High throughput sequencing of fungal communities

Ari Jumpponen
Division of Biology
Ecological Genomics Institute
Kansas State University

Ken Jones
Environmental Health Science
University of Georgia

Outline

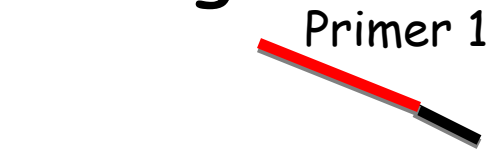
- High thoughput sequencing?
  - DNA tags as sample IDs
  - 454 pyrosequencing platform
- Sequencing of regions on 454 platform
- Trade offs - depth vs. sample numbers
  - Trade offs - costs

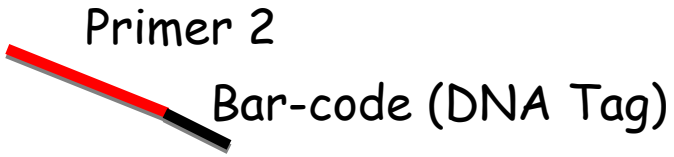
High throughput sequencing

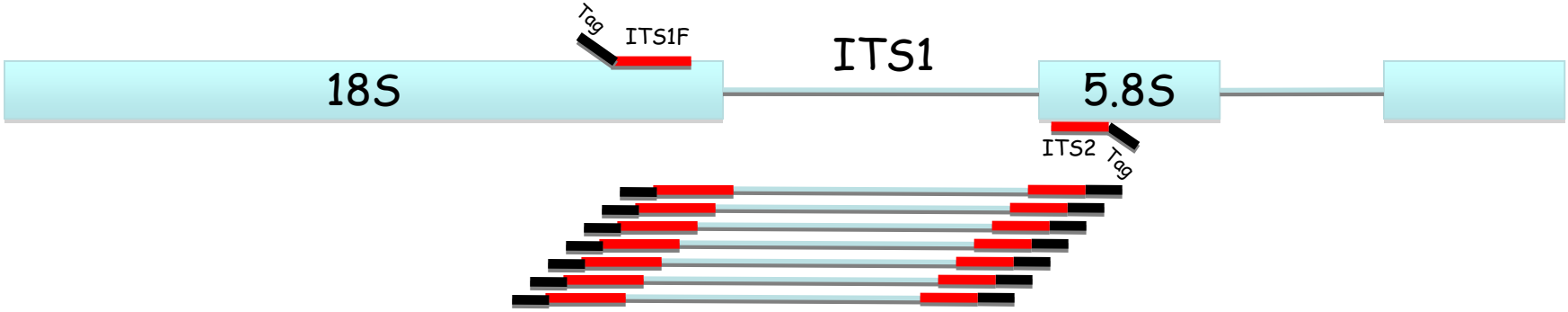
- Sequencing of a large number of templates from heterogeneous samples
  - Clone libraries and Sanger sequencing
    - Automated colony picking
    - 96- or 384-well formats for extension/sequencing
  - Direct sequencing on 454 pyrosequencing platform
    - Incorporation of 454 endemic priming sites
    - emPCR signal amplification and massively parallel sequencing
- Both can use DNA tagging (primer barcoding for sample ID)

Primer tagging - DNA barcoding

- Use a user-determined identifier DNA-tag
- Tagging both primers allows non-directional sequencing with vector-residing primers
- The tags can be added by ligation into A-overhang, directly in PCR or in a nested PCR reaction

