

**Ecosystems and Networks Integrated** with Genes and **SNIGMA** Molecular Assemblies

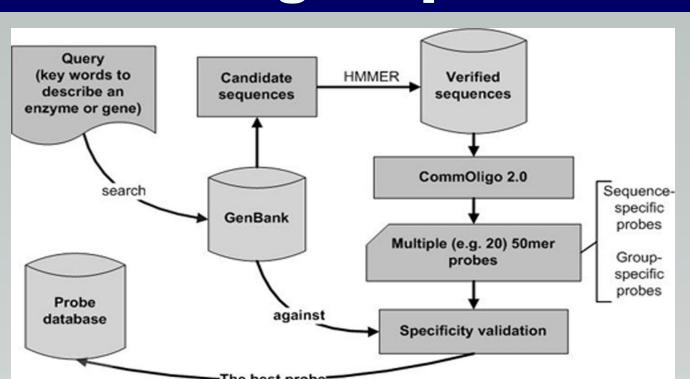
**Zhili He**<sup>1,2</sup>, Ye Deng<sup>1,2</sup>, Joy D. Van Nostrand<sup>1,2</sup>, Meiying Xu<sup>1,3</sup>, Christopher L. Hemme<sup>1,2</sup>, Qichao Tu<sup>1</sup>, Xingyuan Li<sup>4</sup>, Liyou Wu<sup>1,2</sup>, Terry J. Gentry<sup>5</sup>, Yifeng Yin<sup>6</sup>, Jost Liebich<sup>7</sup>, Terry C. Hazen<sup>2,8</sup>, and Jizhong Zhou<sup>1,2,8</sup>

<sup>1</sup>University of Oklahoa, Norman, OK; <sup>2</sup>Virtual Institute for Microbiology, Guangdong Institute for Microbi University, College Station, TX; <sup>6</sup>Trinity Biosystems Inc., Melo Park, CA; <sup>7</sup>Forschungszentrum Jülich, Germany; <sup>8</sup>Lawrence Berkeley National Laboratory, Berkeley, CA **Q-3138** 

# ABSTRACT

A new generation of functional gene arrays (GeoChip 3.0) has been developed, with ~28,000 probes covering approximately 57,000 gene variants from 292 functional gene families involved in carbon, nitrogen, phosphorus and sulfur cycles, energy metabolism, antibiotic resistance, metal resistance and organic contaminant degradation. GeoChip 3.0 also has several other distinct features, such as a common oligo reference standard (CORS) for data normalization and comparison, a software package for data management and future updating, and the gyrB gene for phylogenetic analysis. Computational evaluation of probe specificity indicated that all designed probes would have a high specificity to their corresponding targets. Experimental analysis with synthesized oligonucleotides and genomic DNAs showed that only 0.0036%-0.025% false positive rates were observed, suggesting that the designed probes are highly specific under the experimental conditions examined. In addition, GeoChip 3.0 was applied to analyze soil microbial communities in a multifactor grassland ecosystem in Minnesota, USA, which demonstrated that the structure, composition, and potential activity of soil microbial communities significantly changed with the plant species diversity. As expected, GeoChip 3.0 is a high throughput powerful tool for studying microbial community functional structure, and linking microbial communities to ecosystem processes and functioning.

