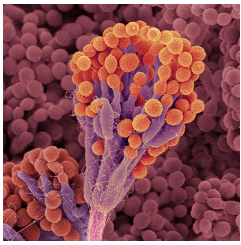
Microbes: drivers of global biogeochemistry

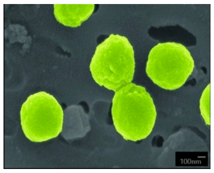
1/24/2014

GEOS 410/510

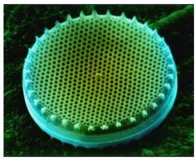
Virginia Rich


forces.si.edu

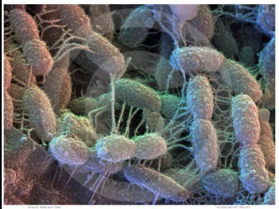
Soil fungus


proportal.mit.edu

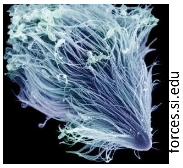
Prochlorococcus marinus


Dennis Kunkel

Diatom


www.bioquest.org

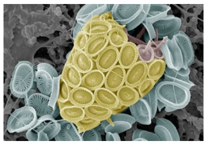
Soil bacteria


forces.si.edu

Termite protist

www.morning-earth.org/Graphic-E/Biosphere/Bios-Microbe-Image/M-PCophryoscolex

Rumen protist *Ophryoscolex*

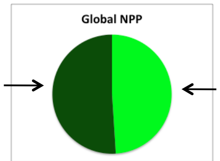

www.nhm.ac.uk

Coccolithophore

I. Big picture: microbes drive biogeochemical cycles

- ~ Half planetary primary production (C fixation):

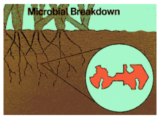
forests & other big green stuff →



← *marine microorganisms (cyanobacteria & phytoplankton)*

- Organic matter degradation:

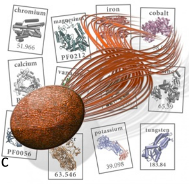
Without microbial recycling, nutrients would be locked up & become unavailable.



courses.worldcampus.psu.edu

- Metabolic diversity:

Microbes perform all major metabolic pathways, and periodically reveal entirely new ones (e.g. proteophodopsin, anaerobic methane oxidation).



http://newscenter.lbl.gov/

- Biomass: ~10^9 microbial cells/ gram surface soil and ~10^6 cells/ml seawater. (You have more microbial cells in your body than human cells). 50-90% of marine biomass is microbial (*Census of Marine Life*).

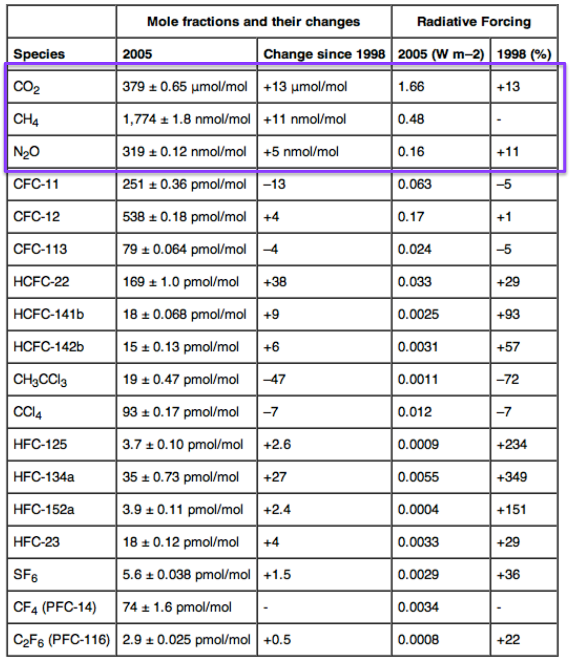
If all multi-cellular life disappeared tomorrow the major biogoechemical cycles would likely proceed with very little change...

And are critical players in GHG cycling…

Right: IPCC table of gases relevant to radiative forcing (Chapter 2, pg 141, Table 2.1. of the IPCC Fourth Assessment Report, 2007)

Of GHGs with both natural AND anthropogenic sources (~CO2, CH4 and N2O), microbes are dominant mediators of their natural cycling:

- CH4 is ~ONLY microbially mediated (then atmospheric half-life of ~10yrs)
- N2O is mainly microbially mediated (some production in atmo. too)
- CO2 is heavily microbially mediated. (~half consumption, >>half of production)



	Mole fractions and their changes		Radiative Forcing	
Species	2005	Change since 1998	2005 (W m–2)	1998 (%)
CO_2	379 ± 0.65 µmol/mol	+13 µmol/mol	1.66	+13
CH_4	1,774 ± 1.8 nmol/mol	+11 nmol/mol	0.48	-
N_2O	319 ± 0.12 nmol/mol	+5 nmol/mol	0.16	+11
CFC-11	251 ± 0.36 pmol/mol	–13	0.063	–5
CFC-12	538 ± 0.18 pmol/mol	+4	0.17	+1
CFC-113	79 ± 0.064 pmol/mol	–4	0.024	–5
HCFC-22	169 ± 1.0 pmol/mol	+38	0.033	+29
HCFC-141b	18 ± 0.068 pmol/mol	+9	0.0025	+93
HCFC-142b	15 ± 0.13 pmol/mol	+6	0.0031	+57
CH_3CCl_3	19 ± 0.47 pmol/mol	–47	0.0011	–72
CCl_4	93 ± 0.17 pmol/mol	–7	0.012	–7
HFC-125	3.7 ± 0.10 pmol/mol	+2.6	0.0009	+234
HFC-134a	35 ± 0.73 pmol/mol	+27	0.0055	+349
HFC-152a	3.9 ± 0.11 pmol/mol	+2.4	0.0004	+151
HFC-23	18 ± 0.12 pmol/mol	+4	0.0033	+29
SF_6	5.6 ± 0.038 pmol/mol	+1.5	0.0029	+36
CF_4 (PFC-14)	74 ± 1.6 pmol/mol	-	0.0034	-
C_2F_6 (PFC-116)	2.9 ± 0.025 pmol/mol	+0.5	0.0008	+22

II. How do microbes make a living?

- "Microbes" can mean several things!!! HERE defined as **single-celled organisms**: bacteria and archaea (together often called the "prokaryotes", also "microbes") plus single-celled eukaryotes

How are microbes involved in all these biogeochemical cycles?

What do microbes – indeed all cells – need to make a living?

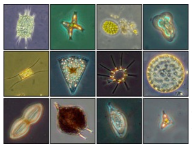
- CARBON for bulk of biomass
- NUTRIENTS (N,P, S) and micronutrients for proteins, nucleic acids, etc.
- WATER as a solvent (and a reactant in biomass production)
- ENERGY to allow them to work against entropy
- ELECTRONS to transfer energy via redox reactions, and perform chemical transformations – so a *source* and a *sink* for electrons

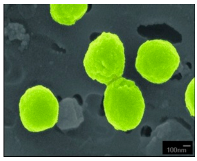
Words we use to describe where organisms get their carbon, energy, and electrons

1. Carbon

- Autotroph *Greek* autos = self, trophe = nutrition. So what is their C source? How do they get it? What are some examples?

wikipedia


earthobservatory.nasa.gov


proportal.mit.edu

- Heterotroph heteros = other, trophe = nutrition. So what is their C source? How do they get it? Examples?

Associated Press

protistuser.uni-frankfurt.de.~schauder.termites.proto7_bg

www.morning-earth.org/Graphic-E/Biosphere/Bios-Microbe-Image/M-PCophryoscolex.jpg

2. Energy

- Phototroph photo = light Energy comes from photons
- Chemotroph chemo = chemical Energy comes from converting energy stored in chemical bonds (via their electrons)

In both cases, captured energy is stored as ATP, carbs, lipids or proteins.

3. Electron Source

- Organotroph organic = C-containing. Use carbon compounds as electron donors. This includes us!
- Lithotroph lithos = rock Use inorganic compounds as electron donors

4. Electron Sink

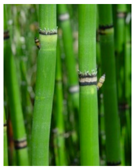
- Aerobic respiration uses O_2 as terminal electron acceptor. When it's available, it gets used because of highly favorable energetics.
- Anaerobic respiration occurs in absence of O_2, using alternate terminal electron acceptor. E.g. *denitrification* uses nitrate (NO_3^-), *sulfate reduction* uses sulfate (SO_4^{2-}).

Examples

http://www.popartuk.com/g/l/lgsb0015+da-vincis-vitruvian-man-homer-simpson-the-simpsons-art-print.jpg

- How would we be classified under this trophic nomenclature?
  - Get C from others
  - Get electrons from C compounds
  - Get energy from bond energy

Therefore we are Chemo organo heterotrophs, as are all multicellular carnivores, herbivores, and many many microbes.



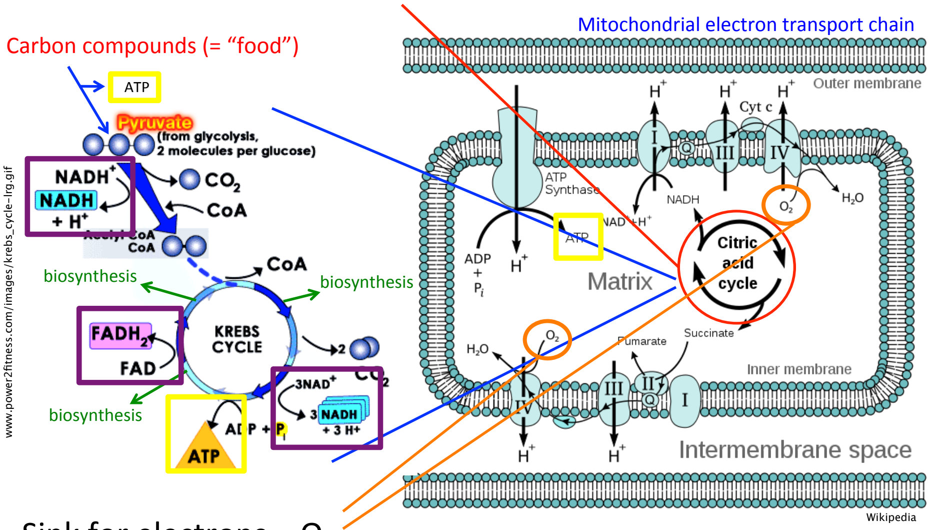
http://www.missouriplants.com/Ferns/Equisetum_hymenale_stems

- How would land plants be classified?
  - Fix CO2
  - Use sun for energy
  - What is their electron source? Is it organic or inorganic

Photo litho autotrophs

not so important to memorize terms as to understand that a diversity of lifestyles exist, & thus a diversity of interactions with biogeochem. cycles

Source for C, energy + electrons = carbon compounds


Carbon compounds (= “food”)

Mitochondrial electron transport chain

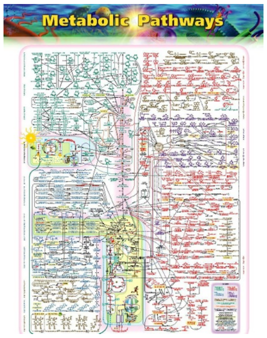
www.power2fitness.com/images/krebs_cycle-lrg.gif

Wikipedia

Sink for electrons = O_2

(producing water)

Many **biogeochemical transformations are unique to Bacteria and Archaea**, and not found in Eukaryotes, e.g.



Nitrogen fixation $N_2 \Rightarrow NH_3$

Nitrification $NH_3 \Rightarrow NO_2^- \Rightarrow NO_3^-$

Anaerobic respiration, the use of electron acceptors other than O_2

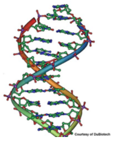
Examples

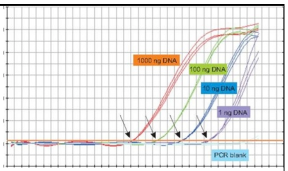
Methanogenesis CO_2 (or CH_3COOH) $\Rightarrow CH_4$

Denitrification $NO_3^- \Rightarrow N_2$

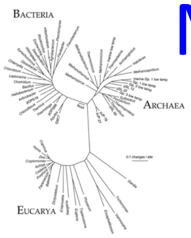
Sulfate reduction $SO_4^{2-} \Rightarrow H_2S$

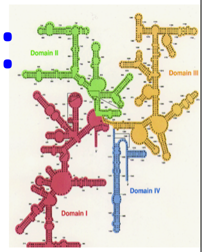
What jumps out just of this brief sampling? The N cycle is dominated by microbial transformations.

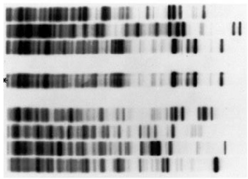


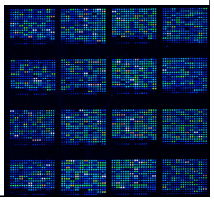


Molecular Microbial Ecology: Primer on Key Concepts & Methods





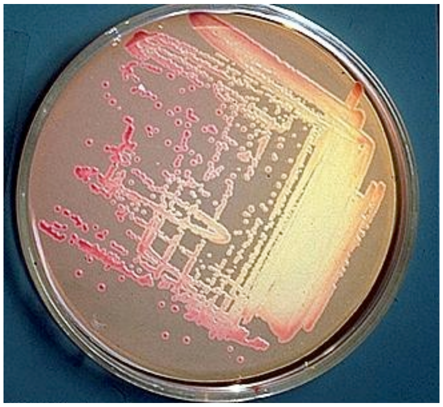




With some material from R. Maier & M. Sullivan

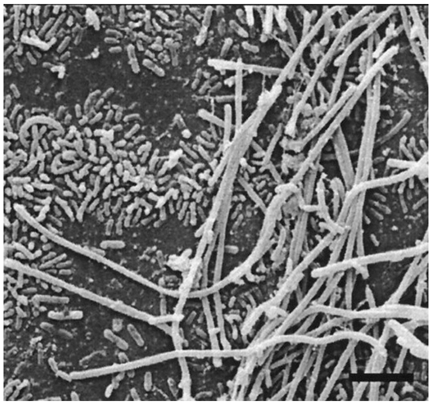
The great plate count anomaly

tame



≠

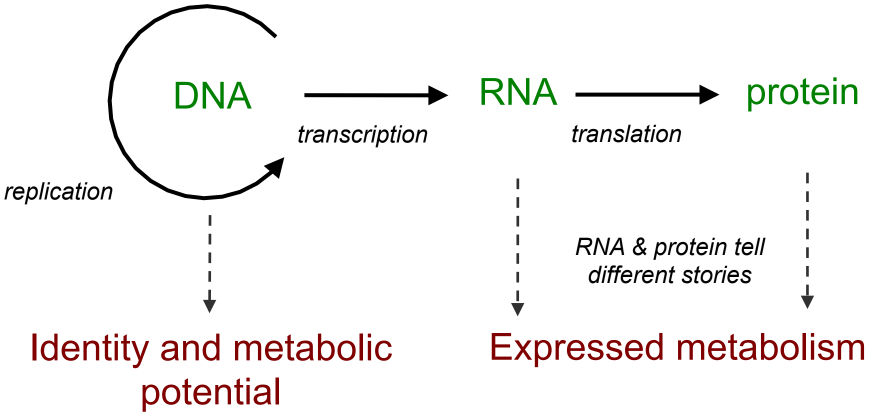
wild



most (>99%) microbes don't grow on plates

Nucleic acids & proteins are a lens into microbial identity, metabolic potential, and expressed activity

The "Central Dogma" of Molecular Biology:



DNA → RNA → protein

replication

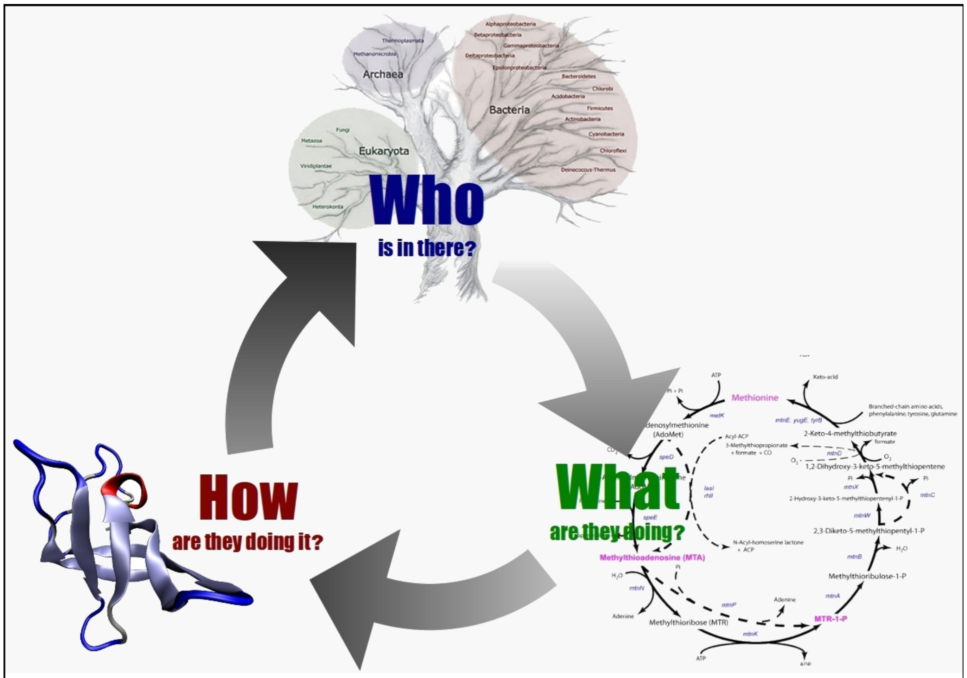
transcription

translation

RNA & protein tell different stories

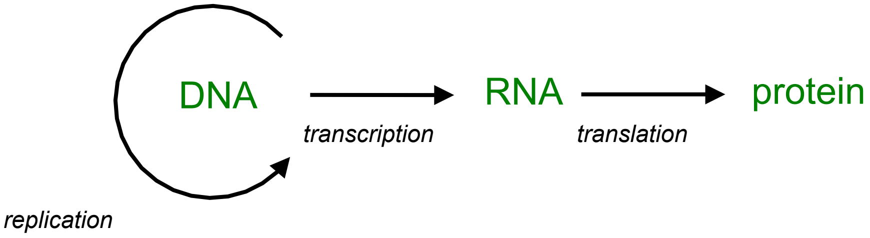
Identity and metabolic potential

Expressed metabolism



http://www.cbs.dtu.dk/researchgroups/metagenomics/mg.jpg

We want to access these 3 biomolecule types: how?



DNA

transcription

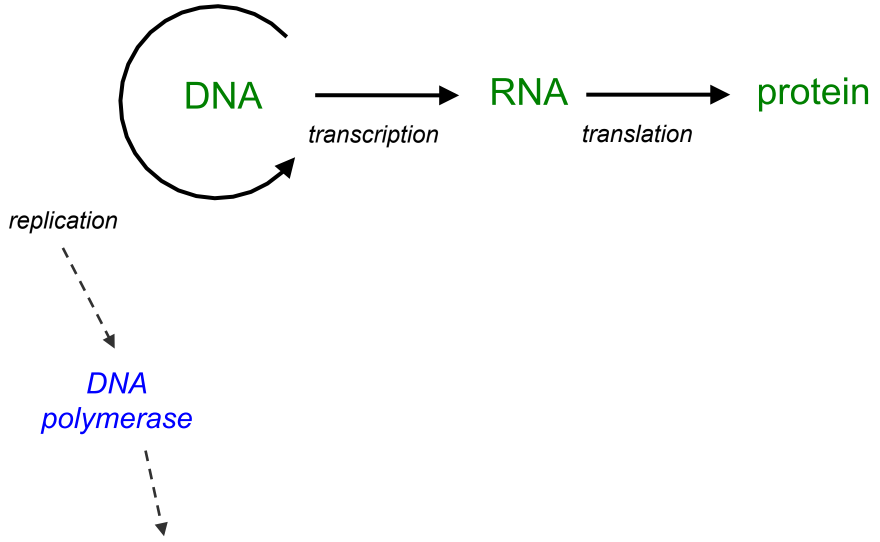
RNA

translation

protein

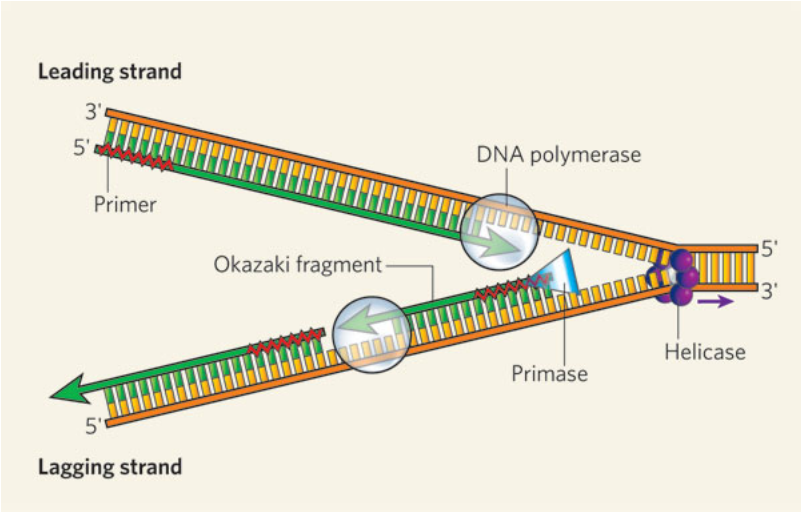
replication

How do you access the information in these molecules?

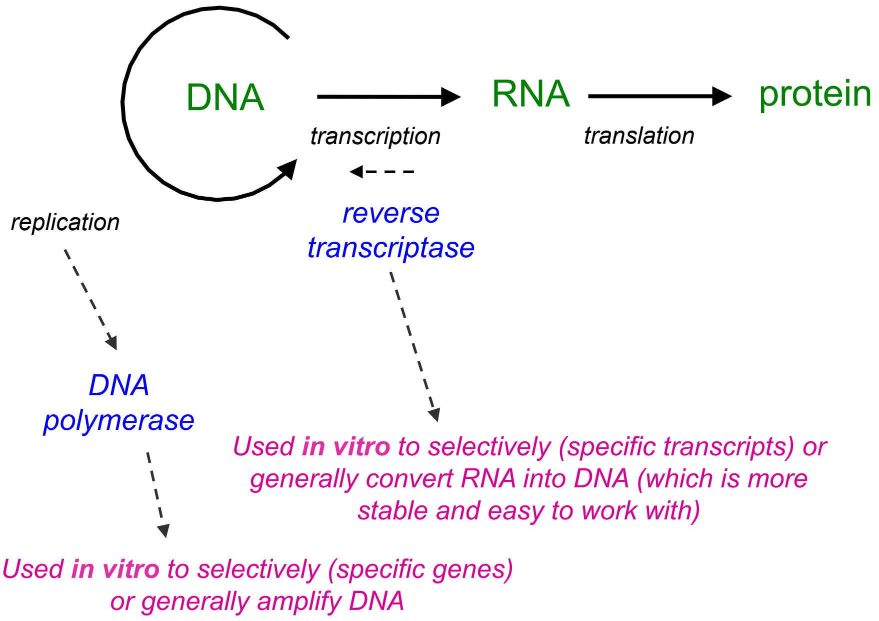


Used ***in vitro*** *to selectively (specific genes)*
or generally amplify DNA

DNA Polymerase copies DNA
but needs a primer to start from, and
needs the double helix to be opened



How do you access the information in these molecules?