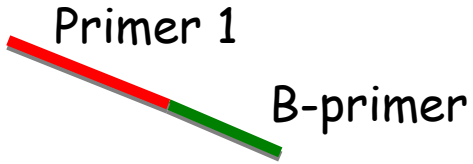


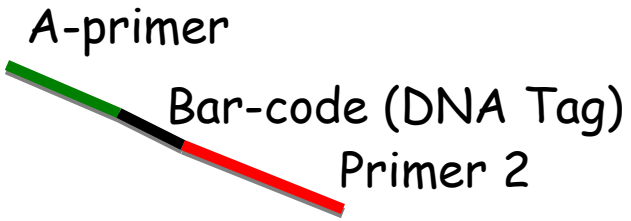


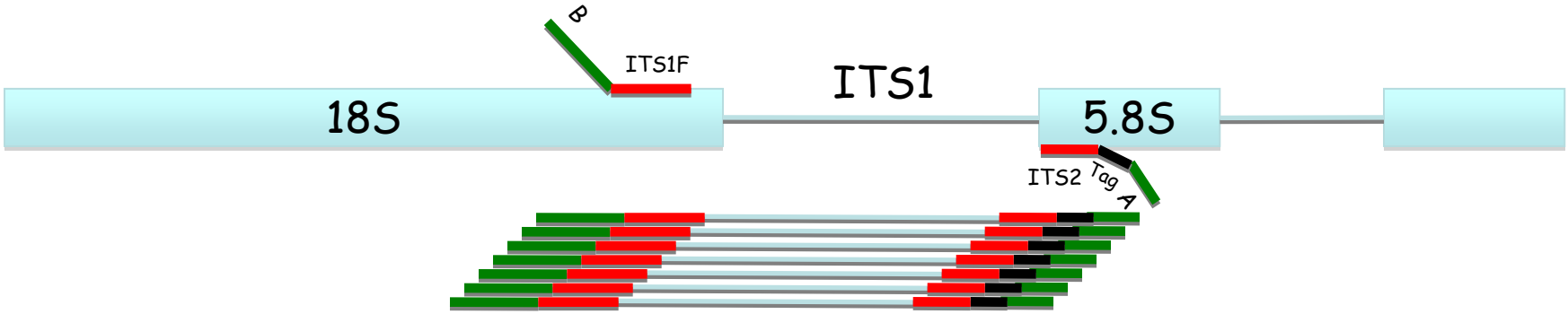


Amplification for 454 pyrosequencing

- Since the sequencing is done with a platform endemic primer, one barcoded primer necessary

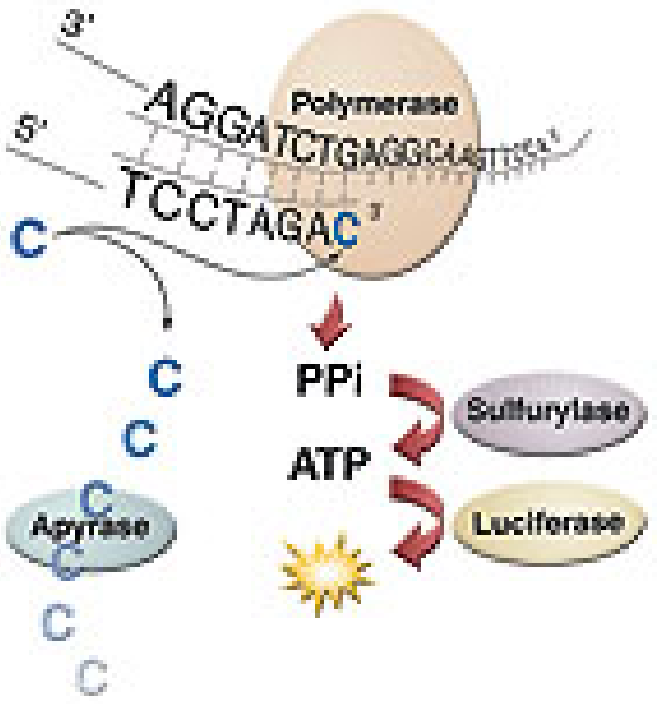






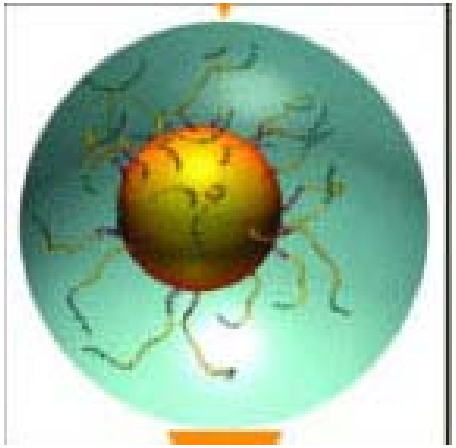
What is pyrosequencing?

- Developed by Ronaghi and Nyrén (Analytical Biochemistry 1996 and Science 1998) for BioTag, licenced now to 454 LifeSciences (bought by Roche 2007)
- In brief, DNA polymerase incorporation of phosphorylated dNTP's emits light.
- Sequencing way faster than with the chain termination method (Sanger).
- Huge throughput: 100M nucleotides in 7 hour run.

