"2nd International Conference on Oceanography"

July 21-23, Las Vegas, Nevada, USA

Single-cell gene expression analysis – technologies and application

Weiwen Zhang

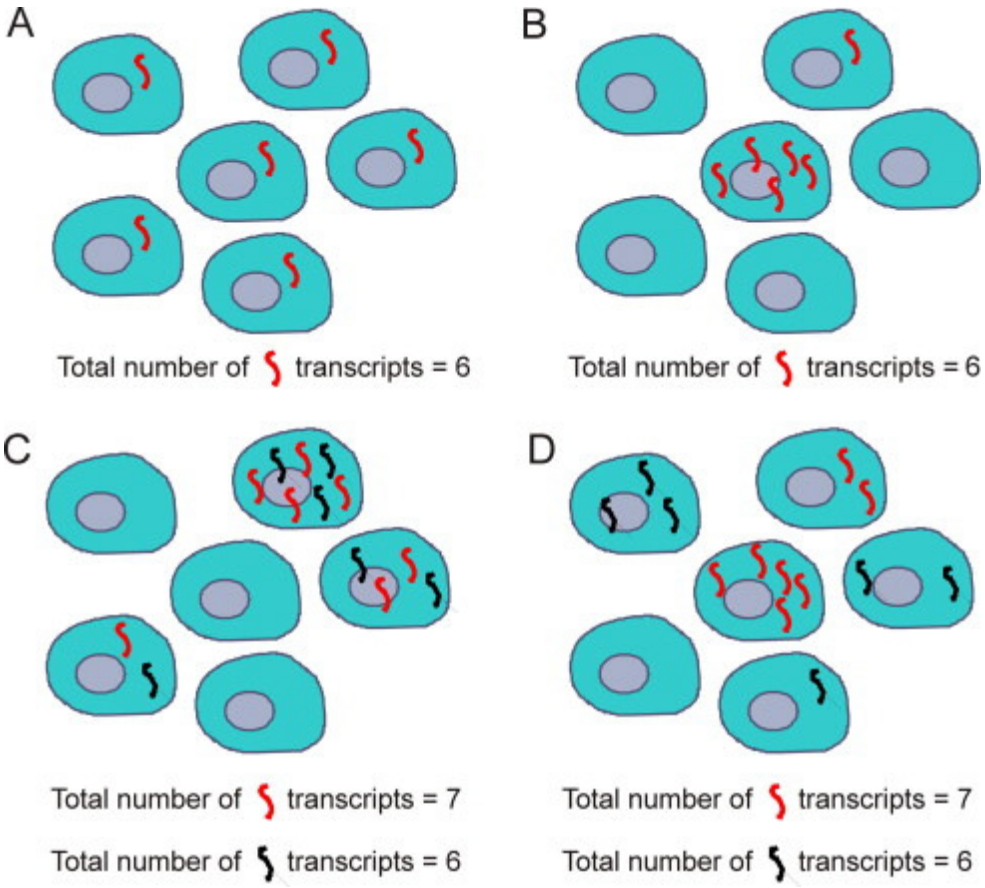
Laboratory of Synthetic Microbiology

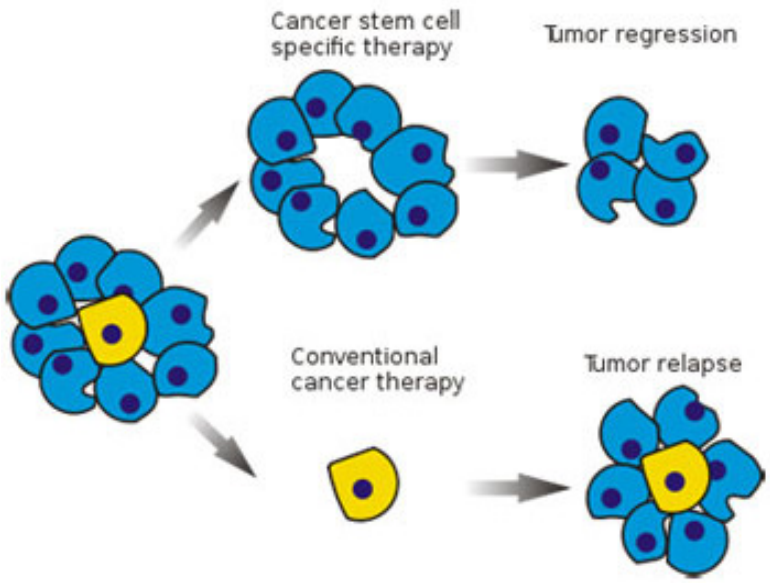
School of Chemical Engineering & Technology

Tianjin University, Tianjin, P.R. China

July 22, 2014

Why analyze gene expression in a single cell?

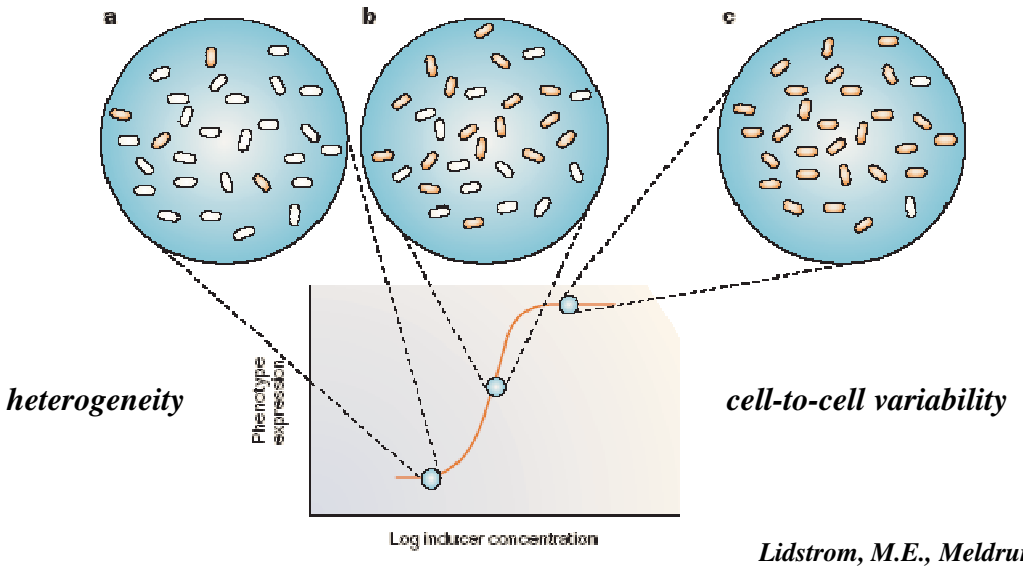




Cancer Stem Cells

Analysis from the population could be misleading!

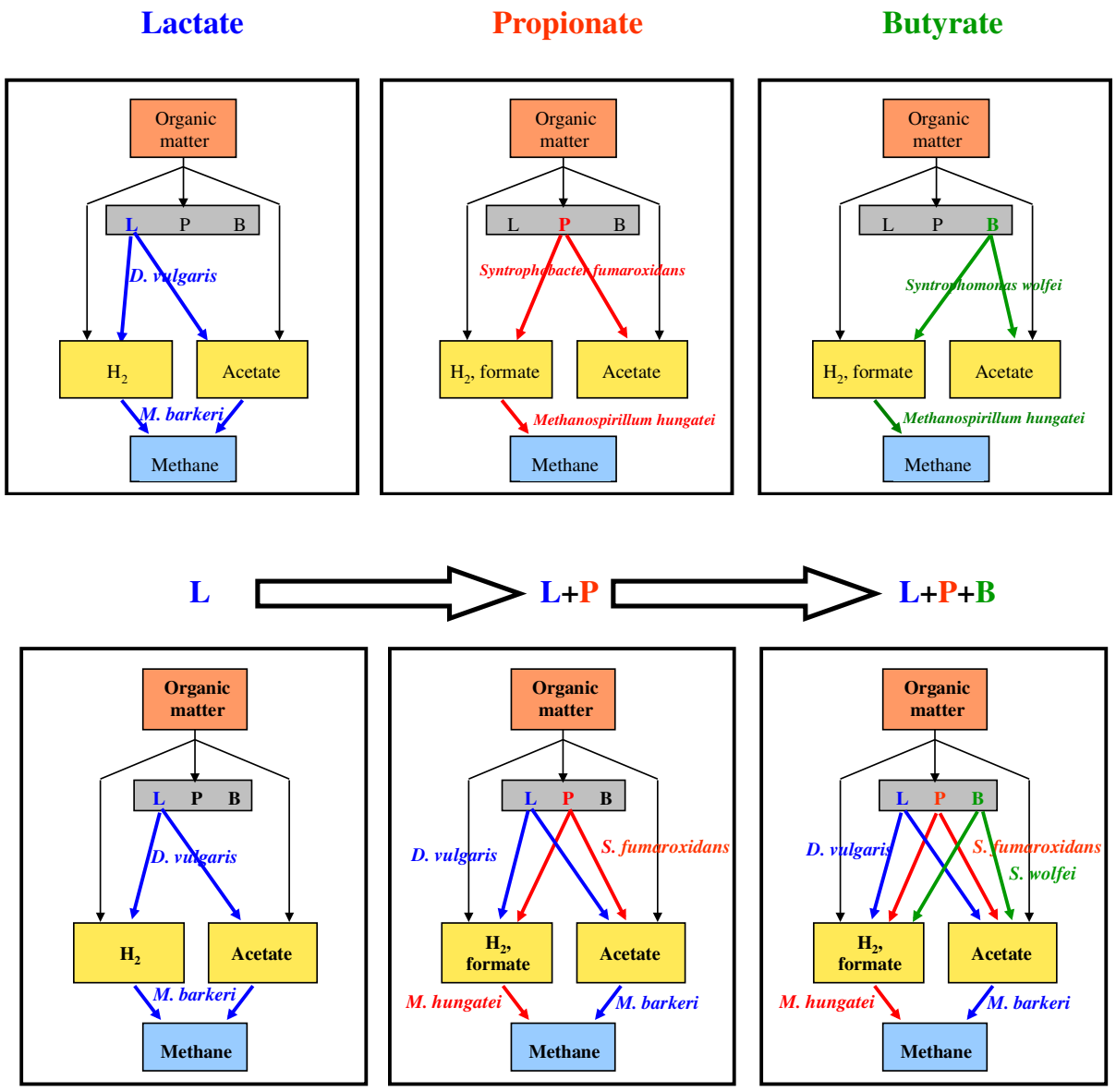
Why analyze gene expression in a single microbial cell?



Lidstrom, M.E., Meldrum, D.R., 2003. Nat Rev. Microbiol, 1:158-164

- **Substantial cell-to-cell heterogeneity even in isogenic populations grown under identical conditions.**
- **Gene expression heterogeneity could cause long-term heterogeneity at the cellular level.**
- **In natural ecosystems, microbial cells with diverse genotypes and phenotypes co-existed.**
- **Only less than 1% of microbial species in natural environments can be cultured and accessed by traditional gene expression analysis methods that typically requires a large number of cells.**

Synthetic ecology, new frontier in synthetic biology!



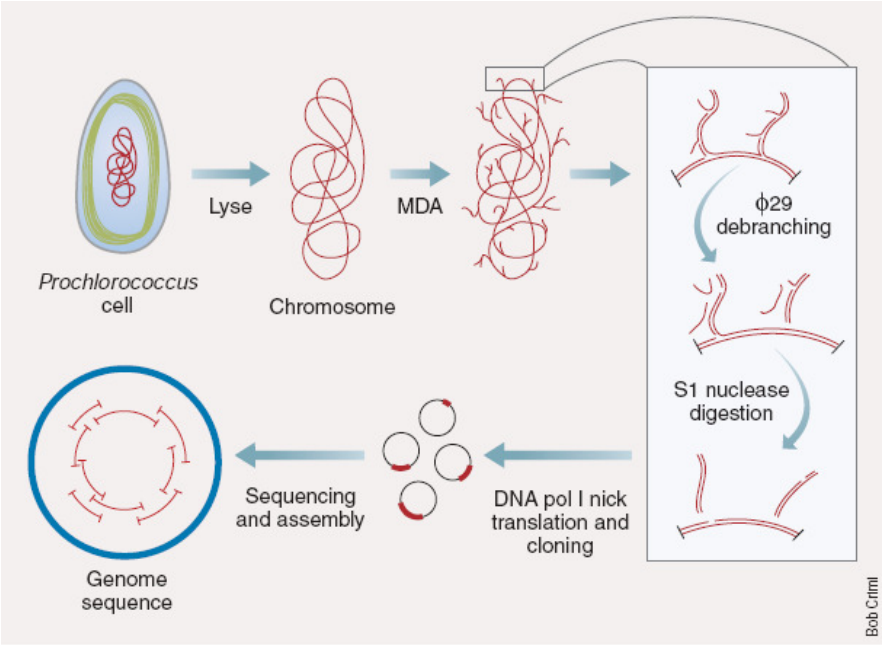
Building more "robust and controllable" eco-systems for biotechnological application

Single-cell Alternatives to Meta-approaches in Environmental Microbiology

- Meta-approaches average cell-cell difference
- Cells with diverse genotypes and phenotypes were found within any community
- Sub-species (strain) level resolution not available
- Single-cell genomics; Single-cell transcriptomics; Single-cell proteomics (?)

Single-cell genomics

A bacterial chromosome = a few femtograms (10^{-15} g) of DNA



The cellular DNA is amplified >10^9-fold by multiple displacement amplification (MDA) using random primers

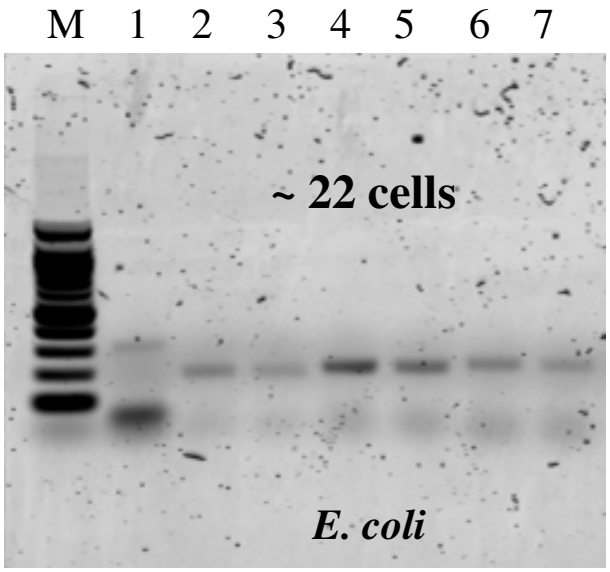
Zhang, *et al*. *Nat. Biotechnol.* 24, 681–687 (2006).

Single bacterial-cell gene expression

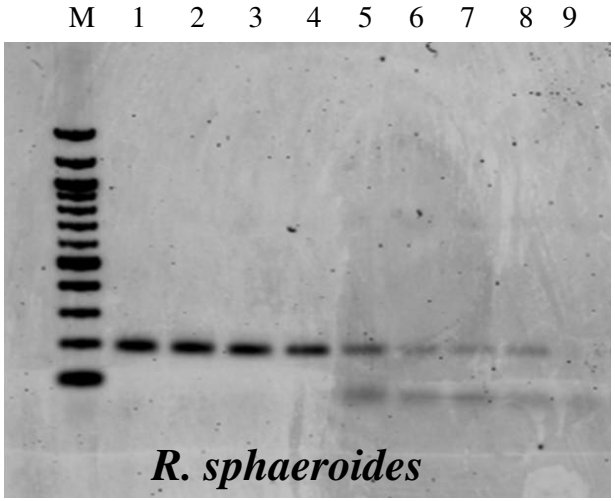
Gene expression analysis at single bacterial cell level, is that possible??

Cell No. in each reaction (When *E. coli* OD_{600} = 1.0, Cell density = $1X10^9$/mL)

Dilution	10	100	1000	10,000	100,000	1,000,000	10,000,000	100,000,000
Cell No.	2.22E+5	2.22E+4	2.22E+3	2.22E+2	22.2	2.22	0.222	0.0222
RNA (ng)	4.26	0.426	4.26E-2	4.26E-3	4.26E-4	4.26E-5	4.26E-6	4.26E-7



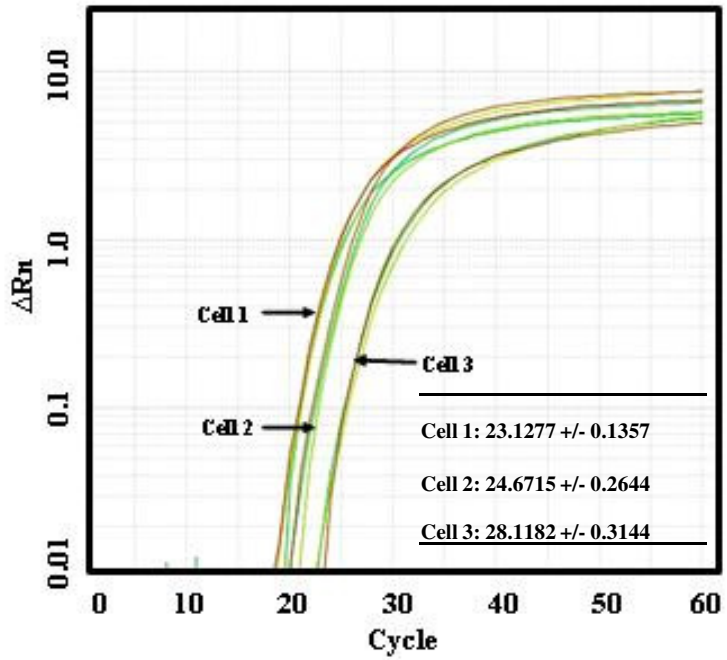
1: *groEL2*
2-3: *rbcL*
4-5: 16S rRNA
6-7: *dnaK*



M: 100 bp marker
1: 500000 cells
2: 50000 cells
3: 5000 cells
4: 1000 cells
5: 500 cells
6: 50 cells
7: 25 cells
8: 5 cells
9: negative control

Two-step RT-qPCR to measuring gene expression in single cell

Amplification of three individual *E. coli* cells from the exponential growing population



16S rRNA gene is the amplification target
Each reaction used 1/20th of the cDNA
Three technical replicates for each cell

Brief protocol:

- **RNA extraction: Carried out using ZR RNA MicroPrep Kit (Zymo Research, Orange, CA) with minor modification.**

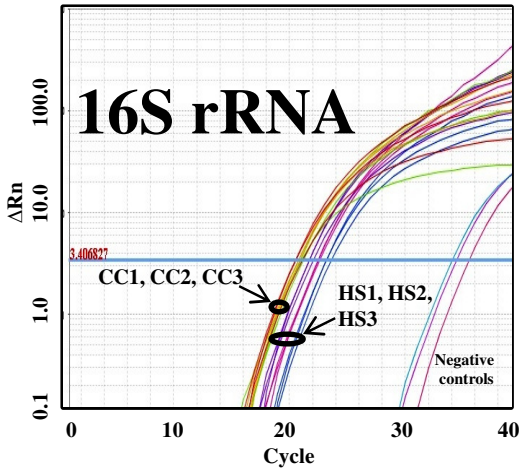
- **cDNA synthesis: SuperScript VILO cDNA Synthesis Kit (Invitrogen)**

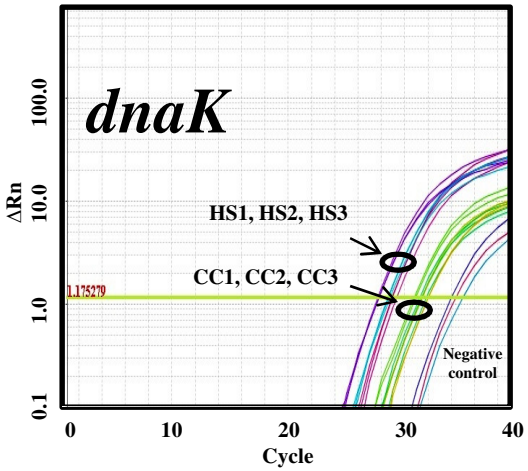
- **qPCR analysis: EXPRESS SYBR GreenER qPCR SuperMixs Kit (Invitrogen, San Diego, CA)**

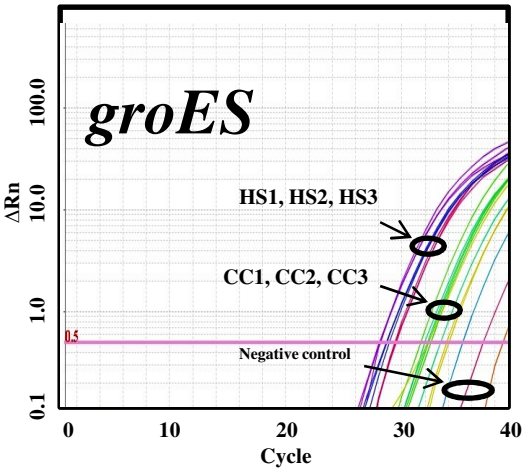
- **Multiple genes each cell**

- **Able to separate technical and biological variation**

Single-cell gene expression analysis of the response to heat shock







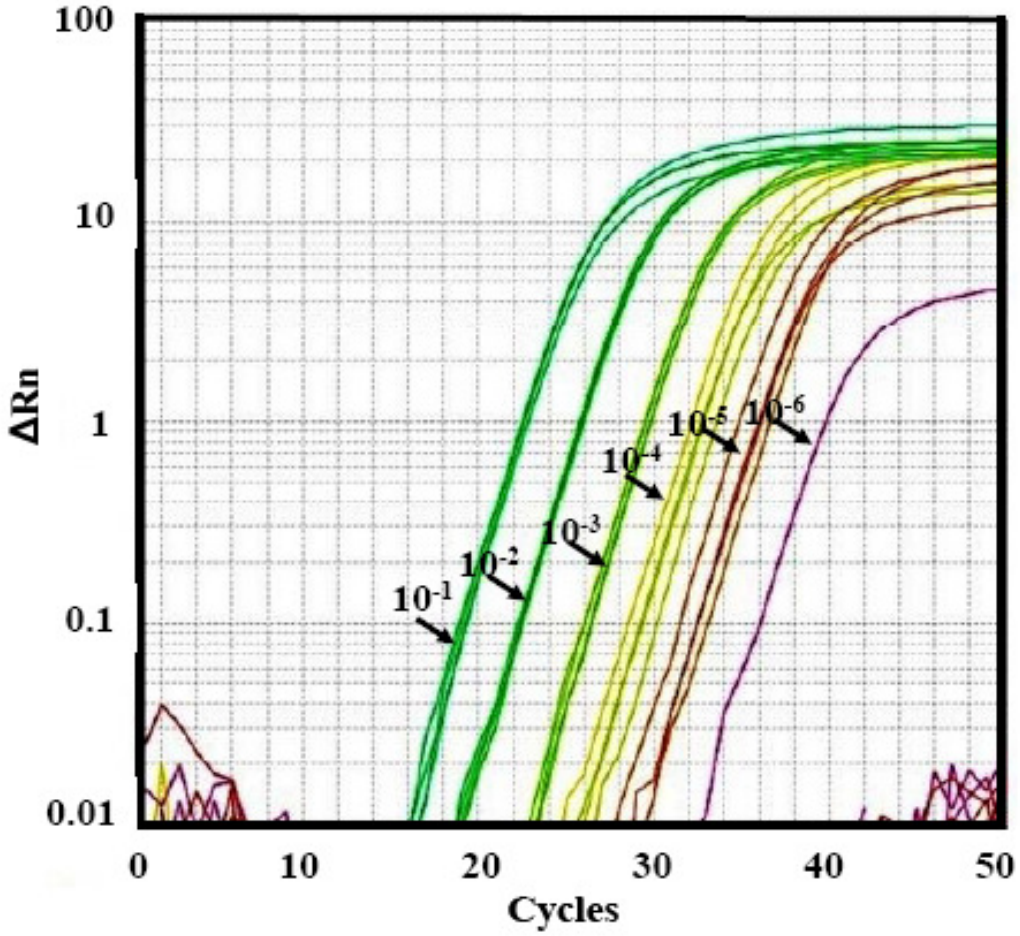
- **Three cells (biological replicates) for each condition (controls vs. heat-shock) were individually isolated**
- **Three genes were analyzed in each cell: 16S rRNA, *dnaK* and *groES***
- **Each reaction used 1/20th of the cDNA**
- **Three technical replicates for each gene**

Average qPCR CT values and standard deviation among all technical and biological replicates

	Control CC (Avg_CT ± StDv)		Heat Shock HS (Avg_CT± StDv)	
16S rRNA	*Cell No. 1*	20.6777 ± 0.3125	*Cell No. 1*	21.7777 ± 0.1864
	Cell No. 2	20.7948 ± 0.0689	*Cell No. 2*	23.2901 ± 0.2512
	Cell No. 3	21.0096 ± 0.1281	*Cell No. 3*	22.4832 ± 0.0818
dnaK	*Cell No. 1*	30.2822 ± 0.1763	*Cell No. 1*	28.6768 ± 0.1008
	Cell No. 2	31.7915 ± 0.3143	*Cell No. 2*	27.7821 ± 0.0468
	Cell No. 3	31.0435 ± 0.3126	*Cell No. 3*	28.7926 ± 0.2161
groES	*Cell No. 1*	31.4224 ± 0.4704	*Cell No. 1*	28.7846 ± 0.1268
	Cell No. 2	32.1555 ± 0.4673	*Cell No. 2*	28.1949 ± 0.0606
	Cell No. 3	32.5109 ± 0.7372	*Cell No. 3*	29.5052 ± 0.0537

Gene expression analysis using diluted cDNA from a single bacterial cell

E. coli contains 10^5-10^6 copies of rRNA molecules!



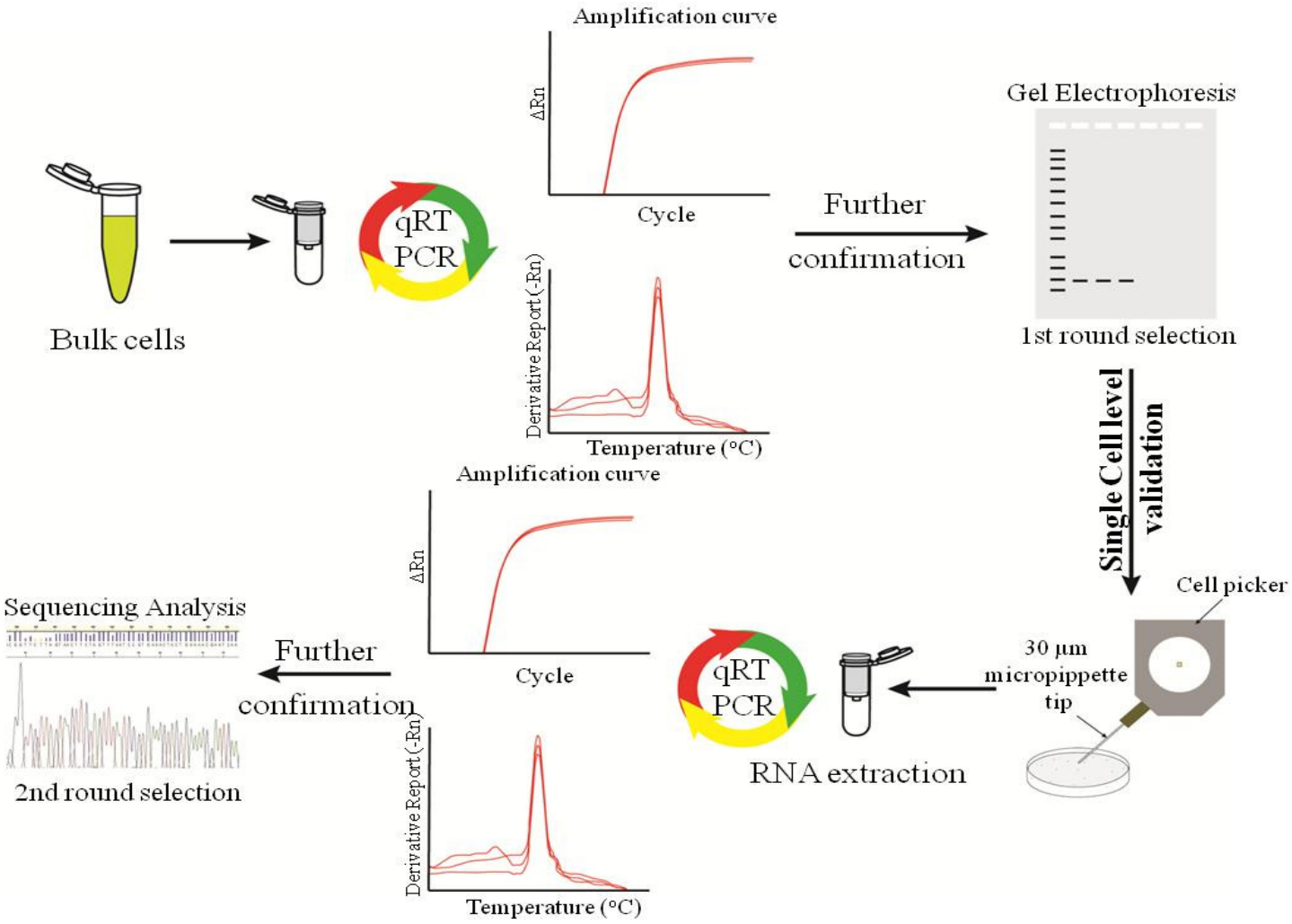
Average qPCR CT values and standard deviation among all technical replicates

Dilution	Avg_CT	StDv
10^{-1}	18.2434	0.0961
10^{-2}	21.6089	0.1713
10^{-3}	25.0732	0.4291
10^{-4}	28.6372	0.5535
10^{-5}	31.9372	0.7767

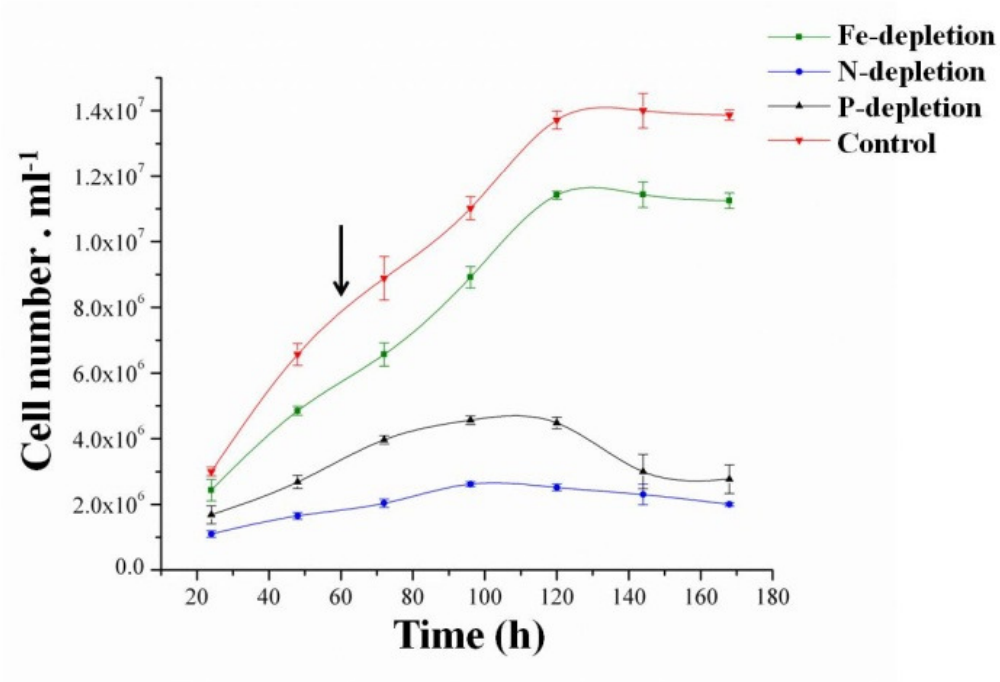
Gao *et al*., 2011, J. Microbiol. Method. 85:221-7.

Scheme of analytical procedure

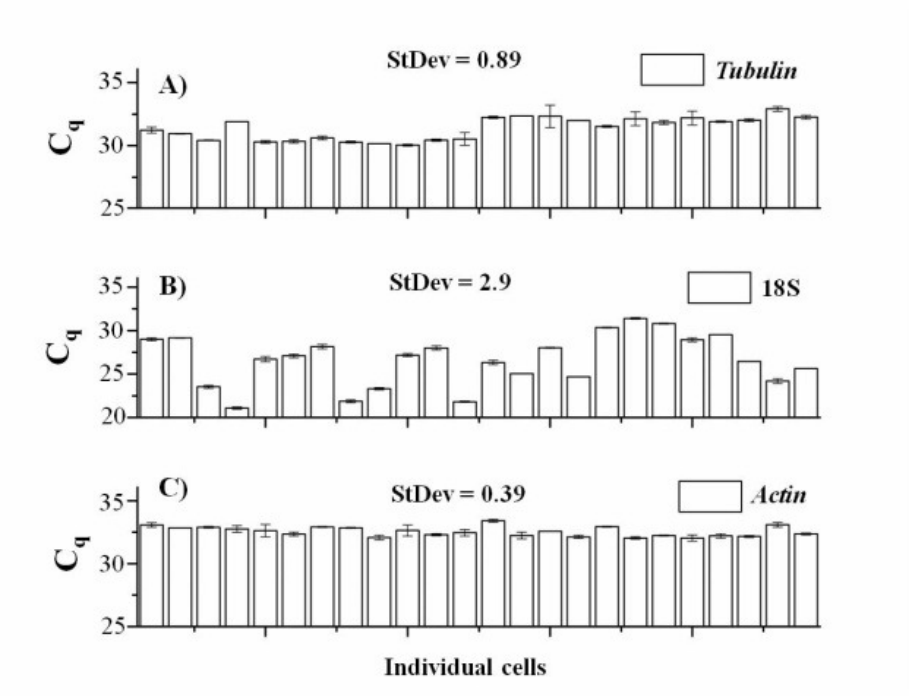
Very tiny amount of total RNA: 1-10 femtogram per *E. coli* cell (1 femtogram = $1e^{-15}$ gram)!



Response heterogeneity of *Thalassiosira pseudonana* to stress

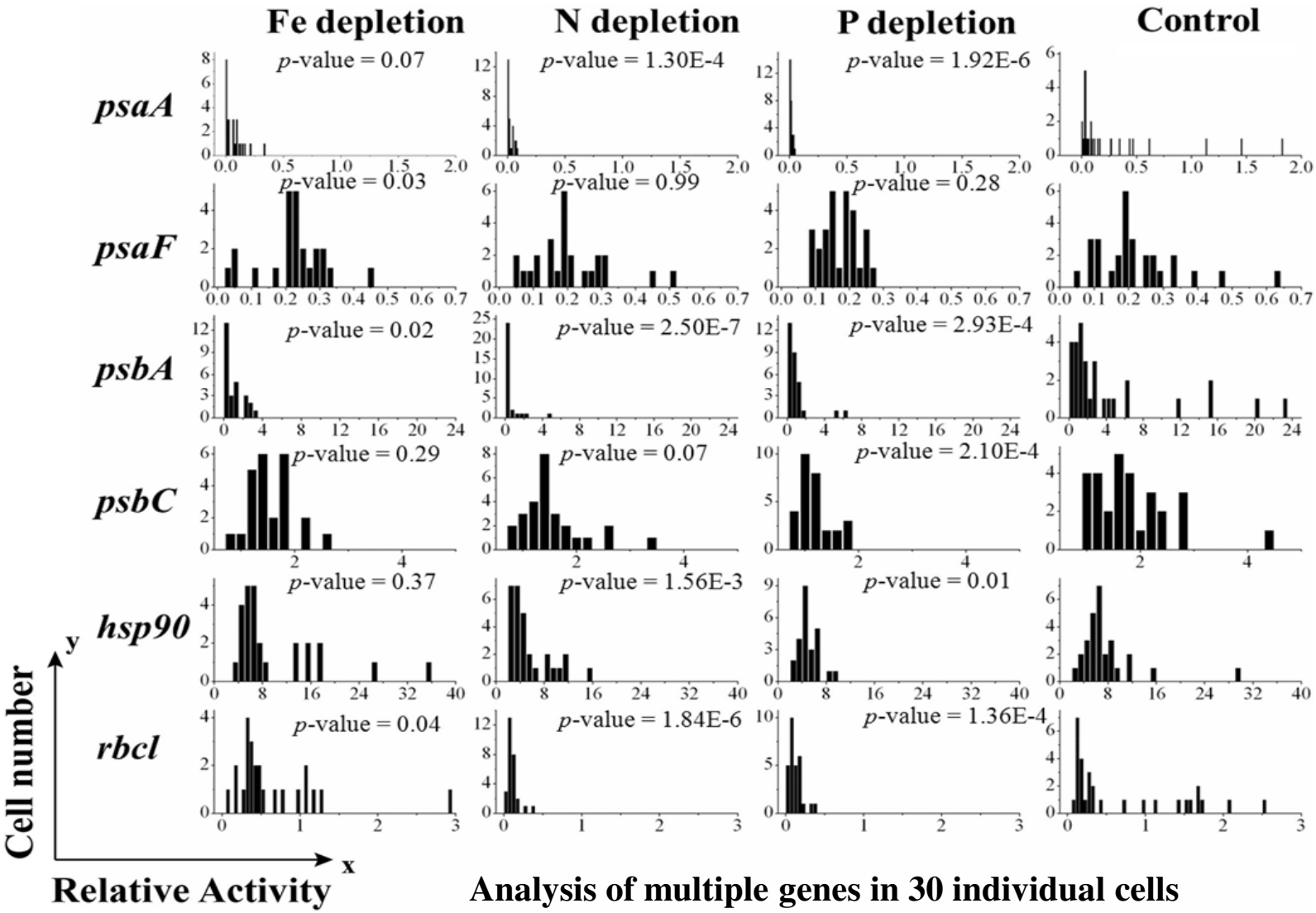


Growth



Selection of internal control

Response heterogeneity of *Thalassiosira pseudonana* to stress



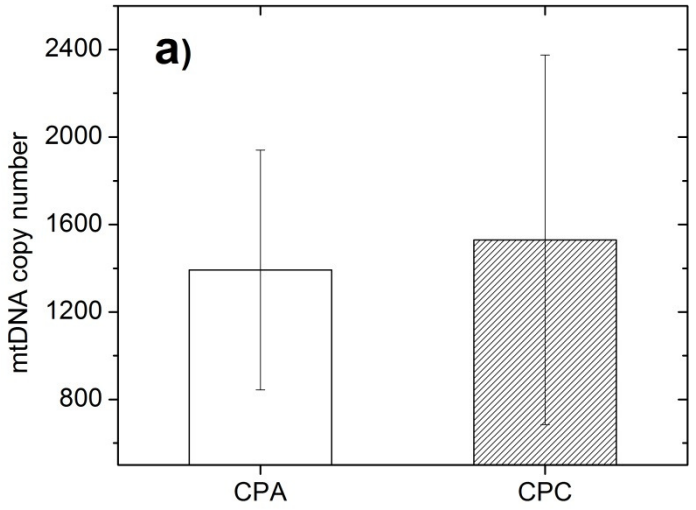
Analysis of multiple genes in 30 individual cells

Shi *et al*., 2013, Appl. Environ. Microbiol. 79:1850-8

Measure mitochondrial gene expression levels in single cells

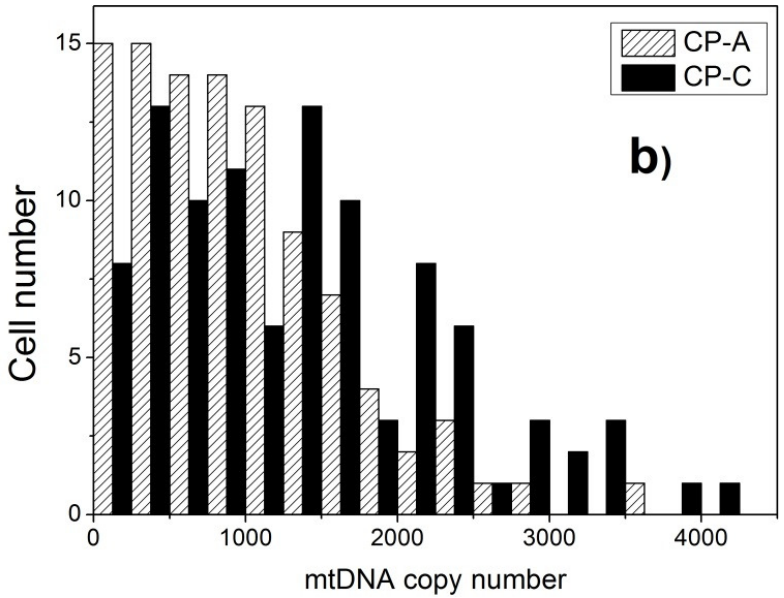
- Cancer progression is a process associated with a series of complex, step-wise changes at the biomolecular level.
- Esophageal adenocarcinoma (EAC) is a highly lethal cancer type and is believed to develop from esophageal epithelial cells.
- Mitochondria found to play a major role in the transformation.
- Single-cell analysis of the differential hypoxia response in two human Barrett's esophageal cell lines, CPA and CPC.