DNA → RNA → protein

transcription

translation

replication

reverse transcriptase

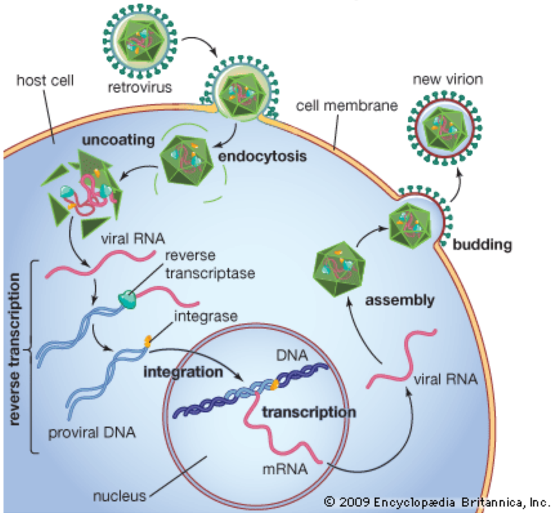
DNA polymerase

*Used **in vitro** to selectively (specific transcripts) or generally convert RNA into DNA (which is more stable and easy to work with)*

*Used **in vitro** to selectively (specific genes) or generally amplify DNA*

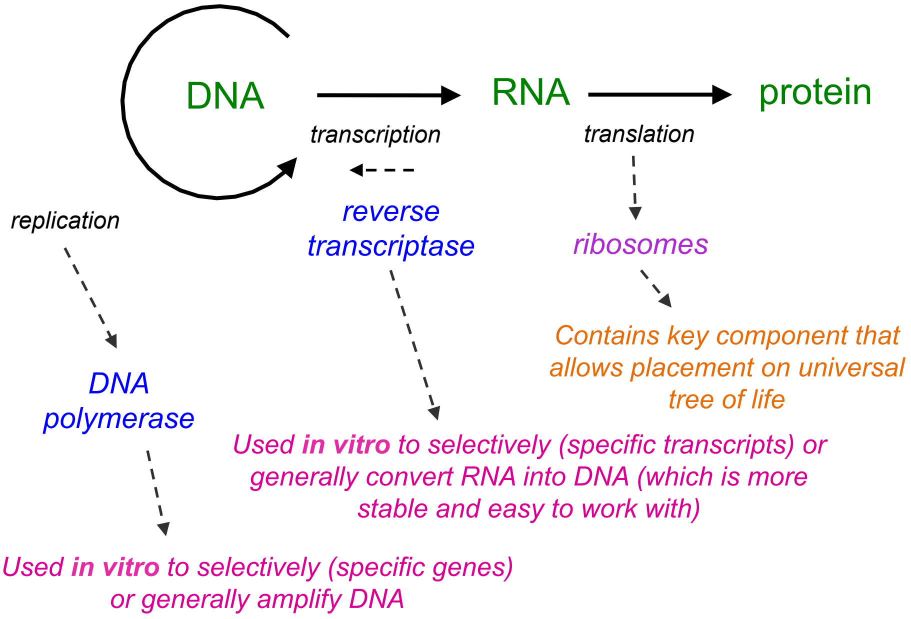
Reverse transcriptase copies RNA into DNA, as an essential step of retrovirus life cycle, telomere maintenance, etc.

Retrovirus infection and reverse transcription



© 2009 Encyclopædia Britannica, Inc.

How do you access the information in these molecules?



DNA

RNA

protein

transcription

translation

replication

reverse transcriptase

ribosomes

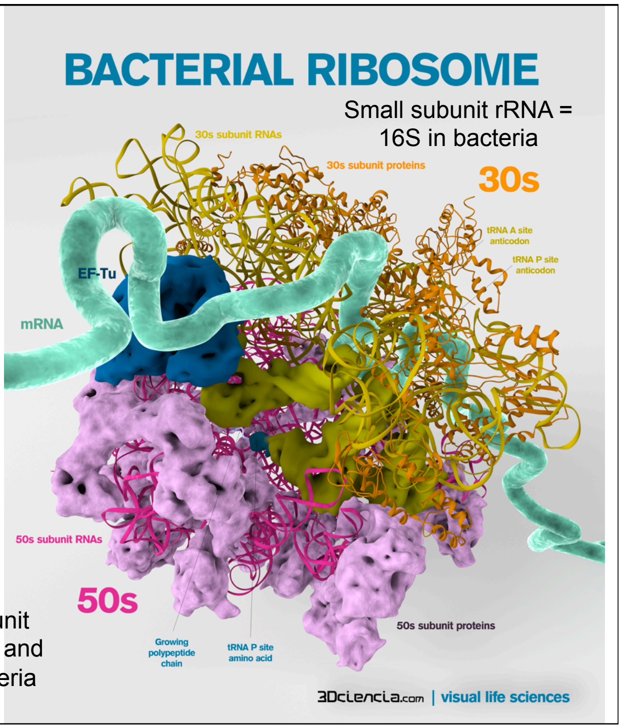
DNA polymerase

Contains key component that allows placement on universal tree of life

*Used **in vitro** to selectively (specific transcripts) or generally convert RNA into DNA (which is more stable and easy to work with)*

*Used **in vitro** to selectively (specific genes) or generally amplify DNA*

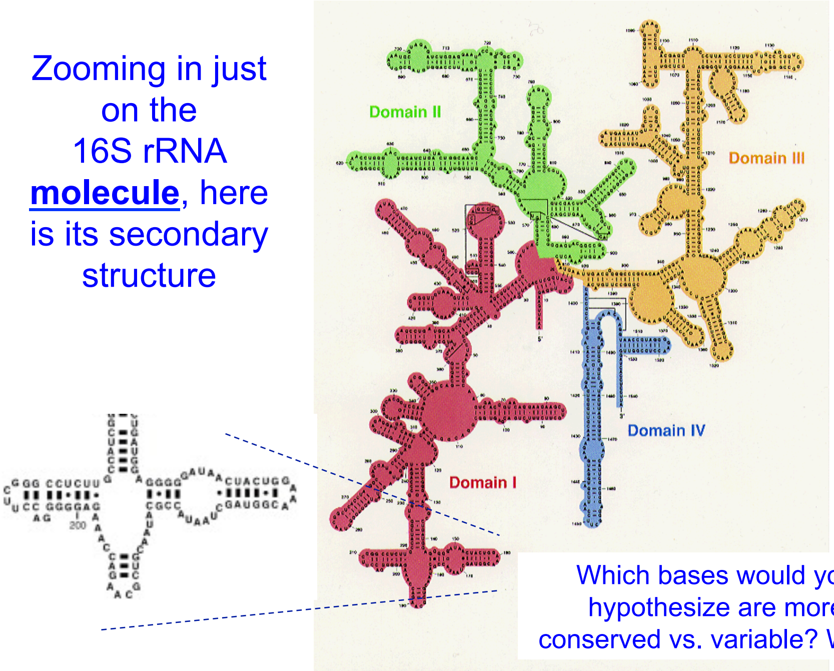
The small subunit ribosomal RNA aka 16S rRNA is central to many environmental molecular microbial investigations



Small subunit rRNA = 16S in bacteria

Large subunit rRNAs = 5S and 23S in bacteria

Zooming in just on the 16S rRNA **molecule**, here is its secondary structure

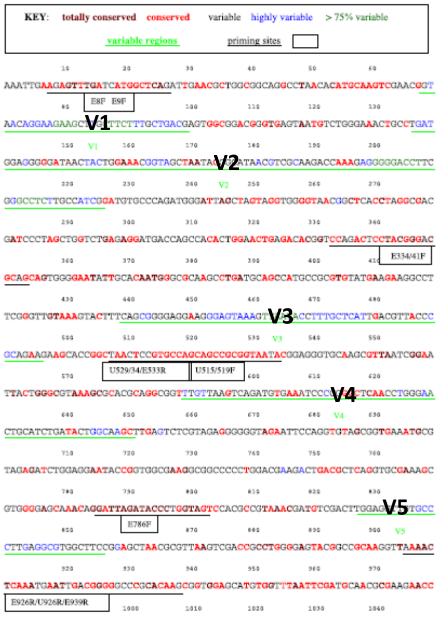


Which bases would you hypothesize are more conserved vs. variable? Why?

Overhead transparencies to accompany Garrett/Grisham: *Biochemistry* page 242
Transparency 33 Figure 7.38 ©1995 Saunders College Publishing

And the primary structure (sequence) of the 16S rRNA **gene**

G.C. Baker et al. / Journal of Microbiological Methods 55 (2003) 541–555



approximately 1500 base pairs
highly conserved regions (red)
hypervariable regions (green underlined)

G.C. Baker et al. / Journal of Microbiological Methods 55 (2003) 541–555 543

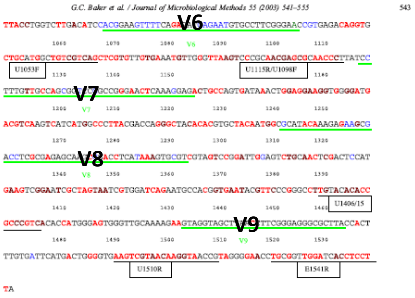
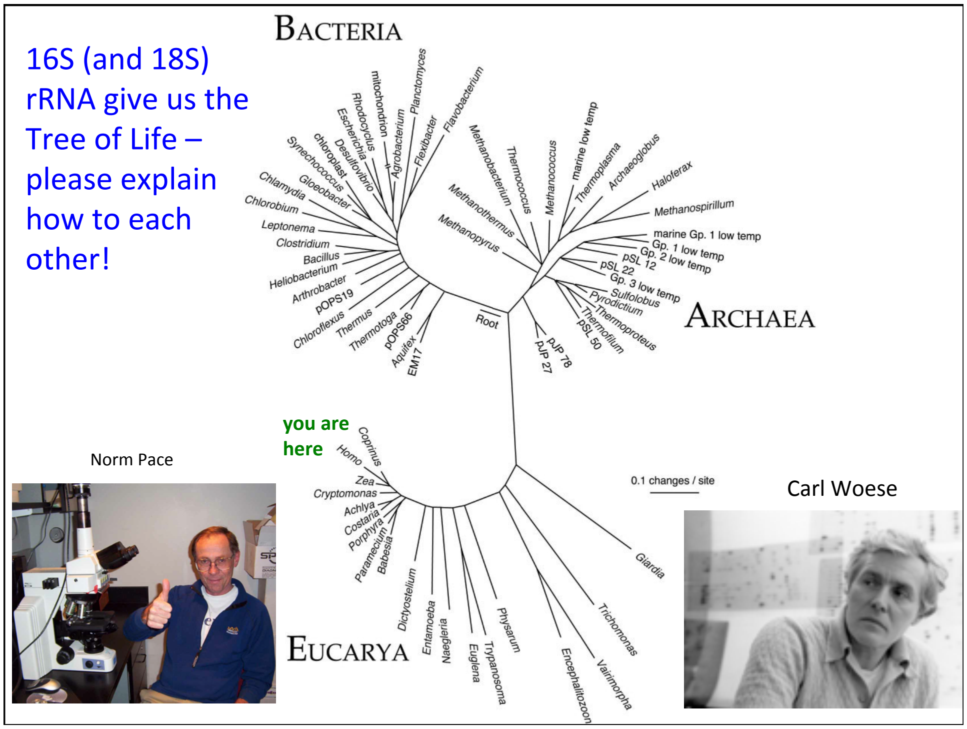


Fig. 1. Bacterial variability map. *E. coli* 16S rRNA gene sequence annotated with bacteria and "universal" priming sites and variable regions V1–V9 (41). The sequence is colour coded to indicate bacterial sequence variability. It is based on the variability map for the 16S rRNA gene produced by Van de Peer et al. (1996), translated into text format (http://www.silk.uia.ac.be/varmaps).

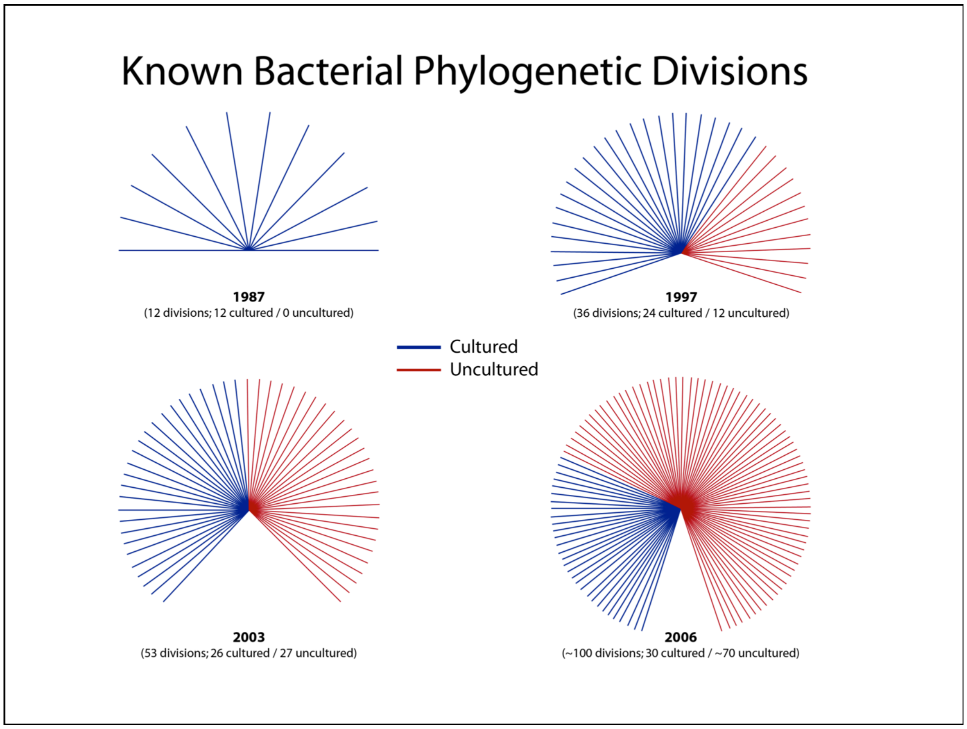
- Conserved regions allow for divergent sequences to be **aligned** for tracking evolutionary relationships
- Hypervariable regions can provide *short* **species-specific** signature sequences useful for identification.



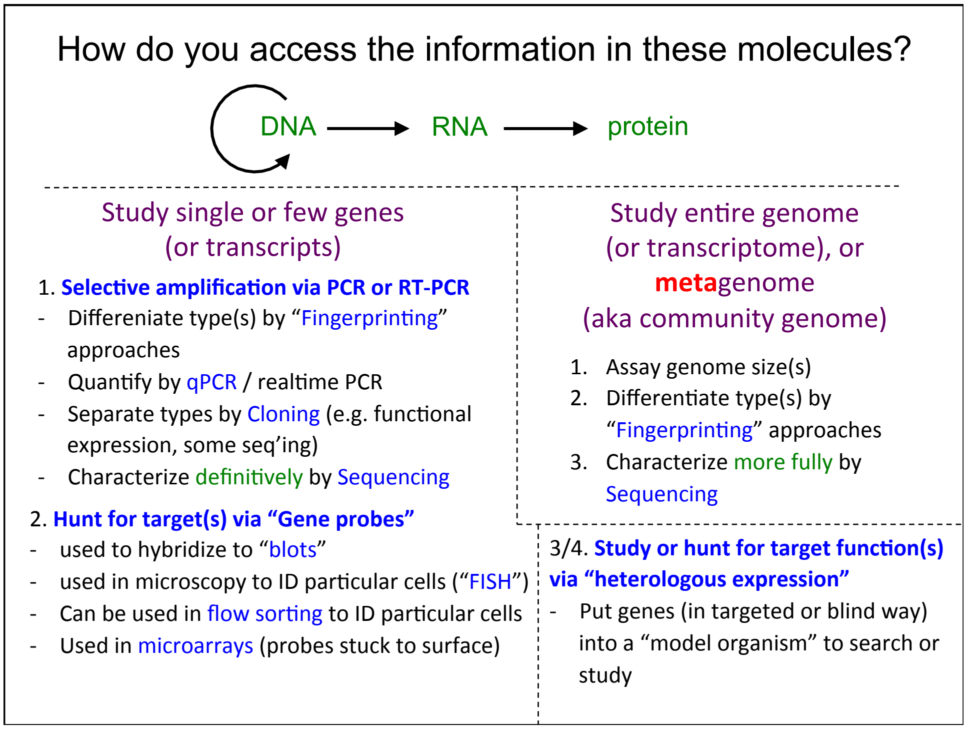
16S (and 18S) rRNA give us the Tree of Life – please explain how to each other!

Norm Pace

Carl Woese



Known Bacterial Phylogenetic Divisions



How do you access the information in these molecules?

DNA → RNA → protein

Study single or few genes (or transcripts)

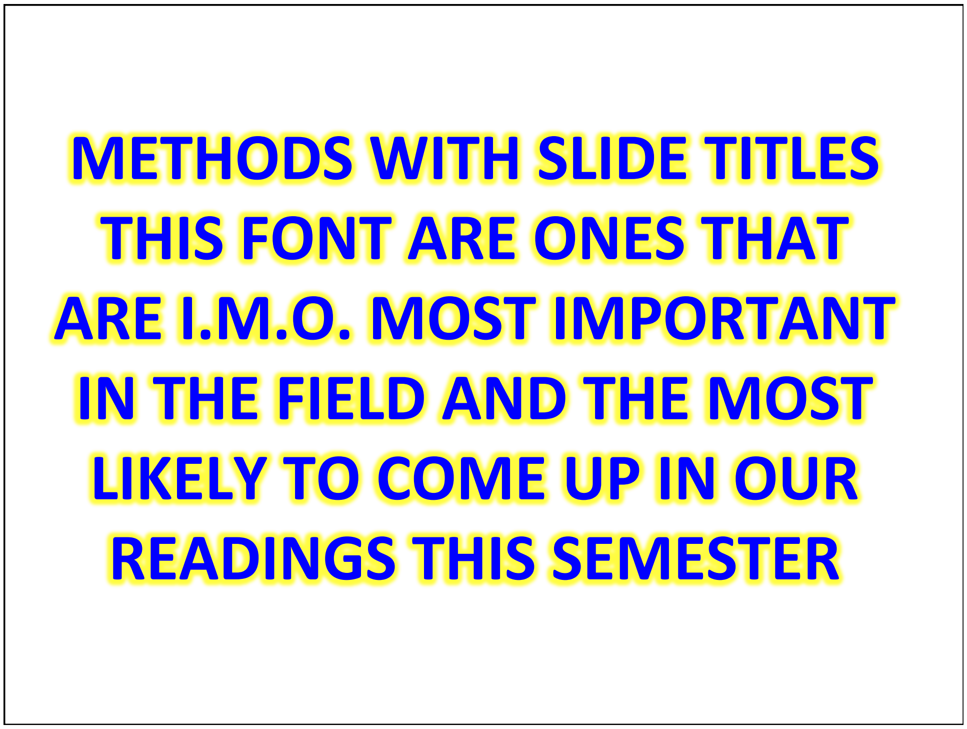
1. **Selective amplification via PCR or RT-PCR**
   - Differeniate type(s) by "Fingerprinting" approaches
   - Quantify by qPCR / realtime PCR
   - Separate types by Cloning (e.g. functional expression, some seq'ing)
   - Characterize definitively by Sequencing

2. **Hunt for target(s) via "Gene probes"**
   - used to hybridize to "blots"
   - used in microscopy to ID particular cells ("FISH")
   - Can be used in flow sorting to ID particular cells
   - Used in microarrays (probes stuck to surface)

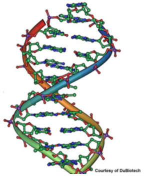
Study entire genome (or transcriptome), or **meta**genome (aka community genome)

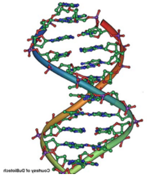
1. Assay genome size(s)
2. Differentiate type(s) by "Fingerprinting" approaches
3. Characterize more fully by Sequencing

3/4. **Study or hunt for target function(s) via "heterologous expression"**
- Put genes (in targeted or blind way) into a "model organism" to search or study

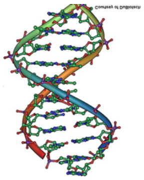


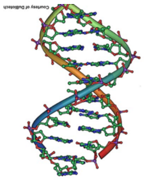
METHODS WITH SLIDE TITLES THIS FONT ARE ONES THAT ARE I.M.O. MOST IMPORTANT IN THE FIELD AND THE MOST LIKELY TO COME UP IN OUR READINGS THIS SEMESTER





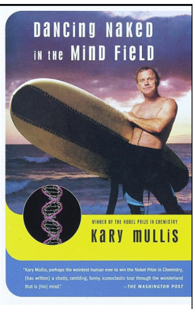
PCR





Kary Mullis, the inventor of PCR

Later became a denier of climate change and of the HIV/AIDS link


Dancing Naked
in the Mind Field
Winner of the Nobel Prize in Chemistry
Kary Mullis
"Kary Mullis, perhaps the weirdest human ever to win the Nobel Prize in Chemistry, [has written] a chatty, rambling, funny, iconoclastic tour through the wonderland that is [his] mind."
—The Washington Post

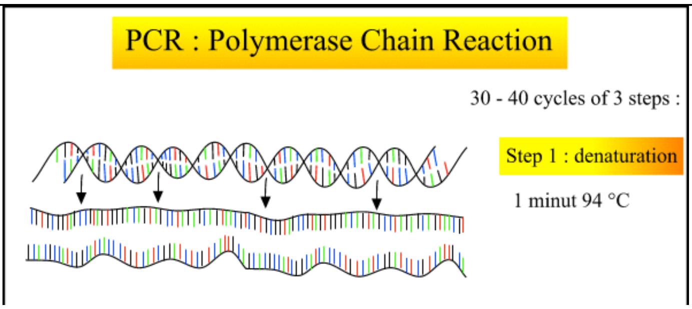
You need 4 ingredients for PCR:

1. Template DNA (that you are copying from)

2. Primers – either specific or random (how do these change what gets amplified?)

3. dNTPs, the building blocks of DNA

4. DNA Polymerase (original was “Taq” polymerase from *Thermus aquaticus*, a hot spring microbe; now there are many other DNA polymerases available)

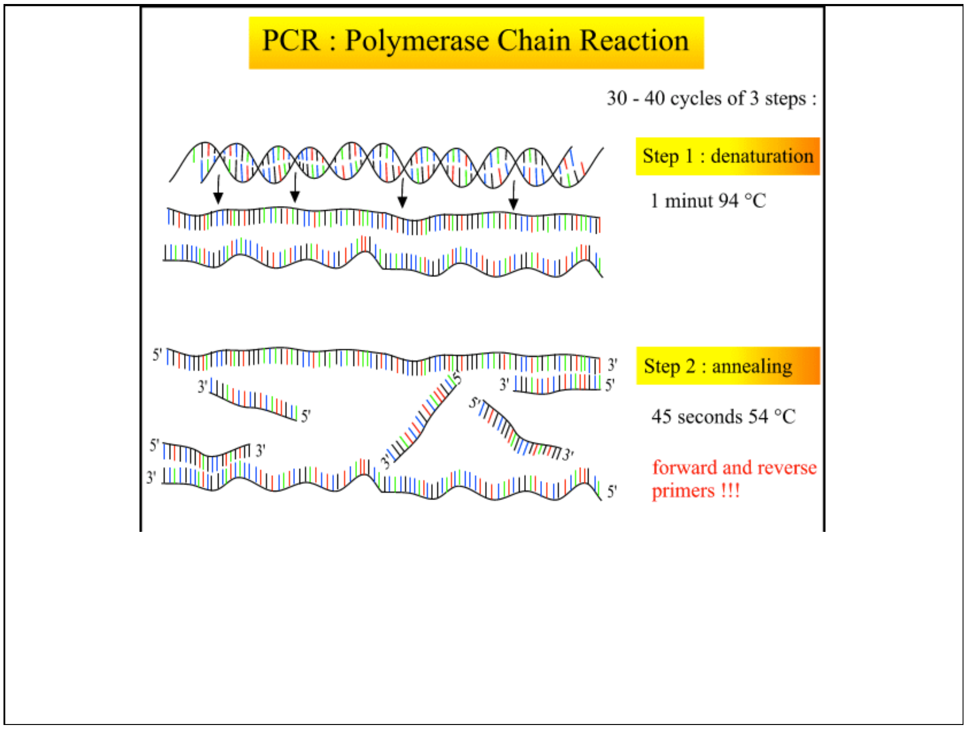


PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :

Step 1 : denaturation

1 minut 94 °C



PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :

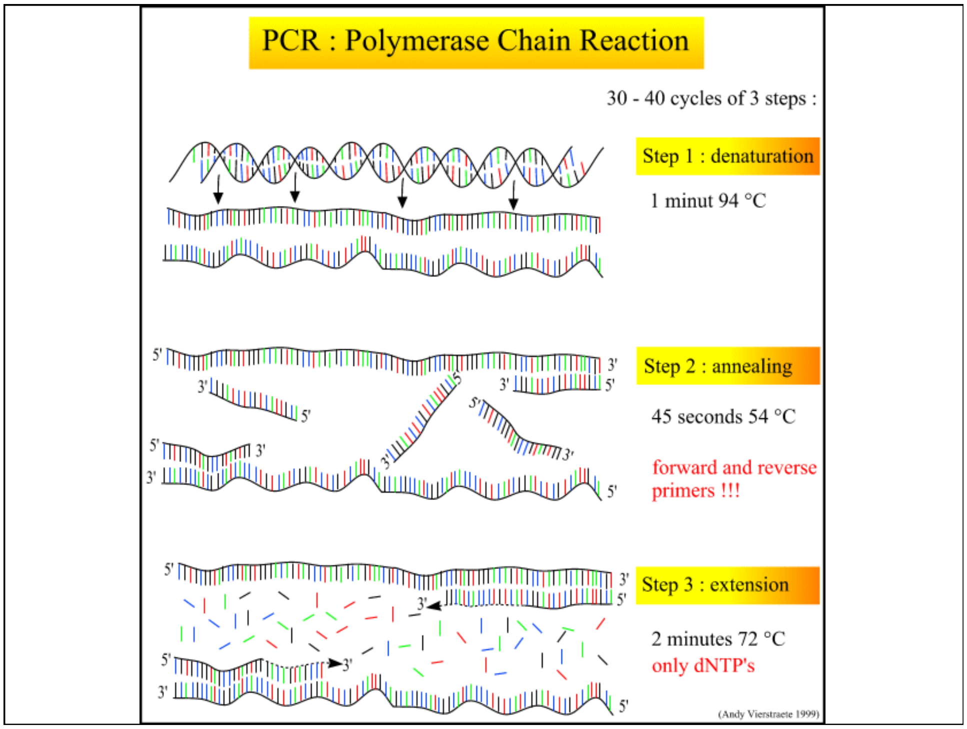
Step 1 : denaturation

1 minut 94 °C

Step 2 : annealing

45 seconds 54 °C

forward and reverse primers !!!



PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :

Step 1 : denaturation

1 minut 94 °C

Step 2 : annealing

45 seconds 54 °C

forward and reverse primers !!!

Step 3 : extension

2 minutes 72 °C

only dNTP's

(Andy Vierstraete 1999)

Setting up and Running your PCR

PCR hood

2. Why prepare the reaction in a PCR hood?

4 Key Ingredients of PCR:

1. **Template DNA**
2. **Primers**
3. **dNTPs**
4. **DNA polymerase**

1. Which of these determines which gene gets amplified?

3 Basic Stages of PCR:

1. **Denaturation**
2. **Annealing**
3. **Elongation**

Thermal cycler

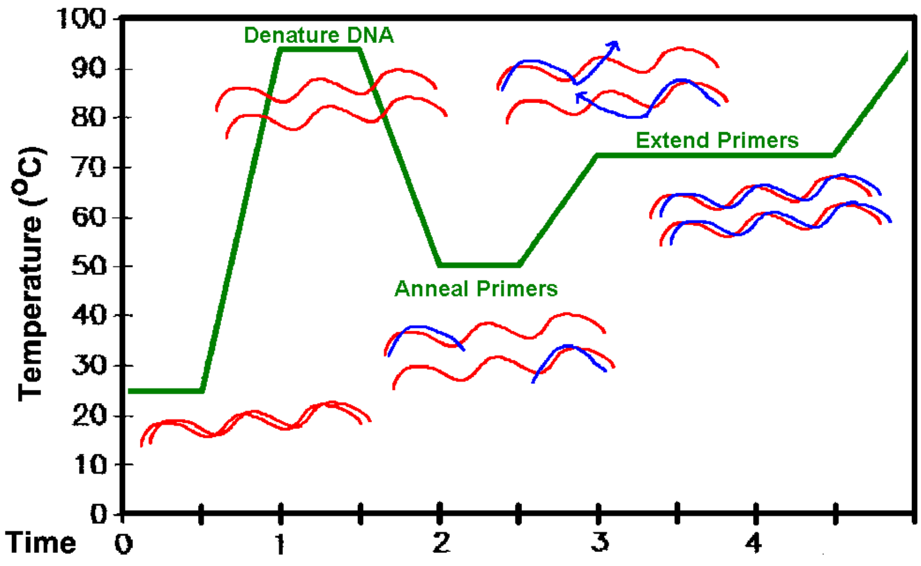
3. Can you draw a typical PCR temperature cycle?

4. Having the temperature too low in which of these stages could lead to non-target amplification?

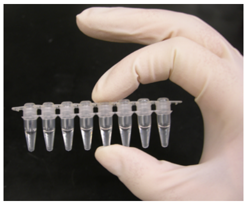
5. Having the temperature too low in which stage might cause poor amplification overall?

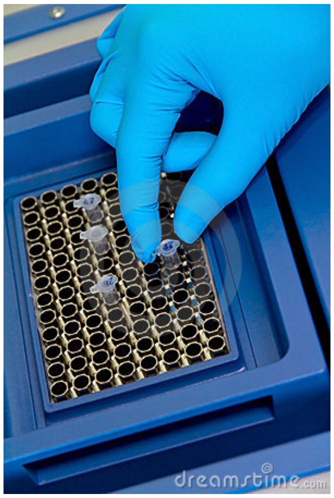
Rest of PCR slides are to review at home if you don’t recall details…

Temperature control in a PCR thermocycler

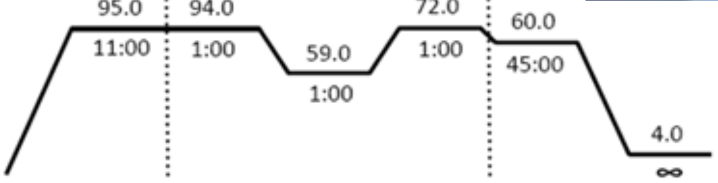


http://www.mun.ca/biology/scarr/PCR_sketch_3.gif

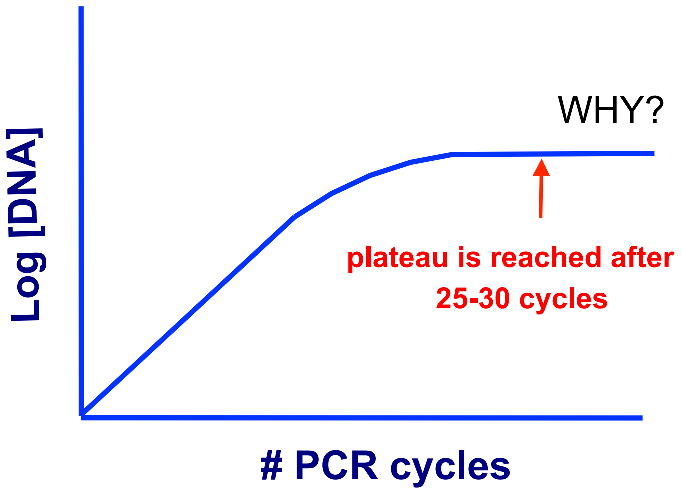




X cycles



After 25 cycles have ~ 3.4 x 10^7 times more DNA



WHY?

plateau is reached after 25-30 cycles

<u>Primer Design</u>

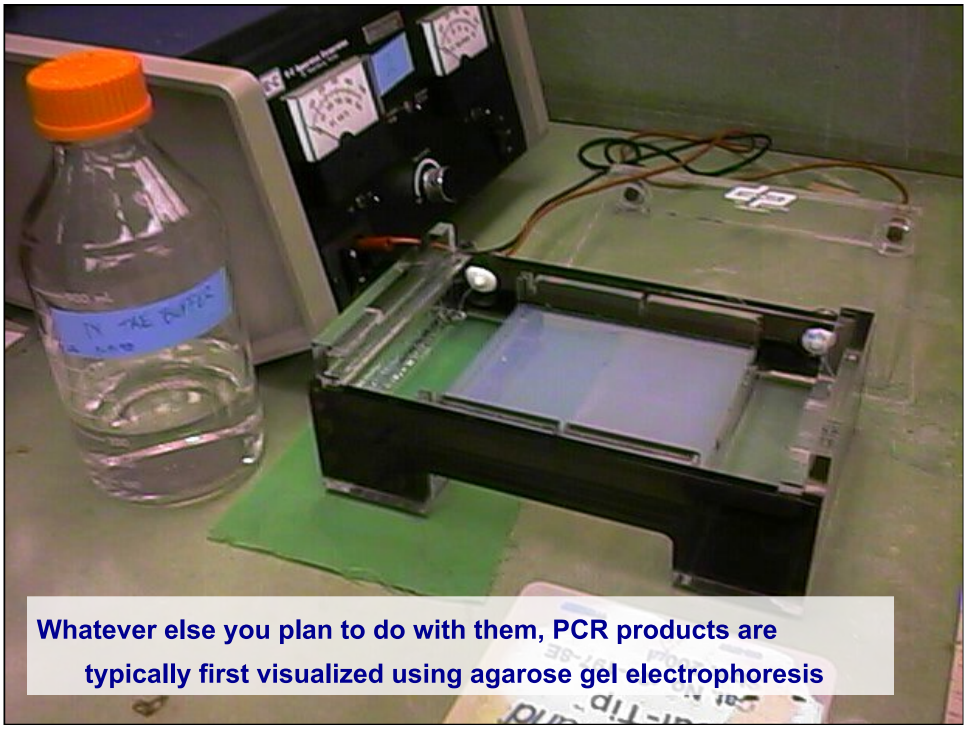
- Primer length – 17 to 30 bp
- GC content > 50%
- Conserved sequences - universal
  16S rDNA
  Dehydrogenase genes
- Conserved sequences – genus level
  Nod genes
  Rhl genes
  LamB genes

+ RANDOM PRIMING FOR NON-SPECIFIC AMPLIFICATION

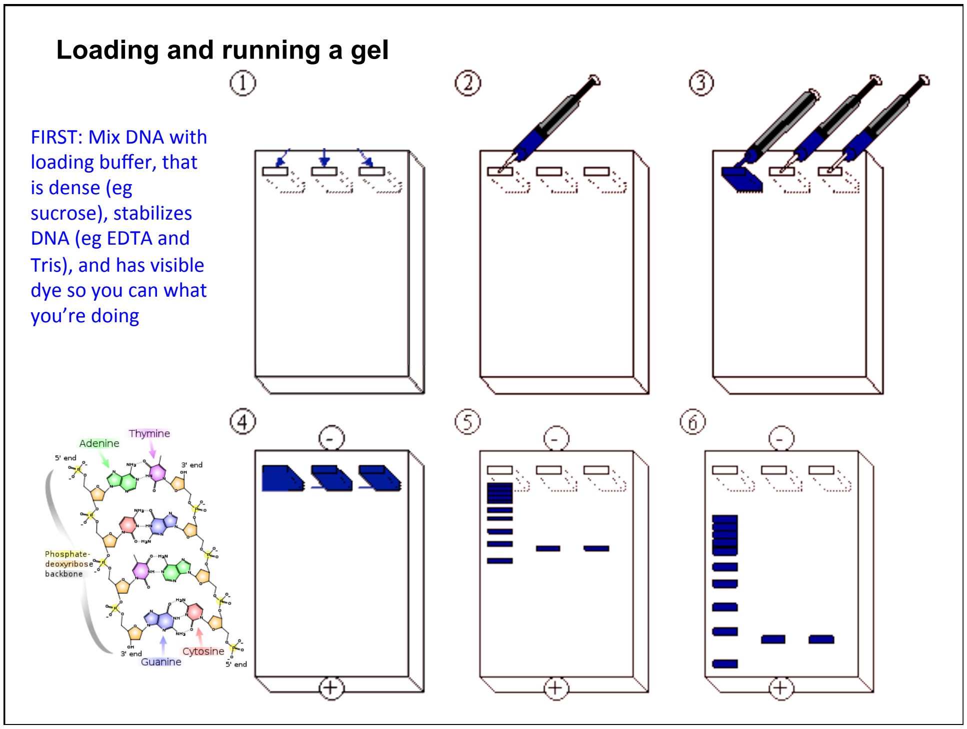
General considerations when doing PCRs:

⚠

- Include no-template negative control!
- Include positive control!
- Identify your limit of detection (sensitivity)
- Caution when using degenerate primers and/or mixed templates
  - all targets may not amplify equivalently… (why not?)
  - some variant primers getting used up before others (→ use “reconditioning PCR”)
  - stochastic variation in early rounds of amplification can have big effect (pool rxns)

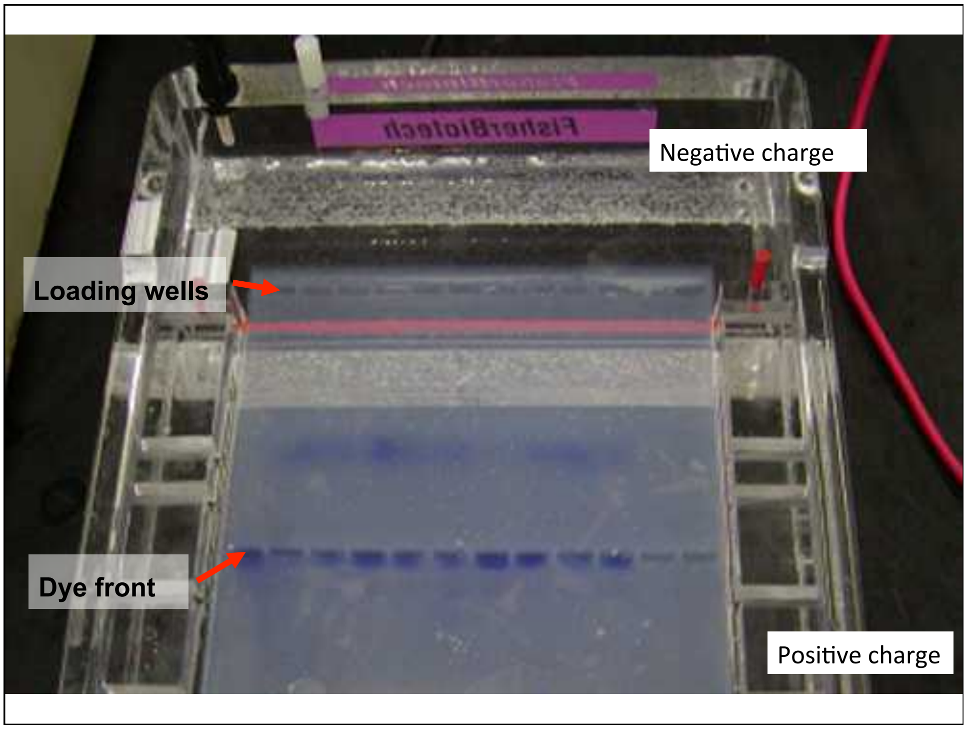


Whatever else you plan to do with them, PCR products are typically first visualized using agarose gel electrophoresis

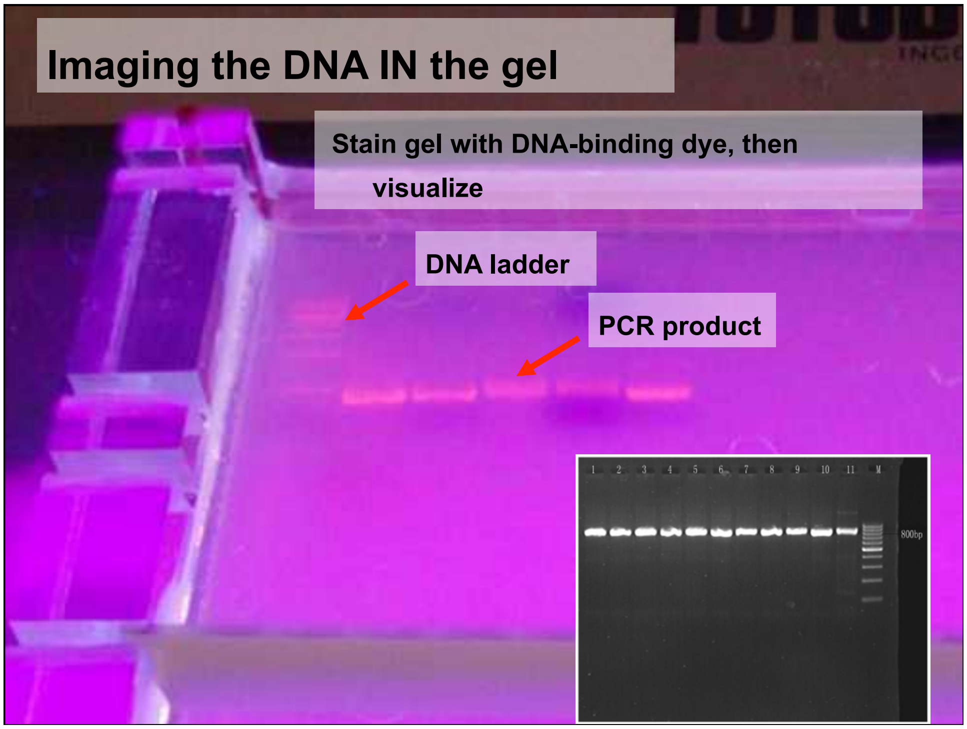


Loading and running a gel

FIRST: Mix DNA with loading buffer, that is dense (eg sucrose), stabilizes DNA (eg EDTA and Tris), and has visible dye so you can what you're doing



Imaging the DNA IN the gel



Stain gel with DNA-binding dye, then visualize

DNA ladder

PCR product

Biorad’s (first) PCR song…

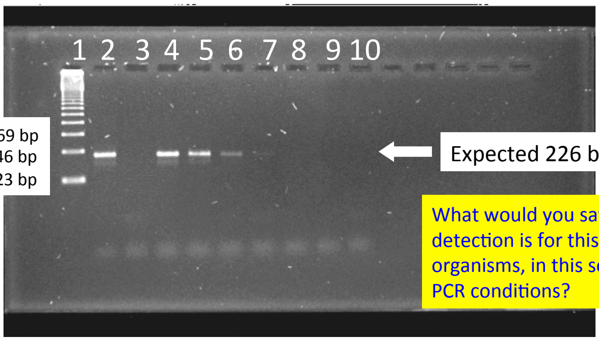
http://youtu.be/x5yPkxCLads

EXAMPLE

GOAL: Identifying limit-of-detection of PCR of the *rhlB* gene in soil DNA extracts (to know how well you can use PCR to detect this gene in the environment).

DESIGN: Sterile Gila soil was inoculated with *P. aeruginosa*, which carries *rhlB*.

rhlB *gene is part of rhamnolipid production pathway…*



Ladder 369 bp
246 bp
123 bp

Expected 226 bp product

What would you say the limit of detection is for this gene, from this organisms, in this soil, under these PCR conditions?

Why include lane 2?

Why include lane 3?

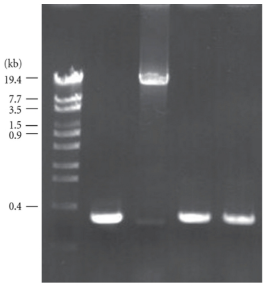
Lane 2 – *P. aeruginosa*
Lane 3 – *E. coli*
Lane 4 – 10^6 cells/g
Lane 5 – 10^5 cells/g
Lane 6 - 10^4 cells/g

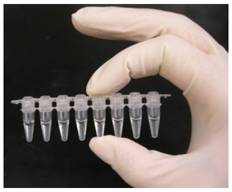
Lane 7 - 10^3 cells/g
Lane 8 - 10^2 cells/g
Lane 9 – sterile soil
Lane 10- water

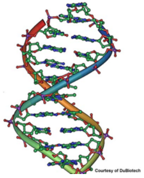
A PCR product should be confirmed to be what you think it is in at least two ways initially.

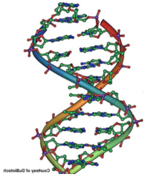
These can include:

1. **Correct product size.**
2. **Sequence the product.**
3. **RFLP analysis (see later).**
4. **Use a gene probe to confirm the product (see later).**
5. **Use alternate PCR approaches… (eg seminested PCR, won't discuss here)**

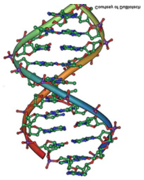


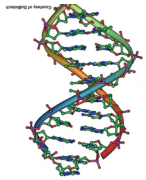






RT-PCR

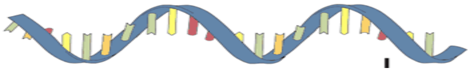




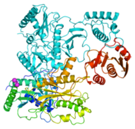
RT-PCR

Reverse transcriptase (RT) is a naturally-occurring enzyme used by VIRUSES and by some regions of eukaryotic and bacterial chromosomes for replication via an RNA stage.

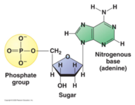
mRNA molecule



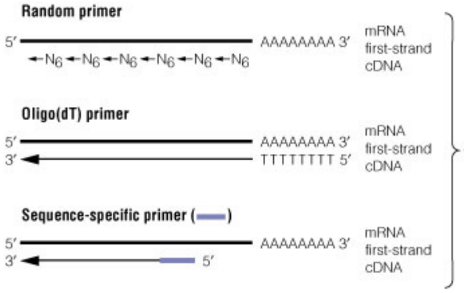
+ reverse transcriptase



+ dNTPs



+ primer, type selected based on mRNA origin & amplification goal

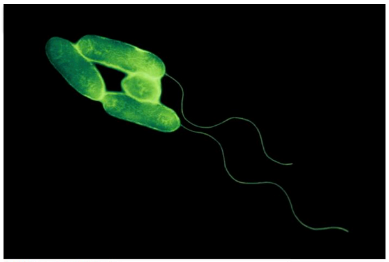


→ **Normal PCR with two primers**

The following RT-PCR example is review at home if you don't know this method well already

Is a gene present vs. it expressed: RT-PCR example

Working to understand microbial hydrocarbon degradation in model lab organism, *Psuedomonas*

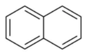


BA2143 [RM] © www.visualphotos.com

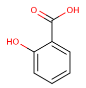
Pseudomonas cells

(visualized by what type of microscopy?)

Pseudomonas napthalene-degradation gene *nahAc* gene, induced by 2 substrates: naphthalene &salicylate.

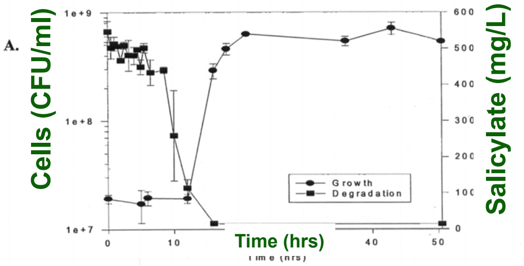


napthalene



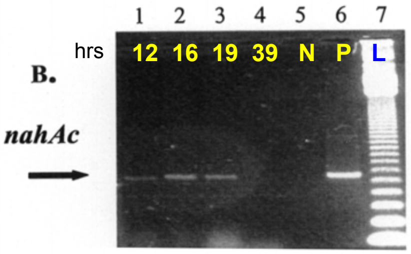
salicylate

Growth of *Pseudomonas* on salicylate over time showing degradation of substrate

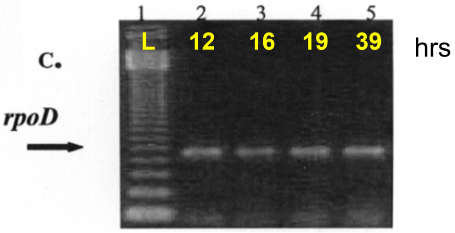


Marlowe, Wang, Pepper and Maier, 2002. AEM mRNA was extracted during growth of a *Pseudomonas* on salicylate

RT-PCRs were performed,
and are visualized here by gel electrophoresis

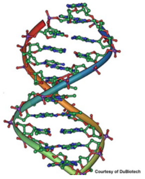
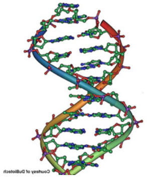


(*nahAc* is napthalene-degradation gene)

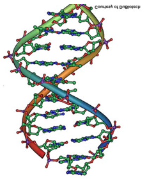
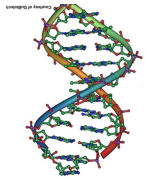


Turn to your neighbor:
1 of you explain panel B,
the other explain panel C

rpoD is a **housekeeping gene**, always turned on, so used as a control in this experiment.

