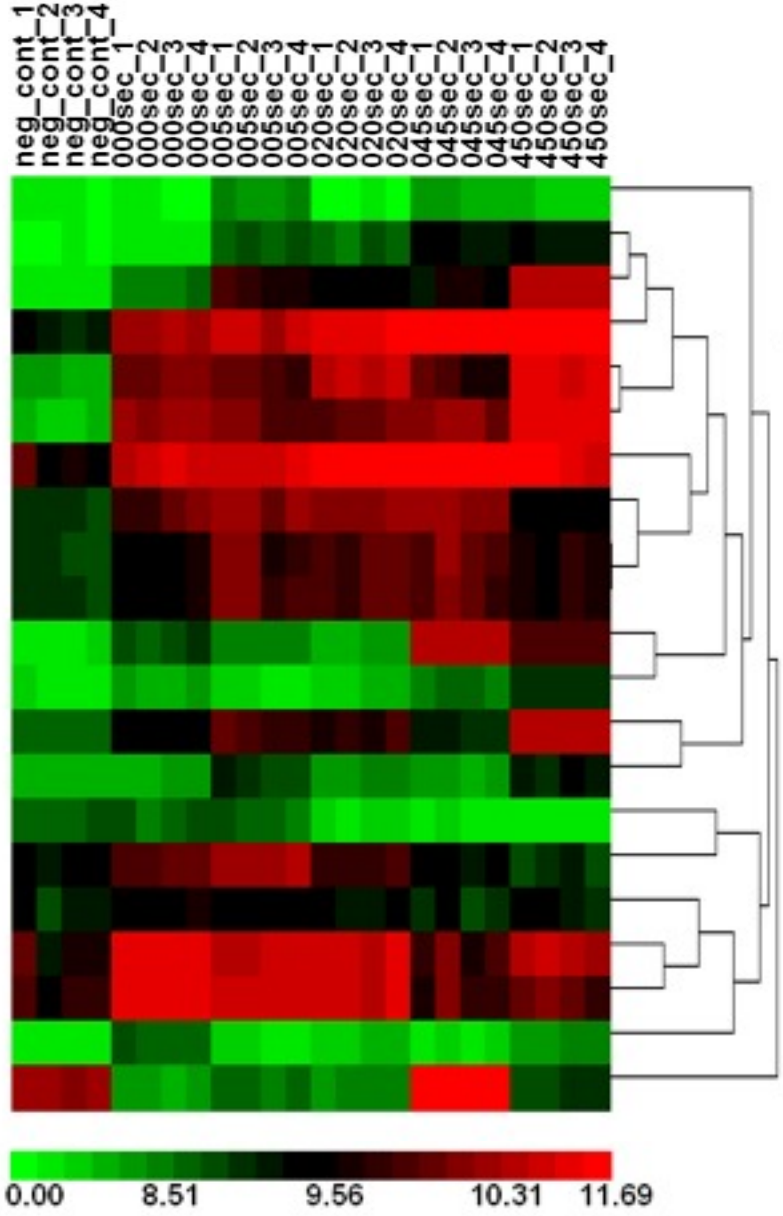






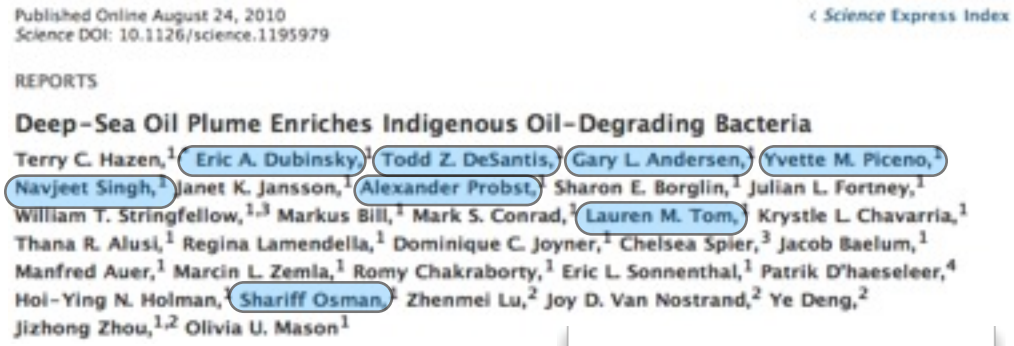


Extraction methods will affect community observations.



