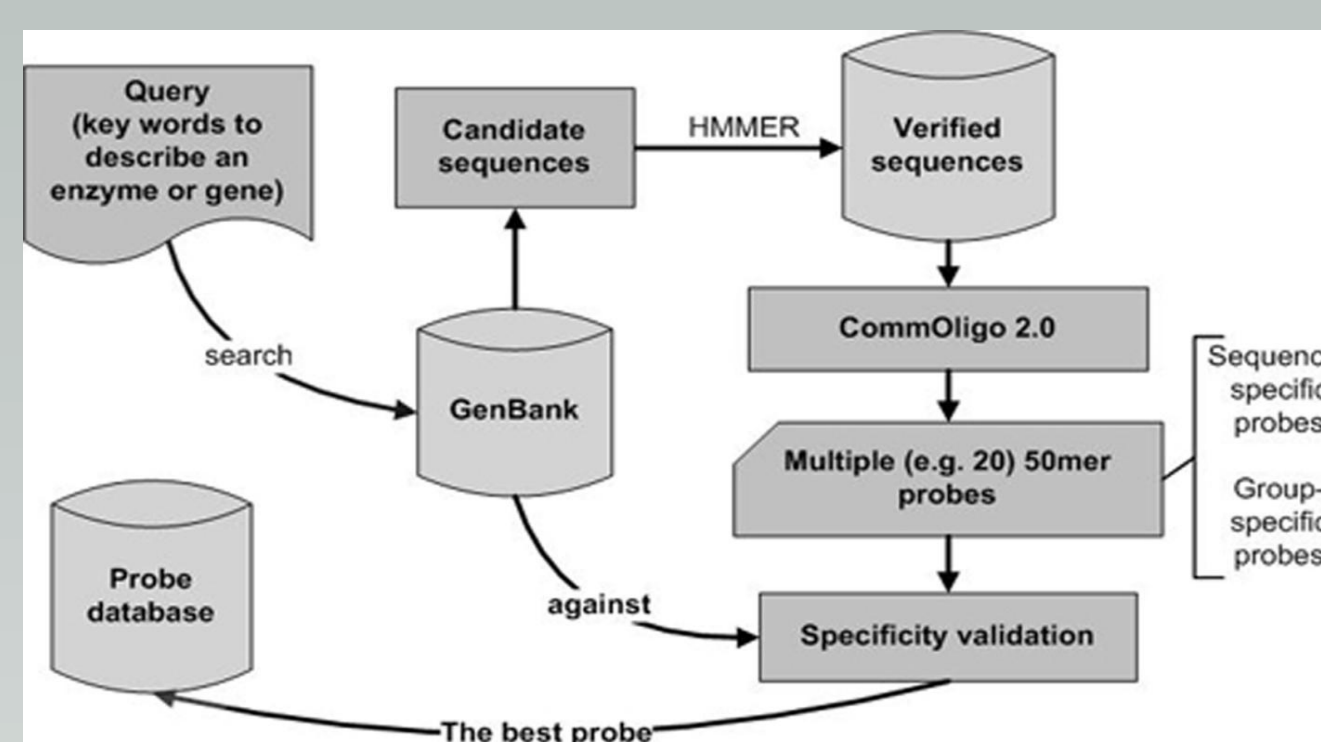


## ABSTRACT

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