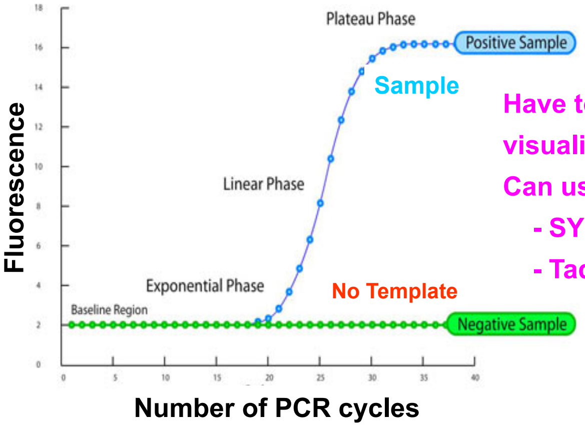



QPCR

Real-Time PCR aka quantitative PCR = qPCR

- allows quantitation of starting template material (DNA or RNA).
- Quantification from cycle # when product is first detected, NOT amount of product accumulated after a fixed number of cycles. Why?
- The higher the starting copy number of the nucleic acid target, the sooner a significant increase in fluorescence is observed.

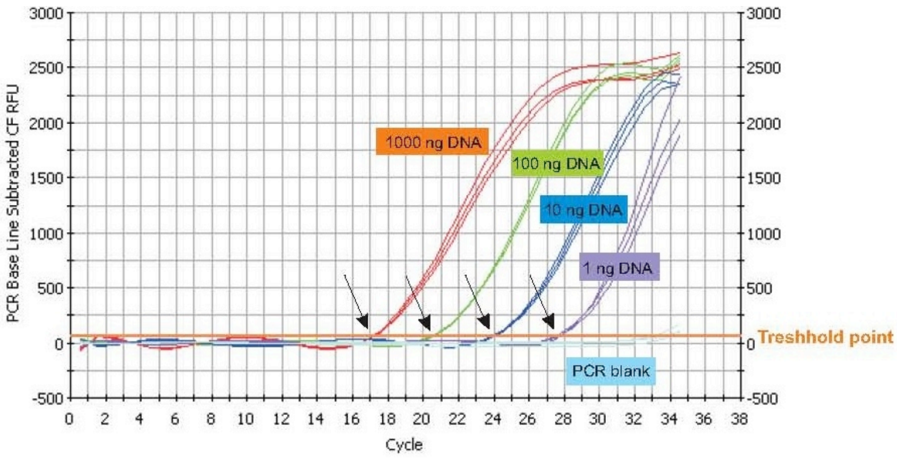


Have to label amplicons to visualize their accumulation.

Can use e.g.:

- SYBR Green (non-specific)

- TaqMan probes (specific)

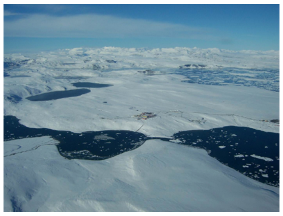


A typical amplification plot generated using a 10-fold dilution series of genomic DNA

The following QPCR example is review at home if you don’t know this method well already

Is a gene present vs. it expressed, *part 2: q*RT-PCR example

- Simulated diesel spill in the Canadian high artic at Ellesmere Island Bioremediation Experimental site
- Compared control site to contaminated site, and time post-nutrient-amendment, for gene copies and gene expression of degradation genes



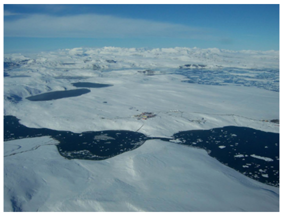
Yergeau E, Sanschagrin S, Beaumier D, Greer CW (2012) Metagenomic Analysis of the Bioremediation of Diesel-Contaminated Canadian High Arctic Soils. PLoS ONE 7(1): e30058. doi:10.1371/journal.pone.0030058
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0030058

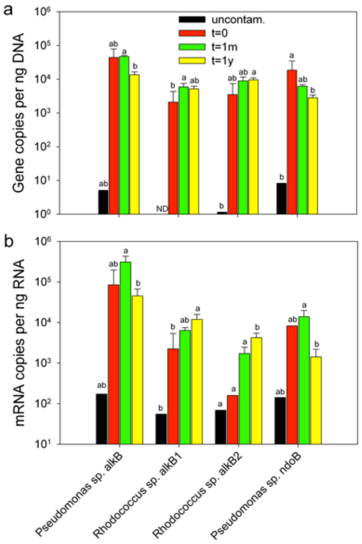
PLoS one

Is a gene present vs. it expressed, *part 2:* *q*RT-PCR example

Figure 2. Hydrocarbon-degrading gene relative abundance and expression.

- Simulated diesel spill in the Canadian high artic at Ellesmere Island Bioremediation Experimental site
- Compared control site to contaminated site, and time post-nutrient-amendment, for gene copies and gene expression of degradation genes



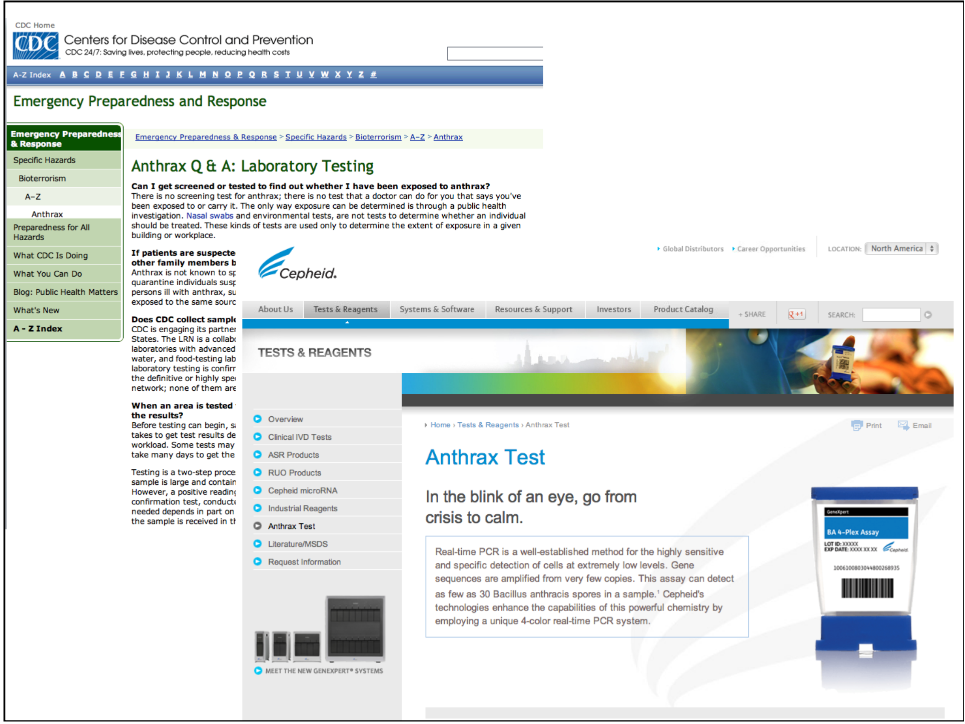


DNA
qPCR

RNA
qRT-PCR

Yergeau E, Sanschagrin S, Beaumier D, Greer CW (2012) Metagenomic Analysis of the Bioremediation of Diesel-Contaminated Canadian High Arctic Soils. PLoS ONE 7(1): e30058. doi:10.1371/journal.pone.0030058
http://www.plosone.org/article/info:doi/10.1371/journal.pone.0030058

PLoS one



CDC Home
Centers for Disease Control and Prevention
CDC 24/7: Saving lives, protecting people, reducing health costs

A-Z Index A B C D E F G H I J K L M N O P Q R S T U V W X Y Z #

Emergency Preparedness and Response

Emergency Preparedness & Response
Specific Hazards
Bioterrorism
A–Z
Anthrax
Preparedness for All Hazards
What CDC Is Doing
What You Can Do
Blog: Public Health Matters
What's New
A - Z Index

Emergency Preparedness & Response > Specific Hazards > Bioterrorism > A–Z > Anthrax

Anthrax Q & A: Laboratory Testing

Can I get screened or tested to find out whether I have been exposed to anthrax?
There is no screening test for anthrax; there is no test that a doctor can do for you that says you've been exposed to or carry it. The only way exposure can be determined is through a public health investigation. Nasal swabs and environmental tests, are not tests to determine whether an individual should be treated. These kinds of tests are used only to determine the extent of exposure in a given building or workplace.

If patients are suspecte
other family members b
Anthrax is not known to sp
quarantine individuals susp
persons ill with anthrax, su
exposed to the same sourc

Does CDC collect sampl
CDC is engaging its partner
States. The LRN is a collab
laboratories with advanced
water, and food-testing lab
laboratory testing is confirr
the definitive or highly spe
network; none of them are

When an area is tested
the results?
Before testing can begin, s
takes to get test results de
workload. Some tests may
take many days to get the

Testing is a two-step proce
sample is large and contair
However, a positive readin
confirmation test, conduct
needed depends in part on
the sample is received in t

Cepheid.

Global Distributors · Career Opportunities · LOCATION: North America

About Us | Tests & Reagents | Systems & Software | Resources & Support | Investors | Product Catalog | + SHARE | +1 | SEARCH:

TESTS & REAGENTS

Overview
Clinical IVD Tests
ASR Products
RUO Products
Cepheid microRNA
Industrial Reagents
Anthrax Test
Literature/MSDS
Request Information

MEET THE NEW GENEXPERT® SYSTEMS

Home › Tests & Reagents › Anthrax Test

Print Email

Anthrax Test

In the blink of an eye, go from crisis to calm.

Real-time PCR is a well-established method for the highly sensitive and specific detection of cells at extremely low levels. Gene sequences are amplified from very few copies. This assay can detect as few as 30 Bacillus anthracis spores in a sample.[1] Cepheid's technologies enhance the capabilities of this powerful chemistry by employing a unique 4-color real-time PCR system.

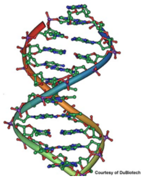
GeneXpert
BA 4-Plex Assay
LOT ID: XXXXX
EXP DATE: XXXX XX XX

How do you access the information in these molecules?

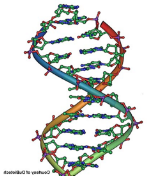
DNA
RNA
protein

Study single or few genes
(or transcripts)

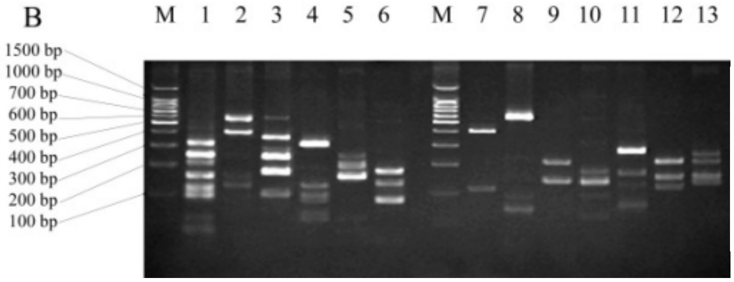
1. **Selective amplification via PCR or RT-PCR**
- Differeniate type(s) by "Fingerprinting" approaches
- Quantify by qPCR / realtime PCR
- Separate types by Cloning (e.g. functional expression, some seq'ing)
- Characterize definitively by Sequencing



"Fingerprinting" methods



- "Fingerprinting" genomic DNA or PCR products to examine whether they are the same or different is a quick inexpensive alternative to sequening that you might read about.
- There are many fingerprinting techniques.
- They do not provide information about the identity or relatedness of the organisms, just an indication of overall differences.



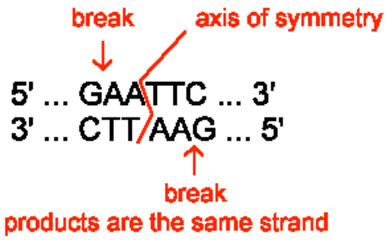
Gel electrophoresis of restriction digested PCR product fragments

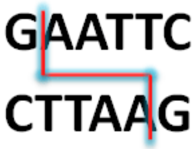
The following explanation of one type of fingerprinting is to read at home if you don't get this general concept and wish to

The brief example of "RFLP" fingerprinting

First review restriction enzymes:

Restriction ***endonucleases***





RFLP Fingerprinting Analysis

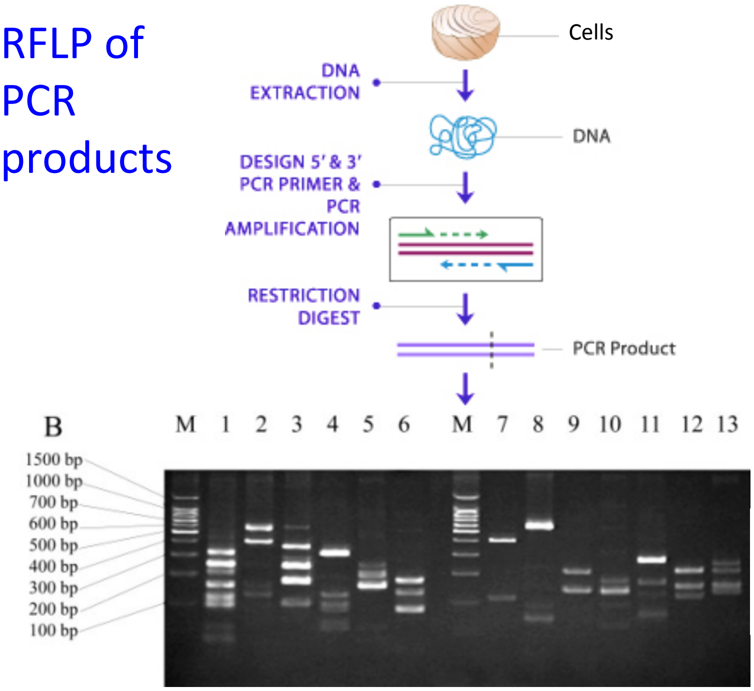
RFLP = restriction fragment length polymorphism

DNA is cut into fragments using one or a set of restriction enzymes.

For PCR products a simple fragment pattern can be distinguished immediately on a gel. This is used to confirm the PCR product or to distinguish between different isolates based on restriction cutting of the 16S-rDNA sequence "ribotyping". Also developed into a diversity measurement technique called "TRFLP".

For chromosomal DNA the RFLP fragments are separated by gel electrophoresis, transferred to a membrane, and probed with a gene probe.

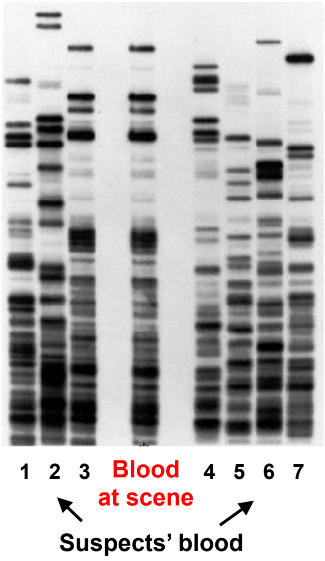
RFLP of PCR products



Gel electrophoresis of PCR product fragments

DNA fingerprinting in forensics

1. DNA is isolated from crime scene, victim, and suspect.
2. DNA in each sample is digested with a restriction enzyme(s).
3. The restriction fragments are separated by agarose gel electrophoresis.
4. The DNA is denatured and transferred to a nylon membrane (Southern blot).
5. The membrane is probed with a radiolabeled probe specific for a single polymorphic VNTR locus.
6. Autoradiography is performed to visualize the fingerprint.



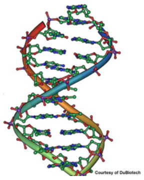
How do you access the information in these molecules?

DNA → RNA → protein

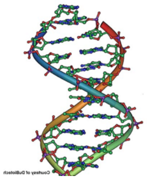
Study single or few genes
(or transcripts)

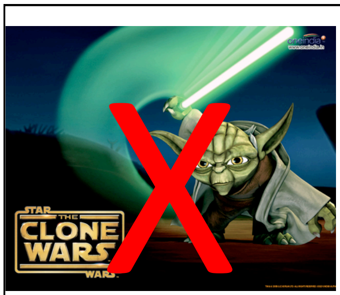
1. **Selective amplification via PCR or RT-PCR**
- Differeniate type(s) by "Fingerprinting" approaches
- Quantify by qPCR / realtime PCR
- Separate types by Cloning (e.g. functional expression, some seq'ing)
- Characterize definitively by Sequencing

3/4. **Study or hunt for target function(s) via "heterologous expression"**
- Put genes (in targeted or blind way) into a "model organism" to search or study



Cloning





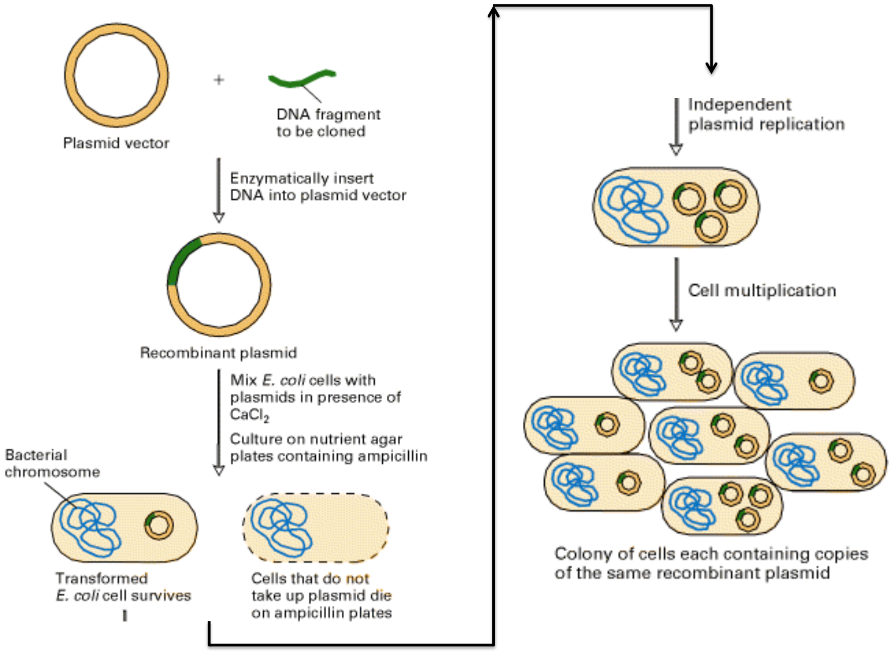
CLONING

DNA cloning
(NOT organismal cloning)
= the process of introducing a foreign piece of DNA into a replication vector and multiplying the DNA – making many many copies (clones) of it...

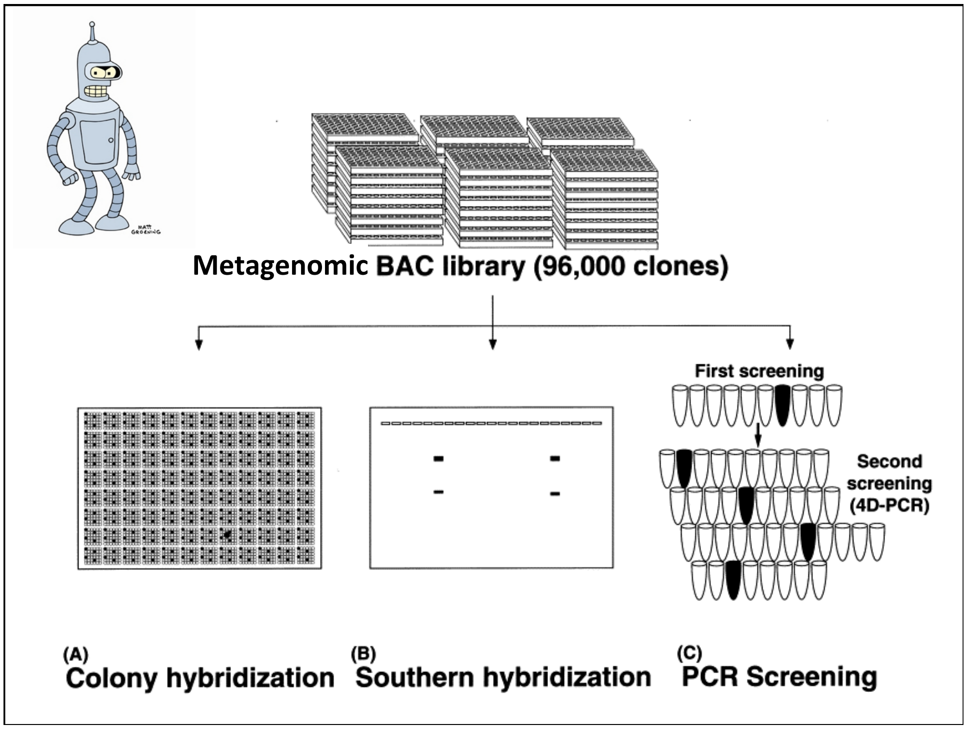
Recombinant DNA = foreign DNA inserted into a vector.