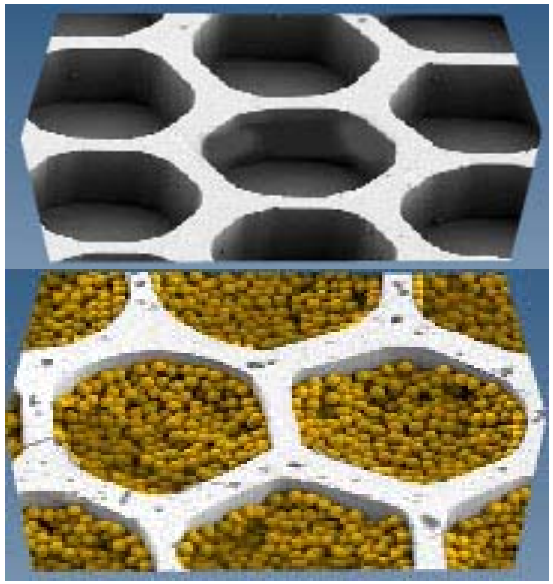


Massively Parallel 454 Sequencing

Emulsion PCR (emPCR) amplification of the PCR products in microreactor bubbles



Dispense beads to picotiter plates and fill packing beads



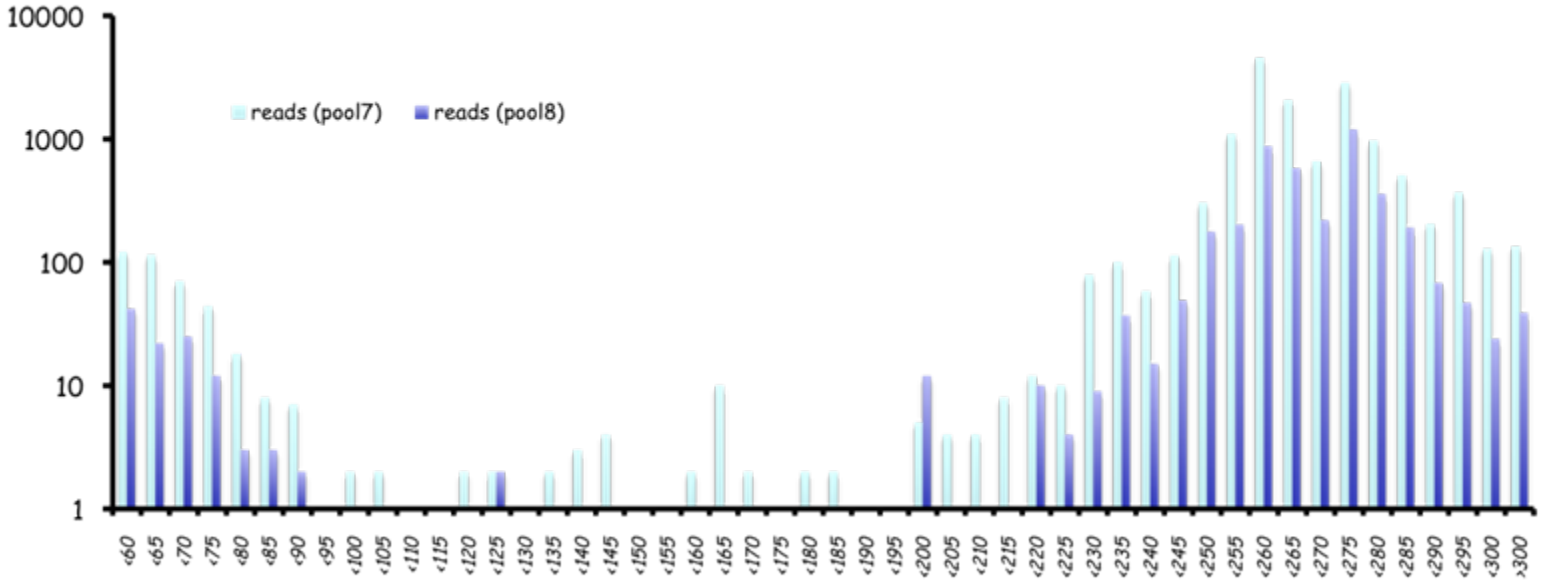
Sequence via pyrosequencing

http://www.454.com/

454 yields

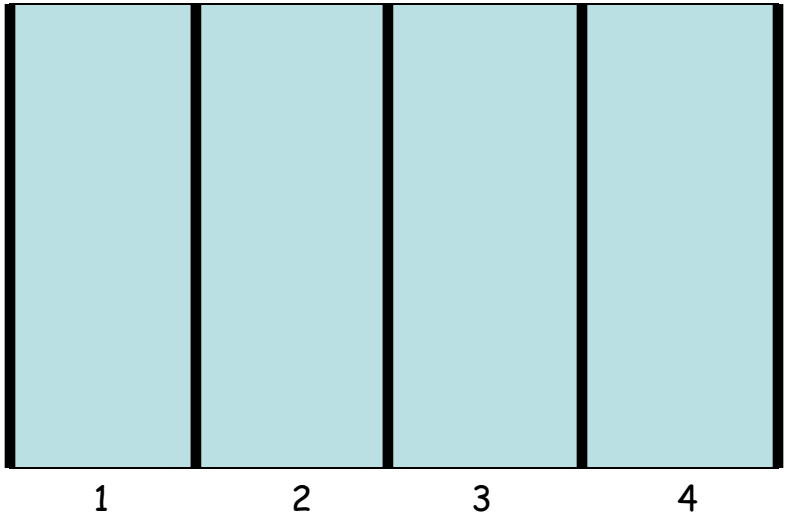
- Up to 250,000 reads, 265bp in length

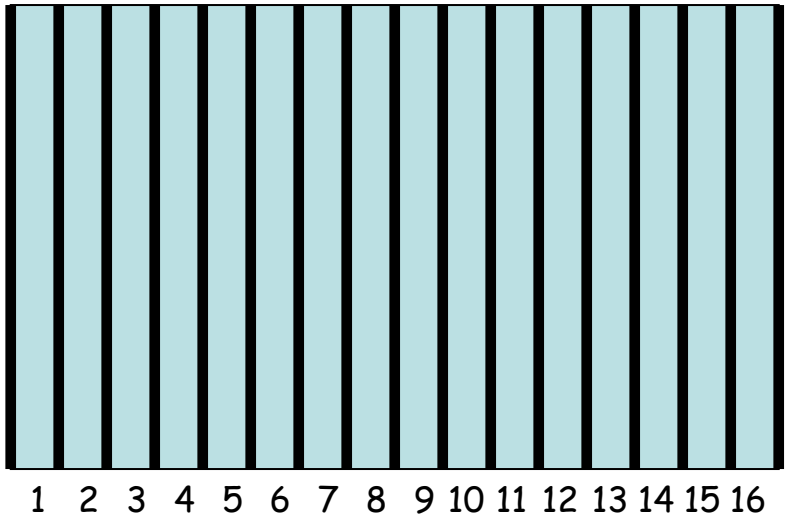