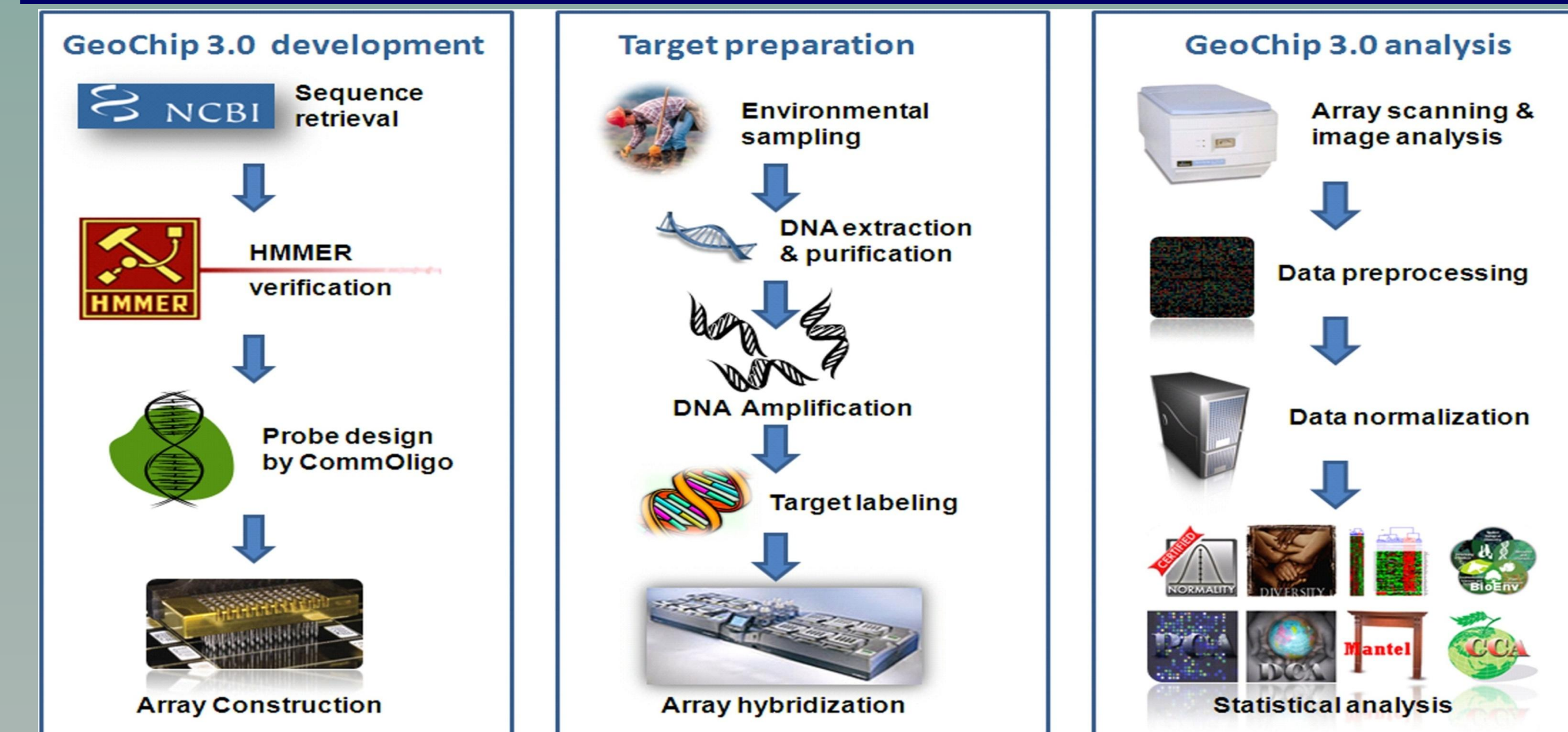
### GeoChip 3.0 design pipeline:

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.