Cloning DNA in environmental microbiology is used to:

1. Make many *identical* copies of a gene as required for “old school” Sanger sequencing.
2. Produce large amounts of a gene PRODUCT (enzyme, etc)
3. Make stable “clone library” of environmental metagenomic DNA pieces, for:
   a. Screening for clones of interest (carrying specific genes)
   b. Sequencing (but not required for newer sequencing methods)
   c. Functional screening (for e.g. bioprospecting)


Plasmid vector
+
DNA fragment to be cloned
Enzymatically insert DNA into plasmid vector
Recombinant plasmid
Mix E. coli cells with plasmids in presence of $CaCl_2$
Culture on nutrient agar plates containing ampicillin
Bacterial chromosome
Transformed E. coli cell survives
Cells that do not take up plasmid die on ampicillin plates
Independent plasmid replication
Cell multiplication
Colony of cells each containing copies of the same recombinant plasmid

Selection for cells containing the *recombinant* plasmid

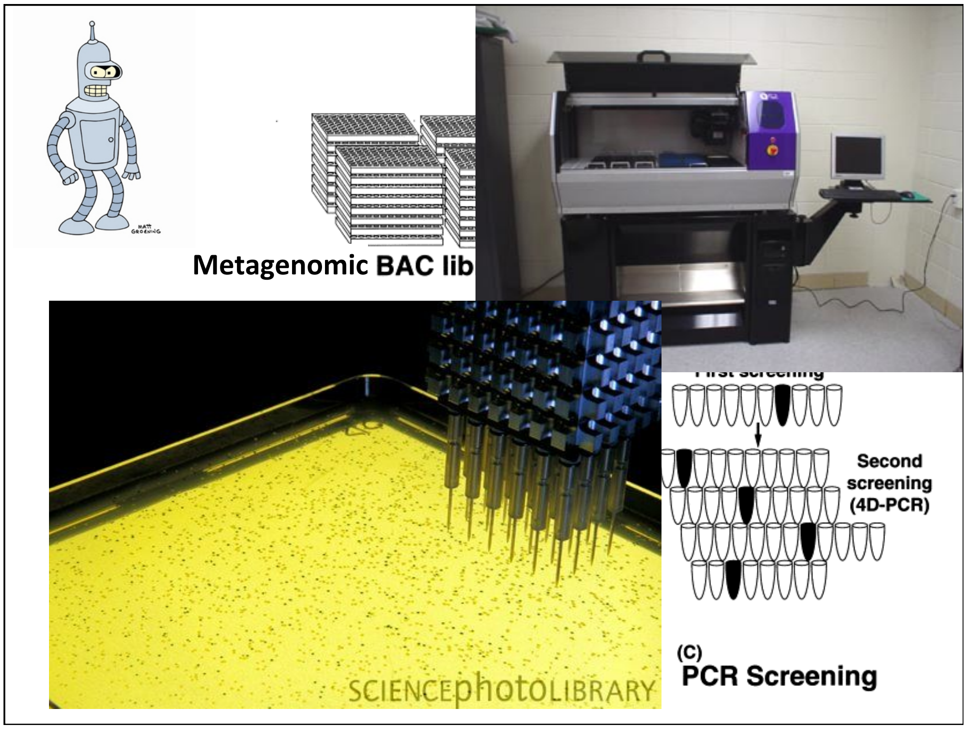


Metagenomic BAC library (96,000 clones)

(A) Colony hybridization

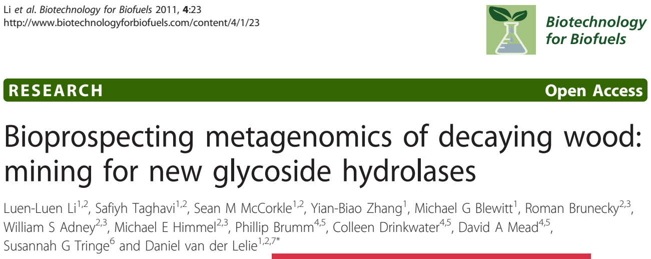
(B) Southern hybridization

(C) PCR Screening



Metagenomic BAC lib

(C) PCR Screening



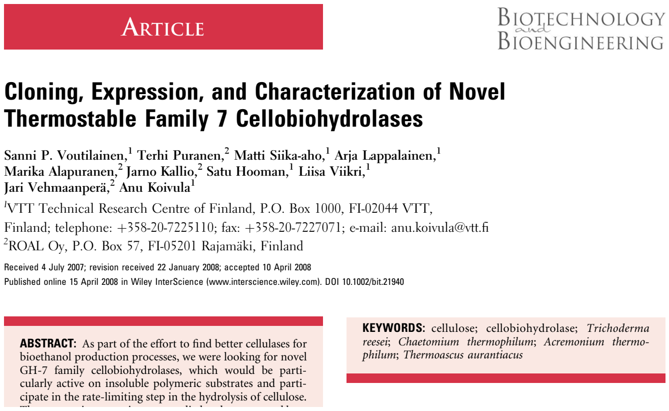
Li *et al. Biotechnology for Biofuels* 2011, **4**:23
http://www.biotechnologyforbiofuels.com/content/4/1/23

Biotechnology for Biofuels

RESEARCH Open Access

Bioprospecting metagenomics of decaying wood: mining for new glycoside hydrolases

Luen-Luen Li[1,2], Safiyh Taghavi[1,2], Sean M McCorkle[1,2], Yian-Biao Zhang[1], Michael G Blewitt[1], Roman Brunecky[2,3], William S Adney[2,3], Michael E Himmel[2,3], Phillip Brumm[4,5], Colleen Drinkwater[4,5], David A Mead[4,5], Susannah G Tringe[6] and Daniel van der Lelie[1,2,7*]



ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Cloning, Expression, and Characterization of Novel Thermostable Family 7 Cellobiohydrolases

Sanni P. Voutilainen,[1] Terhi Puranen,[2] Matti Siika-aho,[1] Arja Lappalainen,[1] Marika Alapuranen,[2] Jarno Kallio,[2] Satu Hooman,[1] Liisa Viikri,[1] Jari Vehmaanperä,[2] Anu Koivula[1]

[1]VTT Technical Research Centre of Finland, P.O. Box 1000, FI-02044 VTT, Finland; telephone: +358-20-7225110; fax: +358-20-7227071; e-mail: anu.koivula@vtt.fi
[2]ROAL Oy, P.O. Box 57, FI-05201 Rajamäki, Finland

Received 4 July 2007; revision received 22 January 2008; accepted 10 April 2008
Published online 15 April 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21940

ABSTRACT: As part of the effort to find better cellulases for bioethanol production processes, we were looking for novel GH-7 family cellobiohydrolases, which would be particularly active on insoluble polymeric substrates and participate in the rate-limiting step in the hydrolysis of cellulose.

KEYWORDS: cellulose; cellobiohydrolase; *Trichoderma reesei*; *Chaetomium thermophilum*; *Acremonium thermophilum*; *Thermoascus aurantiacus*

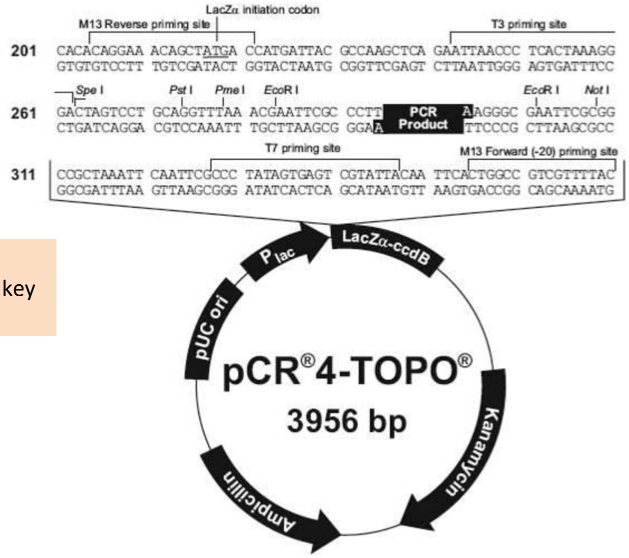
A little more on cloning to read at home if you want to know more…

Cloning vectors differ generally by:

- size (of vector itself and amount of foreign DNA they can carry)
- host organism
- copy # in that host
- whether foreign DNA gets expressed or not, and if so how much…

Vector Type	Host Type	Insert size (kb)
Plasmid	Bacteria e.g. *E. coli*	<10 kb
Phage	Bacteria e.g. *E. coli*	9-20 kb
Fosmid	Bacteria e.g. *E. coli*	40kb
BAC (Bacterial Artificial Chromosome)	Bacteria e.g. *E. coli*	75-150 kb
YAC (Yeast Artificial Chromosome)	Bacteria and Yeast	100-1000 kb

Cloning Vector Map


LacZα initiation codon
M13 Reverse priming site
T3 priming site
201 CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTCA GAATTAACCC TCACTAAAGG
GTGTGTCCTT TGTCGATACT GGTACTAATG CGGTTCGAGT CTTAATTGGG AGTGATTTCC
Spe I
Pst I
Pme I
EcoR I
EcoR I
Not I
261 GACTAGTCCT GCAGGTTTAA ACGAATTCGC CCTT PCR Product AAGGGC GAATTCGCGG
CTGATCAGGA CGTCCAAATT TGCTTAAGCG GGAA TTCCCG CTTAAGCGCC
T7 priming site
M13 Forward (-20) priming site
311 CCGCTAAATT CAATTCGCCC TATAGTGAGT CGTATTACAA TTCACTGGCC GTCGTTTTAC
GGCGATTTAA GTTAAGCGGG ATATCACTCA GCATAATGTT AAGTGACCGG CAGCAAAATG
P lac
LacZα-ccdB
pUC ori
pCR®4-TOPO®
3956 bp
Kanamycin
Ampicillin

Restriction enzymes play key role here too!

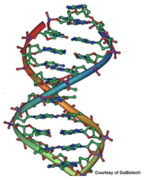
How do you access the information in these molecules?

DNA
RNA
protein

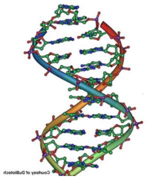
Study single or few genes
(or transcripts)

1. **Selective amplification via PCR or RT-PCR**

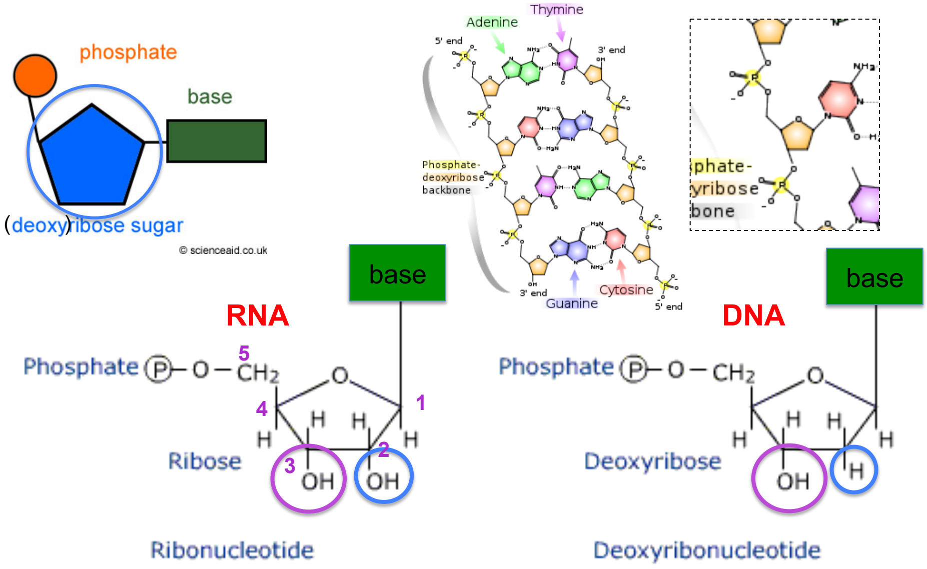
- Differeniate type(s) by "Fingerprinting" approaches
- Quantify by qPCR / realtime PCR
- Separate types by Cloning (e.g. functional expression, some seq'ing)
- Characterize definitively by Sequencing


Courtesy of DuBiotech

SEQUENCING

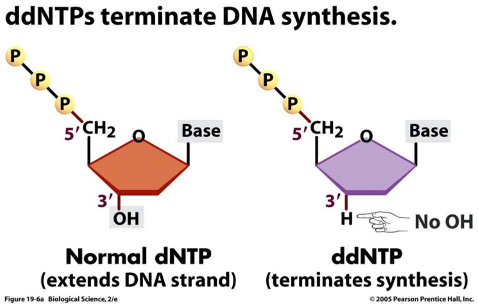


Molecular techniques are based on the structure of these biomolecules, let's focus here on DNA and RNA

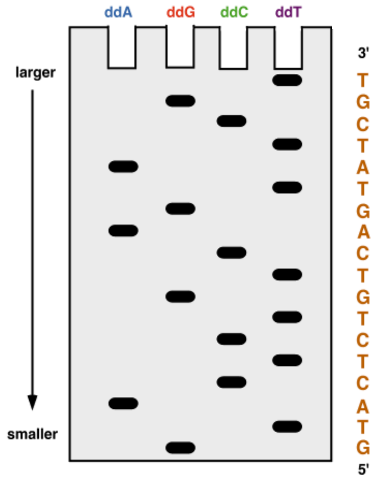


Sanger Sequencing:

“chain terminating” nucleotide analogs sprinkled among normal dNTPs

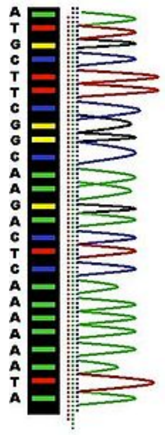

ddNTPs terminate DNA synthesis.
P
P
P
5′ CH2
O
Base
3′
OH
Normal dNTP
(extends DNA strand)
P
P
P
5′ CH2
O
Base
3′
H
No OH
ddNTP
(terminates synthesis)
Figure 19-6a Biological Science, 2/e
© 2005 Pearson Prentice Hall, Inc.

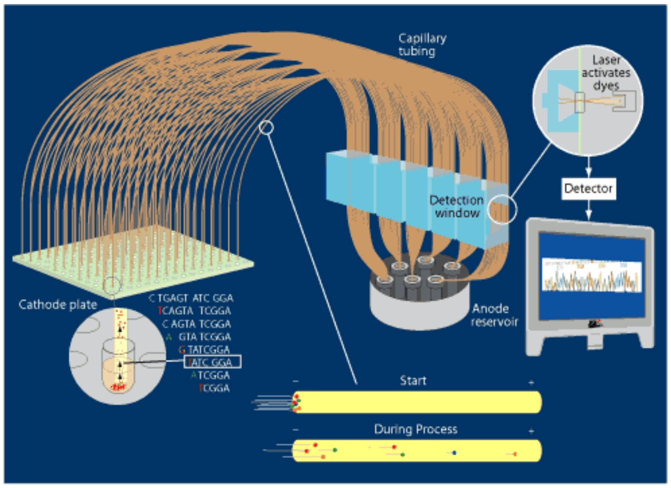
4 separate reactions, each run on separate lane of gel…



Today’s version of Sanger:

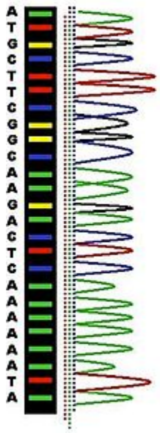
- separates fragments through capillary electrophoresis not gel
- performed 96 to 384 “lanes” at a time
- using fluorescently-labeled ddNTPs (so can use 1 lane per sample instead of 4)
- ~750bp / read

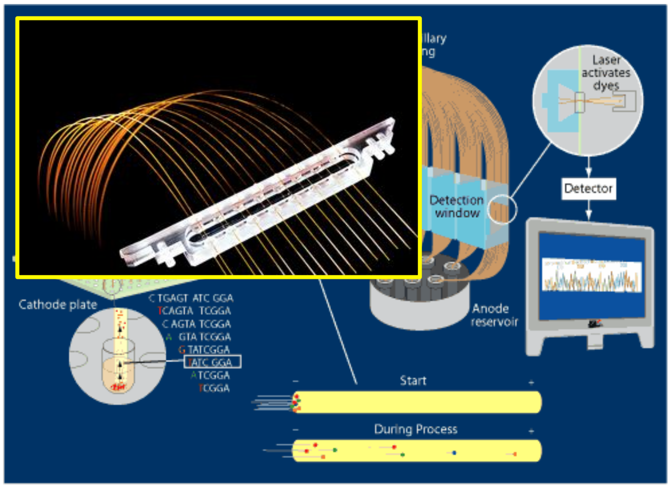




Today's version of Sanger:

- separates fragments through capillary electrophoresis not gel
- performed 96 to 384 "lanes" at a time
- using fluorescently-labeled ddNTPs (so can use 1 lane per sample instead of 4)
- ~750bp / read



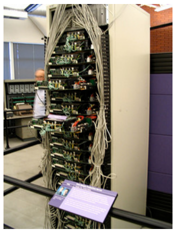


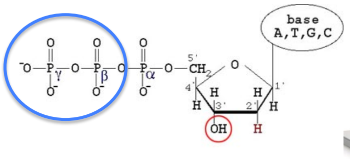
"Next Generation" High Throughput Sequencing Technologies

– 454 Pyrosequencing
– Illumina sequencing
– NUMEROUS others

"Next-gen" sequencing is:
- Much cheaper per base
- MUCH higher-throughput – thousands to millions of reads per sample
- Gives shorter reads (400bp for 454, 150 for Illumina)
- Does not require cloning first

Almost all are still "sequencing by synthesis"
- they "read" the sequence of DNA as it is copied from the template
- the signal they see is based on either
  a. labeled dNTPs, like in modern Sanger (Illumina)
  b. Detection of the successful addition of a dNTP and release of PPi
  (c. others)





dNTP
deoxyribonucleotide triphosphate

Many options here at UofA …

Sanger, Illumina, Pyrosequencer @ Arizona Genomics Institute (Rod Wing)

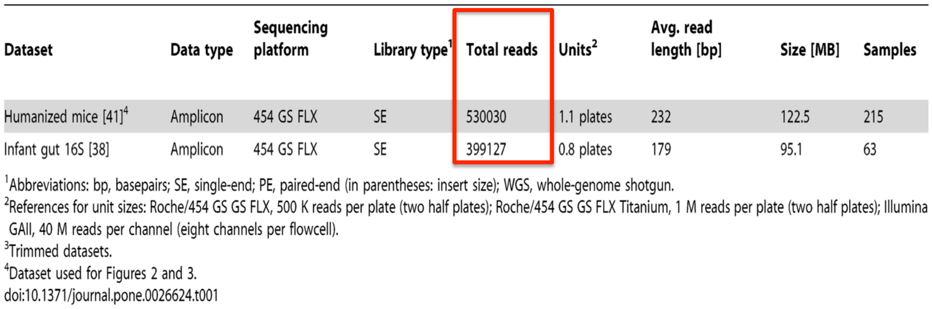
Sanger, Pyrosequencer @ Arizona Research Labs

Proteomics + Flow cytometry cores

Special case of 1-gene sequencing: high-throughput 16S rRNA amplicon sequencing

Amplicons generated first by PCR; sequenced by pyrosequencing called "pyrotags", sequenced by Illumina called "iTags"

Angiuoli et al., 2011, PLoS ONE, Evaluated different seq'ing methods for different applications; excerpt from table here to show high # of sequences recovered per human gut habitat



Dataset	Data type	Sequencing platform	Library type[1]	Total reads	Units[2]	Avg. read length [bp]	Size [MB]	Samples
Humanized mice [41][4]	Amplicon	454 GS FLX	SE	530030	1.1 plates	232	122.5	215
Infant gut 16S [38]	Amplicon	454 GS FLX	SE	399127	0.8 plates	179	95.1	63

[1] Abbreviations: bp, basepairs; SE, single-end; PE, paired-end (in parentheses: insert size); WGS, whole-genome shotgun.

[2] References for unit sizes: Roche/454 GS GS FLX, 500 K reads per plate (two half plates); Roche/454 GS GS FLX Titanium, 1 M reads per plate (two half plates); Illumina GAII, 40 M reads per channel (eight channels per flowcell).

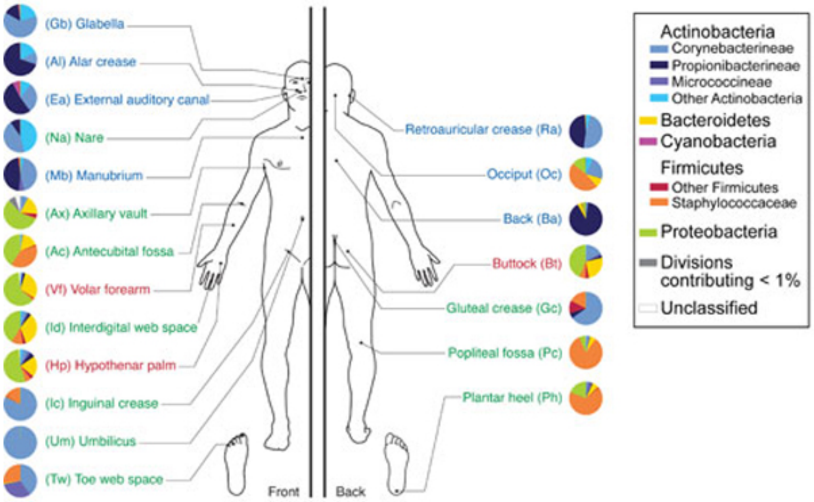
[3] Trimmed datasets.

[4] Dataset used for Figures 2 and 3.

doi:10.1371/journal.pone.0026624.t001

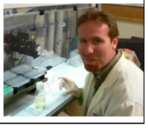
- Multiplexing! Short DNA "barcodes" allow multiple samples to be run together

Example of skin microbiome



Elizabeth Grice & Discover Magazine. In a thorough survey of our skin microbiome, Elizabeth Grice identified species from at least 205 different genera. Your forearm has the richest community with an average of 44 species, while your nostril, ears and inguinal crease (between leg and groin) are the most stable habitats. Grice also found at bacteria from a specific body part have more in common than those from a specific person. Your butt microbes have more in common with mine than they do with your elbow microbes.

Here are a number of slides on different types of sequencing technologies to read at home / refer back to when interested (courtesy of MBS)

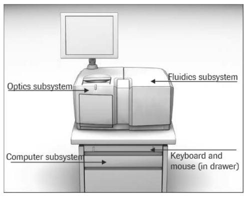


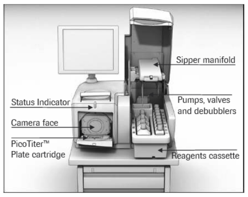
454 Pyrosequencing - the generations

Stats/ run	GS20	FLX	Titanium
Total sequence (Mb)	40	100	1,000
Read length (bp)	100	>200	>400
# reads	400,000	400,000	1M
Paired Ends?	NO	Y, 50%	Y, 50%

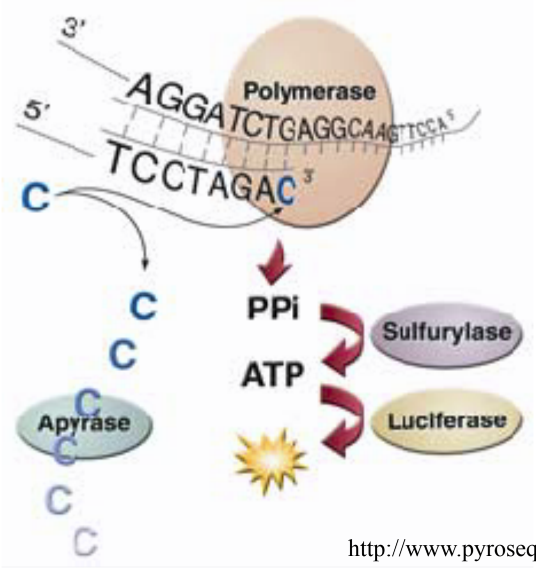
Cost / bp --> *0.03 ¢* *0.01 ¢* *0.003 ¢*

(Sanger is currently 0.1 ¢)



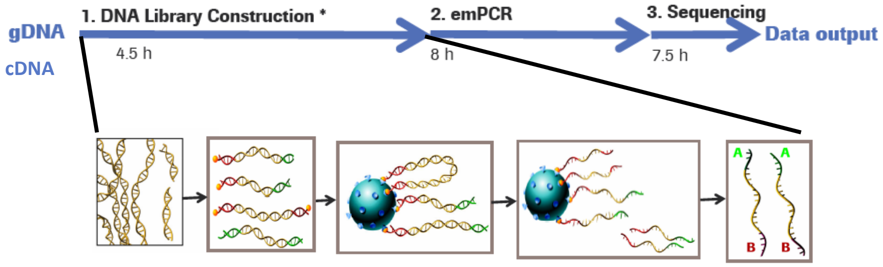


The "pyro" in pyrosequencing



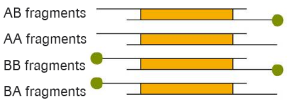
http://www.pyrosequencing.com/

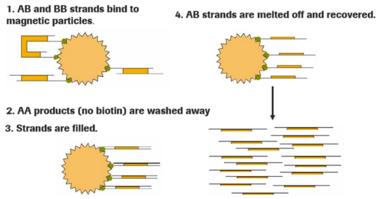
Pyrosequencing - Library construction



- Library is created from any dsDNA
- Genome fragmentation by nebulization
- Ligation of adapters A & B

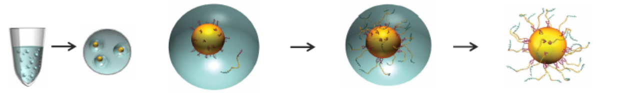
- A/B fragments selected using streptavidin-biotin purification
- Denaturation to select for sstDNA library with A/B adaptors
- No cloning; no colony picking




1. AB and BB strands bind to magnetic particles.
4. AB strands are melted off and recovered.
2. AA products (no biotin) are washed away
3. Strands are filled.