- Sequence quality control
  - ID Primer
  - ID Barcode
  - Ambiguous bases



454 sequencing in regions

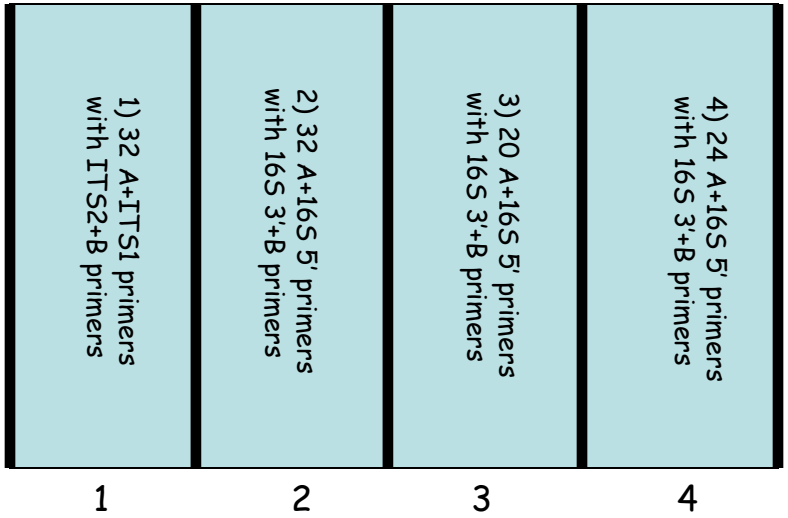
- RXN blocking
  - Each 454 rxn can be divided to 4 or 16 blocks
  - Loss of 10% (4 blocks) or 15-20% (16 blocks) of the throughput
  - Independent amplicon/data generation
- Combination of 16 regions and 64 tagged primers allows sequencing of 1,064 templates at ~150 reads per sample