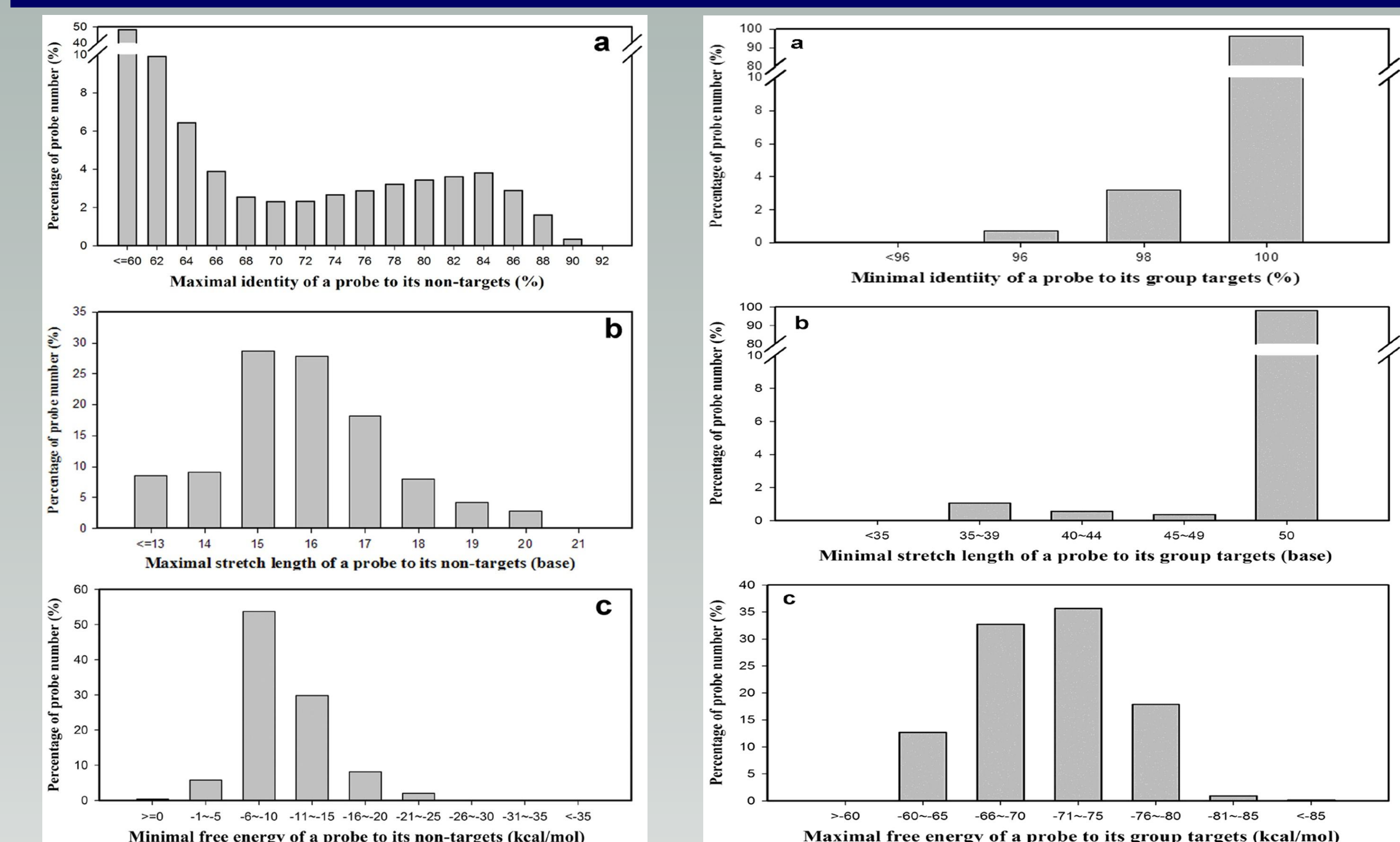
Summary of GeoChip 3.0 probes by covered microbial domain and phylum

Domain	Phylum	No. species	No. probes	No. CDS covered	
Archaea		140	886	1807	
	Crenarchaeota	19	237	560	
	Euryarchaeota	107	577	1090	
	Thaumarchaeota	1	7	10	
	Unclassified archaea	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
Dietyoglovi		2	2	2	
Firmicutes		371	2155	4881	
Fusobacteria		3	12	24	
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
		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*	26	228	583
		<b>Total</b>	<b>3172</b>	<b>27812</b>	<b>56990</b>