## **Design Pipeline and Probe Information**



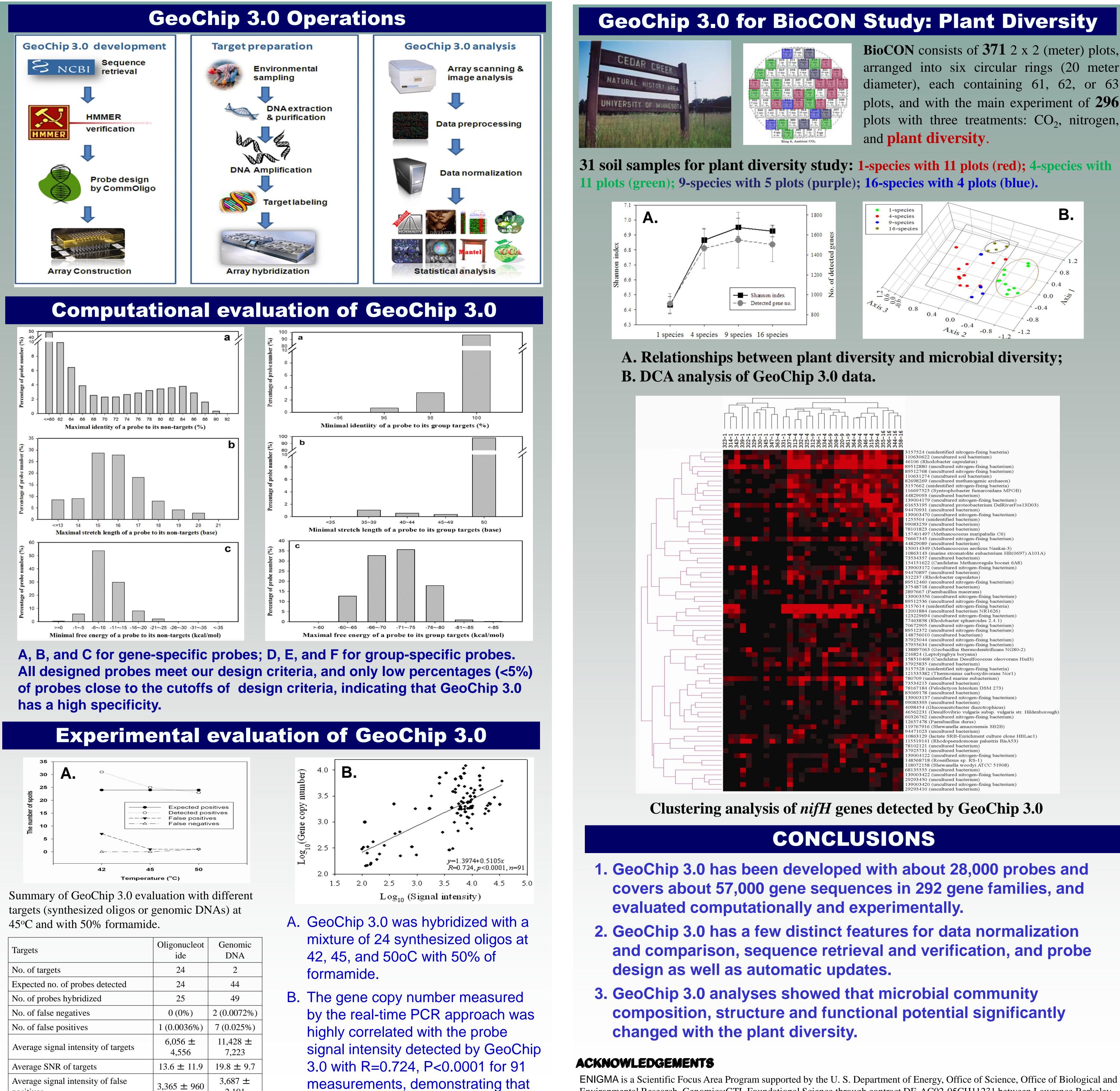
## **GeoChip 3.0 design pipeline:** The whole pipeline runs on a web-based Common Gateway Interface (CGI) server and all scripts were written in Perl, including four modules: sequence retrieval and verification, oligo probe design, probe validation, and automatic update.

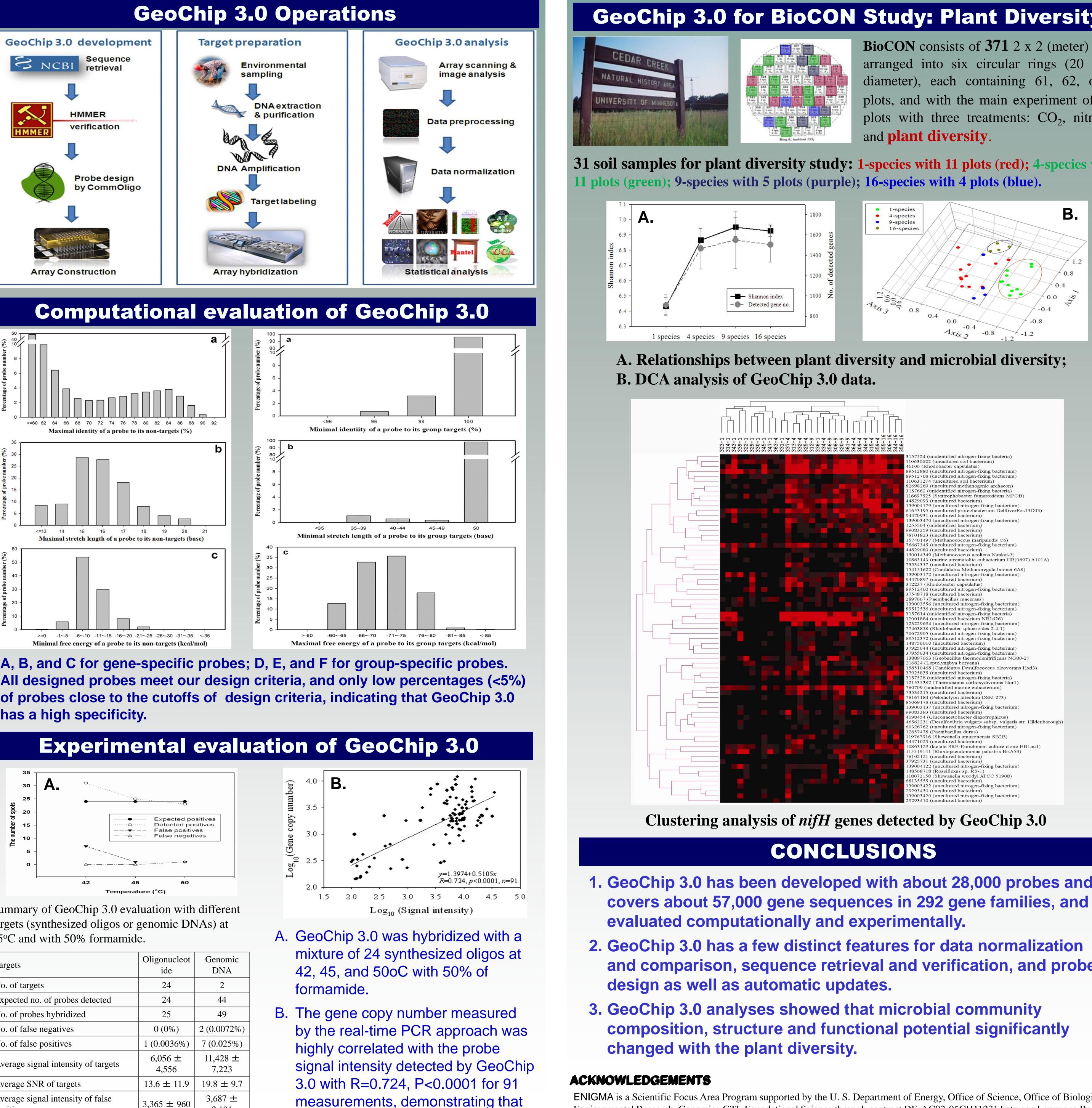
## Summary of GeoChip 3.0 probes by covered microbial domain and phylum