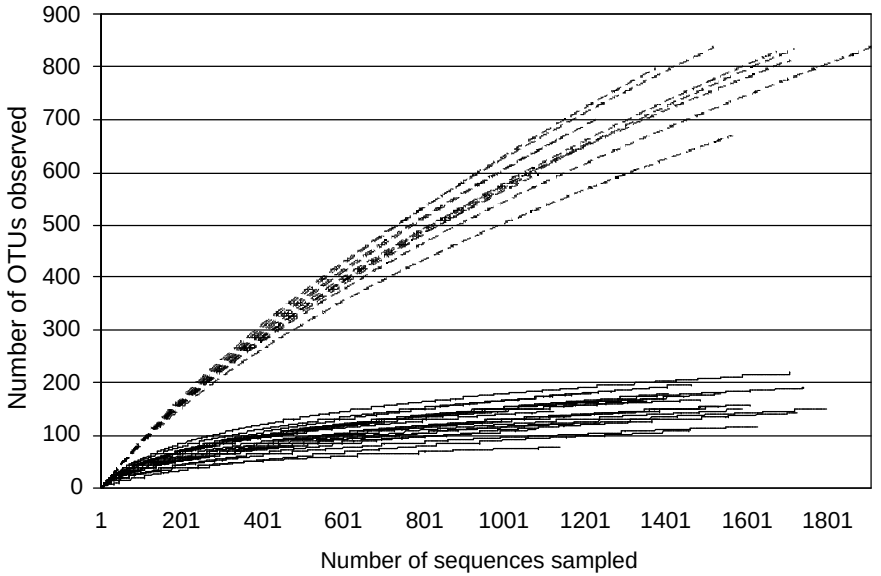
Example of region separation

- 4 experiments in 4 block + barcoding
  - 1) 32 barcoded (5bp barcodes) samples to study microeukaryon responses to rainfall manipulation and warming.
  - 2) 32 barcoded (5bp barcodes) samples to study bacterial responses to rainfall manipulation and warming
  - 3) 20 barcoded (5bp barcodes) samples Peromyscus gut microbes in contaminated and control sites
  - 4) 24 barcoded samples to study bacterial resposes to N deposition
  - = 108 samples for a total of ≥160,000 reads; ~1,500 reads per sample



Acquired Data – Numbers

- Rainfall Manipulation Plots (RaMPs)
- 1,200-1,800 reads per soil sample
- Eubacteria
  - 44,363 16S reads
  - 7,190 non-singletons (72% of the reads) + 12,439 singetons
  - Total 19,629 OTUs
- Eukarya
  - 41,512 ITS1 reads
  - 1,802 non-singletons (92% of the reads) + 3,264 singetons
  - Total 5,066 OTUs
- Eukarya nearly saturated, bacteria likely require 10x more sampling…



OTU richness per sample. Dashed line - bacteria; solid line - fungi.