## GeoChip 3.0 Operations

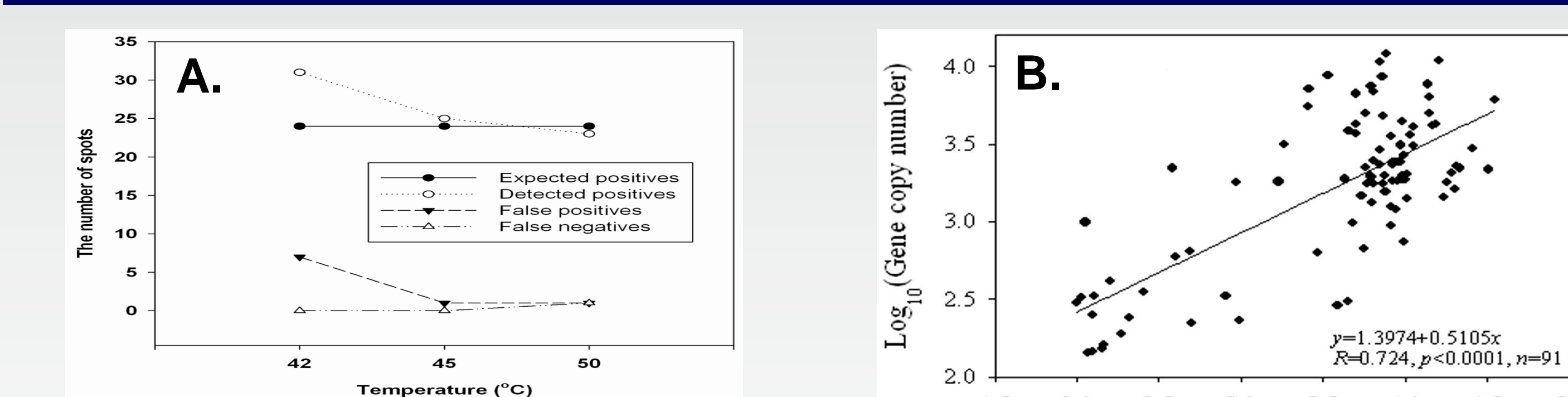


## Computational evaluation of GeoChip 3.0



**A, B, and C for gene-specific probes; D, E, and F for group-specific probes. All designed probes meet our design criteria, and only low percentages (<5%) of probes close to the cutoffs of design criteria, indicating that GeoChip 3.0 has a high specificity.**

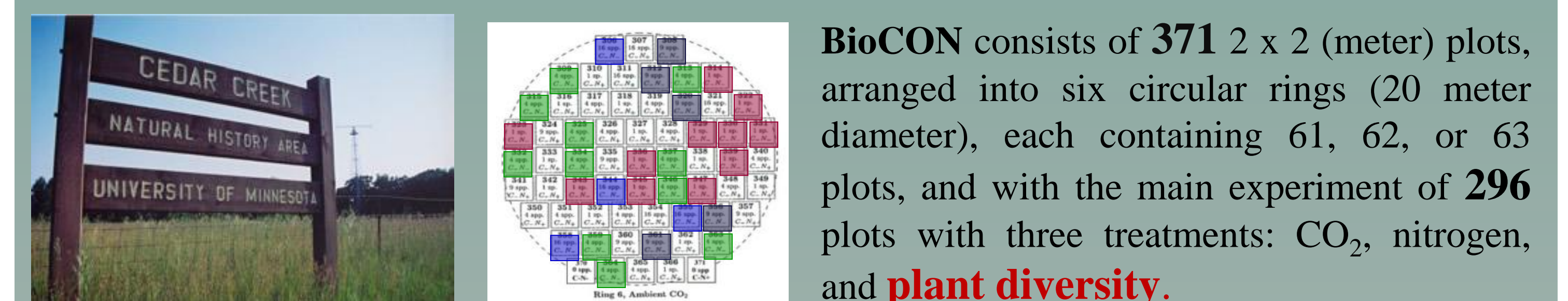
## Experimental evaluation of GeoChip 3.0