Domain	Phylum	No. species	No. probes	No. CDS covered	
Archaea		140	886	1807	
	Crenarchaeota	19	237	560	
	Euryarchaeota	107	577	1090	
	Thaumarchaeota	1	1	1	
	Unclassified archaeon	13	71	156	
Bacteria		2744	24939	52228	
	Acidobacteria	6	107	212	
	Actinobacteria	312	2468	5181	
	Aquificae	18	29	56	
	Bacteroidetes	166	684	1305	
	Chlamydiae	9	31	68	
	Chlorobi	13	178	361	
	Chloroflexi	15	302	642	
	Cyanobacteria	115	585	1259	
	Deinococcus-Thermus	8	100	215	
	Dictyoglomi	2	2	2	
	Firmicutes	371	2155	4881	
	Fusobacteria	3	12	34	
	Lentisphaerae	2	20	40	
	Nitrospirae	3	12	16	
	Planctomycetes	11	156	277	
	Proteobacteria	1428	13937	30107	
	Spirochaetes	24	46	128	
	Synergistetes	1	2	4	
	Tenericutes	36	47	88	
	Thermodesulfobacteria	4	8	10	
	Thermotogae	12	67	135	
	Verrucomicrobia	180	3956	7148	
	Unclassified bacterium	5	35	59	
Fungi		262	1759	2372	
C	Ascomycota	153	1377	1879	
	Basidiomycota	81	279	381	
	Glomeromycota	1	2	2	
	Microsporidia	1	1	2	
	Neocallimastigomycota	4	8	8	
	Unclassified fungus	22	92	100	
Others <sup>*</sup>		26	228	583	
Total		3172	27812	56990	