Depth vs. number of samples

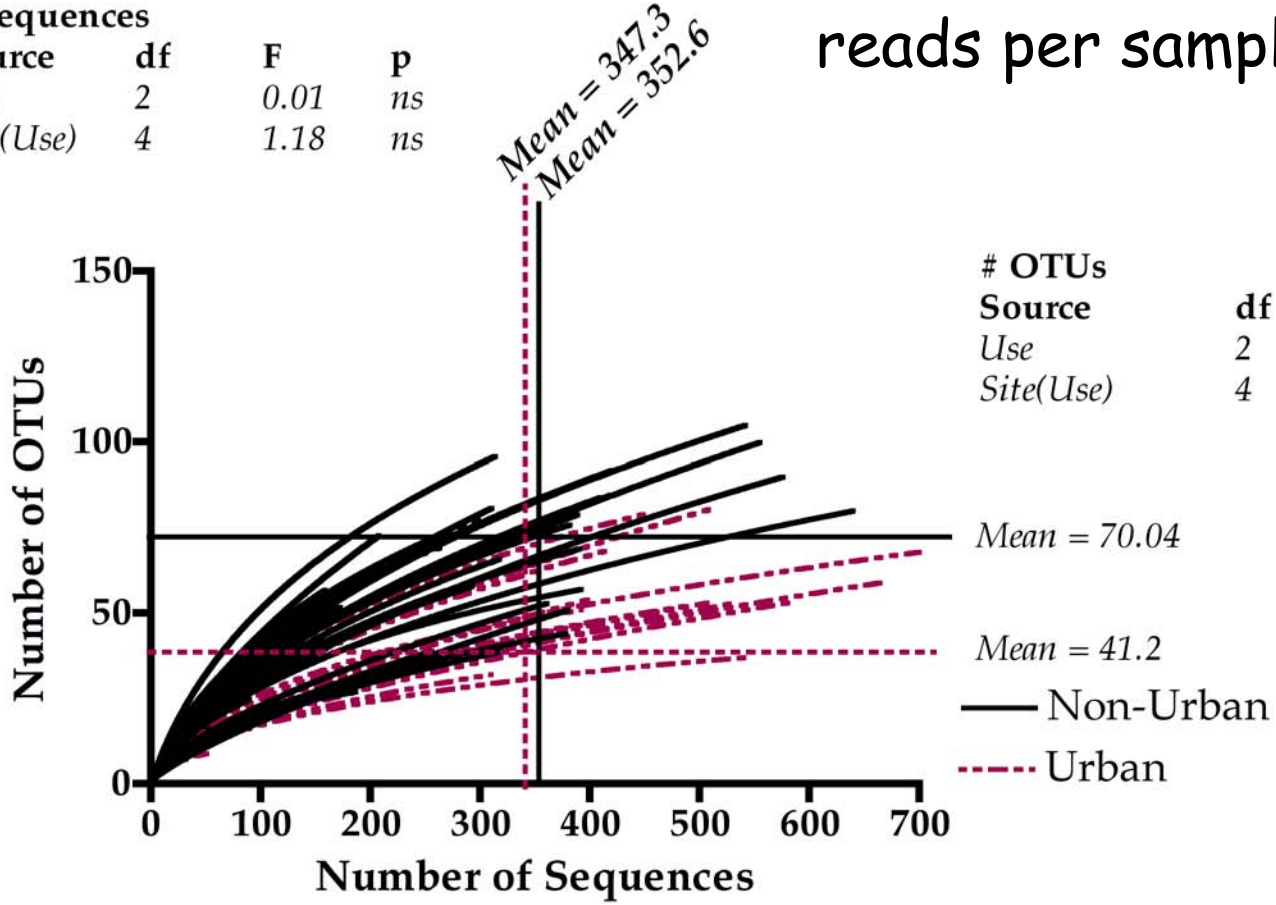
Foliar endophyte communities: no saturation at 300-700 reads per sample

 # Sequences

Source	df	F	p
Use	*2*	*0.01*	*ns*
Site(Use)	*4*	*1.18*	*ns*

 # OTUs

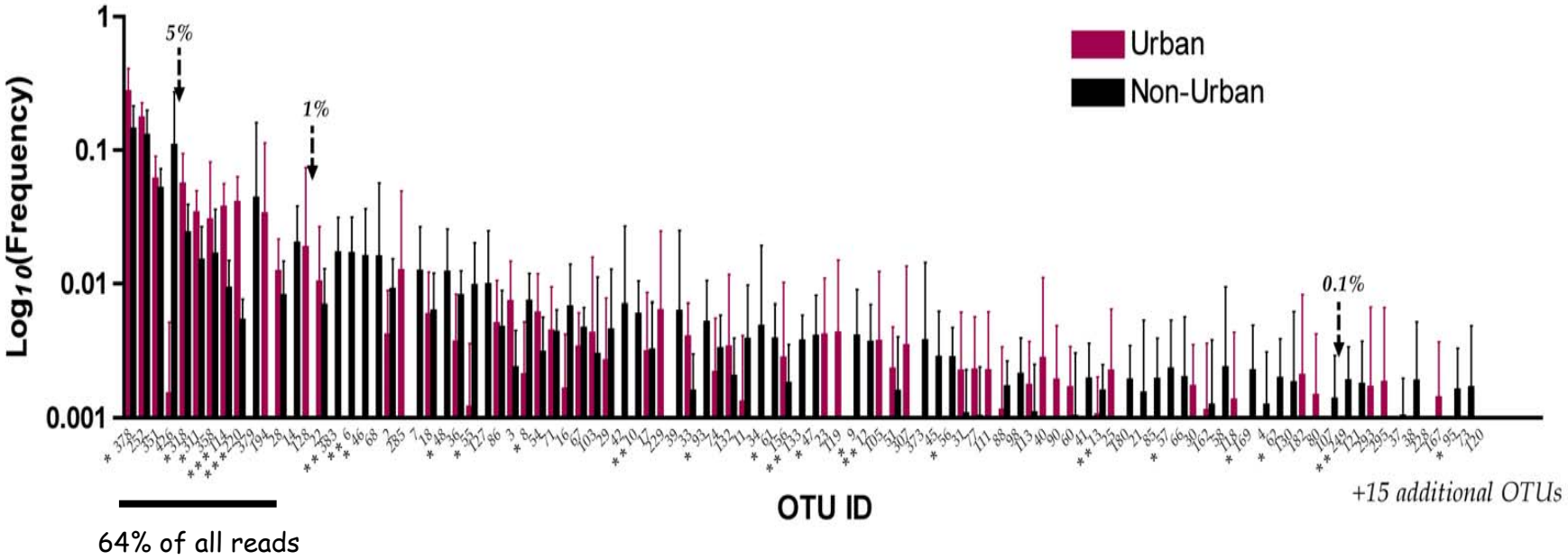
Source	df	F	p
Use	*2*	*13.08*	*<.0001*
Site(Use)	*4*	*1.40*	*ns*



Depth vs. number of samples

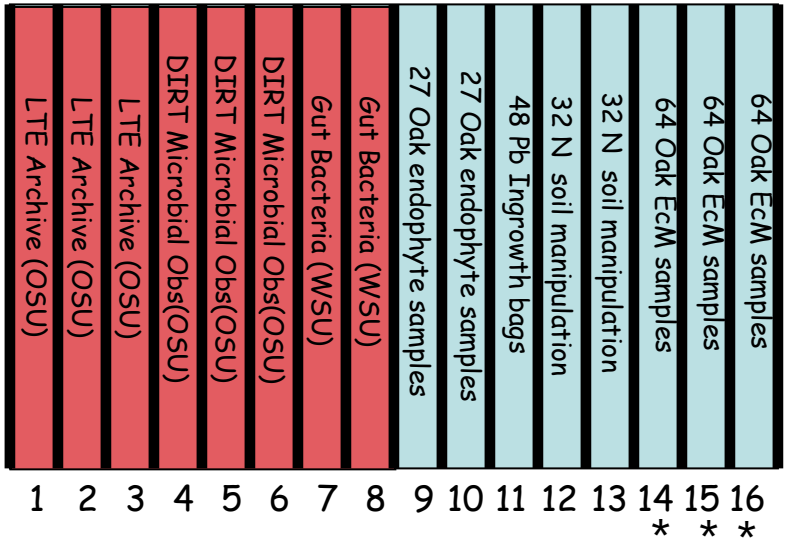
Foliar endophyte communities:

1) Top 10 most abundant OTUs comprised 64% of all reads
2) Only 14 most abundant OTUs ≥ 1% of all reads
3) Only 4 most abundant OTUs ≥ 5% of all reads



Costs - Urban Ectomycorrhizas

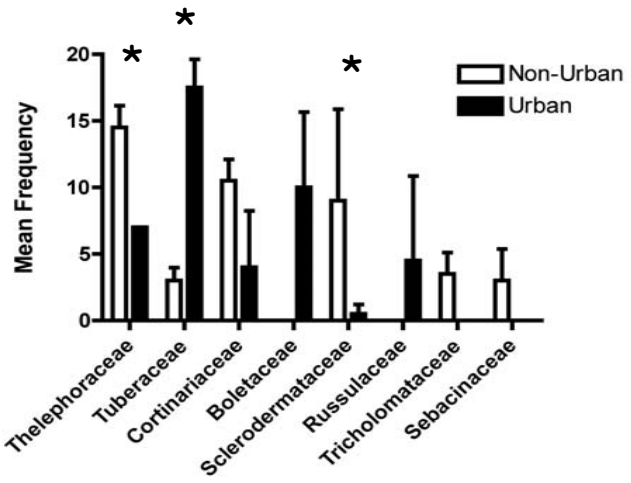
- 16 region RXN divided between 3 labs
- 6 microbial expts in 14 regions
  - 14-16) 64 barcoded EcM samples for each of 3 seasonal samples
  - 12-13) 2x32 barcoded LTER soil samples
  - 11) 48 barcoded fungal samples from Pb contaminated and control sites
  - 9-10) 54 barcoded samples for oak foliar endophytes
  - (7-8) Peromyscus gut microbiome)
  - 4-6) 20 soil archive samples
  - 1-3) 12 DIRT microbial observatory samples
  - = total of 385 samples for ~100,000 reads; average of ~260 reads per sample
- Total 454 cost $12,500; ~$780 per region.

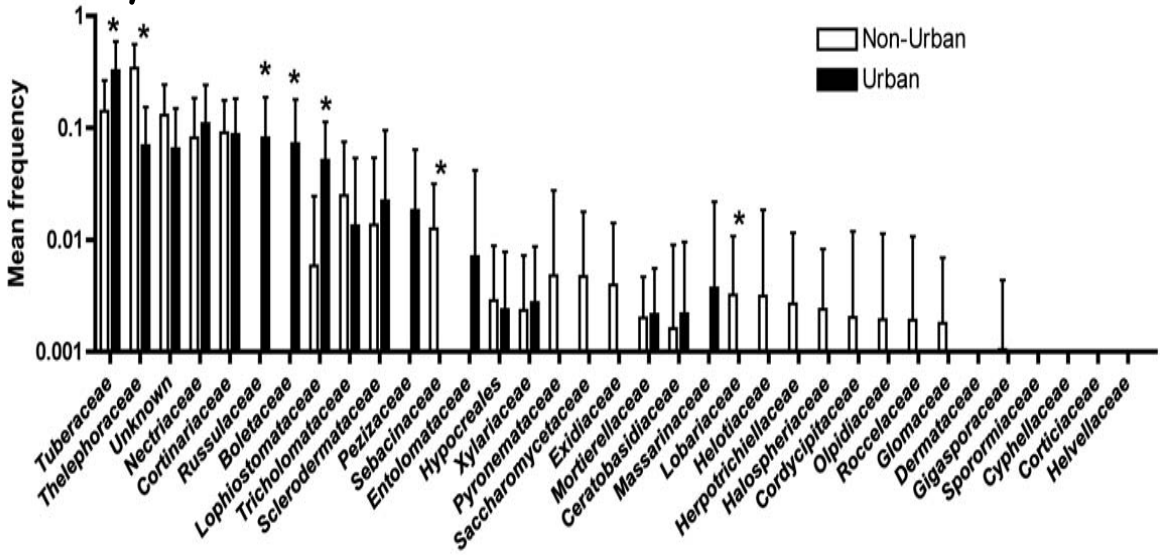


Costs - Urban Ectomycorrhizas

Clone and sequence:
384 reads (4 libraries, 96 reads)
$3.65 per read ~ $1,400
64 samples: 16 pooled per library

454 sequencing:
~27K reads (3 regions)
192 samples
$780 per region ~ $2,340





Conclusions:

Thelephoraceae
Tuberaceae
Sclerodermataceae
Boletaceae

Clone
Non-Urban (2x)*
Urban (6x)*
Non-Urban (10x)*
Absent in Non-Urban

454
Non-urban (6x)*
Urban (6x)*
Non-Urban (1.5x)^{ns}
Absent in Non-Urban

Thanks to…

- People
  - J. David Mattox - City of Manhattan
  - Gary Kilner - KSU
  - Chulee Yaege – KSU
  - Charles Kramer - KSU
  - Nicholas Simpson – KSU
  - Ken Jones - UGA
- Funding
  - BRIEF - Division of Biology, KSU
  - Ecological Genomics Institute, KSU
  - National Science Foundation
- Facilities/Infrastructure
  - City of Manhattan
  - Konza Prairie Biological Station
  - Ecological Genomics Institute (KSU)