Images courtesy of Roche (Technical presentation)

Pyrosequencing - emPCR



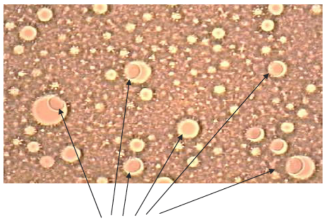
Anneal sstDNA to an excess of 28 μm DNA Capture beads

Emulsify DNA Capture beads and PCR reagents in water-in-oil microreactors

Clonal amplification occurs inside microreactors

Break microreactors and enrich for DNA-positive beads

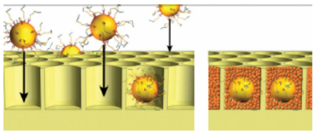
sstDNA library → **Clonally-amplified sstDNA attached to bead**



DNA Capture Beads

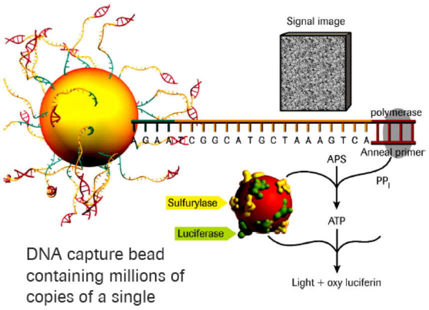
“polonies” generated by bead-associated emulsion PCR

Pyrosequencing - sequencing



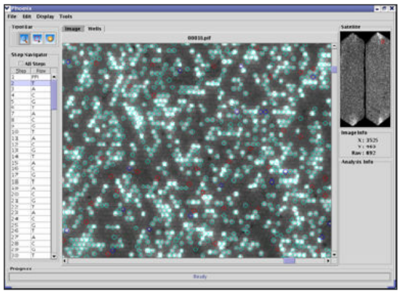
- Well diameter average for PicoTiterPlate is 44 µm
- A single clonally amplified sstDNA bead is deposited per well.
- A layer of packing and enzyme beads are deposited
- Plate is loaded into instrument for sequencing

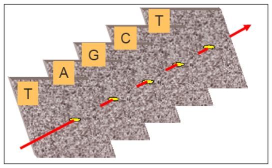
Amplified sstDNA library beads ⟶ **Packed PTP**

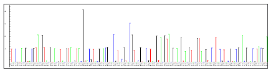


DNA capture bead containing millions of copies of a single clonal fragment

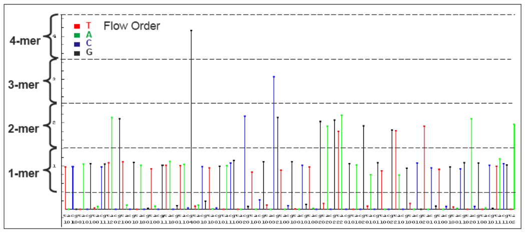
Pyrosequencing - image processing







Signal strength is determined by homopolymer length.



TCAG

Illumina Genome Analyzer
(bought Solexa in 2006)

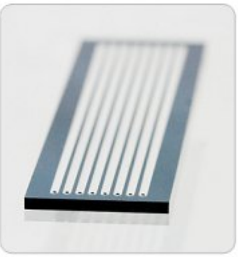
Sequencing-by-synthesis using “bridged” amplification to generat
“polonies”

~30-35bp (50bp) reads, 2GB, $4K / run

Not strong for “denovo” genomic sequencing

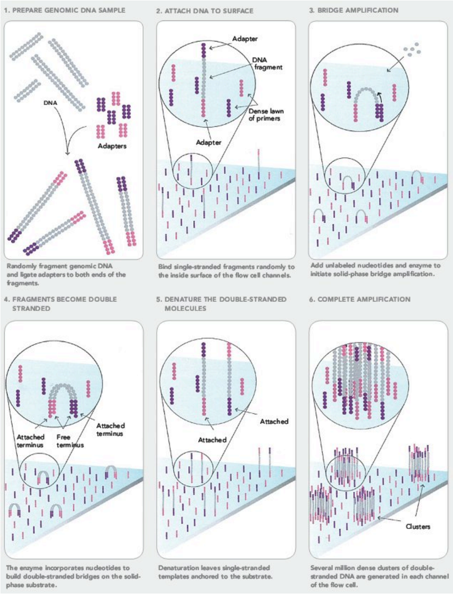
Useful for
-“resequencing” genome projects
- gene expression in model systems (replace the microarray?)

FIGURE 1: ILLUMINA GENOME ANALYZER FLOW CELL



Up to eight samples can be loaded onto the flow cell for simultaneous analysis on the Illumina Genome Analyzer.

Cluster station 1G sequencer



1. PREPARE GENOMIC DNA SAMPLE

Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

2. ATTACH DNA TO SURFACE

Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

3. BRIDGE AMPLIFICATION

Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

4. FRAGMENTS BECOME DOUBLE STRANDED

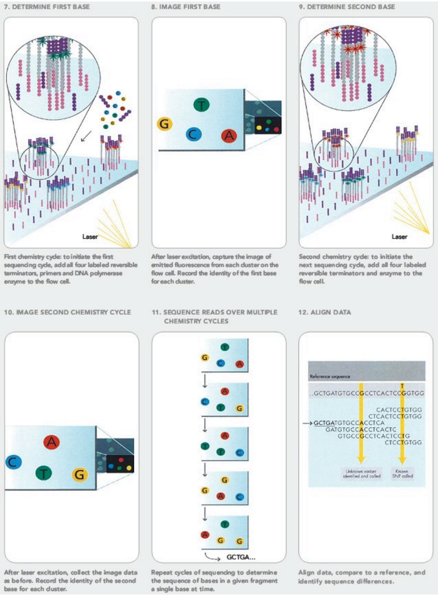
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

5. DENATURE THE DOUBLE-STRANDED MOLECULES

Denaturation leaves single-stranded templates anchored to the substrate.

6. COMPLETE AMPLIFICATION

Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.



7. DETERMINE FIRST BASE

First chemistry cycle: to initiate the first sequencing cycle, add all four labeled reversible terminators, primers and DNA polymerase enzyme to the flow cell.

8. IMAGE FIRST BASE

After laser excitation, capture the image of emitted fluorescence from each cluster on the flow cell. Record the identity of the first base for each cluster.

9. DETERMINE SECOND BASE

Second chemistry cycle: to initiate the next sequencing cycle, add all four labeled reversible terminators and enzyme to the flow cell.

10. IMAGE SECOND CHEMISTRY CYCLE

After laser excitation, collect the image data as before. Record the identity of the second base for each cluster.

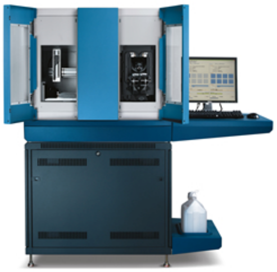
11. SEQUENCE READS OVER MULTIPLE CHEMISTRY CYCLES

Repeat cycles of sequencing to determine the sequence of bases in a given fragment a single base at time.

12. ALIGN DATA

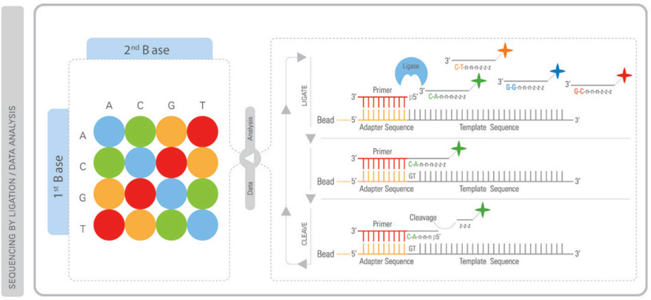
Align data, compare to a reference, and identify sequence differences.

ABI SOLiD



SOLiD 2.0 = 2GB/run
SOLiD 3.0 = 20Gb from 400M reads
(~35bp reads, 25bp for PE reads, but 600bp-10kb insert-size PE reads

Fragment, ligate linkers, emPCR, deposit beads on slide … then … sequence by ligation with TWO-base calling

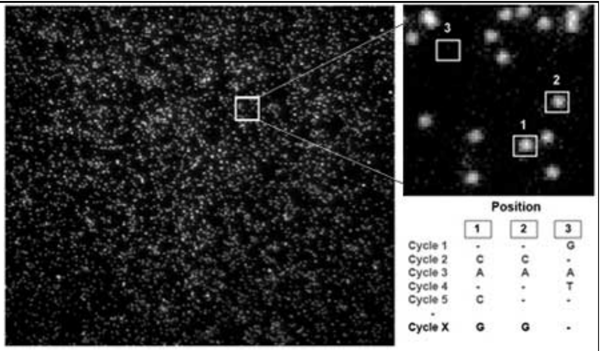


HeliScope

Sequencing-by-synthesis

tSMS = single DNA molecules are captured on an application-specific proprietary surface

See web page video



	1	2	3
Cycle 1	-	-	G
Cycle 2	C	C	-
Cycle 3	A	A	A
Cycle 4	-	-	T
Cycle 5	C	-	-
-			
Cycle X	G	G	-

25GB/run = 1 billion reads x 25bp each

NOW: 25 Mb / hr

SOON: 90Mb/hr (improve efficiency / error) to 360Mb/hr
(increased spot density)

- 1 wk data acquisition, server holds 2 runs of data

A 2000 pound, 32-CPU, $1.35M jalopy?

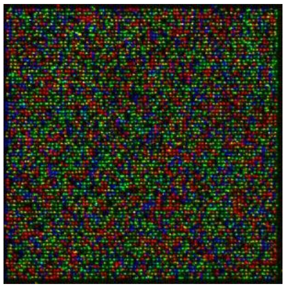
PacBio

The Polonator: open-source
- novel, low-cost PCR polymerase and ligase, + license-free fluors — "freedom fluors!"

Now: 1 run = 80hours, 10Gb @ 28bp / read (14bp from each PE)
By 2013: all "real-time" runs to yield a draft human genome in < 3 minutes, and finished human genome in 15 minutes

Images each fluor-labelled nucleotide as it is incorporated into growing DNA strand by tethering polymerase to a 20-zeptoliter well (10^{-21} = "the world's smallest detection volume") and visualizing ~10 base additions per second

Complete Genomics: Service only



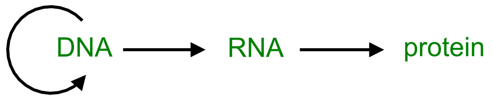
32,000 ft^2 facility = 1,000 human genomes in 2009
+ 20,000 genomes in 2010

Data not published, but $4,000 human genome sequenced July '08
"The speed of the instrument is about 10 times faster than SOLiD and Illumina," Reid claims. "This [genome] ran 4 instruments for a 7 day run -- a 28-instrument-day experiment. By the launch of our product in Q2 [of 2009], it will be a 4-instrument-day experiment."

$1,000 human genome in "Spring 2009"

(sequencing-by-hybridization using ligation + gridded arrays to 1 billion DNA "nanoballs" = cPAL or combinatorial probe-anchor ligation)
40bp "reads" (linkers)

How do you access the information in these molecules?



Study single or few genes
(or transcripts)

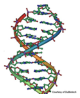
1. **Selective amplification via PCR or RT-PCR**

- Differeniate type(s) by "Fingerprinting" approaches
- Quantify by qPCR / realtime PCR
- Separate types by Cloning (e.g. functional expression, some seq'ing)
- Characterize definitively by Sequencing

2. **Hunt for target(s) via "Gene probes"**

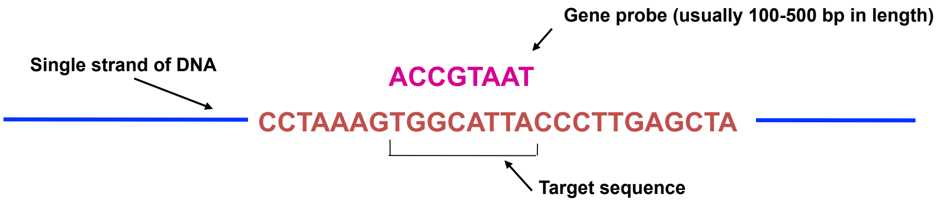
- used to hybridize to "blots"
- used in microscopy to ID particular cells ("FISH")
- Can be used in flow sorting to ID particular cells
- Used in microarrays (probes stuck to surface)

GENE
PROBES



Gene probes

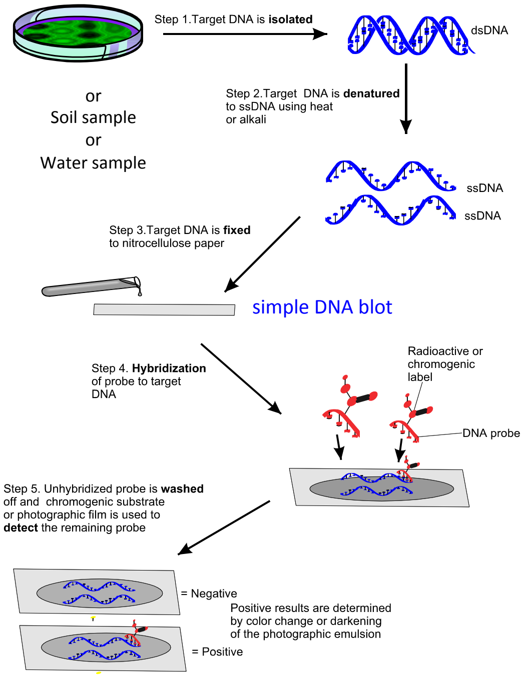
A gene probe is a short specific sequence of DNA that is used to query whether a sample contains "target" DNA, or DNA complementary to the gene probe.


Gene probe (usually 100-500 bp in length)
Single strand of DNA
ACCGTAAT
CCTAAAGTGGCATTACCCTTGAGCTA
Target sequence

The target sequence can be a universally conserved region such as the **16S-rDNA gene** or it can be in a region that is conserved within a specific genus or species such as the *nod* genes for nitrogen fixation by *Rhizobium* or the *rhl* genes for rhamnolipid biosurfactant production by *Pseudomonas aeruginosa*, or the *mcrA* gene of methanogenesis in various *Archaea*.

GENE PROBES AND BLOTTING

General concept: Making “blots” and testing the presence / relative abundance of a given gene by hybridizing a gene probe…



or
Soil sample
or
Water sample

Step 1.Target DNA is **isolated**

Step 2.Target DNA is **denatured** to ssDNA using heat or alkali

Step 3.Target DNA is **fixed** to nitrocellulose paper

simple DNA blot

Step 4. **Hybridization** of probe to target DNA

Step 5. Unhybridized probe is **washed** off and chromogenic substrate or photographic film is used to **detect** the remaining probe

= Negative

Positive results are determined by color change or darkening of the photographic emulsion

= Positive

Simple DNA blotting

Example: Gene probes + microscopy = FISH, fluorescent in situ hybridization

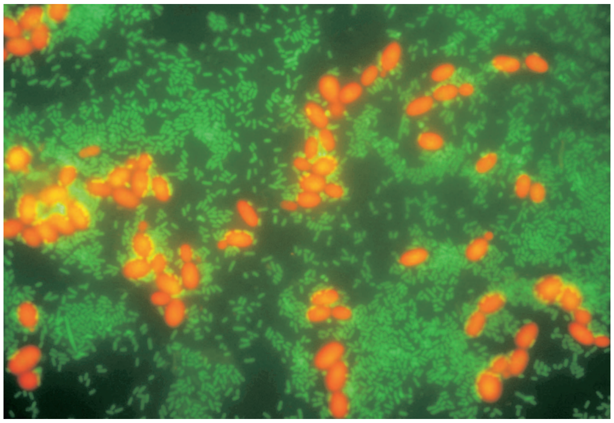


FIG. 1. Tape-FISH for detection of *Salmonella* strains in mixed culture from tomato surfaces. Tomatoes were spiked with a mixture of *S. enterica* serovar Typhimurium (10^7 CFU cm^{-2}) and *R. glutinis* (10^6 CFU cm^{-2}) and then sampled with adhesive tape after drying. Tapes were hybridized for 30 min with a combination of probes targeting *Salmonella* cells (Sal3/Salm-63 cocktail, green label) and eukaryotic cells (EUK 516, red label). These results demonstrate the utility of tape-FISH for simultaneous visualization of the distribution and interactions between multiple phylotypes occurring together on produce surfaces.

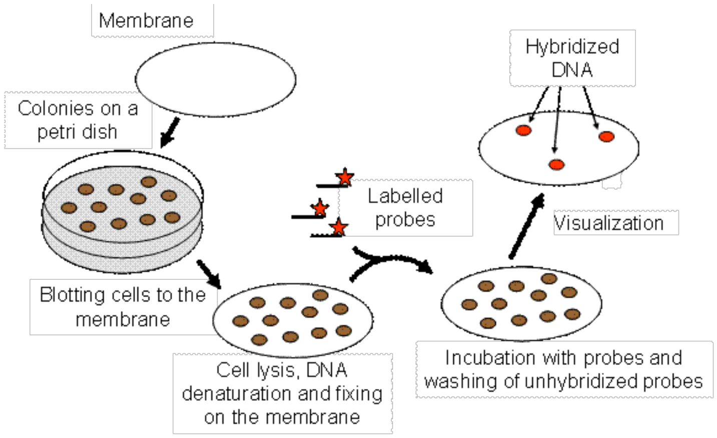
Bisha & Brehm-Stecher, 2009. AEM.

Cells lifted from tomato surface, hybridized with fluorescently-labeled DNA probes targeting 16S (or 18S) rRNA. Probes for eukaryotic cells and for Salmonella bacteria. Which color is for which?

Two more examples to read at home if you want to know more…

Example 1: Using a PCB-degrading gene probe to examine whether there are PCB-degraders in a given soil sample.

Colony blotting

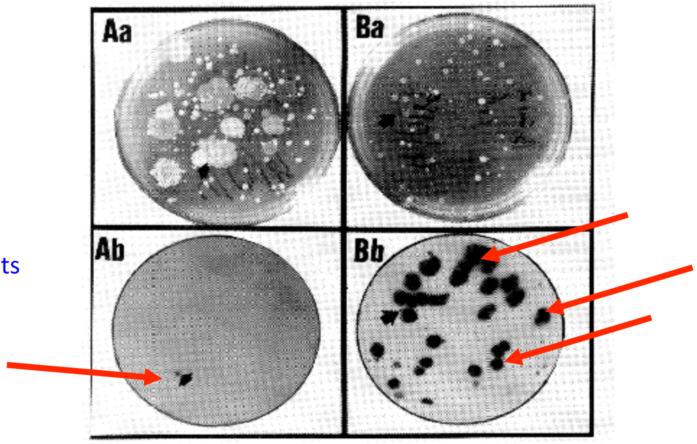

Membrane
Colonies on a petri dish
Blotting cells to the membrane
Cell lysis, DNA denaturation and fixing on the membrane
Labelled probes
Incubation with probes and washing of unhybridized probes
Visualization
Hybridized DNA

Example 1: Using a PCB-degrading gene probe to examine whether there are PCB-degraders in a given soil sample.

Colonies

vs.

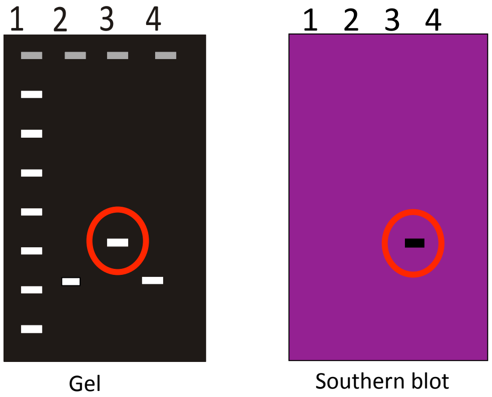
Colony blots



Aa – Bacterial colonies from garden soil
Ab – Colony hybridization with gene probe from PCB-degrading genes
Ba – Bacterial colonies from a PCB-contaminated landfill site
Bb – Colony hybridization with gene probe from PCB-degrading genes

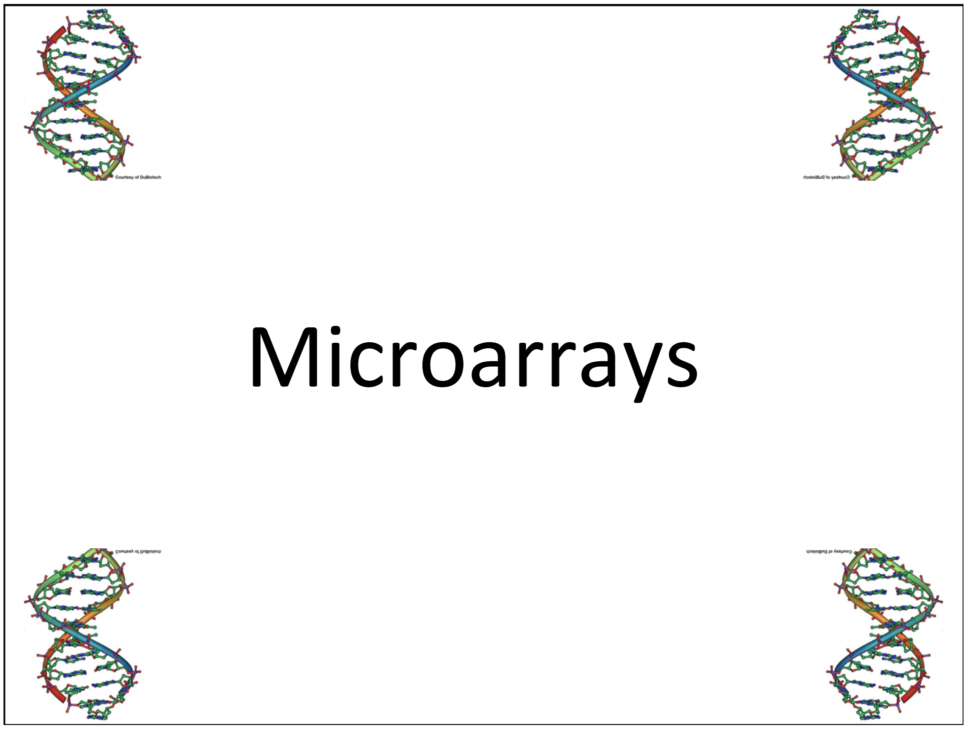
Example 2: Southern blot (developed by EM Southern)

Electrophoresed DNA is transferred onto a membrane and probed.

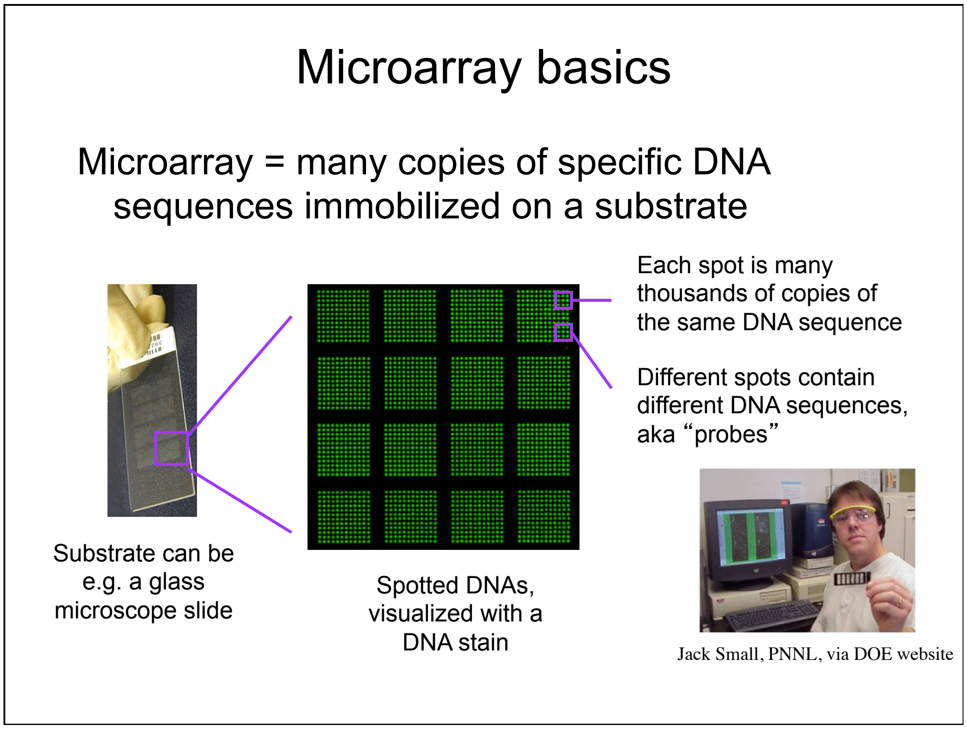


Forensics!