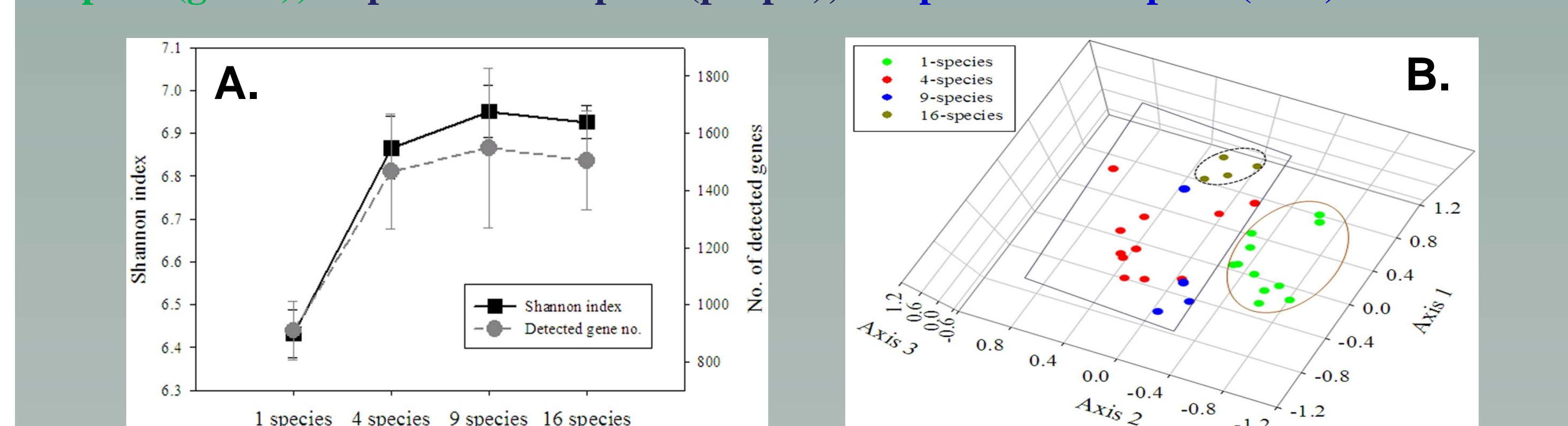
Summary of GeoChip 3.0 evaluation with different targets (synthesized oligos or genomic DNAs) at 45°C and with 50% formamide.

Targets	Oligonucleotide 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 (0.0%)	2 (0.0072%)
No. of false positives	1 (0.0036%)	7 (0.025%)
Average signal intensity of targets	6,056 ± 4,556	11,428 ± 7,223
Average SNR of targets	13.6 ± 11.9	19.8 ± 9.7
Average signal intensity of false positives	3,365 ± 960	3,687 ± 2,191
Average SNR of false positives	4.3 ± 1.5	6.7 ± 4.3