Family	Class	Peak Duration (sec)
Phaeophyceae (phylum)	Stramenopiles (no rank)	5
Basidiomycota (phylum)	Fungi (kingdom)	45
Family_1.1	Cyanobacteria	450
Ascomycota (phylum)	Fungi (kingdom)	450
Vibrionaceae	Gammaproteobacteria	450
Flavobacteriaceae	Flavobacteria	450
Clostridiaceae	Clostridia	45
Methylocystaceae	Alphaproteobacteria	45
Acetobacteraceae	Alphaproteobacteria	45 n.s.
Rhodospirillaceae	Alphaproteobacteria	45 n.s.
Lactobacillaceae	Bacilli	45
Staphylococcaceae	Bacilli	450
Family_3.1	Cyanobacteria	450
Family_2.2	Cyanobacteria	450
Mycoplasmataceae	Mollicutes	5
Crenotrichaceae	Sphingobacteria	5
Xanthomonadaceae	Gammaproteobacteria	0 n.s.
Burkholderiaceae	Betaproteobacteria	0
Sphingomonadaceae	Alphaproteobacteria	0
Sphingobacteriaceae	Sphingobacteria	0
Acholeplasmataceae	Mollicutes	45

DeSantis, 2005, FEMS Let



Published Online August 24, 2010
Science DOI: 10.1126/science.1195979

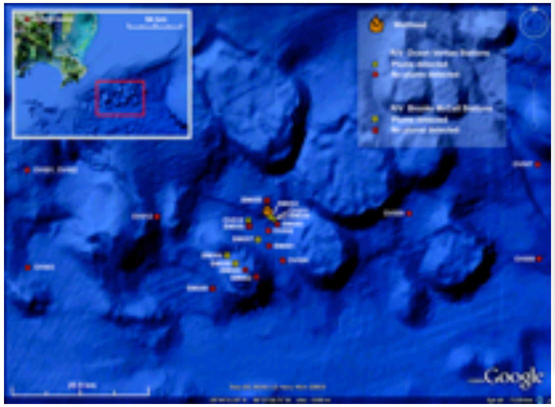
‹ Science Express Index

REPORTS

Deep-Sea Oil Plume Enriches Indigenous Oil-Degrading Bacteria

Terry C. Hazen,[1] Eric A. Dubinsky,[1] Todd Z. DeSantis,[1] Gary L. Andersen,[1] Yvette M. Piceno,[1] Navjeet Singh,[1] Janet K. Jansson,[1] Alexander Probst,[1] Sharon E. Borglin,[1] Julian L. Fortney,[1] William T. Stringfellow,[1,3] Markus Bill,[1] Mark S. Conrad,[1] Lauren M. Tom,[1] Krystle L. Chavarria,[1] Thana R. Alusi,[1] Regina Lamendella,[1] Dominique C. Joyner,[1] Chelsea Spier,[3] Jacob Baelum,[1] Manfred Auer,[1] Marcin L. Zemla,[1] Romy Chakraborty,[1] Eric L. Sonnenthal,[1] Patrik D'haeseleer,[4] Hoi-Ying N. Holman,[1] Shariff Osman,[1] Zhenmei Lu,[2] Joy D. Van Nostrand,[2] Ye Deng,[2] Jizhong Zhou,[1,2] Olivia U. Mason[1]

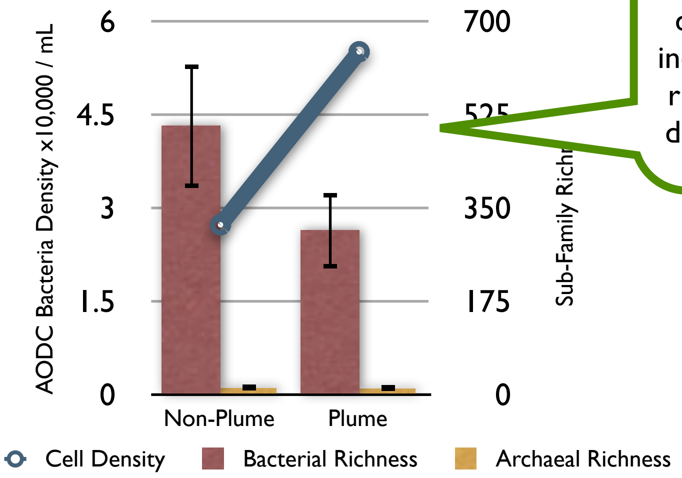
- Although the blowout is one of the largest oil spills in history …
- An ultra low-concentration hydrocarbon plume formed ~1150 m below the surface.
- Is a microbial community shift detectable?
- Which taxa are enriched by the hydrocarbons?



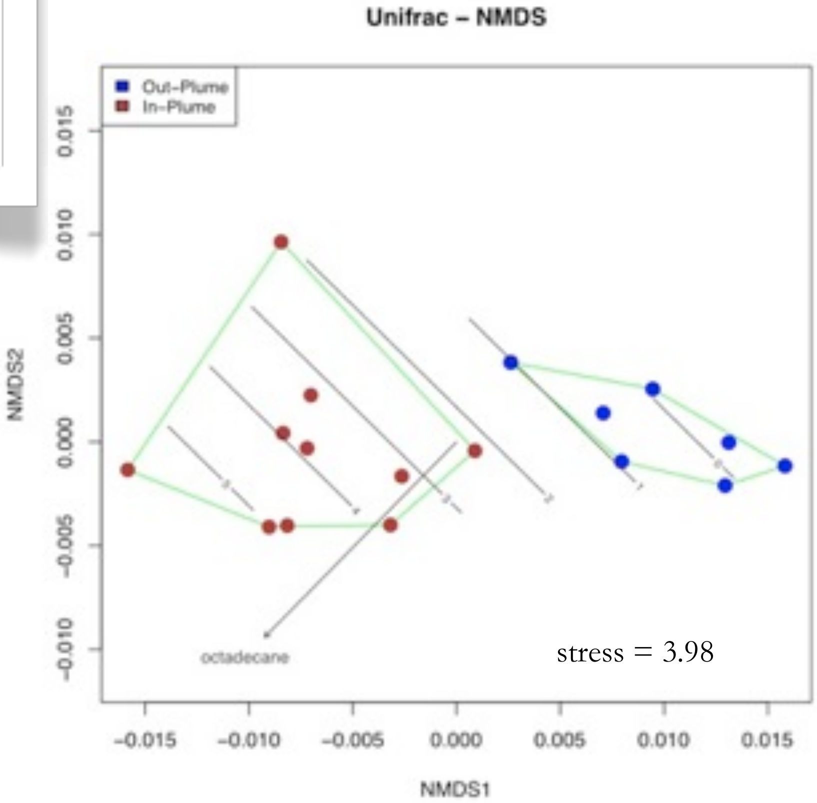
Sampling sites around the ruptured MC252 well head from May 25 to June 7, 2010.

- Hydrocarbon Increase Above Background:
  - EPA Gulf of Mexico Hydrocarbon Concentration Allowance* = 29,000 parts per billion (29.000 mg/L)
  - Deep Horizon Oil Spill Plume Hydrocarbon Concentration = 139 parts per billion (00.139 mg/L)

*The EPA NPDES (National Pollutant Discharge Effluent Standard) permits for Offshore Gulf of Mexico installations contain a NOT TO EXCEED limit of 29 ppm 42 mg/L (42 parts per million) on a daily basis AND a NOT TO EXCEED limit of 29 mg/L per day (29 parts per million) on a monthly basis.



Cell density increased, richness declined.


Unifrac – NMDS
Out-Plume
In-Plume
NMDS2
NMDS1
octadecane
stress = 3.98