Northern blot ?
Electrophoresed RNA is transferred onto a membrane & probed.
Western blot ?
You guessed it, separated proteins transferred to membrane & probed.



Microarrays



Microarray basics

Microarray = many copies of specific DNA sequences immobilized on a substrate

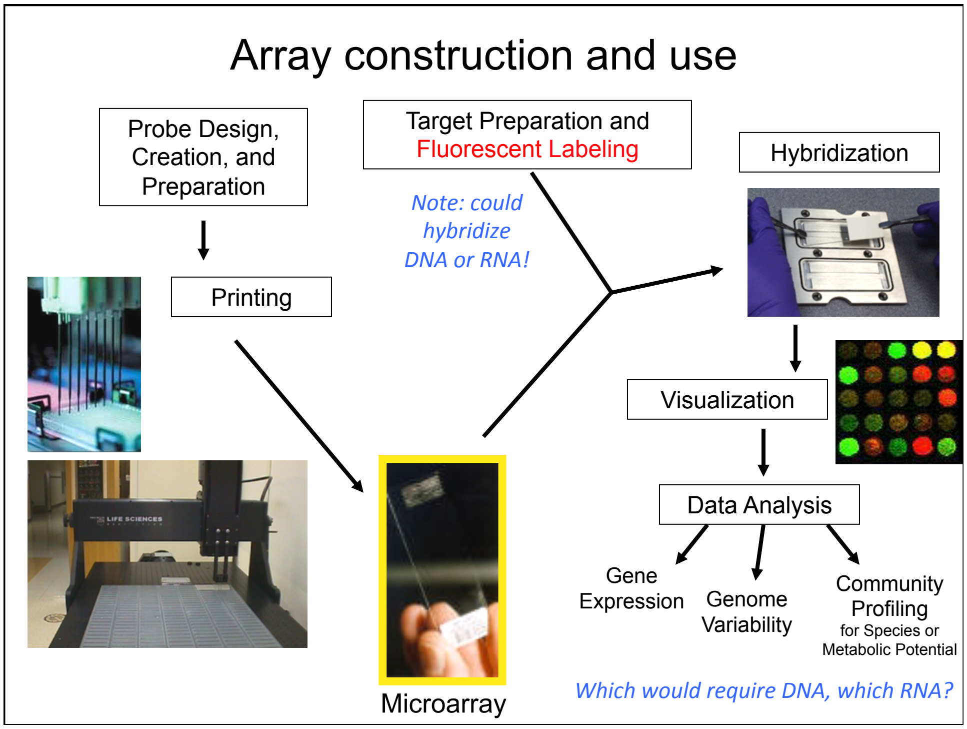
Each spot is many thousands of copies of the same DNA sequence

Different spots contain different DNA sequences, aka “probes”

Substrate can be e.g. a glass microscope slide

Spotted DNAs, visualized with a DNA stain

Jack Small, PNNL, via DOE website



Array construction and use

Probe Design, Creation, and Preparation

Printing

Target Preparation and Fluorescent Labeling

Note: could hybridize DNA or RNA!

Hybridization

Visualization

Data Analysis

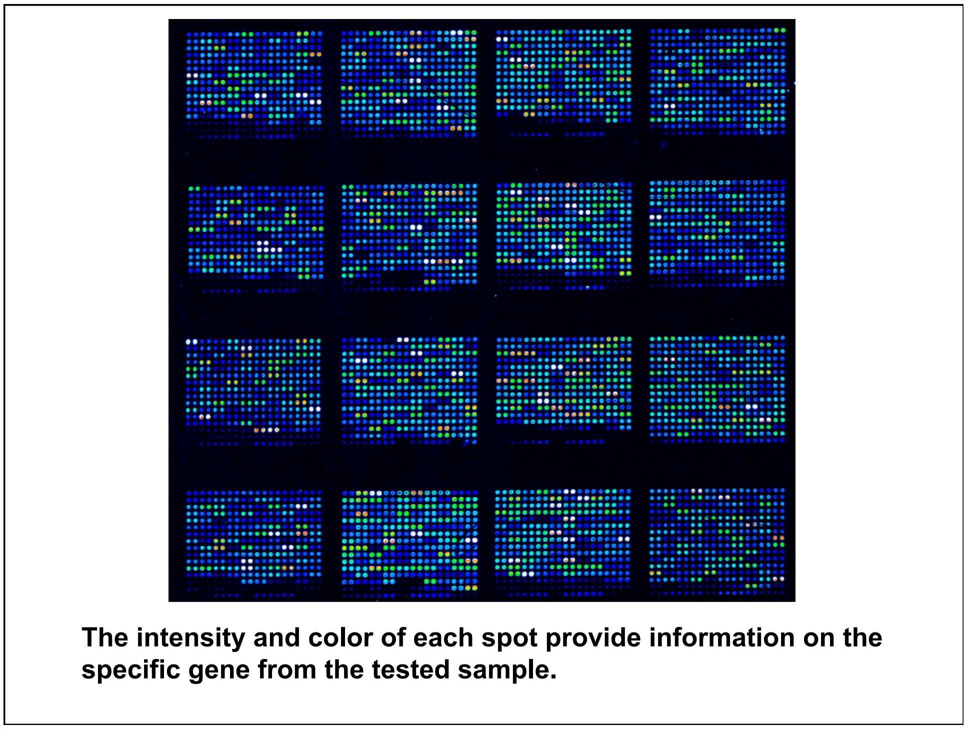
Gene Expression

Genome Variability

Community Profiling for Species or Metabolic Potential

Microarray

Which would require DNA, which RNA?



The intensity and color of each spot provide information on the specific gene from the tested sample.

Three *Main* types of Arrays used in Environmental Microbiology:

1. Functional Gene Arrays: Target known "functional genes" (products mediate a process of interest) (e.g. "Geochip"

2. "Phylochips": Target 16S rRNA "fingerprint" genes

3. Organism-specific arrays (usually for transcriptomics): e.g. *Bacillus subtilis* Genome Array, *E. coli* Genome Arrays…

How do you access the information in these molecules?

DNA
RNA
protein

Study single or few genes (or transcripts)

1. **Selective amplification via PCR or RT-PCR**
- Differeniate type(s) by "Fingerprinting" approaches
- Quantify by qPCR / realtime PCR
- Separate types by Cloning (e.g. functional expression, some seq'ing)
- Characterize definitively by Sequencing

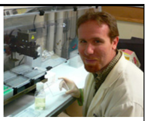
2. **Hunt for target(s) via "Gene probes"**
- used to hybridize to "blots"
- used in microscopy to ID particular cells ("FISH")
- Can be used in flow sorting to ID particular cells
- Used in microarrays (probes stuck to surface)

Study entire genome (or transcriptome), or **meta**genome (aka community genome)

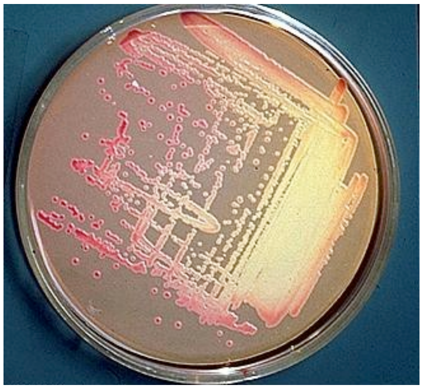
1. Assay genome size(s)
2. Differentiate type(s) by "Fingerprinting" approaches
3. Characterize more fully by Sequencing

3/4. **Study or hunt for target function(s) via "heterologous expression"**
- Put genes (in targeted or blind way) into a "model organism" to search or study

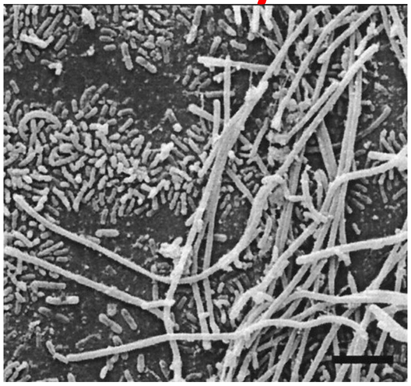
What is a (meta)genome?



isolate



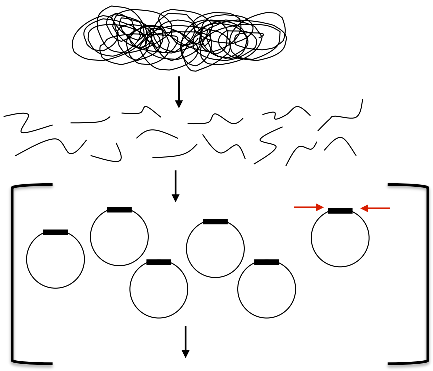
community



Genomics

Metagenomics

Shotgun sequencing (WGS)



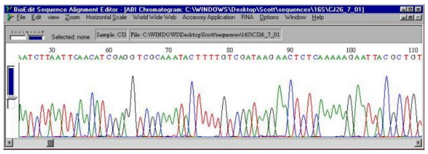
genomic DNA

sheared

Cloning OPTIONAL

depends on seq'ing method and goals

clone library (insert sizes of 1-2, 3-4, 30-40, 100kb)



BioEdit Sequence Alignment Editor - [ABI Chromatogram: C:\WINDOWS\Desktop\Scott\sequences\16S\CJ26_7_01]

Sample: CJ2 File: C:\WINDOWS\Desktop\Scott\sequences\16S\CJ26_7_01

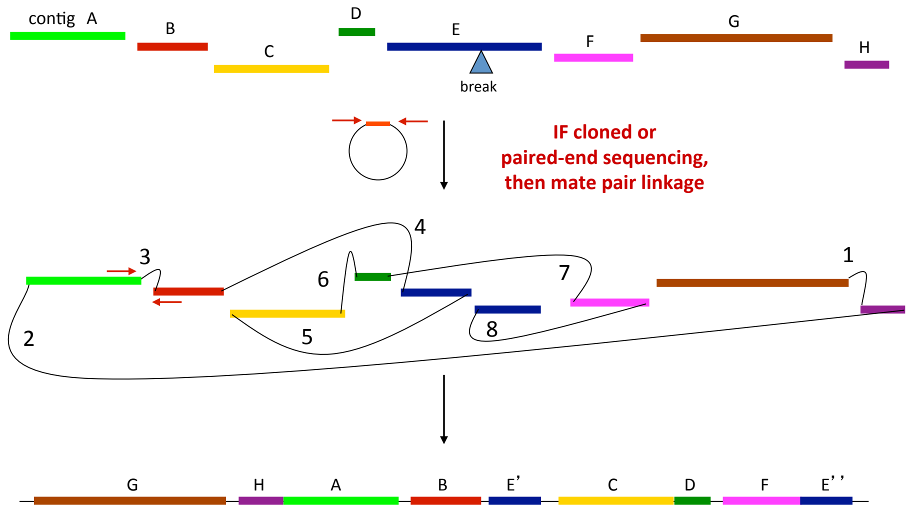
50
60
70
80
90
100
110
AATCTTAATTCAACATCGAGGTCGCAAATACTTTTGTCGATAAGAACTCTCAAAAGAATTACGCTGT

sequencing

assemble reads by alignment identity

```
...ACGGCTGCGTTACATCGATCAT
                    ACATCGATCATTTACGATACCATTG...
```

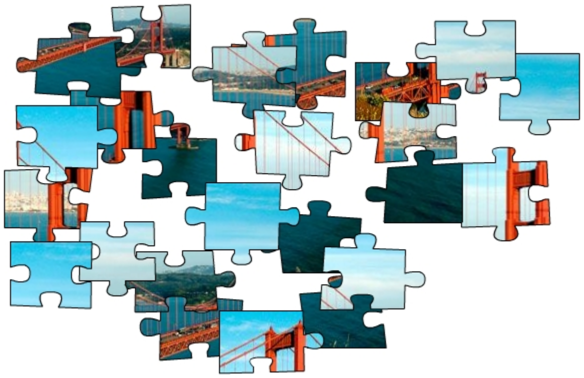
Genome scaffolding



IF cloned or paired-end sequencing, then mate pair linkage

"composite" genome scaffold

Genome assembly

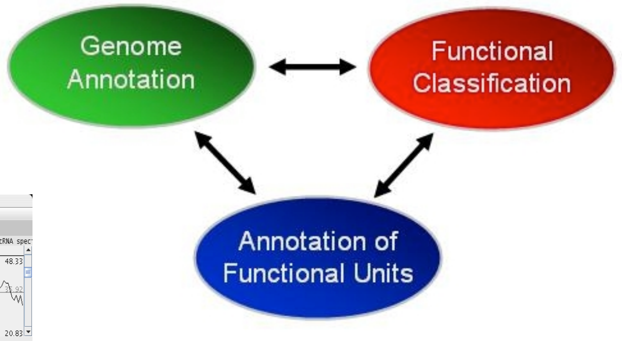


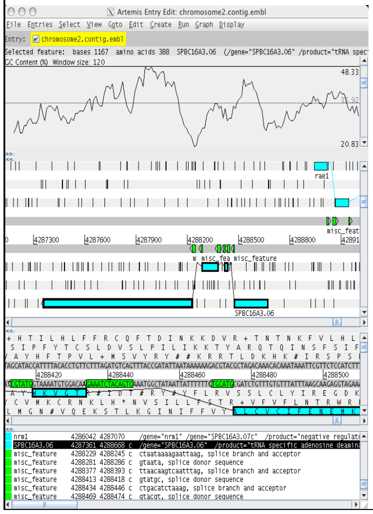
Genome assembly

When is a genome "finished"?
(by Poisson Calculations)

Fold coverage	Percent of genome sequenced
0.25 x	22%
0.50 x	39%
0.75 x	53%
1 x	63%
2 x	88%
3 x	95%
4 x	98%
5 x	99.4%
6 x	99.75%
7 x	99.91%
8 x	99.97%
9 x	99.99%
10 x	99.995%

Genome annotation is never done …

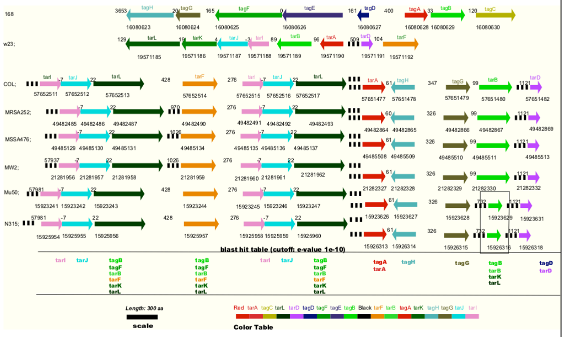




```
Selected feature:  bases 1167  amino acids 388  SPBC16A3.06  (/gene="SPBC16A3.06" /product="tRNA spec
GC Content (%)  Window size: 120
```

```
+ H T I L H L F F R C Q F T D I N K K D V R + T N T N K F V L H L
 S I P F Y T C S L D V S L P I L I K K T Y A R Q T Q I N S F S I F
A Y H F T P V L + M S V Y R Y # # K R R T L D K H K # I R S P S
TAGCATACCATTTTACACCTGTTCTTTAGATGTCAGTTTACCGATATTAATAAAAAAGACGTACGCTAGACAAACACAAATAAATTCGTTCTCCATCTT
ATCGTATGGTAAAATGTGGACAAGAAATCTACAGTCAAATGGCTATAATTATTTTTTCTGCATGCGATCTGTTTGTGTTTATTTAAGCAAGAGGTAGAA
 A Y W K V G T R # I D T # R Y # Y F L R V S S L C L Y I R E G D K
 C V M K C R N K L H * N V S I L L F S T R + V F V F L N T R W R
L M G N # V Q E K S T L K G I N I F F V Y A L C V C I F E N E M K
```

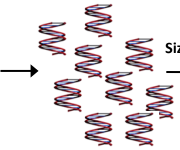
```
nrm1          4286042 4287070    /gene="nrm1" /gene="SPBC16A3.07c"  /product="negative regulat
SPBC16A3.06   4287361 4288668 c  /gene="SPBC16A3.06"  /product="tRNA specific adenosine deamina
misc_feature  4288229 4288245 c  ctaataaaagaattaag, splice branch and acceptor
misc_feature  4288281 4288286 c  gtaata, splice donor sequence
misc_feature  4288377 4288393 c  ttaacaagtcaatttag, splice branch and acceptor
misc_feature  4288413 4288418 c  gtatgc, splice donor sequence
misc_feature  4288434 4288446 c  ctgacatctaaag, splice branch and acceptor
misc_feature  4288469 4288474 c  gtacgt, splice donor sequence
```



Community genomics (a.k.a. metagenomics)

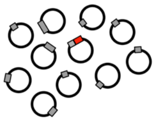
Environmental Sample

Extract DNA



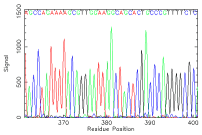
Sheared
Size selection

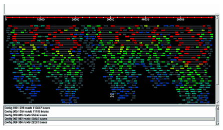
Clone: OPTIONAL, depends on seq'ing method and goals:



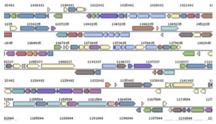
Library Type:
Shotgun (small-insert) 3kb
Fosmid (large-insert) 40 kb
BAC (large-insert) BIG STUFF!
--- transcription-free? ---

High throughput sequence





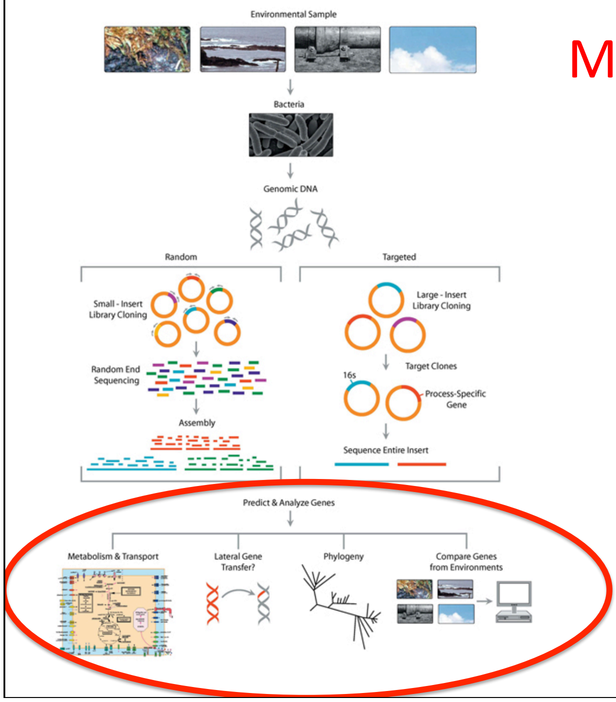
Assemble reads



Call genes

Annotate potential function

Metagenomics



Environmental Sample

Bacteria

Genomic DNA

Random

Targeted

Small - Insert Library Cloning

Random End Sequencing

Assembly

Large - Insert Library Cloning

Target Clones

16s

Process-Specific Gene

Sequence Entire Insert

Predict & Analyze Genes

Metabolism & Transport

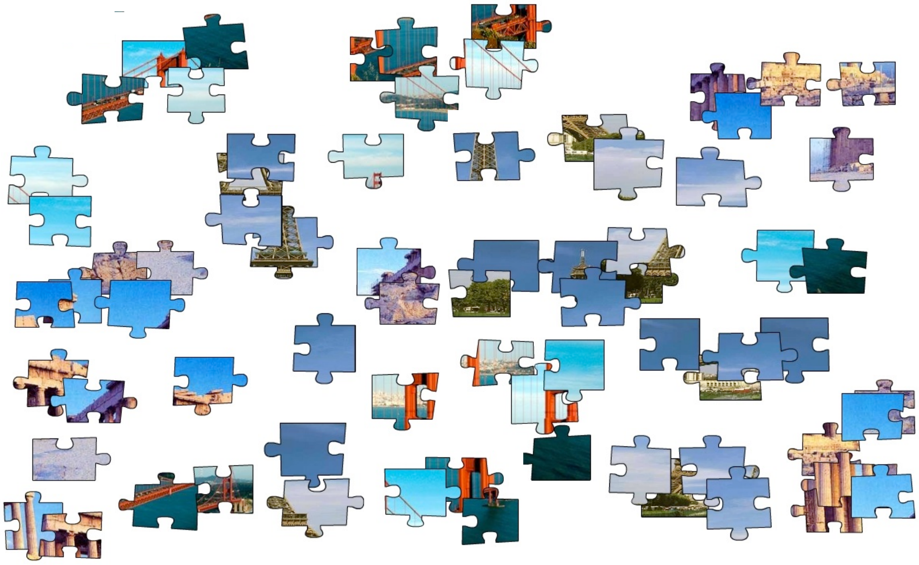
Lateral Gene Transfer?