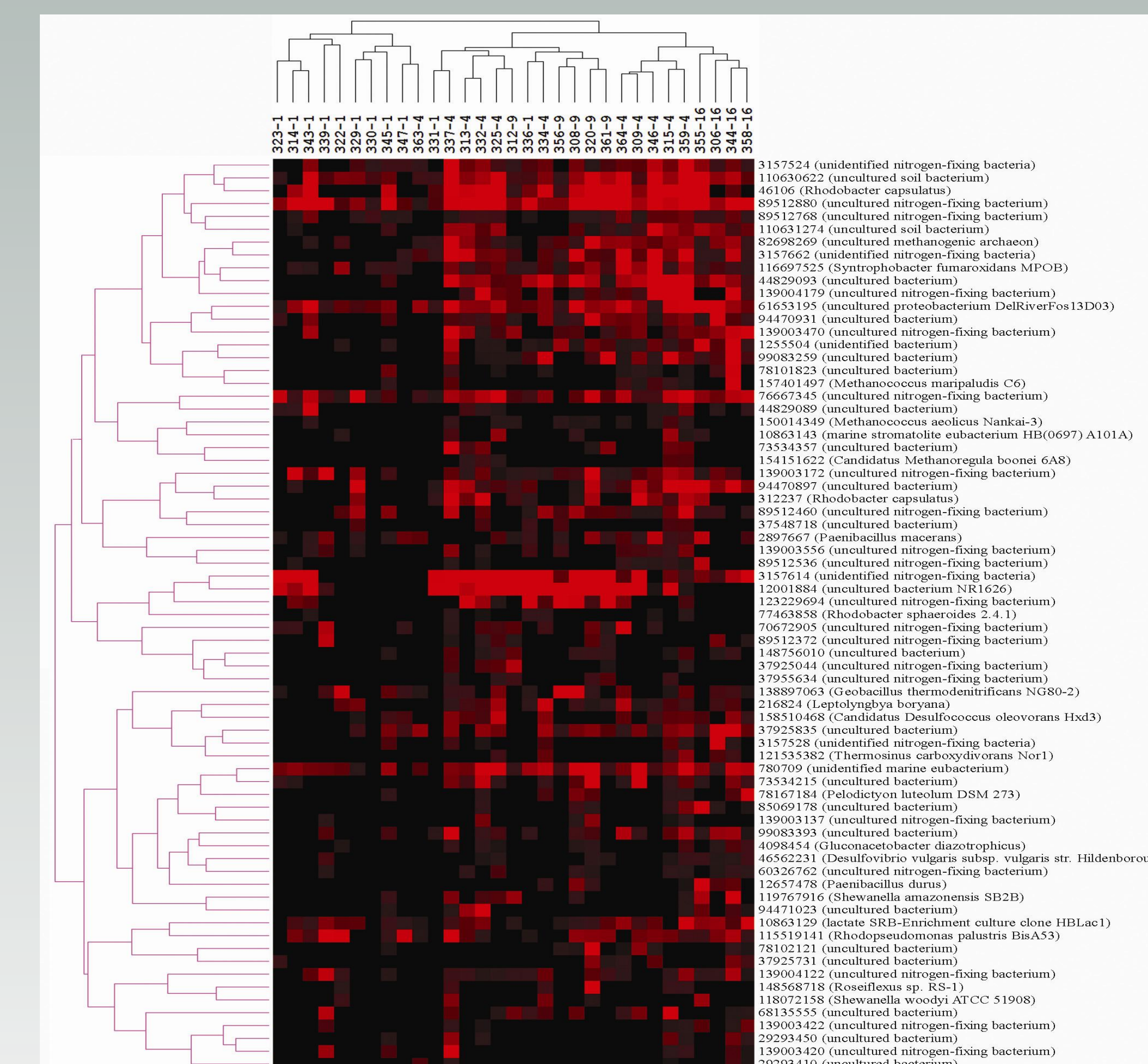
## GeoChip 3.0 for BioCON Study: Plant Diversity



**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, and plant diversity.**



**A. Relationships between plant diversity and microbial diversity; B. DCA analysis of GeoChip 3.0 data.**



Clustering analysis of *nifH* genes detected by GeoChip 3.0

## CONCLUSIONS

1. GeoChip 3.0 has been developed with about 28,000 probes and covers about 57,000 gene sequences in 292 gene families, and evaluated computationally and experimentally.
2. GeoChip 3.0 has a few distinct features for data normalization and comparison, sequence retrieval and verification, and probe design as well as automatic updates.
3. GeoChip 3.0 analyses showed that microbial community composition, structure and functional potential significantly changed with the plant diversity.

## ACKNOWLEDGEMENTS

ENIGMA is a Scientific Focus Area Program supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Foundational Science through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy.