Metagenomic Analysis of the Feedback Responses of Soil Microbial Communities to Elevated CO_2

Joy Van Nostrand, Ph.D.

Joy.vannostrand@ou.edu
http://ieg.ou.edu

Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019

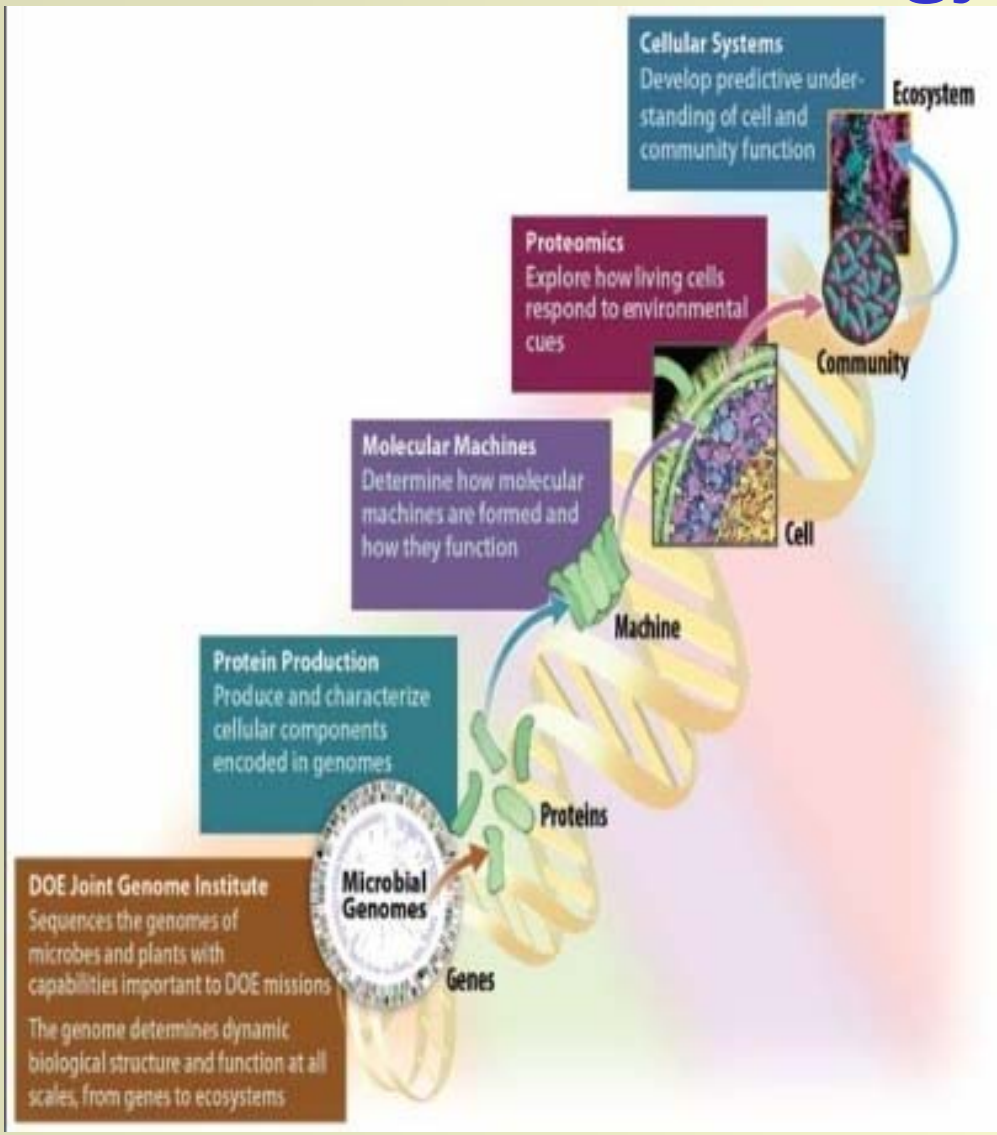
Advances in Microbial Diagnostic Microarrays Workshop
Teramo, IT
March 15-16, 2010

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

Outline

- **GeoChip development**
  - **GeoChip 3.0**
  - **GeoChip 4.0**
- **GeoChip applications**
  - Responses of microbial community to elevated CO_2
  - Effects of plant species diversity on microbial communities
  - Effects of plant functional groups on microbial communitiues
- **Ecological network analysis**
  - **Random matrix theory**

Some Grand Challenges in 21st Century Biology



- **Linking genomics to ecology**
  - Linking genomics to ecological processes and functions
  - Responses to CO2, global warming and water precipitation
- **Linking biodiversity to ecosystem functions**
- **Informational scaling**
  - From cells to individuals, populations, communities, ecosystems and biosphere.
  - Spatial, temporal

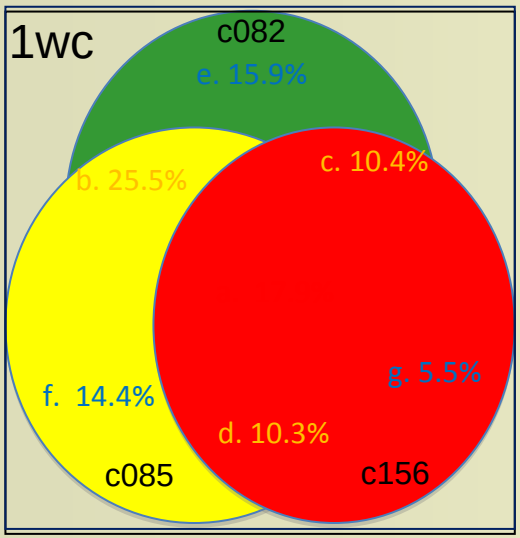
High throughput approaches

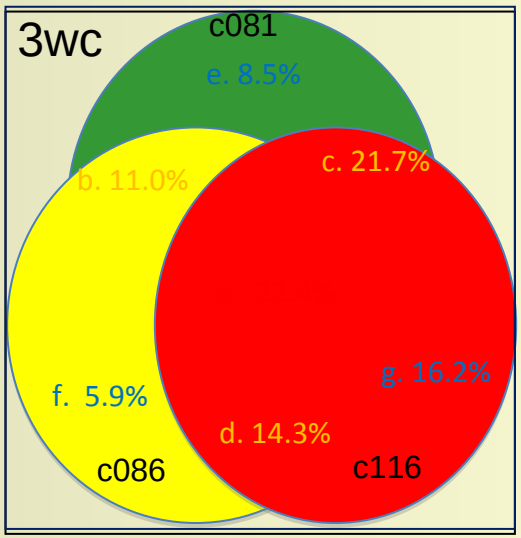
- **Open format detection**
  - Cannot assure the same genes/proteins/organisms will be compared across different samples. The results can not be expected and thus are open.
  - High throughput Sequencing
    - 454 sequencing, 250 bp, 60-100 mb/run
    - Solexa, SOLiD: 35 bp, 1-2 gb/run
  - Proteomics
  - Metabolomics
- **Closed format --- Microarrays**
  - **Ensure that the same genes/proteins/organisms can be compared across different samples. The results can be expected, and thus are closed.**
  - **PhyloChip: 16S genes**
  - **GeoChip: functional genes**

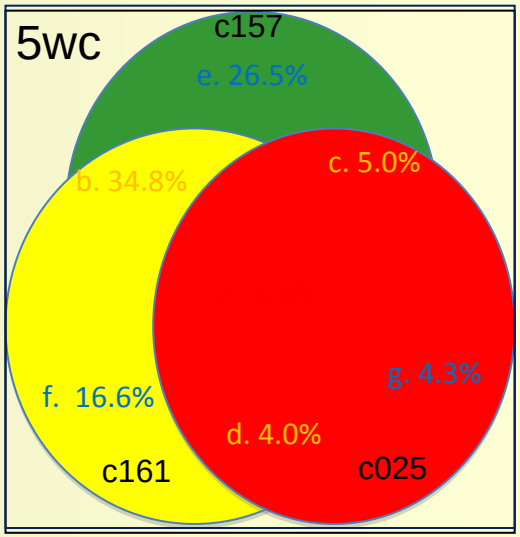
Comparisons between open format and closed format detection

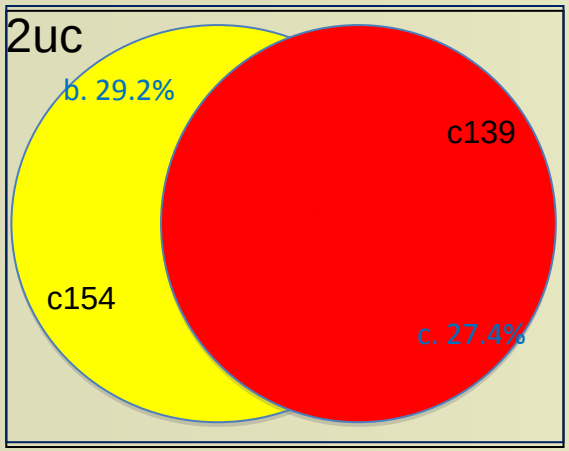
	Open format	Closed format
Sensitivity to random sampling errors	High	Low
Effects by dominant organisms	Yes	No
Finding new things	Yes	No
Sensitivity to contaminated DNA	Yes	No
Comparison across samples	?	Yes

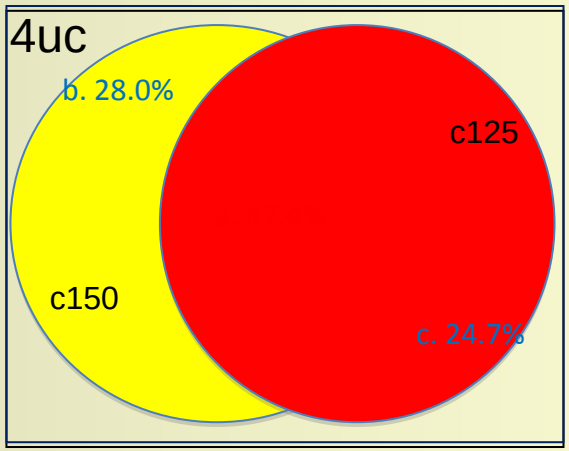
OTU Overlaps among Tags of Each Individual Sample (Treatment: uc)

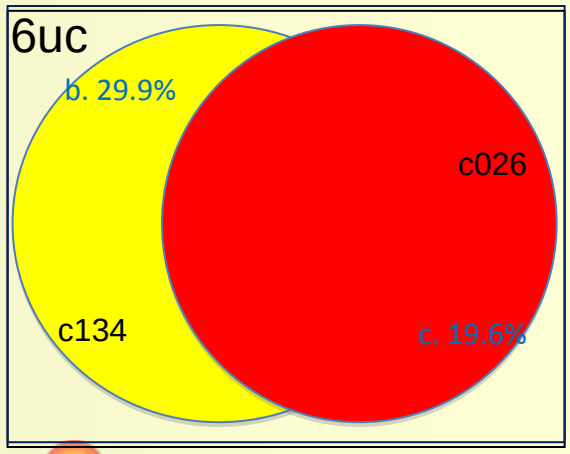








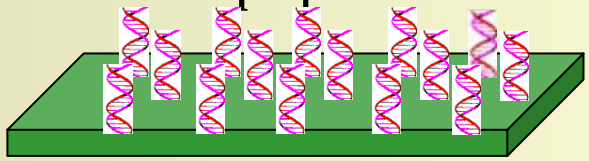


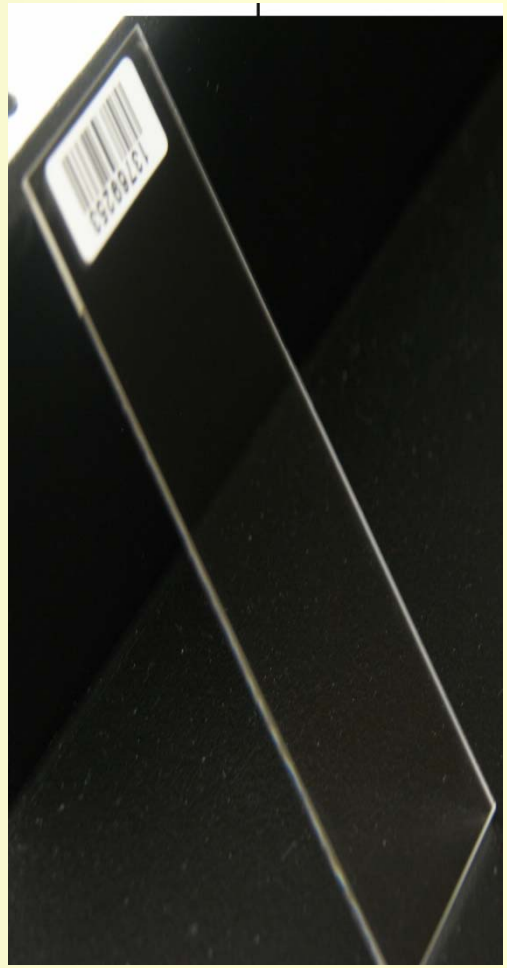


INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

GeoChip or Functional Gene Arrays (FGAs)

- Microarrays: Glass slides or other solid surface containing thousands of genes arrayed by automated equipment.





- FGAs contain probes from the genes involved in various geochemical, ecological and environmental processes.
  - C, N, S, P cycling
  - Organic contaminant degradation
  - Metal resistance and reduction
- Typical format: 50mer oligonucleotide arrays
- Useful for studying microbial communities
  - Functional gene diversity and activity
  - Limited phylogenetic diversity.

Main advantages of GeoChip compared to other approaches (e.g., 16S-based 454 sequencing)

- **Detecting functions: Geochemical processes**
- **Higher resolution: Species-strain level resolution**
- **Quantitative: no PCR is involved**

GeoChip: A high throughput tool for linking community structure to functions

The ISME Journal (2007) 1, 67–77

www.nature.com/ismej

ORIGINAL ARTICLE

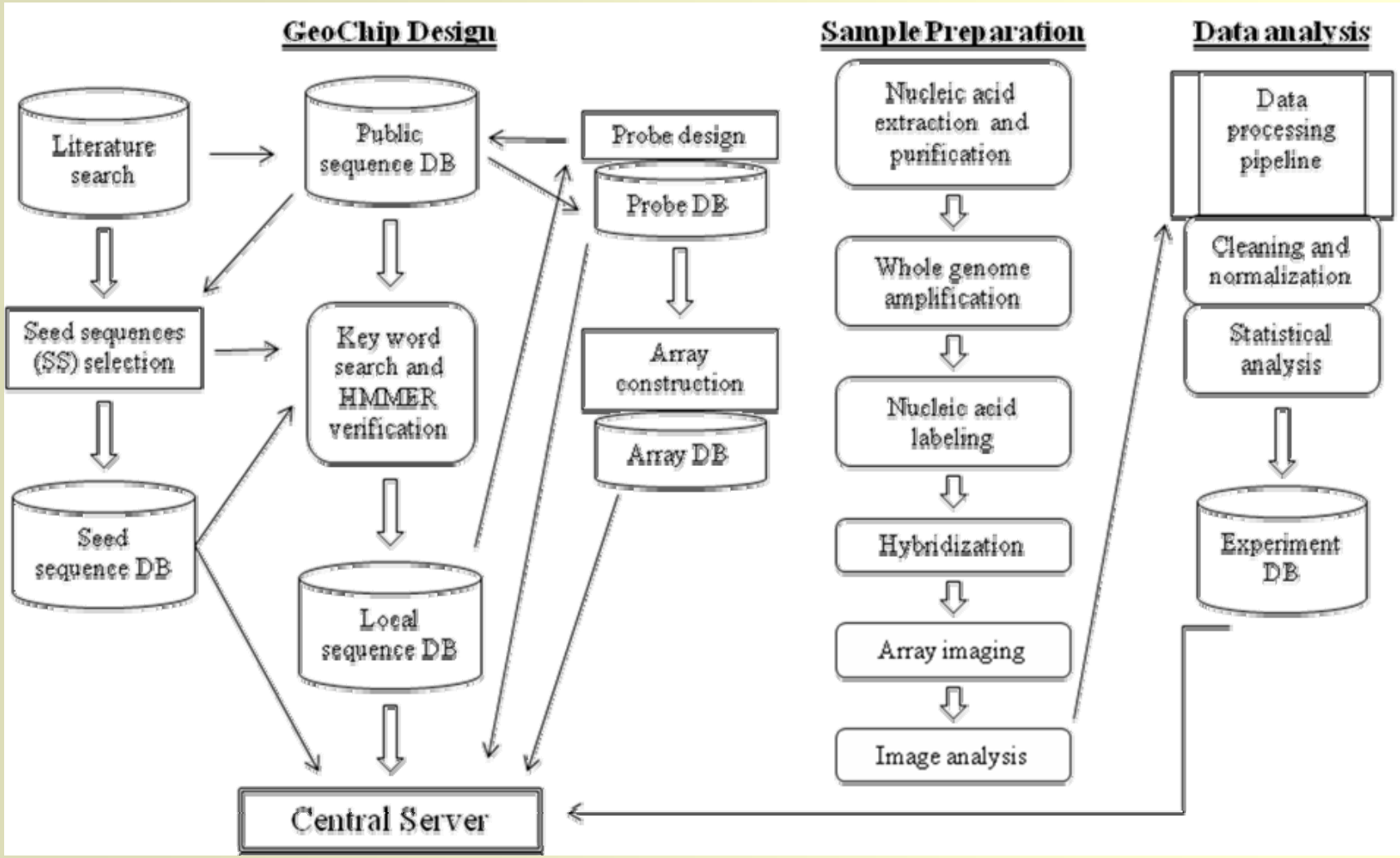
GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes

Zhili He[1,2], Terry J Gentry[2,3], Christopher W Schadt[2], Liyou Wu[1,2], Jost Liebich[2,5], Song C Chong[2], Zhijian Huang[2,6], Weimin Wu[4], Baohua Gu[2], Phil Jardine[2], Craig Criddle[4] and Jizhong Zhou[1,2]

Highlighted by:

- A press release by Nature Press Office
- Reported by many Newspapers
- National Ecology Observatory Networks (NEON), Roadmap
- National Academy of Sciences, Metagenomics report
- R&D 100, among most outstanding 100 technological innovations and breakthrough in 2009

Overview of GeoChip 3.0 development and analysis



Summary of GeoChip 3.0 probe and sequence information by functional gene category

Functional process	No. of gene categories	No. sequences retrieved	No. of probes designed	No. CDS covered
Antibiotic resistance	11	7571	1710	**2904**
Carbon degradation	31	9839	2720	**4737**
Carbon fixation	5	3378	898	**1806**
Methane metabolism	3	4182	254	**434**
Nitrogen cycling	13	27162	3561	**6892**
Phosphorus utilization	3	1441	599	**1212**
Sulfur cycling	3	4296	1328	**1773**
Metal remediation	41	16825	4917	**10458**
Contaminant degradation	190	31236	8815	**16948**
Energy process	2	901	413	**449**
Others (e.g., GyrB)	3	9359	1860	**3897**
Total	**305**	**116,190**	**27,075**	**51,510**

- **> 300 functional gene categories**
- **Universal standards to allow data comparison across different experiments & times**

Carbon degradation

Gene/category	Unique probe	Group probe	Total probe	Total covered CDS
Carbon degradation				
acetylglucosaminidase	32	75	107	214
amyA	61	170	231	467
amyX	0	5	5	12
apu	4	2	6	8
ara	21	65	86	174
ara_fungi	23	10	33	50
cda	11	6	17	25
cellobiase	36	41	77	145
endochitinase	199	168	367	606
endoglucanase	64	24	88	109
exochitinase	15	16	31	63
exoglucanase	54	9	63	83
glucoamylase	23	35	58	111
glx	17	4	21	33
isopullulanase	0	1	1	2
lip	25	4	29	39
mannanase	20	9	29	45
mnp	17	2	19	22
nplT	4	16	20	39
pectinase	27	2	29	33
phenol_oxidase	126	81	207	272
pulA	21	88	109	231
xylA	18	72	90	188
xylanase	60	67	127	221
Subtotal	**878**	**972**	**1850**	**3192**

Carbon fixation and methane metabolism

Gene/category	Unique probe	Group probe	Total probe	Total covered CDS
Carbon fixation				
aclB	20	13	33	53
CODH	13	63	76	138
FTHFS	68	126	194	323
pcc	8	249	257	585
rubisco	139	146	285	515
Subtotal	**248**	**597**	**845**	**1614**
Methane metabolism				
mcrA	104	106	210	392
mmoX	22	22	44	90
pmoA	85	39	124	270
Subtotal	**211**	**167**	**378**	**752**

Nitrogen cycling

Gene/category	Unique probe	Group probe	Total probe	Total covered CDS
Nitrogen cycling				
amoA	100	95	195	528
gdh	26	19	45	94
hao	2	4	6	18
napA	11	22	33	83
narG	289	160	449	656
nasA	67	86	153	259
nifH	885	333	1218	2467
nirK	255	143	398	1005
nirS	351	155	506	923
norB	55	25	80	102
nosZ	191	119	310	596
ureC	57	218	275	603
Subtotal	**2289**	**1379**	**3668**	**7334**

Phosphorus utilization and sulphur cycling

Gene/category	Unique probe	Group probe	Total probe	Total covered CDS
Phosphorus				
ppk	47	67	114	237
ppx	44	296	340	832
Subtotal	**91**	**363**	**454**	**1069**
Sulphur				
dsrA	595	155	750	954
dsrB	371	131	502	685
sox	47	52	99	161
Subtotal	**1013**	**338**	**1351**	**1800**

Metal reduction and resistance

Gene/category	Total probe	Total covered CDS
Metal reduction and resistance		
Arsenic resistance	396	803
Cadmium resistance	1254	2808
Chromium resistance	543	1292
Mercury resistance/reduction	292	594
Nickel resistance	42	88
Zinc resistance	1044	2197
Other metals and metalloids	1803	4135
Other metal reduction	413	449
Subtotal	**5,787**	**12,366**

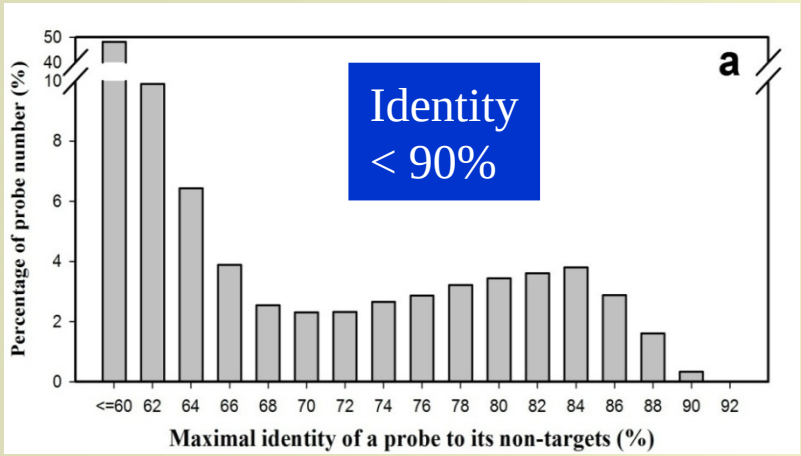
Organic contaminant degradation

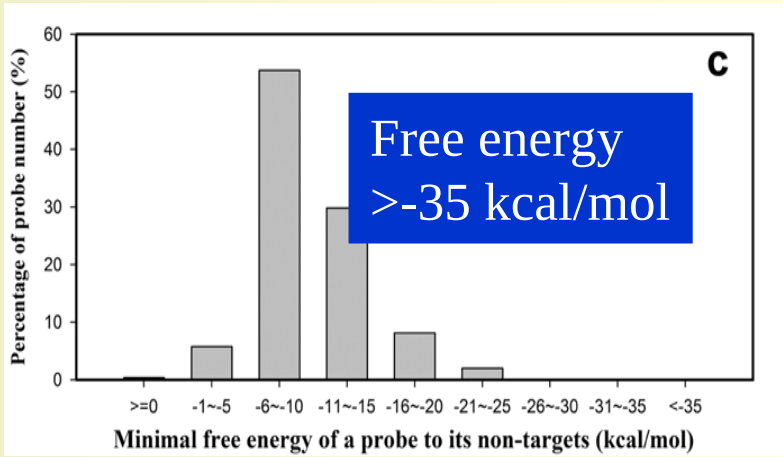
Gene/category	Total probe	Total covered CDS
Contaminant degradation		
BTEX and related aromatics	423	3084
Chloronated aromatics	250	473
Nitroaromatics	122	489
Heterocyclic aromatics	38	66
Hydrocarbons (e.g., PAHs)	179	2089
Chloronated solvents	180	360
Pesticides	1258	3083
Other chemicals and by-products	3936	7855
Subtotal	**6,386**	**17,499**

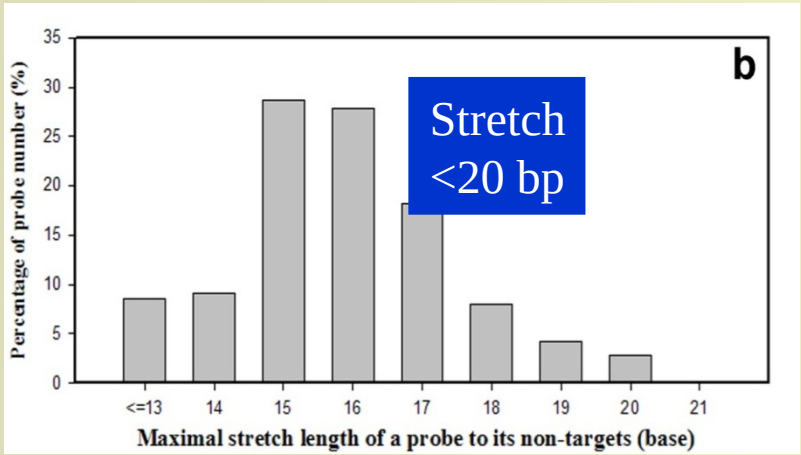
Energy-related metabolism processes

Gene/category	Total probe	Total covered CDS
Energy-related metabolism processes		
Cytochromes	365	365
Hydrogenase	48	85
Subtotal	**413**	**450**

Computational evaluation of GeoChip 3.0





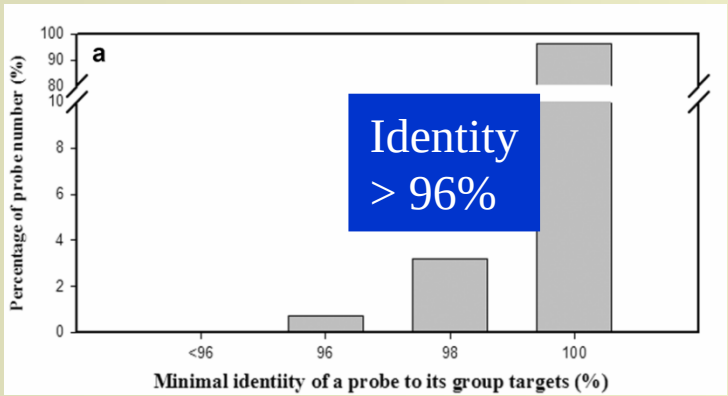


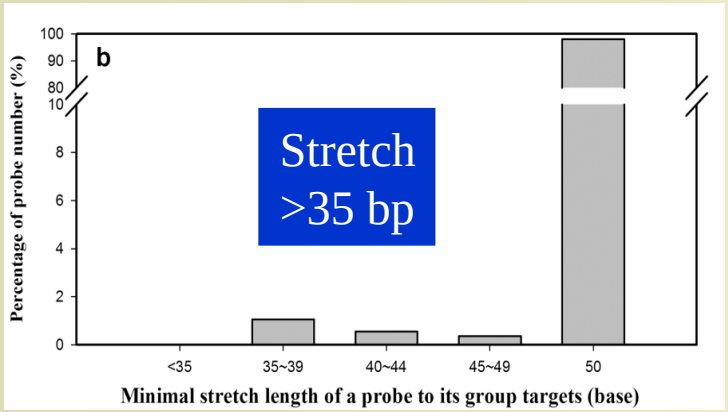
Sequence-specific probes (SSP):

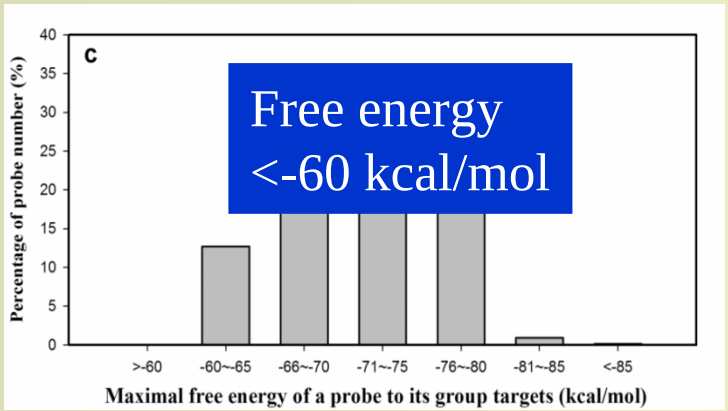
A majority (90-95%) of probes on the array were far away from the thresholds of probe design criteria, indicating that they should be highly specific to their corresponding targets.

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

Computational evaluation of GeoChip 3.0







Group-specific probes (GSP):

- GSP are very important for environmental studies since functional genes are highly homologous.
- GeoChip 3.0 has more GSP (66.7%) and covers more than 47,000 sequences in comparison with GeoChip 2.0 with 17.7% for 3,000 sequences.
- 95% of GSP have 100% sequence homology to their corresponding targets.

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

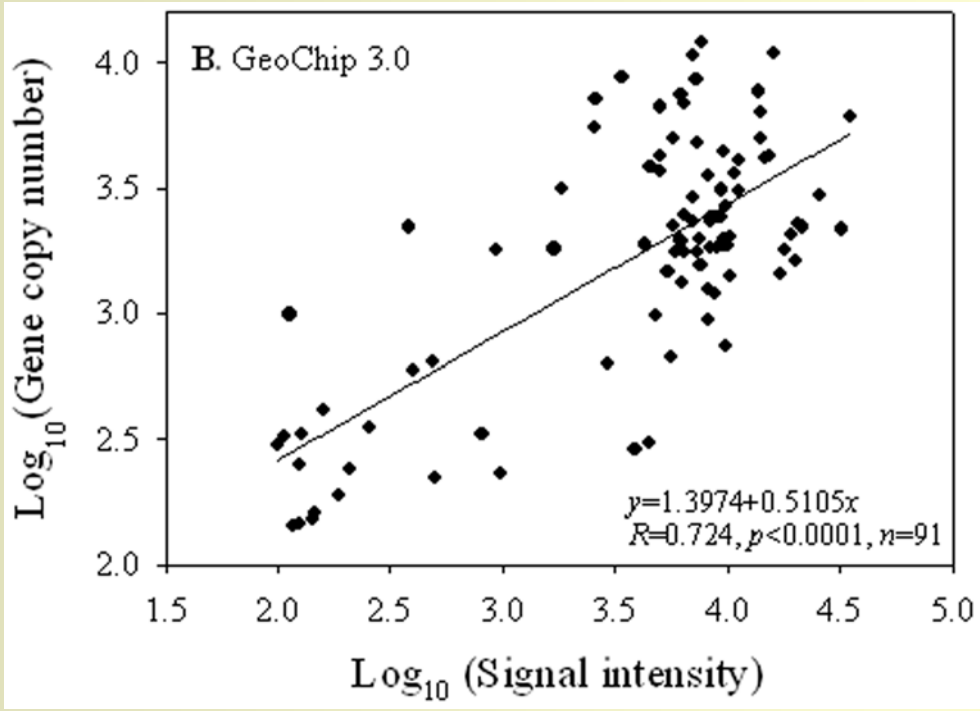
Experimental evaluation of GeoChip specificity

Summary of GeoChip 3.0 hybridization with different targets (oligonucleotides or genomic DNA at 45°C and 50% formamide

Targets	Oligonucleotide	genomic DNA
No. of targets	24	2
Expected no. of probes detected	24	44
No. of probes hybridized	25	53
No. of false negatives	0 (0%)	2 (0.0072%)
No. of false positives	1 (0.0036%)	7 (0.025%)
Average signal intensity of targets	6056±4556	11428±7223
Average SNR of targets	13.6±11.9	19.8±9.7
Average signal intensity of false positives	3365±960	3687±2191
Average SNR of false positives	4.3±1.5	6.7±4.3

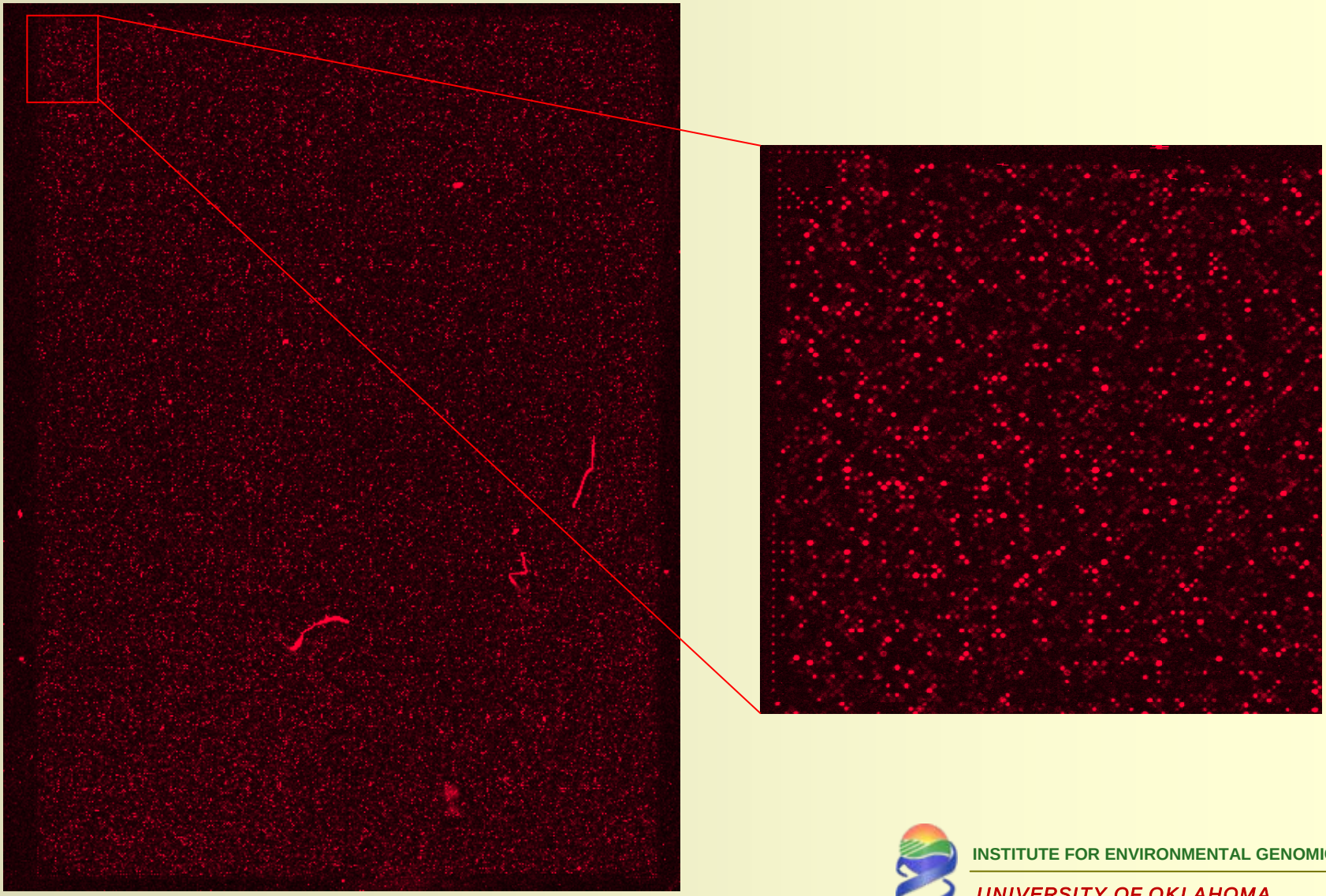
Only very low percentages of false positives (0.0036 ~ ~0.025%) were observed when synthesized oligos and *Shewanella* genomic DNA were used as targets.

GeoChip 3.0 data are quantitative



Real-time PCR showed that GeoChip 3.0 data were quantitative.

GeoChip 4.0—Nimblegen layout



Examples of most recent applications

- Groundwaters
  - Monitoring bioremediation processes: Ur, Cr
  - Impacts of contaminants on microbial communities
- Soils
  - Grass land soils: effects of plant diversity and climate change on soil communities
  - Forest soil: spatial scaling
  - Agricultural soils: tillage, no tillage
  - Oil-contaminated soils
- Aquatic environments
  - Hydrothermal vents
  - Marine sediments
  - River sediments
- Bioreactors
  - Wastewater treatments
  - Biohydrogen
  - Microbial fuel cell
- Bioleaching

Issues related to specificity, sensitivity and quantitation

- **Specificity, sensitivity, quantitation**
  - Wu et al. 2001; AEM:67: 5780-5790
  - Rhee et al. 2004, AEM 70:4303-4317
  - Tiquia et al. 2004. BioTechniques 36, 664-675
  - Wu et al. 2004; EST, 38: 6775-6782
  - He et al, 2007; The ISME J, 1: 67-77
  - He and Zhou, 2008, AEM, 74: 2957–2966
- **Probe design criteria**
  - He et al. 2005. AEM. 71:3753-3760
  - Liebich et al. AEM, 72:1688-1691
  - **Deng et al. 2009, BMC Genomics**
- **New probe designing software: CommOligo**
  - Li et al. 2005. Nucl. Acids Res. 33:6114-6123
- **Whole community genome amplification (WCGA)**
  - Wu et al. 2006. AEM: 72:4931-4941.
- **Whole community RNA amplification (WCRA)**
  - Gao et al, 2007, AEM: 73: 563-571.
- **Review:**
  - Gentry et al. 2006, Microbial Ecology, 52: 159-175.
  - Zhou and Thompson, 2002, Curr Opion Biotech: 13:204-207
  - Zhou, 2003; Curr Opion. Microbiol, 6:288-294

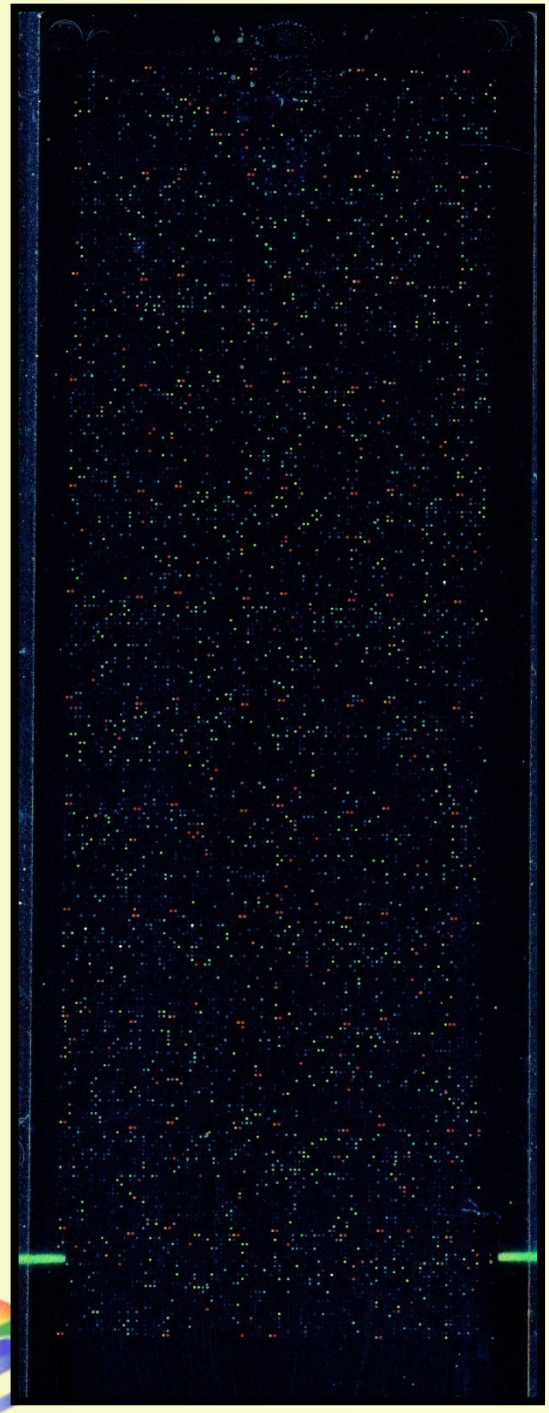
- **Applications**
  - He et al, 2007; The ISME J, 1: 67-77 , Leigh et al, 2007, The ISME J, 1: 163-179
  - Yergeau et al, 2007, The ISME J, 1: 134-148.
  - Zhang et al. 2007. FEMS Microbiology Letters 266: 144-151.
  - Wu et a., 2008, AEM, **74: 4516-4529**
  - **Zhou et al. 2008. PNAS, 105: 7768-7773**
  - **Wang et al. 2009. PNAS, 106: 4840-4845**
  - Liang et al, 2009. Chemosphere, 75: 193-199
  - Liang et al. 2009, Chemosphere, 3: 231-242
  - Mason et al. 2009. ISME J., 3: 231-242
  - Waldron et al. 2009. EST, 43: 3529-3534
  - Van Nostrand et al. 2009, EM, 11:2611-2626
  - He et al. Ecol Letter
  - Xu et al., ISME J, in press

INSTITUTE FOR ENVIRONMENTAL GENOMICS

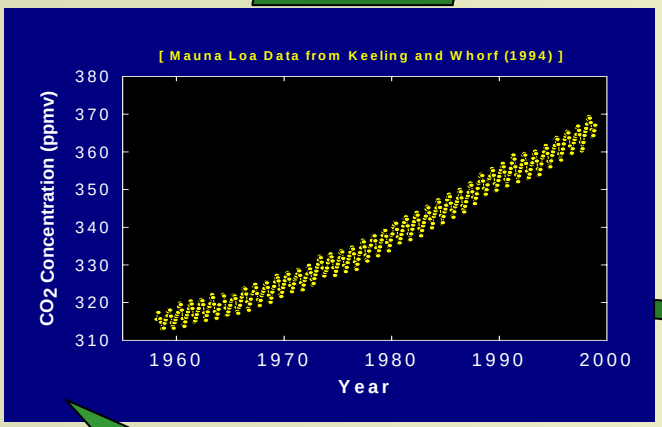
UNIVERSITY OF OKLAHOMA

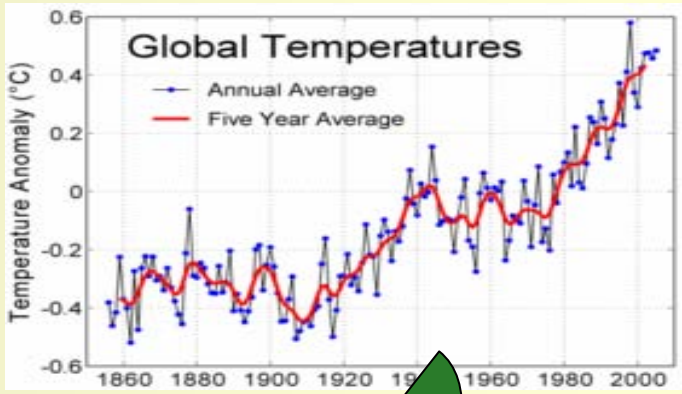
Overview of Microarray Analysis

- **DNA extraction from environmental samples, multiple samples, times**
- **Whole Community Rolling Circle Amplification (1-100ng DNA)**
- **Label DNA with Cy5**
- **Hybridization to GeoChip at 42, 45 or 50C with 50% formamide**
- **Data processing with automatic pipeline**
- **Statistical analysis**
- **Data interpretation**



Grand Questions: Positive or Negative Feedbacks?





+

+

+

-

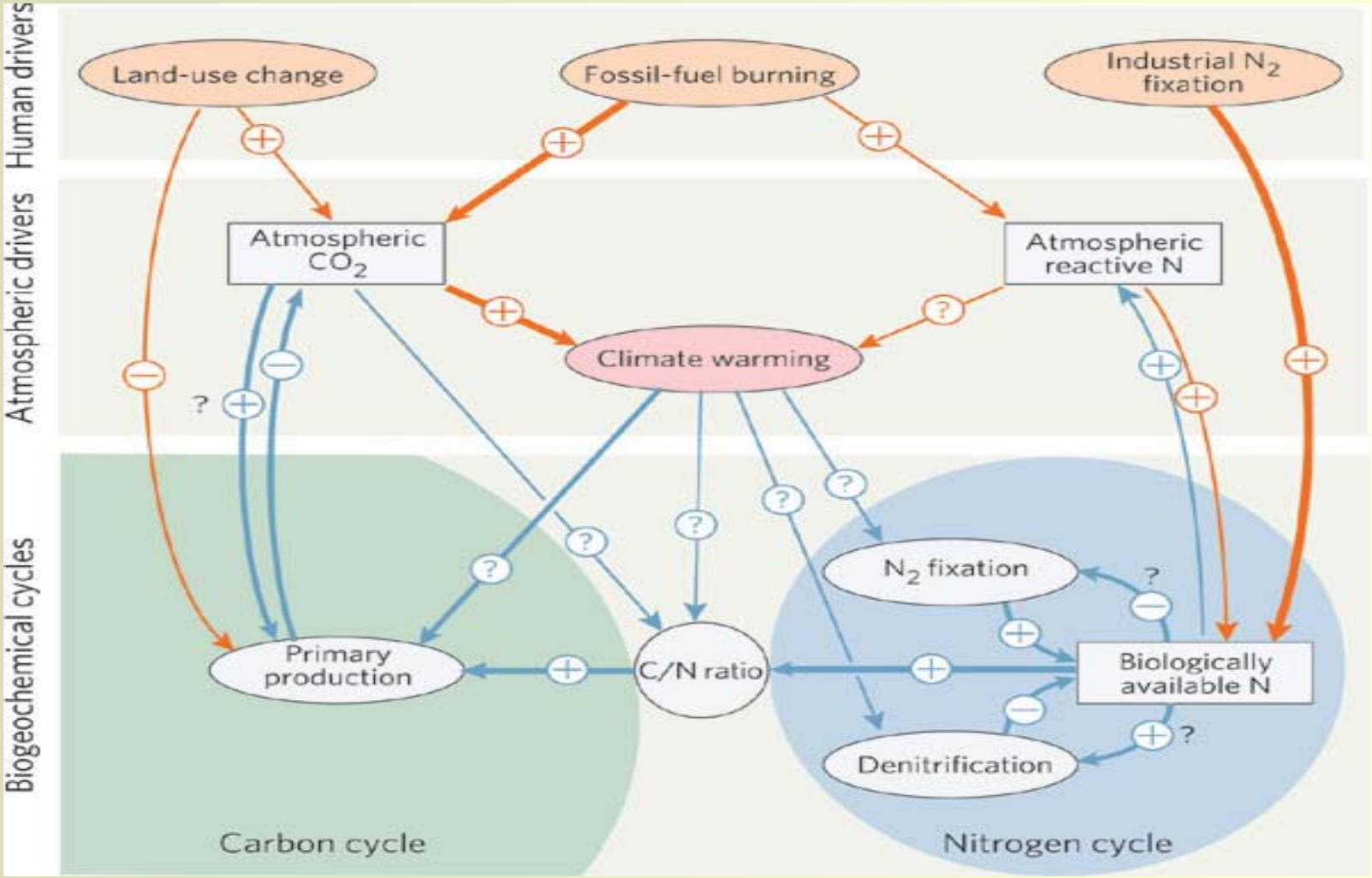
+

INSTITUTE FOR ENVIRONMENTAL GENOMICS

UNIVERSITY OF OKLAHOMA

Nitrogen–Carbon–Climate Interactions

From Cruber et al, 2008. Nature 451, 293-296



It is not clear how rising CO_2 and temperature will affect various C and N cycling processes.

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

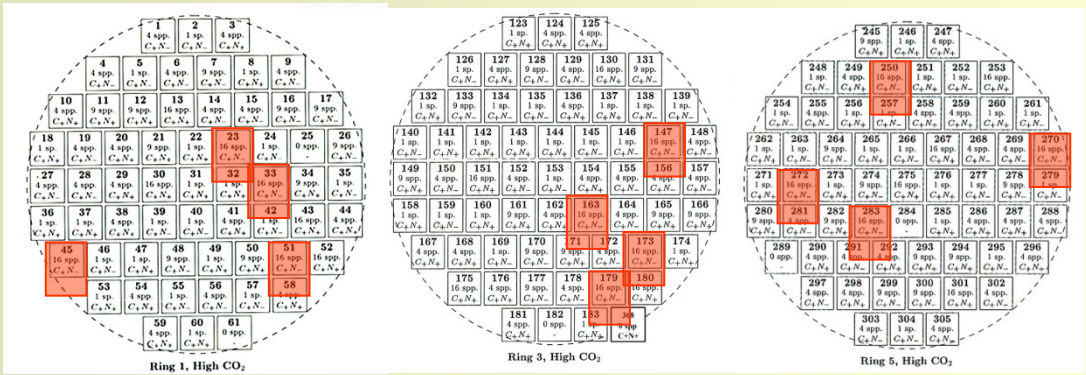
Response of grass ecosystems to elevated CO_2

BioCON (Biodiversity, CO_2, and Nitrogen) at the University of Minnesota by Peter Reich

Picture 17. Aerial photo of an elevated CO_2 ring in 2005

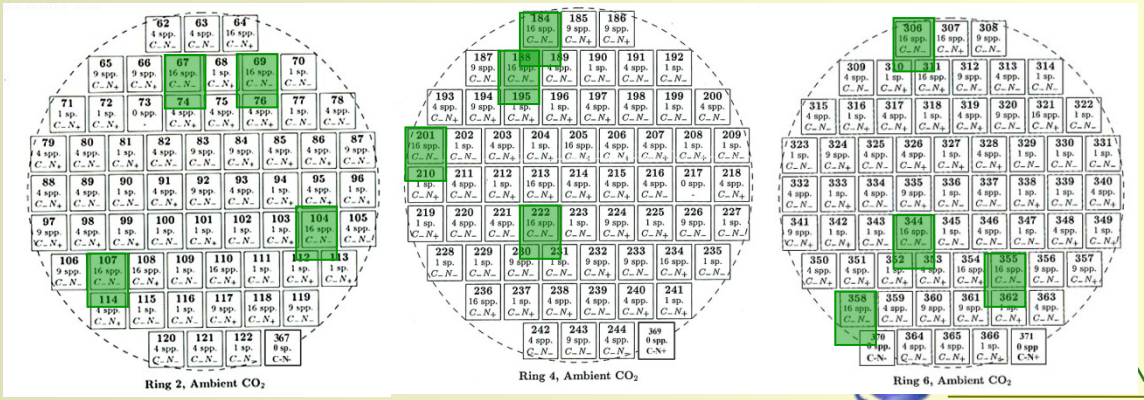
Each ring: 20m diameter

Elevated CO_2



There are 296 plots in total.

Ambient CO_2



Each plot: 2×2m

Summary of the experiments

- **Response of microbial communities to elevated CO_2,**
  - 24 samples, 16 species plots, with or without CO2
  - GeoChip
  - 454 sequencing
  - PhyloChip
  - PLFA, BioLog
  - Network analysis
- **Effects of Plant diversity on microbial communities**
  - 31 samples, 1, 4, 9, 16 species
  - GeoChip
  - 454 sequencing
- **Effects of plant functional groups on microbial communities**
  - GeoChip
  - BioLog
- **Microbial community temporal dynamics**
  - 48 samples, 6 rings, 16 species plots, each ring has a composite sample, 8 time points
  - GeoChip
  - BioLog
- N and C fixation
- Interactive effects
  - 296 samples
  - GeoChip
  - Sequencing

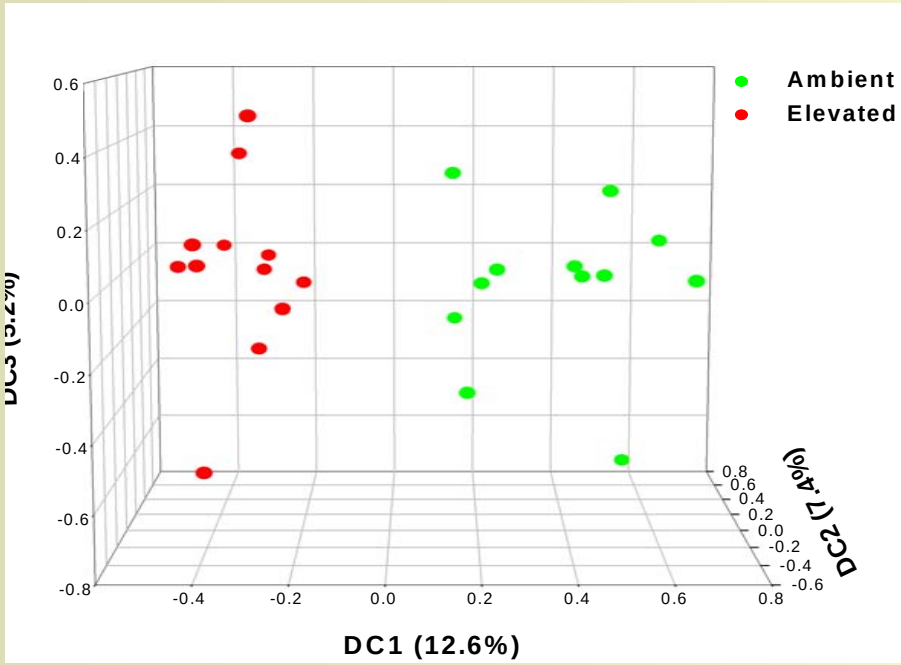
Microbial diversity and gene number detected by GeoChip and pyrosequencing

Sample	GeoChip		454 pyrosequencing	
	Elevated CO_2	Ambient CO_2	Elevated CO_2	ambient CO_2
1/D	1778.23±179. 17	1811.70±271.64	425.55±133.78	423.74±155.41
H'	11.79±0.08	11. 65±0.14	9.63±0.25	9.67±0.21
Evenness	0.31±0.02	0.29±0.03	0.81±0.02	0.81±0.02
Total gene detected	2850±410	2541±441	2501±553	2524±368

No significant difference in terms of gene number/OTU and diversity were observed.

Effects of elevated CO_2 on soil microbial communities

DCA analysis of GeoChip data

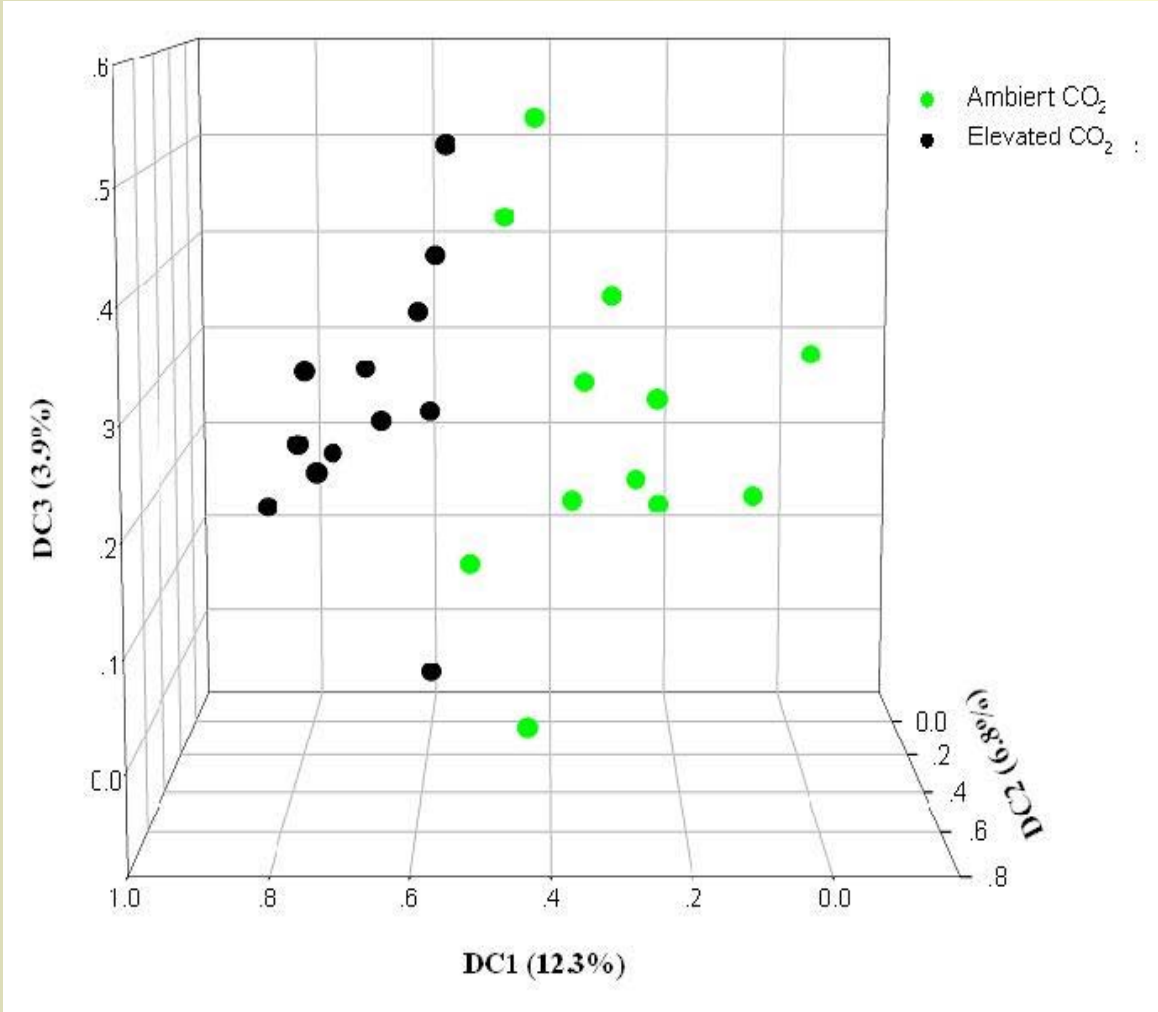


- 24 soil samples: 12 from ambient CO_2 (368 µmol/mol, **green**) rings, and 12 from high CO_2 (560 µmol/mol, **red**) rings.
- 16-species plots
- Ambient nitrogen

5038 functional genes were detected with at least 3 samples out of 12 for ambient CO_2 and elevated CO_2 samples, which were well separated by DC1, suggesting a significant difference between both communities.

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

DCA analysis of 454 sequence data at the genus level



- The communities under CO_2 are somewhat separated better from those under no CO_2 than PhyloChip, but worse than GeoChip
- These are due to resolution
- GeoChip: species-strain level
- 454 sequencing: genus level.
- PhyloChip: family and subfamily

At least 3-sample shared genus were used.

The effects of CO_2 regimes measured by different analytical methods

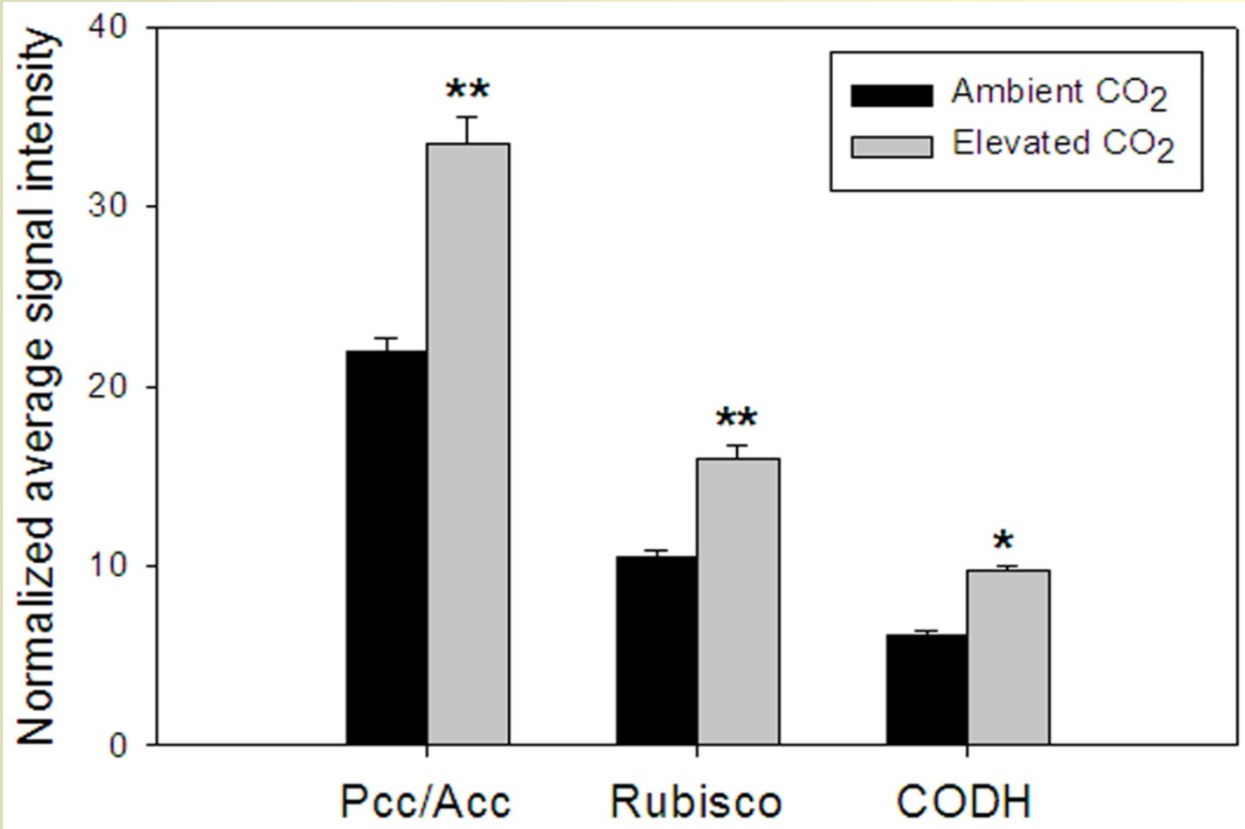
		GeoChip		454 sequencing		PLFA	BiOLOG
		2005	2007	Genus (0.95)	Species (0.97)		
N		1212	5038	15847	23184	34	93
ANOSIM	*R*	0.514	0.141	0.081	0.148	0.209	0.014
	P	<0.001	0.023	0.072	0.019	0.003	0.315
adonis*	*F*	7.132	1.753	1.312	1.537	6.712	0.911
	P	<0.001	0.028	0.017	0.002	0.009	0.593
mrpp**	*δ*	27.1	0.507	0.617	0.602	0.223	0.268
	P	<0.001	0.030	0.022	0.003	0.009	0.356

- The entire communities under elevated CO_2 are significantly different from those under no CO_2 based on GeoChip, 454 sequencing and PFLA, but not the data from EcoPlates.

* non-parametric MANOVA

** multi response permutation procedure

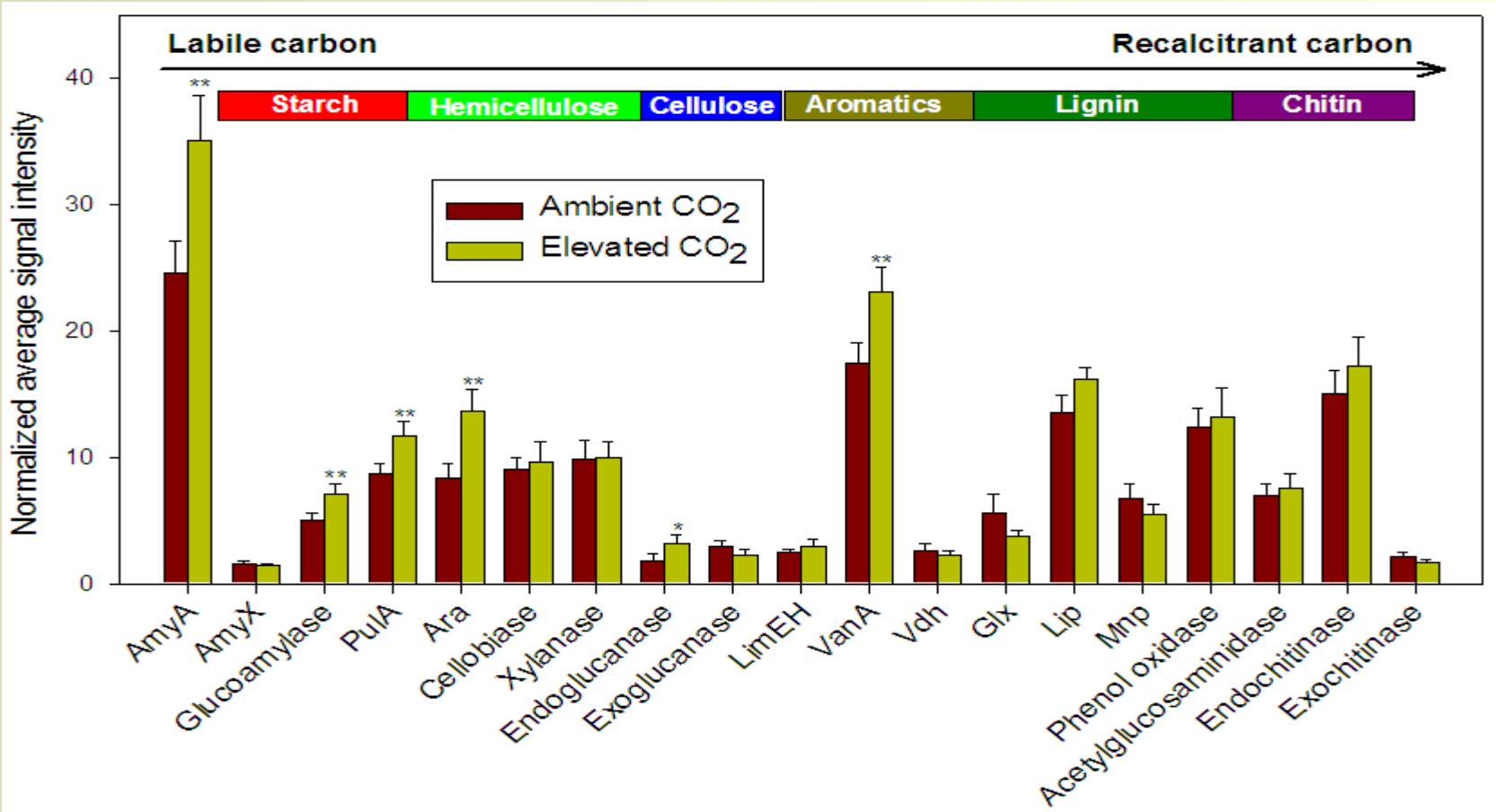
Abundance of detected genes involved in carbon fixation



- Three pathways involved in C fixation increased significantly
- Indicating potential negative feedback

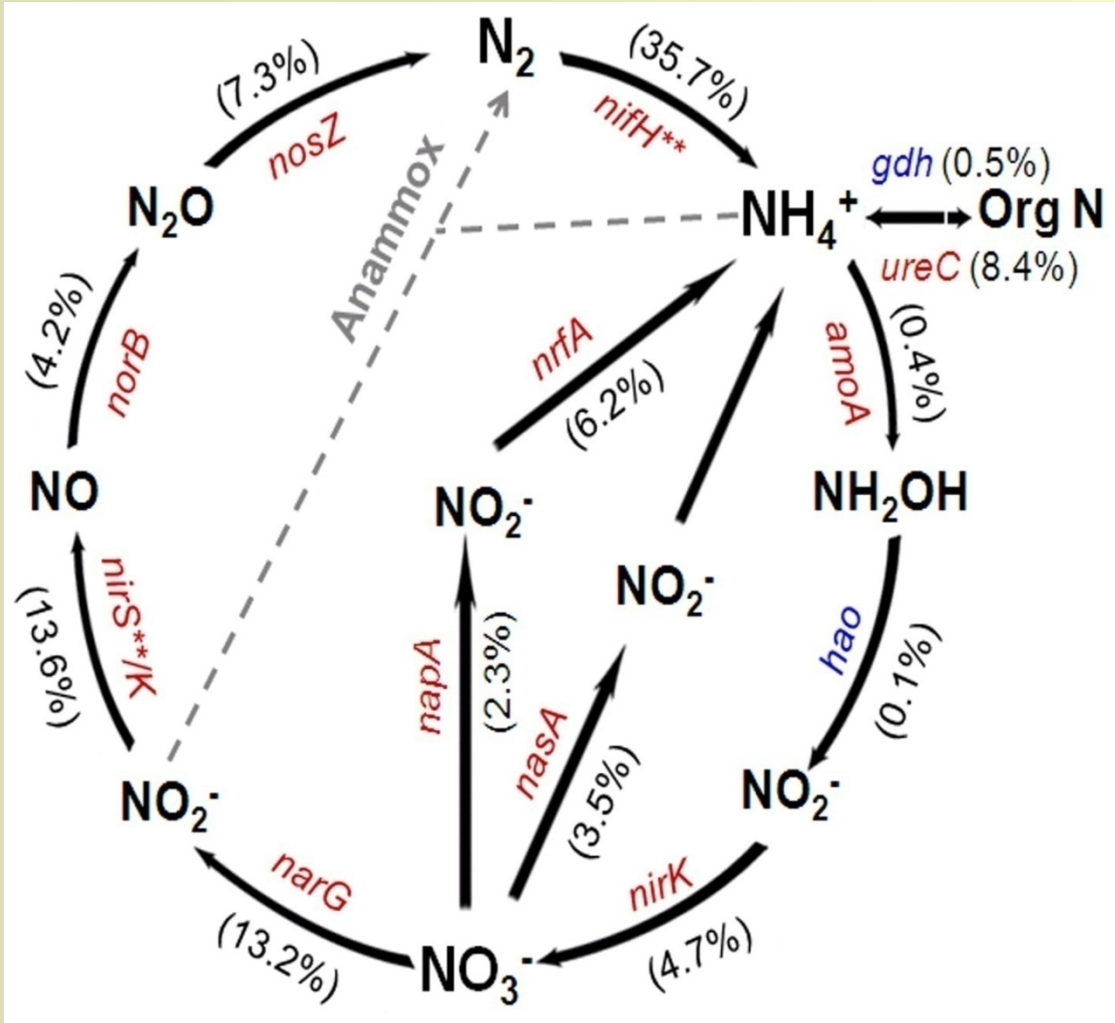
INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA

Abundance changes of key genes involved in carbon degradation



- Genes involved in labile C degradation increased, but no changes for recalcitrant C
- Indicating potential negative feedback

Relative signal intensity of detected genes involved in the N cycle



- NifH genes increased significantly at eCO_2
- nirS gene increased significantly
- No changes for other genes involved in N cycling.
- Indicating potential negative feedback to atmospheric CO_2

N fixation

Each soil sample (30g, ODW) was incubated for 41 days with two kinds of synthetic air:

- **20% O_2 + 80% $^{15}N_2$ (99 atom %):**

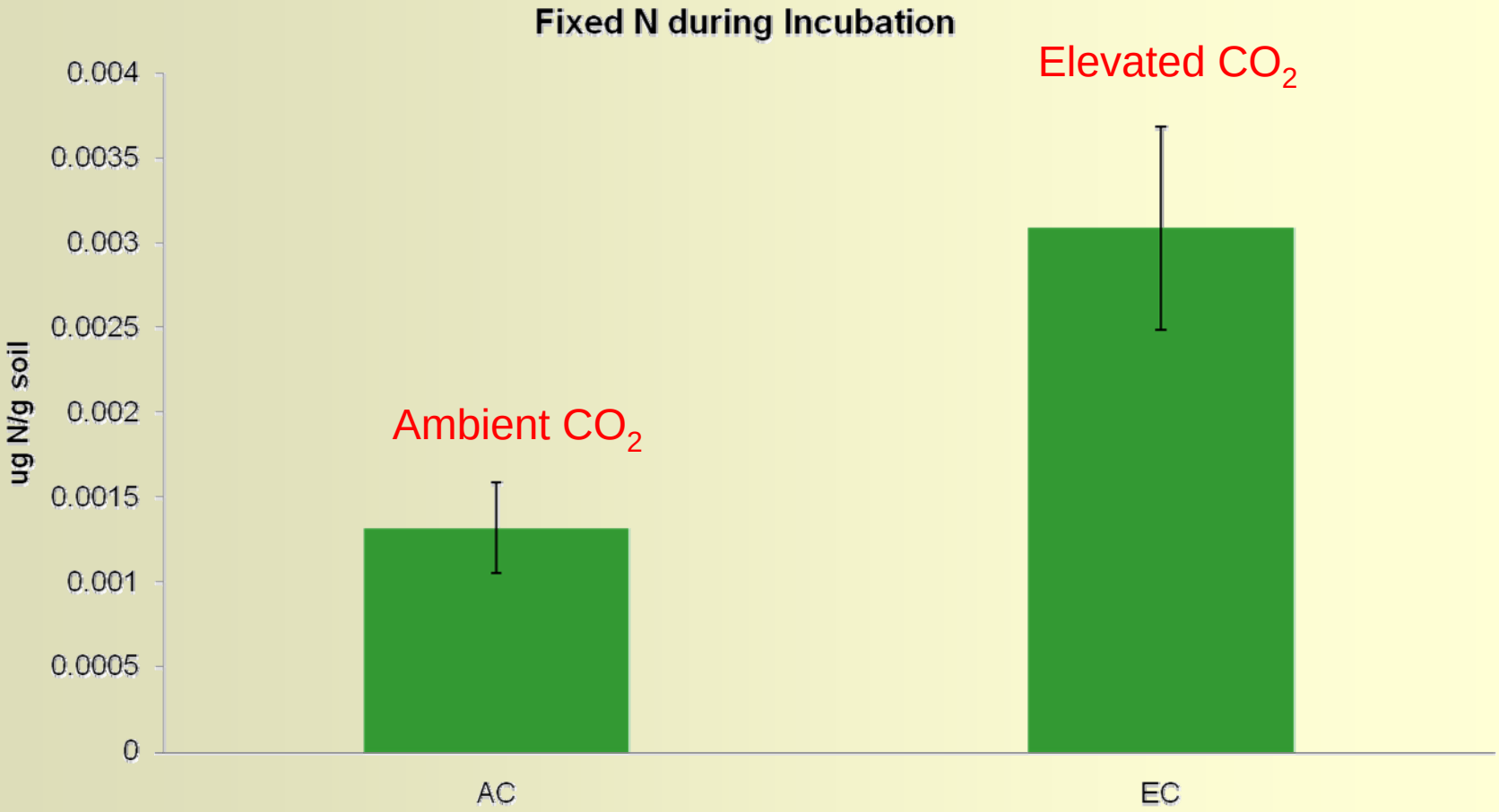
  the soil ^{15}N content increased due to N fixation, nitrification and denitrification

- **20% O_2 + 80% $^{14}N_2$. as a control**

  the soil ^{15}N content increased due to nitrification and denitrification

- **Each sample: the difference of soil ^{15}N content between the above two incubations--N fixation.**

N fixation by isotope



T-test: p-value=0.0208
Error bar: standard error

Ecological networks

- Conventional methods: Shannon diversity indices
  - Species number
  - Abundance of each species
  - **Ignore interactions among different species**
- Network methods
  - Interactions
- Questions
  - Can random matrix theory be used to describe ecological networks?
  - Does a microbial community show a general network behaviors?
  - Whether and how does CO2 affect ecological network?

Novel approach for network identification by Random Matrix Theory

Yang et al. 2008. *BMC Genomics* 2008, **9**:S11
Luo et al, 2007, BMC Bioinformatics, 8:299
Luo et. 2006, Physics Letters A: 357: 420-423.
Luo et al. 2005. Physical Review E, 73, 1-5

Random Matrix Theory and Level Statistics

Poisson Distribution:

$$P(s) = \exp(-s)$$

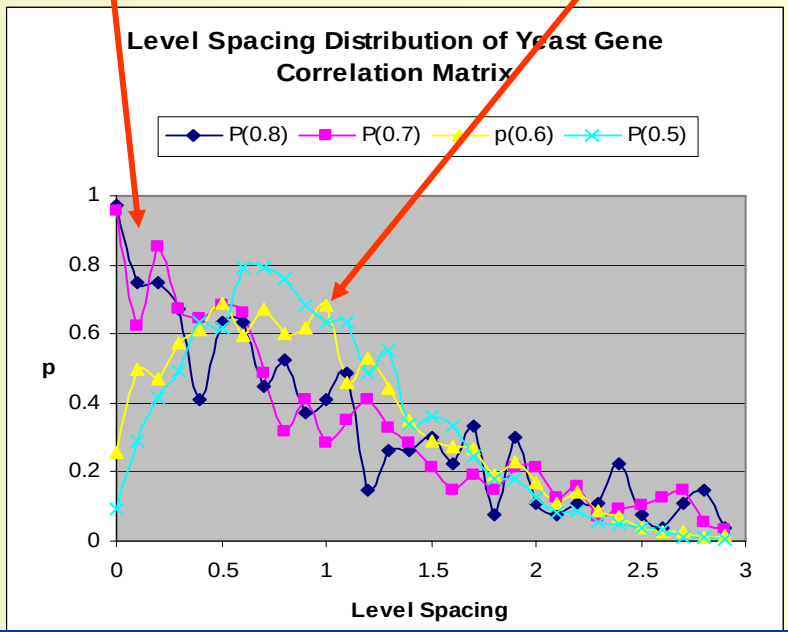
Wigner-Dyson Distribution:

$$P(s) = \frac{\pi}{2} s \exp\left(-\frac{\pi s^2}{4}\right)$$

- Random properties: Wigner-Dyson distribution
- Nonrandom properties: Poisson distribution

Poisson Distribution (cutoff >0.7)

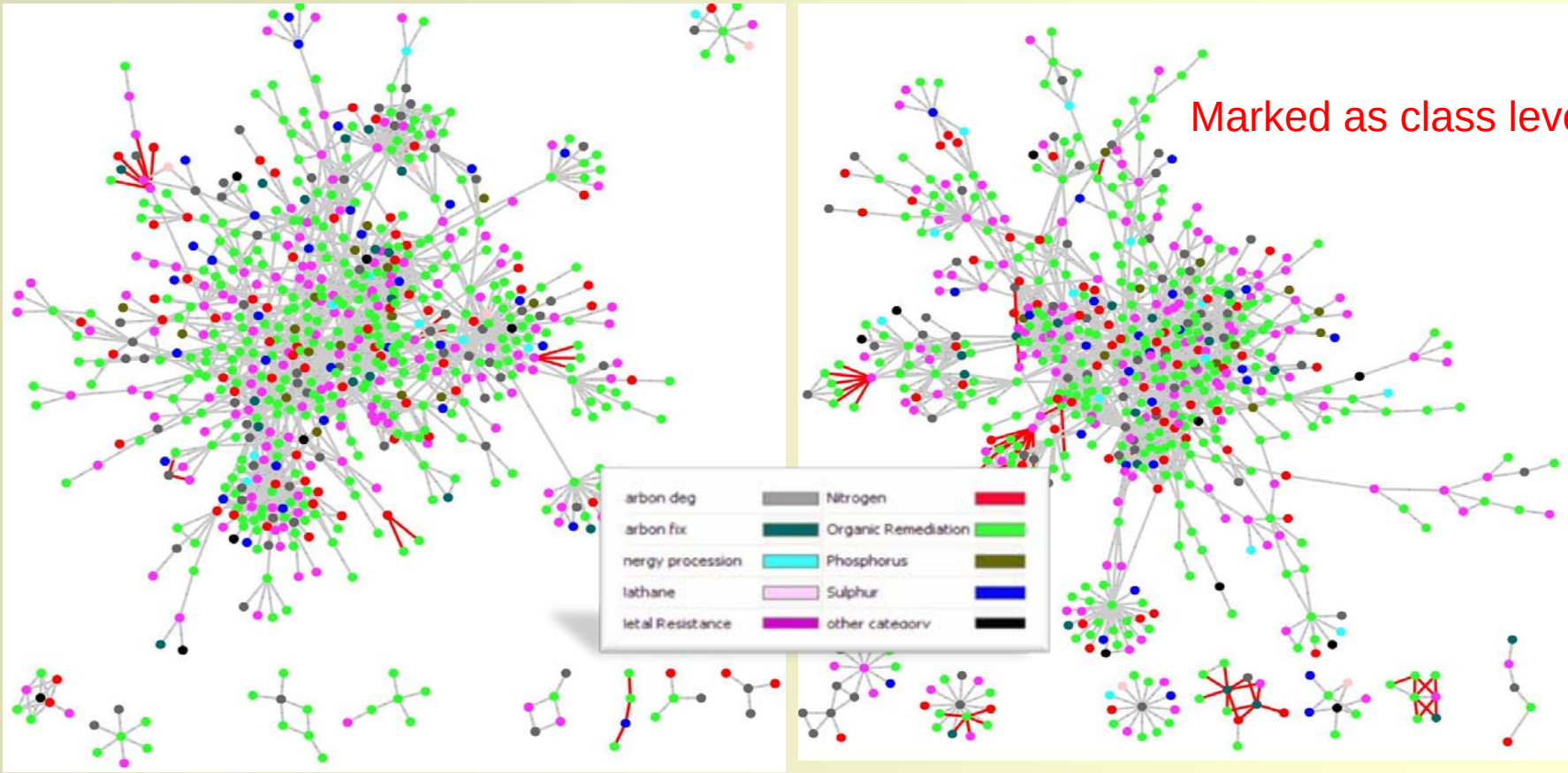
Wigner-Dyson Distribution (cutoff < 0.7)



Main advantages:

- Universal laws support
- Automatic cutoff
- Reliable, sensitive, robust

Networks at high and ambient CO_2

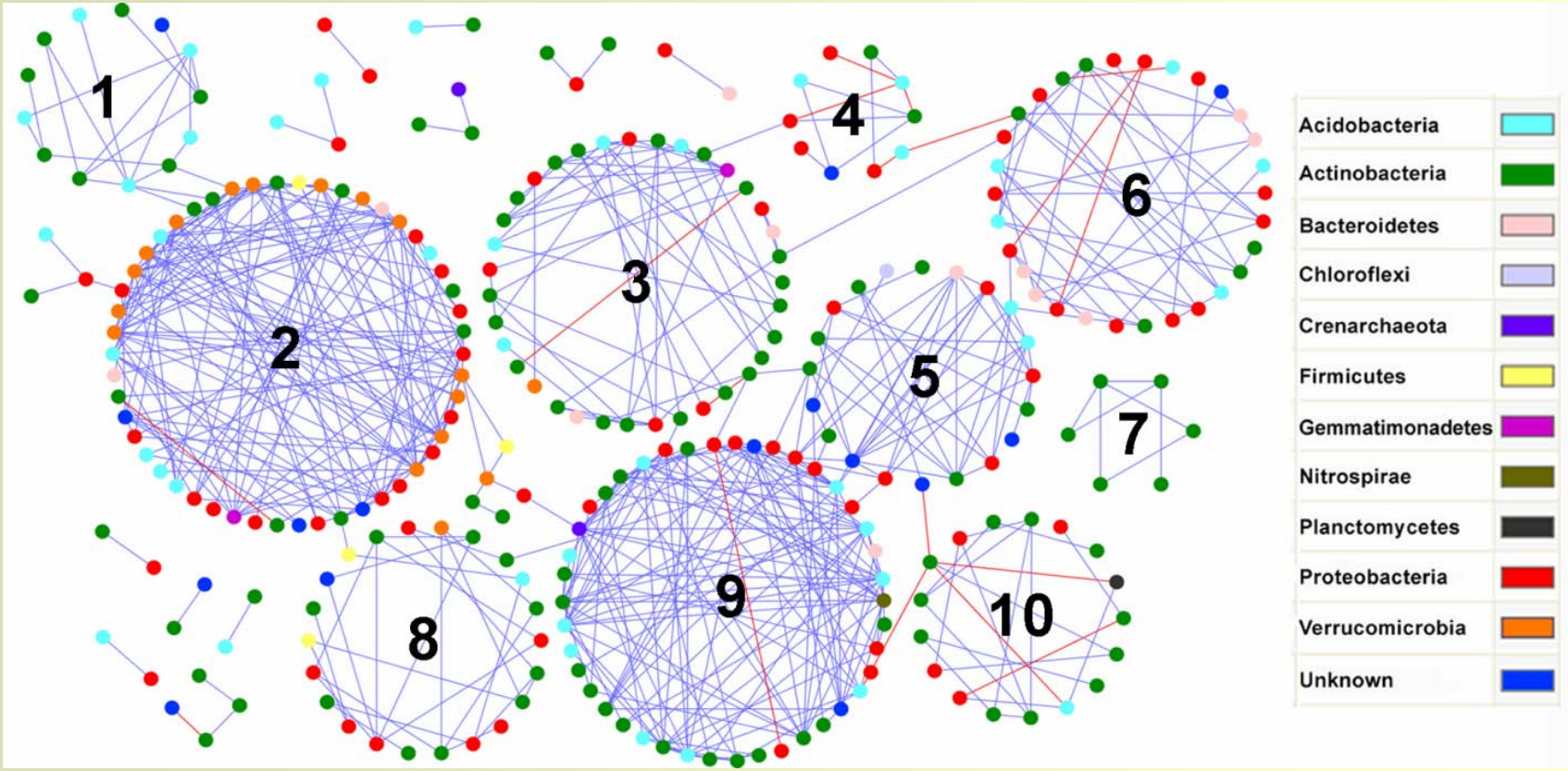


Ambient CO2

High CO2

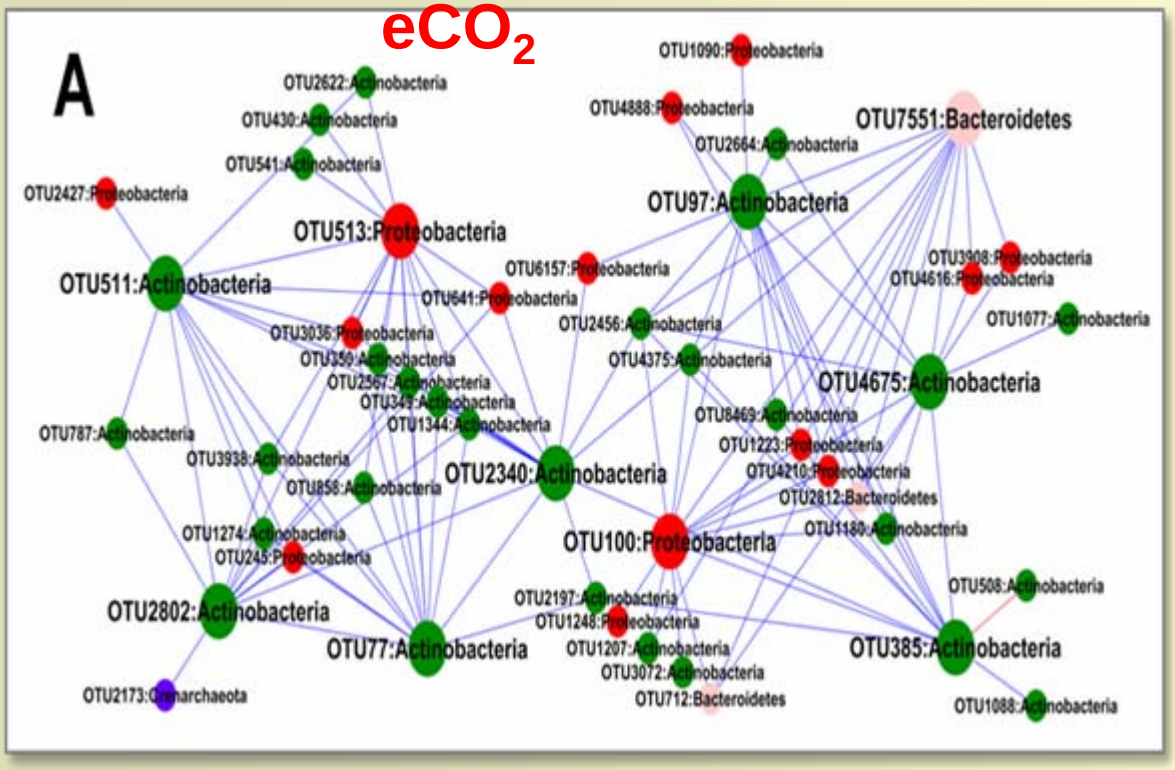
- Based on GeoChip data
- Community interaction network can be identified using RMT method

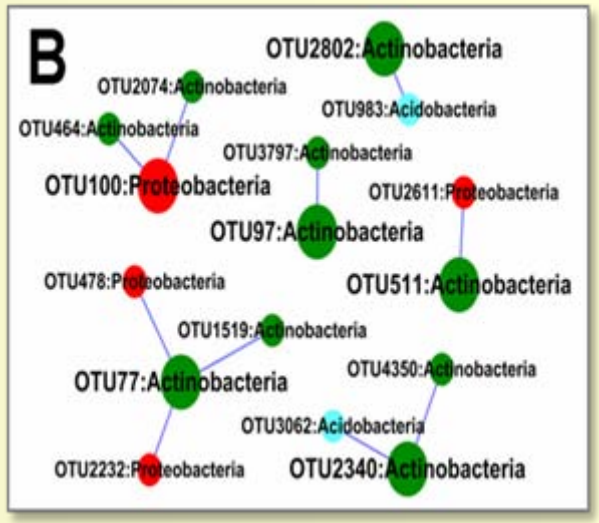
Modularity of Molecular Ecological Networks (MENs)



- Based on 454 sequencing data at ambient CO2.
- All the MENs examined were modular, with distinct modules.
- A module is a group of OTUs/functional genes that are highly correlated among themselves, but have few connections with OTUs/functional genes belonging to other modules.

Effects of elevated CO2 on community phylogenetic structure

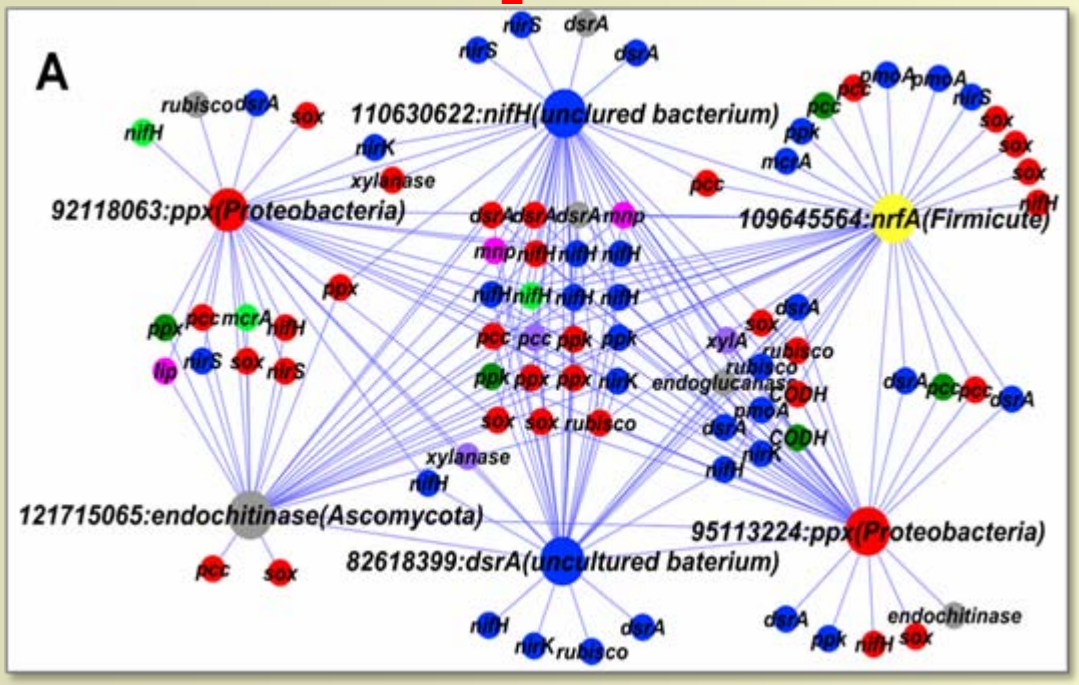




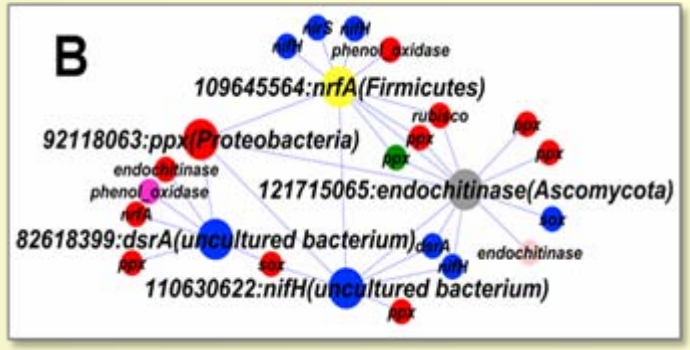
- Top 10 OTUs with the highest connectivity at eCO_2 and their corresponding OTUs at aCO_2.
- More complicated network interactions at eCO_2 than aCO_2

Effects of elevated CO_2 on community functional structure

eCO_2



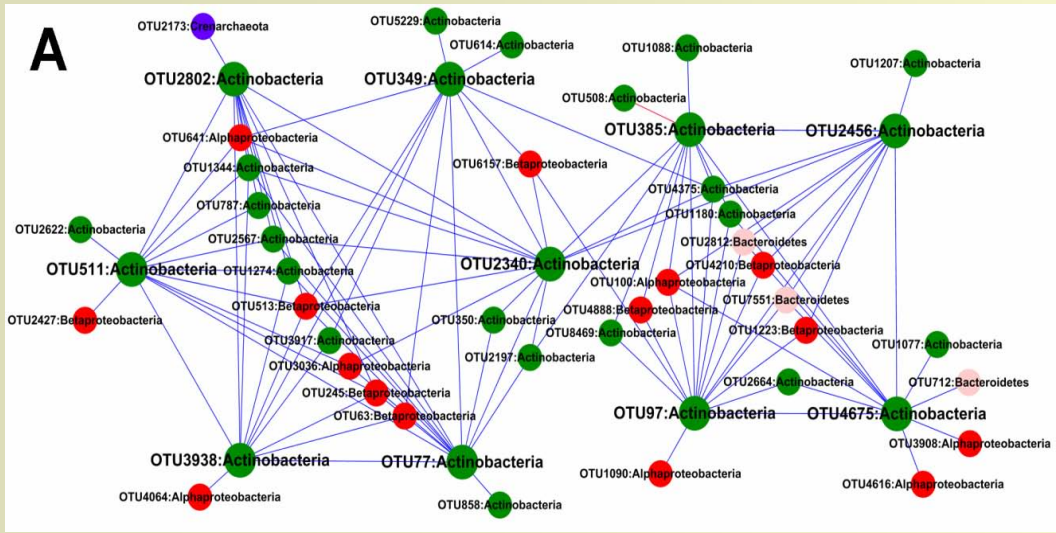
aCO_2



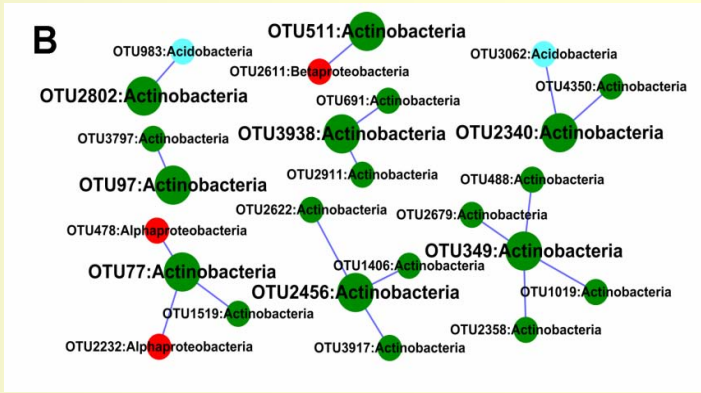
- Top 10 functional genes with the highest connectivity at eCO_2 and their corresponding OTUs at aCO_2.
- More complicated network interactions at eCO_2 than aCO_2

Network interactions of Actinobacteria

eCO_2



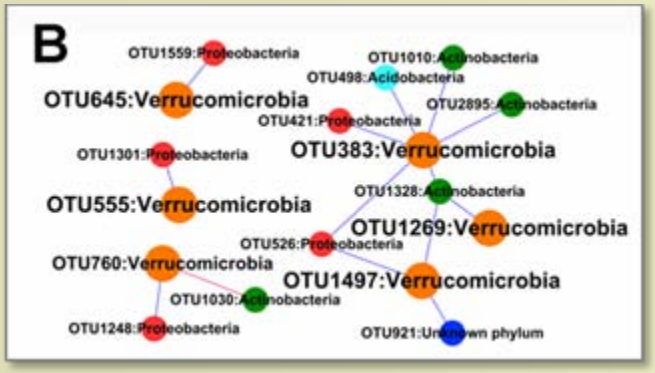
aCO_2



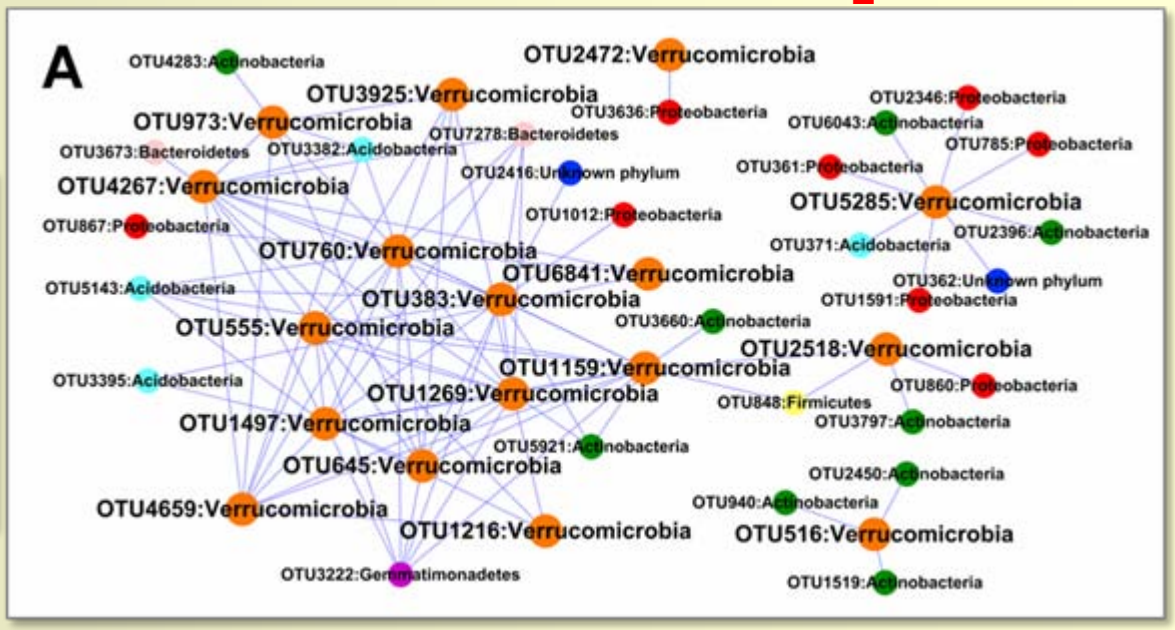
- Top 10 OTUs of Actinobacteria at eCO_2 and their corresponding OTUs at aCO_2
- More complicated network interactions at eCO_2 than aCO_2

Network interactions of Verrucomicrobia

eCO_2



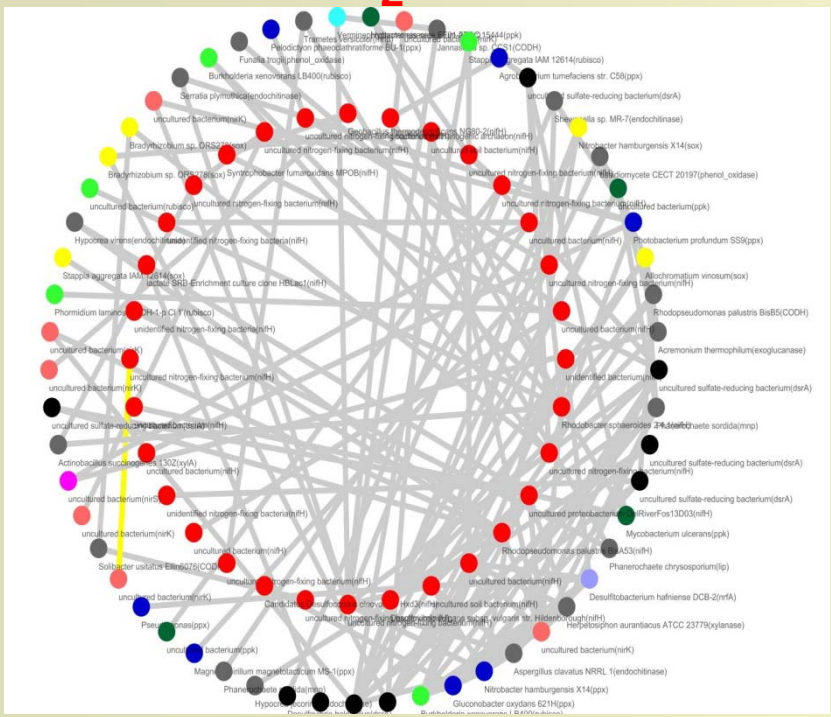
aCO_2



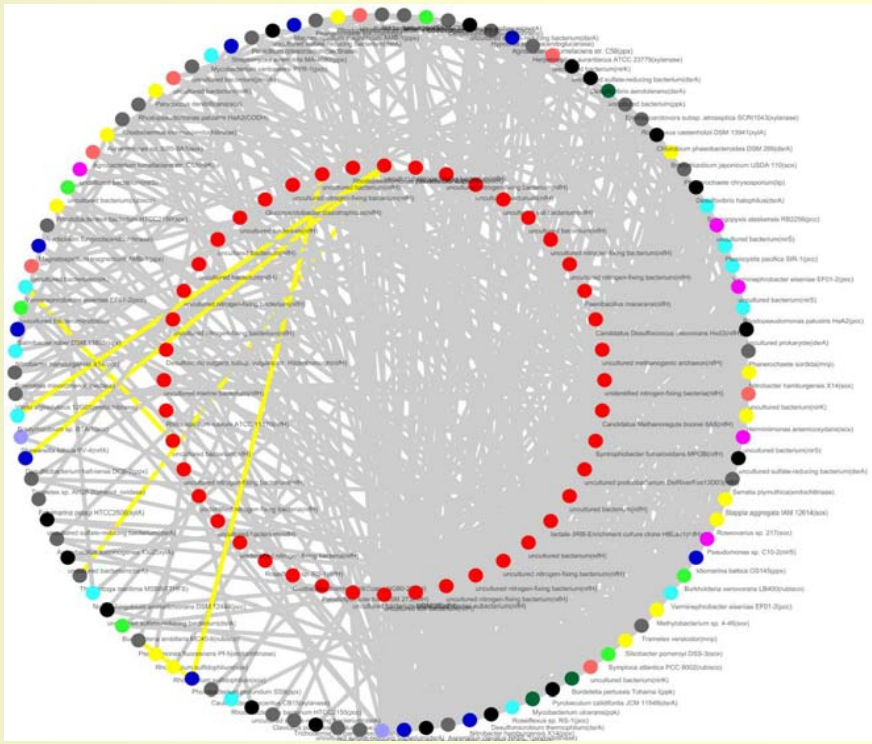
- All OTUs of Verrucomicrobia at eCO_2 and aCO_2
- More complicated network interactions at $aeCO_2$ than eCO_2

Networks is complicated, nifH genes as an example

aCO_2



eCO_2

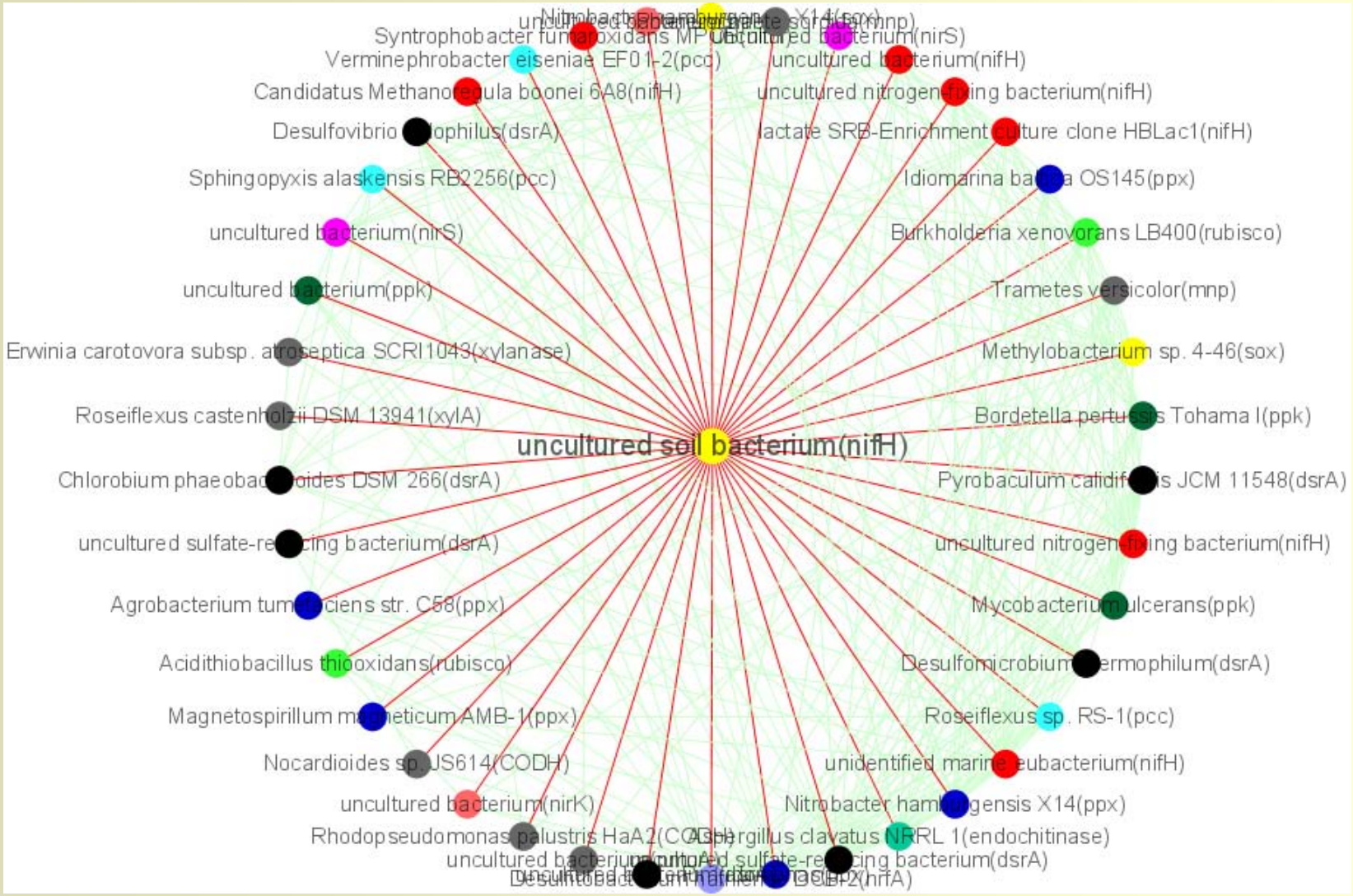


- Elevated CO_2 significantly changed community structure
- The network of nifH genes is much more complex under elevated CO_2 than ambient CO_2.

INSTITUTE FOR ENVIRONMENTAL GENOMICS

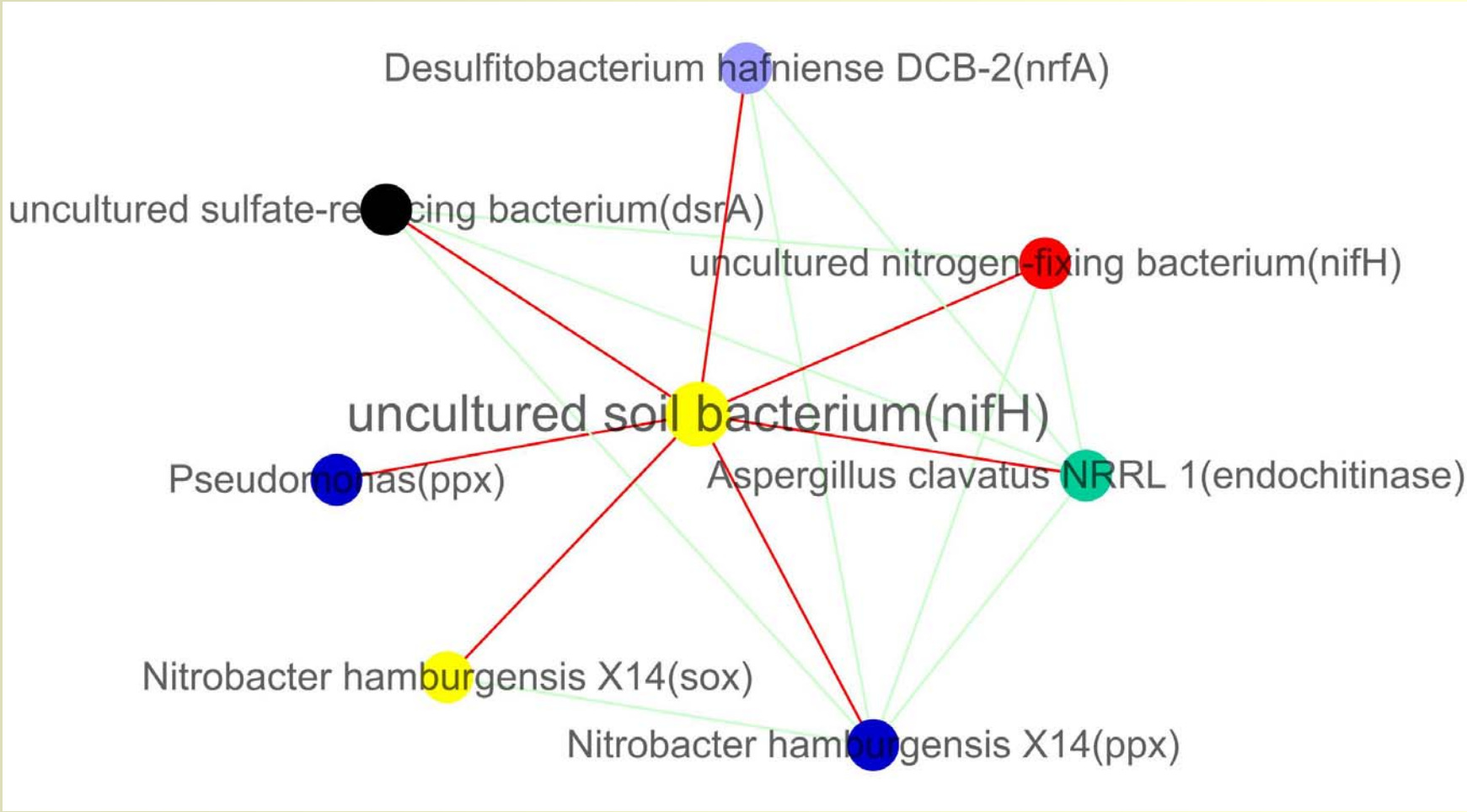
UNIVERSITY OF OKLAHOMA

nifH gene with the highest connectivity under eCO_2



- **The uncultured N fixing bacterium interacts with a variety of functional groups such as, denitrifiers, sulfate-reducers, photosynthetic bacteria, cellulose-degrading bacteria and other N fixers.**

The same nifH gene under aCO2



- **The same nifH gene has much fewer links at aCO_2**

Acknowledgement

OU

Jizhong Zhou
Zhili He
Liyou Wu
Meiying Xu
Ye Deng
Sanghoon Kang
Qichao Tu
Kai Xue
Feifei Liu

Univ of Minnesota

Peter Reich
Sarah Hobbie

Funding

- **USDA**
  - ✓ **NSF-USDA Microbial Observatories**

INSTITUTE FOR ENVIRONMENTAL GENOMICS
UNIVERSITY OF OKLAHOMA