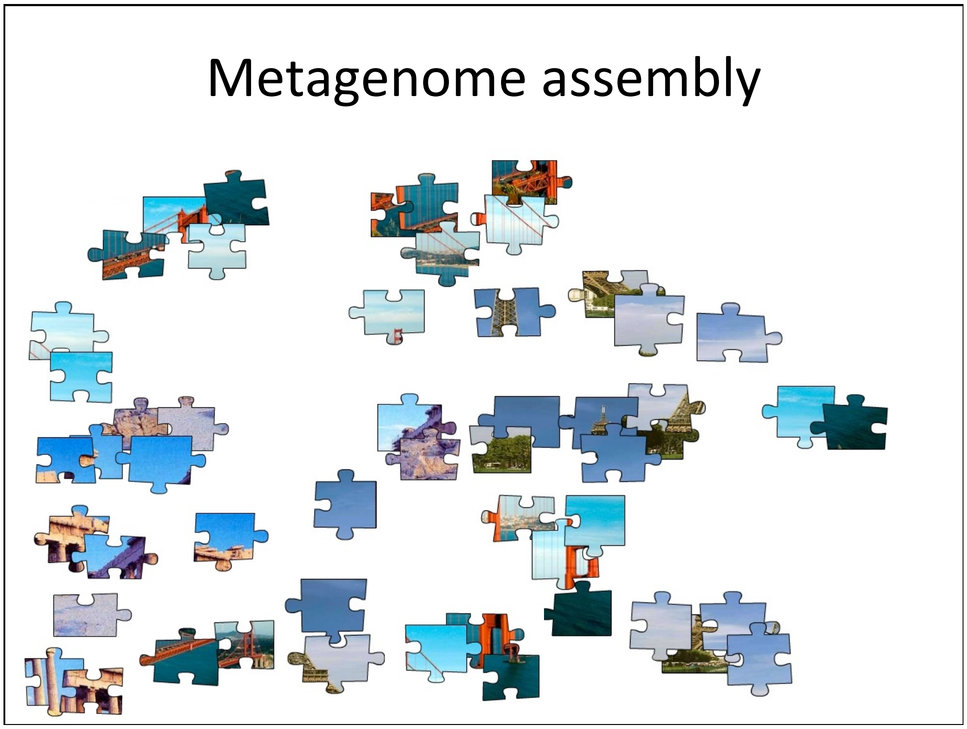
Phylogeny

Compare Genes from Environments

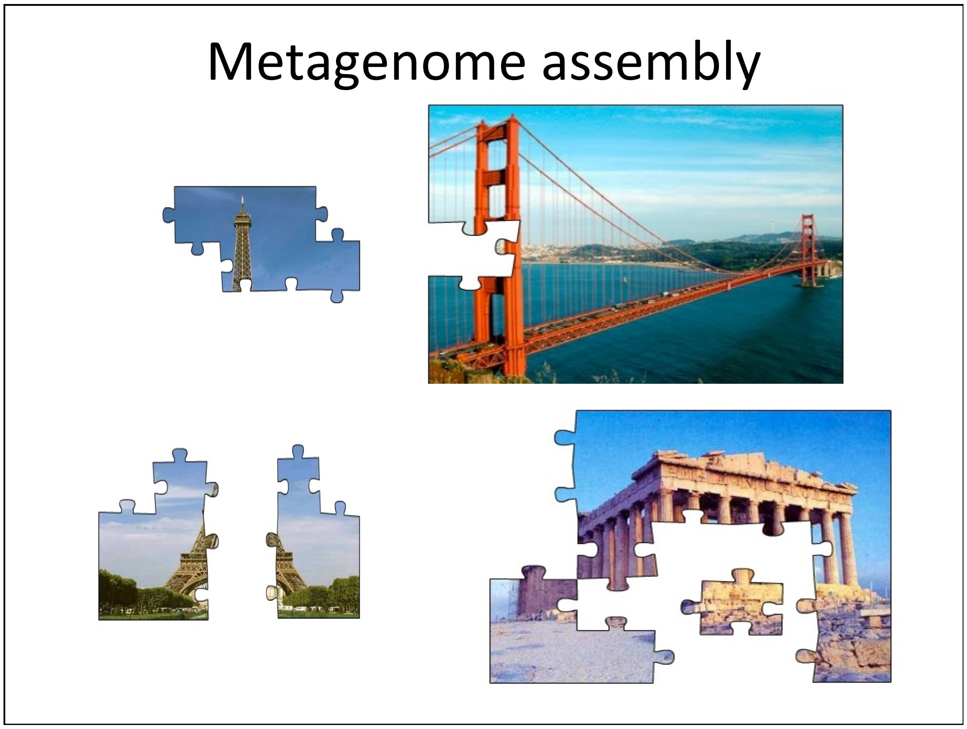
PLoS Biology Ocean Metagenomics Collection

Metagenome assembly

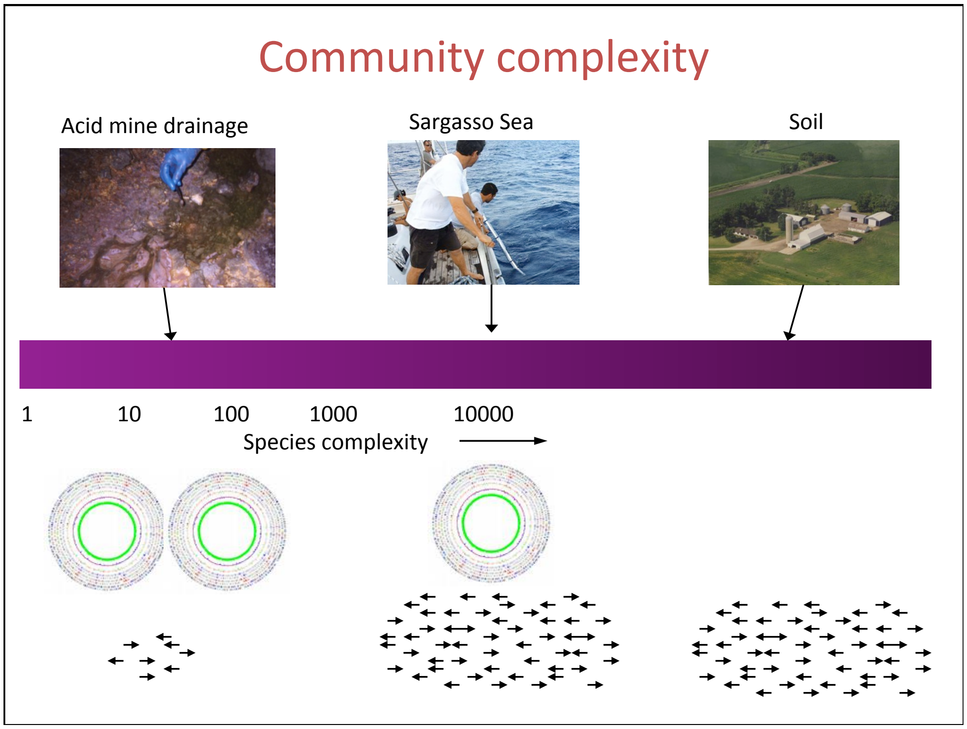




Metagenome assembly



Metagenome assembly



Community complexity

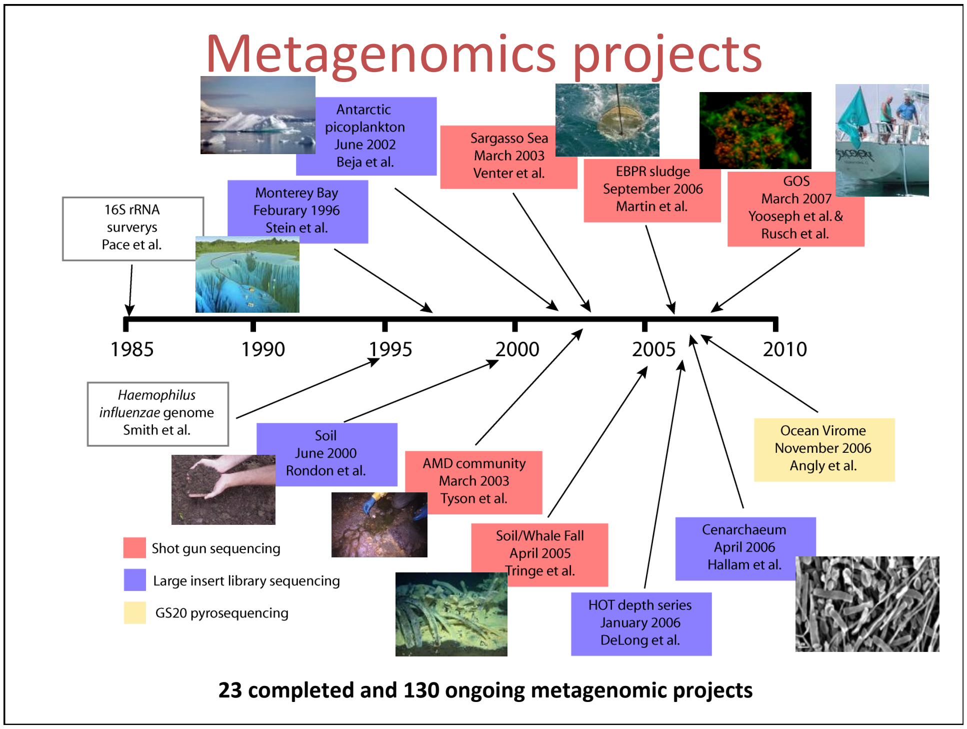
Acid mine drainage

Sargasso Sea

Soil

1 10 100 1000 10000

Species complexity



Metagenomics projects

Antarctic picoplankton
June 2002
Beja et al.

Sargasso Sea
March 2003
Venter et al.

EBPR sludge
September 2006
Martin et al.

GOS
March 2007
Yooseph et al. &
Rusch et al.

16S rRNA
surverys
Pace et al.

Monterey Bay
Feburary 1996
Stein et al.

Haemophilus influenzae genome
Smith et al.

Soil
June 2000
Rondon et al.

AMD community
March 2003
Tyson et al.

Soil/Whale Fall
April 2005
Tringe et al.

HOT depth series
January 2006
DeLong et al.

Cenarchaeum
April 2006
Hallam et al.

Ocean Virome
November 2006
Angly et al.

🟥 Shot gun sequencing

🟦 Large insert library sequencing

🟨 GS20 pyrosequencing

23 completed and 130 ongoing metagenomic projects

What to do with the data?

EGTs = Environmental Gene Tags

Predict ORFs (genes) in sequence data

Assign a function to ORFs

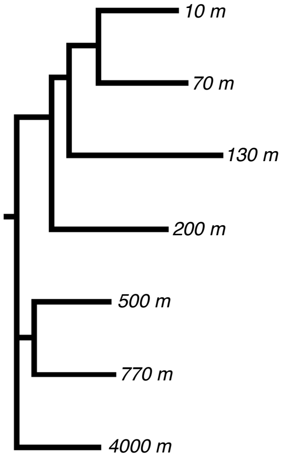
Compare relative abundance across habitats

Community Genomics Among Stratified Microbial Assemblages in the Ocean's Interior

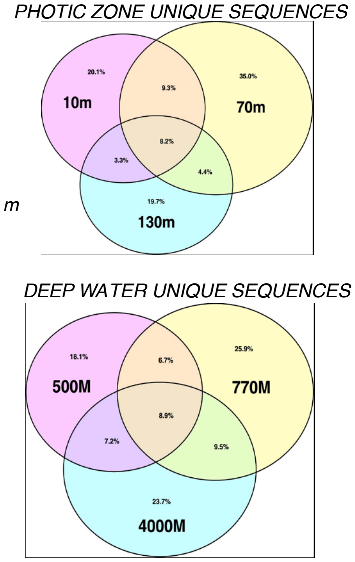
Edward F. DeLong,[1*] Christina M. Preston,[2] Tracy Mincer,[1] Virginia Rich,[1] Steven J. Hallam,[1] Niels-Ulrik Frigaard,[1] Asuncion Martinez,[1] Matthew B. Sullivan,[1] Robert Edwards,[3] Beltran Rodriguez Brito,[3] Sallie W. Chisholm,[1] David M. Karl[4]

SCIENCE VOL 311 27 JANUARY 2006

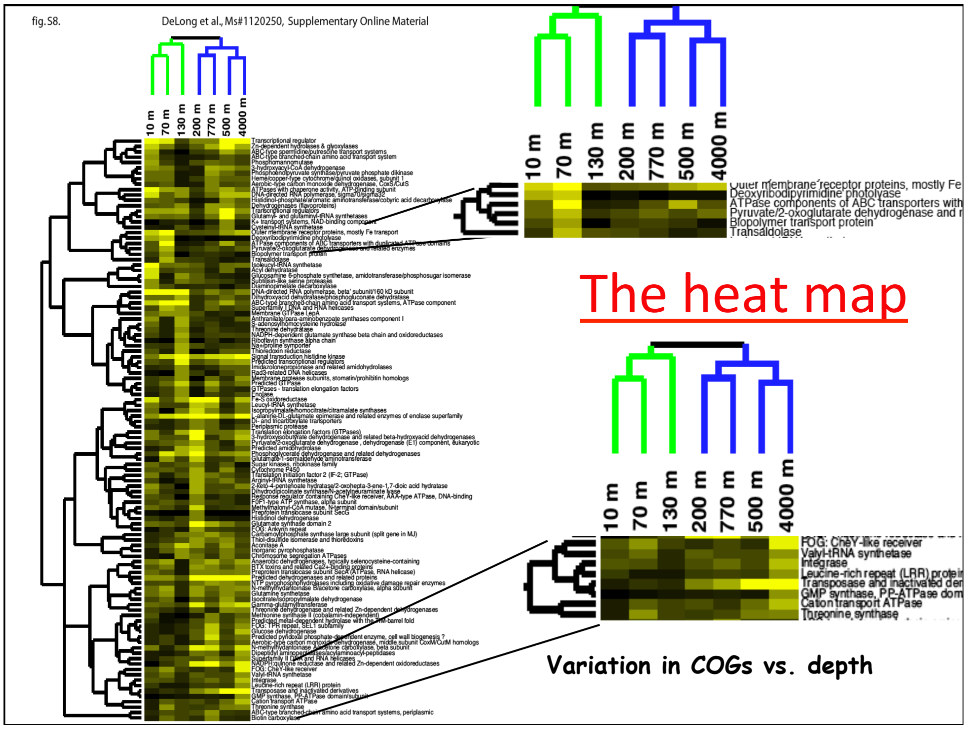
Cladogram



Venn diagram


PHOTIC ZONE UNIQUE SEQUENCES

DEEP WATER UNIQUE SEQUENCES



The heat map

Variation in COGs vs. depth

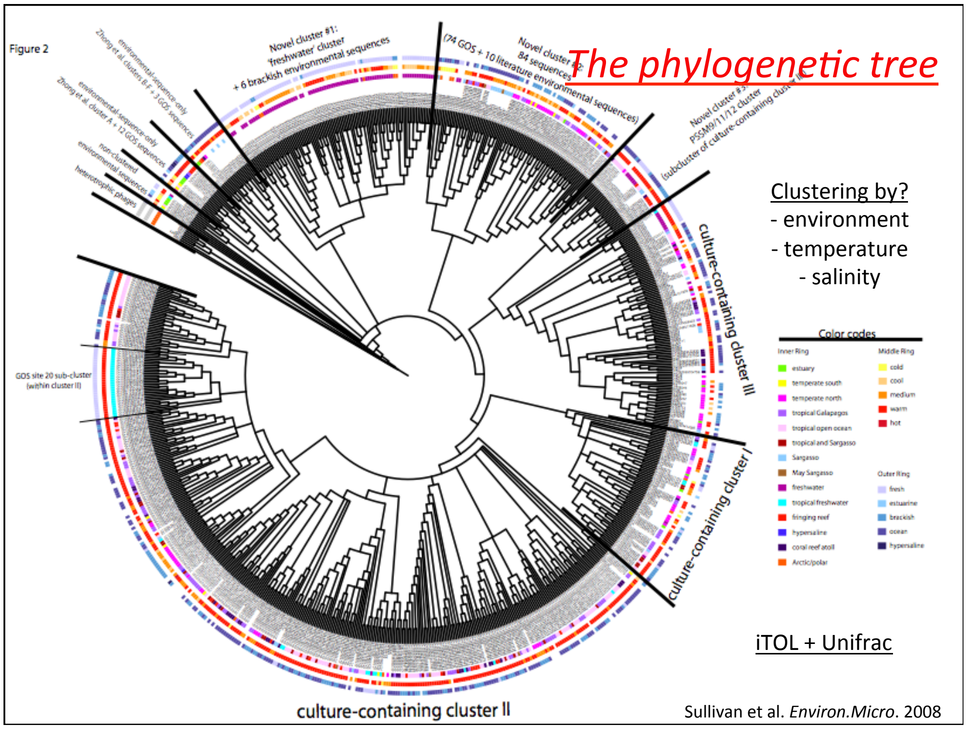


Figure 2
Novel cluster #1: 'freshwater' cluster + 6 brackish environmental sequences
environmental-sequence-only Zhong et al. clusters B-F + 3 GOS sequences
environmental-sequence-only Zhong et al. cluster A + 12 GOS sequences
non-clustered environmental sequences
heterotrophic phages
(74 GOS + 10 literature environmental sequences)
Novel cluster #2: 84 sequences
Novel cluster #3: P-SSM9/11/12 cluster (subcluster of culture-containing cluster III)
culture-containing cluster III
culture-containing cluster I
culture-containing cluster II
GOS site 20 sub-cluster (within cluster II)
Color codes
Inner Ring
estuary
temperate south
temperate north
tropical Galapagos
tropical open ocean
tropical and Sargasso
Sargasso
May Sargasso
freshwater
tropical freshwater
fringing reef
hypersaline
coral reef atoll
Arctic/polar
Middle Ring
cold
cool
medium
warm
hot
Outer Ring
fresh
estuarine
brackish
ocean
hypersaline

The phylogenetic tree

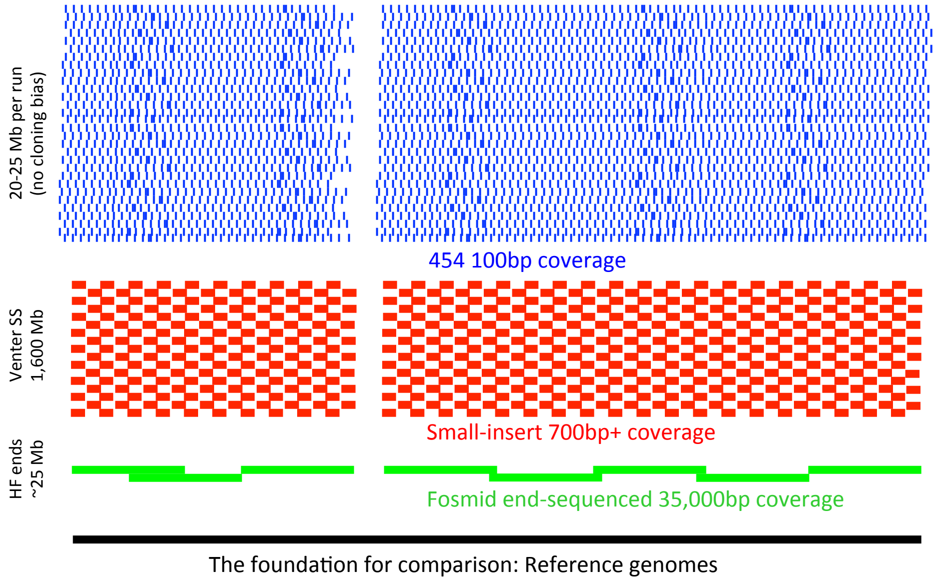
Clustering by?

- environment
- temperature
- salinity

iTOL + Unifrac

Sullivan et al. *Environ.Micro*. 2008

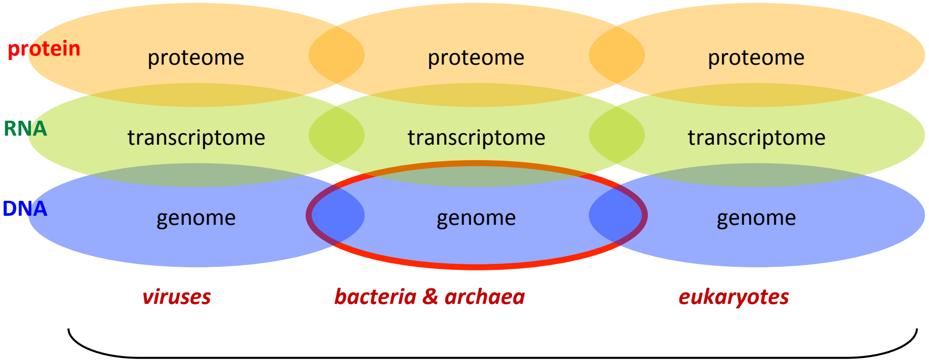
Comparison to reference genomes


20-25 Mb per run
(no cloning bias)
454 100bp coverage
Venter SS
1,600 Mb
Small-insert 700bp+ coverage
HF ends
~25 Mb
Fosmid end-sequenced 35,000bp coverage
The foundation for comparison: Reference genomes

Genomic analyses are strong
(tho' functional database limited)

comparative genomic analyses are limited
+ metagenomic analyses are still fairly primitive

Metagenomics is just the first level



protein
RNA
DNA

proteome | proteome | proteome

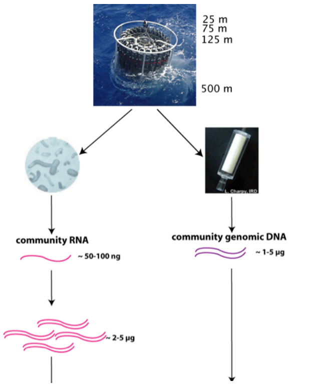
transcriptome | transcriptome | transcriptome

genome | genome | genome

viruses
bacteria & archaea
eukaryotes

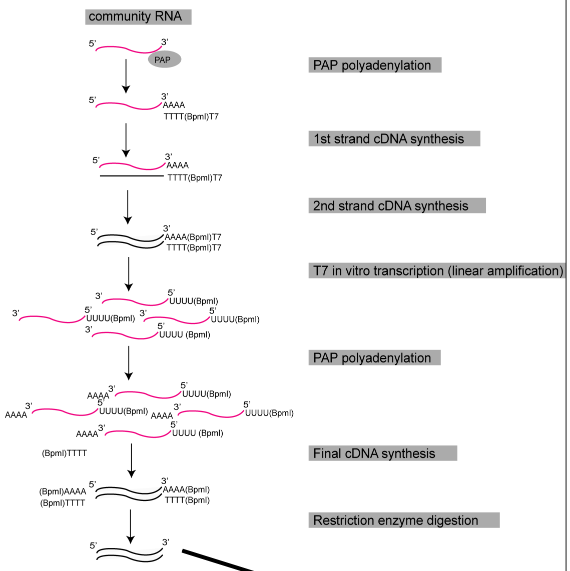
microbial communities

Metatranscriptomics



ng RNA --> µg cDNA

gDNA required for norm’ n



community RNA

PAP polyadenylation

1st strand cDNA synthesis

2nd strand cDNA synthesis

T7 in vitro transcription (linear amplification)

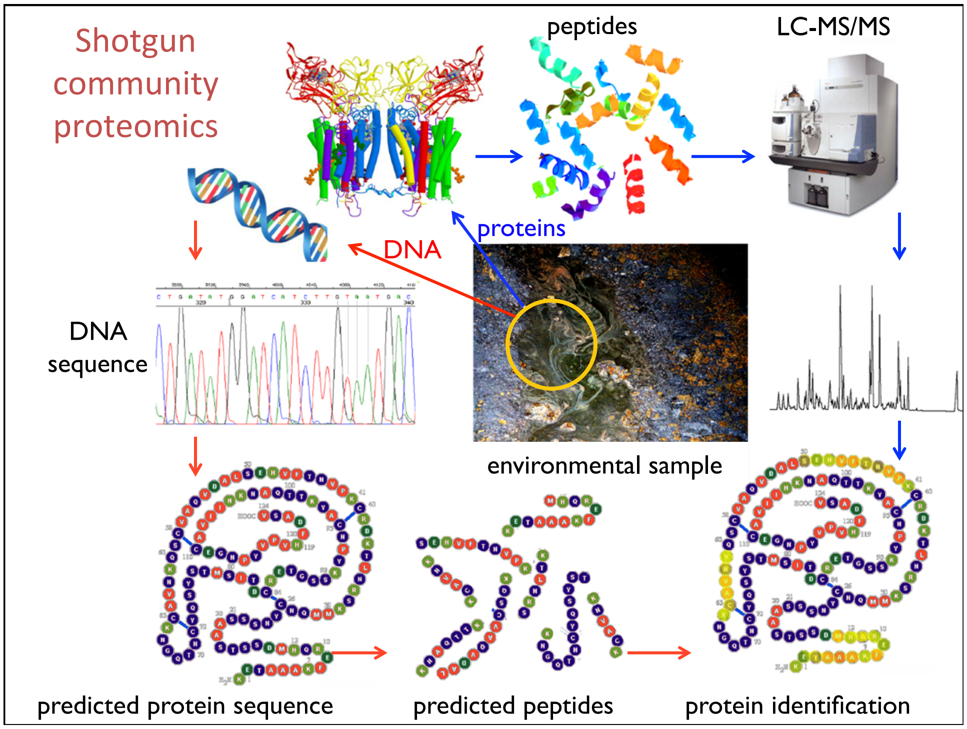
PAP polyadenylation

Final cDNA synthesis

Restriction enzyme digestion

Pyrosequence

Modified method optimized in DeLong/Chisholm labs



Shotgun community proteomics

peptides

LC-MS/MS

proteins

DNA

DNA sequence

environmental sample

predicted protein sequence

predicted peptides

protein identification

Handling the data

“We keep the sequences and quality values, and throw away pretty much everything else almost immediately. It's cheaper to re-sequence than to store the raw data."

“We’ re burning data on hard drives to ship between sites.”

“Nothing is backed up to tape anymore.”

… and the metadata … and data analyses …

Metadata: sample information must be databased and linked to each sequence

Data analysis:
Two main work-flows: reference-guided assemblies (for variant analysis) and all-against-all (for metagenomics and transcriptomics applications)

The emergence of "Proxy" data in database searches

What does it mean to you?

No matter the organism -- time to think about genomics

- genomes, metagenomes, gene + protein expression
- population studies through phylogenies, association studies, population genOMICs

(Clinical researchers soon will include a personal genome within a patient's electronic medical record)