Gene category	No. of genes or enzymes	No. of probes	Sequence -specific probes	Group- specific probes	Covere d CDS
1. Carbon cycling	41	5196	1765	3431	10573
1.1 Carbon degradation1.1.1 Starch	33 8	<u>3777</u> 772	<u>1324</u> 263	2453 509	7337 1429
1.1.2 Cellulose	4	305	151	154	484
1.1.3 Hemicellulose	5	505	182	323	940
1.1.4 Lignin	4	330	274	56	419
1.1.5 Chitin	3	574	227	347	1179
1.1.6 Pectin           1.1.7 Others	<u> </u>	<u>33</u> 1258	<u>31</u> 196	$\frac{2}{1062}$	<u>35</u> 2851
1.2 Carbon fixation	4	1043	235	808	2544
1.3 Acetogenesis	1	122	35	87	250
1.4 Methane metabolism	3	254	171	83	442
1.3.1 Methane oxidation	2	118	87	31	193
1.3.2 Methane production	<u> </u>	136	84	52	249
2. Nitrogen cycling 2.1 Nitrogen fixation	16	<u>3763</u> 1224	<u>2148</u> 764	<u>1615</u> 460	7839 2277
2.2 Nitrification	2	111	<u> </u>	42	257
2.3 Denitrification	5	1543	1061	482	2995
2.4 Ammonification	2	315	87	228	742
2.5 Dissimilatory N reduction	2	262	85	177	568
2.6 Assimilatory N reduction	3	266	80	186	749
2.7 Anammox 3. Phosphorus utilization	3	42 <b>599</b>	183	40 <b>416</b>	251 <b>1220</b>
3.1 Organic P utilization	<u> </u>	31	26	5	36
3.2 Inorganic P biosynthesis	1	182	90	92	318
3.3 Inorganic P degradation	1	386	67	319	866
4. Sulphur cycling	4	1504	1083	421	2042
4.1 Sulphur reduction	2	1123	881	242	1438
4.2 Sulphur oxidation	1	195	66	129	339
4.3 Others	1	186	136	50	265
5. Energy process	2	<b>508</b>	410	<b>98</b>	<u>671</u>
5.1 <i>c</i> -type cytochromes 5.2 Hydrogenases	1	<u>384</u> 124	<u> </u>	<u>14</u> 84	<u>398</u> 273
6. Metal Resistance	41	4870	<u>603</u>	4267	10962
6.1 Aluminum	1	81	13	68	128
6.2 Arsenic	3	392	59	333	814
6.3 Cadmium	2	252	33	219	558
6.4 Cadmium and cobalt 6.5 Chromium	3	<u>1002</u> 543	<u>99</u> 65	<u>903</u> 478	2289 1292
6.6 Cobalt	1	49	6	43	96
6.7 Cobalt and nickel	3	7	1	6	17
6.8 Copper	5	804	71	733	1797
6.9 Lead	3	33	3	30	65
6.10 Mercury	6	291	77	214	620
6.11 Nickel6.12 Selenium	<u> </u>	<u>35</u> 3	3 2	32	<u>71</u> 4
6.13 Silver	4	213	17	196	419
6.14 Tellurium	4	558	50	508	1504
6.15 Zinc	2	589	103	486	1246
6.16 Miscellaneous	1	18	1	17	42
7. Organic Remediation	173	8614	2165	<u>6449</u>	17441
7.1 Aromatics 7.1.1 Aromatic carboxylic acids	<u>    124    </u> 38	6192 3302	<u>1654</u> 830	<u>4538</u> 2472	<u>12416</u> 6554
7.1.2 BTEX and aromatics	18	532	147	385	1105
Chlorinated aromatics	10	460	147	313	869
7.1.4 Heterocyclic aromatics	97.1.3	70	35	35	126
7.1.5 Nitoaromatics	9	500	70	430	1115
7.1.6 Polycyclic aromatics	18	241	119	122	416
7.1.7 Other aromatics 7.2. Chlorinated solvents	<u>22</u> 6	<u>1087</u> 355	<u> </u>	781 244	2231 747
7.3 Herbicide related compounds	12	816	135	<u> </u>	1682
7.4 Pesticide related compounds	4	288	37	251	604
7.5 Other Hydrocarbons	15	454	155	299	894
7.6 Others	12	509	73	436	1098
8. Antibiotic resistance	11	1594	265	1329	3944
8.1 Transporters	5	1181	104	1077	2913
	4	351	147	204	858
8.2 Beta-lactamases				٨٥	172
8.2 Beta-factamases 8.3 Others 9. Phylogenetic marker (GyrB)	2 1	62 1164	14 <b>629</b>	48 <b>535</b>	173 <b>2298</b>

# GeoChip 3.0: a high throughput tool for analyzing microbial community composition, structure, and functional activity





Targets	Oligonucleot ide	Genomic DNA	
No. of targets	24	2	
Expected no. of probes detected	24	44	
No. of probes hybridized	25	49	
No. of false negatives	0 (0%)	2 (0.0072%)	
No. of false positives	1 (0.0036%)	7 (0.025%)	
Average signal intensity of targets	6,056 ± 4,556	$11,428 \pm 7,223$	
Average SNR of targets	$13.6 \pm 11.9$	$19.8 \pm 9.7$	
Average signal intensity of false positives	$3,365 \pm 960$	3,687 ± 2,191	
Average SNR of false positives	$4.3 \pm 1.5$	$6.7 \pm 4.3$	

GeoChip 3.0 is quantitative.



**BioCON** consists of **371** 2 x 2 (meter) plots, arranged into six circular rings (20 meter diameter), each containing 61, 62, or 63 plots, and with the main experiment of 296 plots with three treatments: CO<sub>2</sub>, nitrogen,

Environmental Research, Genomics: GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.