Microarrays – (Wilson et al 2002)

60,000 probes/chip now 500,000/chip - hybridize fluorescently labeled PCR pool, or cDNA pool
May miss total unknowns
Reading differences in hybridizations within array may be tricky
Currently expensive

Environmental Genomics

Terminology:

**Shotgun cloning** - undirected cloning effort were everything is cloned and sequenced

**contig** - assembled continuous sequence derived from sequence reads of a single clone

**Scaffold** - assembled continuous sequences derived from multiple overlapping contigs whose physical connection is inferred from sequence identity

**nX coverage** - mean number of times a region was sequenced from independent clones.

**mini-scaffold** - scaffolds assembled only by paired ends of overlapping contigs (i.e., 1X coverage)


Iron Mt. biofilm (Tyson et al 2004)

Pink biofilm growing at pH 0.87, simple system composed of 6 16S rRNA types.

*Leptospirillum* gr II: High GC scaffolds with 10X coverage - very low polymorphism (0.08%), interpreted as evidence for one strain or species.

*Leptospirillum* gr III High GC scaffolds with 3X coverage, contains the only nitrogen fixation genes in the system

*Ferroplasma* Type II 10X coverage, scaffolds were 22% different from to FerI isolate even though the 16S gene was 99% identical. average polymorphism within scaffolds was 2.2% - interpreted as several strains with recombination
**Sulphobacillus** partial 16 S recovered - scaffolds for it (if any) may have been assigned to *Leptospirillium* group III

**Sargasso Sea** (Venter et al. 2004.)

**Some numbers:**
- 200 L of filtered surface water (0.1 - 3 um), shotgun cloning of 2-6 kB inserts (no PCR),
- 1.66 million sequences obtained,
- 246 mpb of unique sequence assembled into 64,398 scaffolds ranging from 826 bp to 2.1 Mbp., 169.9 Mbp mini-scaffolds and unpaired singleton reads;
- 1.2 million protein coding genes (10X more that currently in databases). 69,901 novel, conserved open reading frames.
- 60,000 16S rRNA sequences, 148 of which are at least 3% different from known sequences.

**Problem of assembly** - usually all unique sequences have equal probability of being sampled, but in an environmental sample depth of coverage of a particular genome should be related to abundance of the organism.

Larger scaffolds sorted into organism bins based on similarity to known sequence and oligonucleotide frequency.

**Estimates of diversity sampled**
- at least 300 species/sample

Assuming homologous sequence > 6% different belong to different species, and using models based on Poisson distributions and 3 coverage-based models estimates, they estimate 1800, to 47,000 unique genomes in the combined sample, and minimum of 12X greater effort would be needed to have a 95% of the unique sequences.

**Populations are sequenced rather than an individual** Scaffolds with 14X coverage contained about 1 SNP/10,000 bases, also insertions of bacteriophage sequences were common.

SAR 11 - like sequences are abundant but very polymorphic - suggestive of a population of related taxa that share a common 16S sequence

Other results of ecological interest:

Evidence for patchiness in marine samples Sample 1 had and abundance of a *Burkholderia* and *Shewanella* not found in sample 2

Copy number bias of 16S sequences obvious (figure 6).
Estimations of species present based on coverage frequency classes (Table 3)

soil is still out of reach - but maybe not for long

**Quantification via oligonucleotide probing**

hybridization to filters with blotted rRNA (Stahl fig.)
Specific hybridization/universal
what about unknown probe specificities?

**Quantification via realtime PCR - based on measure product as it accumulates** (see Mackay et al 2002) for a review,(Landeweert et al 2003) for ecological example.

Based on assaying the accumulation of product as it occurs. Thermocycler excites and detects fluorophores.

simplest way involves intercalary dyes (ethidium, SYBER green, YO-PRO-1) - accumulation of double-stranded product detected. Only works for a quantification of one specific template at a time.

accumulation of specific product can be monitored with specific probes
Usually based on **FRET**: fluorescence resonance energy transfer
either two fluorescent labels or one fluorescent label and a **NFQ** (non-fluorescent quencher)

several variations (more all the time) (see fig 3 of McKay).

5'nuclease oligoprobes or "TaqMan" probes based on 5'-3' nuclease activity of Taq polymerase (A in figure)

hairpin probes (or "molecular beacons") inactive until bound to template (B) - more sensitive to mismatch than linear probes

two adjacent probes (C) - inactive until bound next to each other on template

Sunrise primers (D) - used as amplification primer; hairpin is removed when primer is incorporated.

Scorpion primer - similar to above but second stand synthesis is not required.

SNIP monitoring via melt profile

**Landeweert et al.** paper uses quantitative PCR to assay competition between two mycorrhizal fungi: *Paxillus* and *Suillus*. Probe design based on two head-to-head
probes that fluoresce when both are bound, and specific amplification primers for ITS 2 regions of both.

Other quantification, included visual counts of mycelium (in single experiments), PLFA estimates of mycelium (which does not differentiate these species),

Three molecular methods used: DGGE analysis of ITS region, cloning and clone counts of ITS region, real-time quantification

Bottom line - all three show the same pattern *Suillus* increases over time as *Paxillus* declines.

**References**