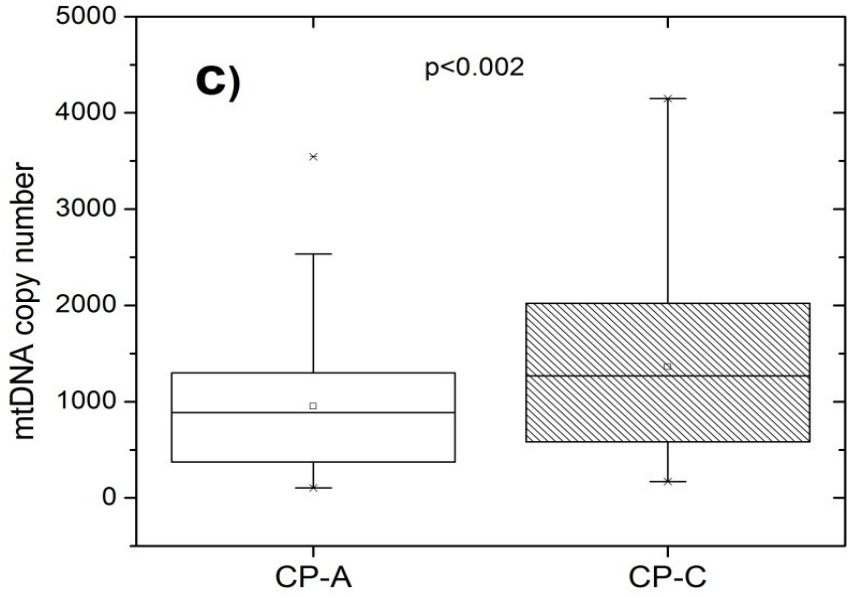
Mt copy number difference



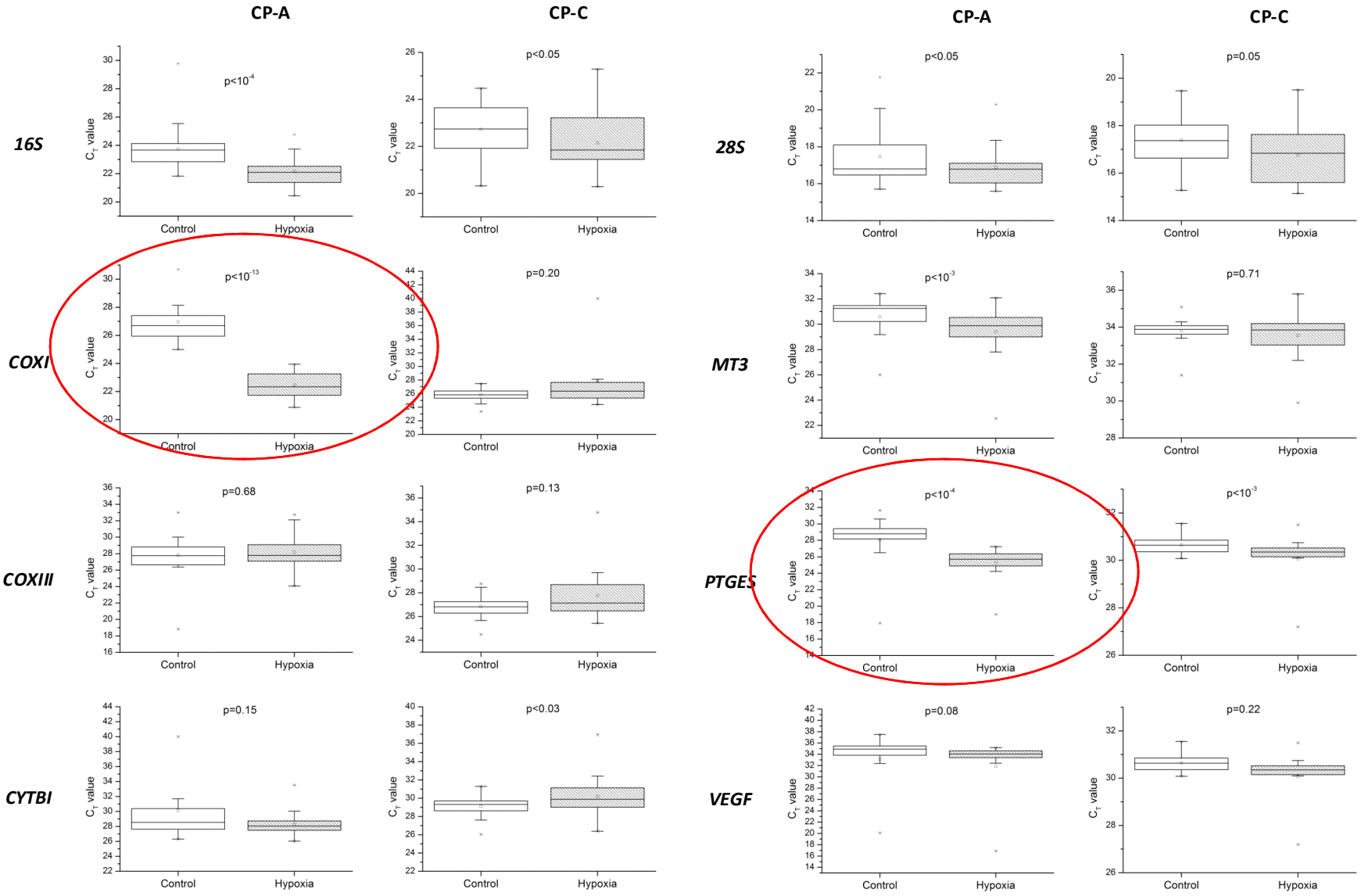
Bulk cells based

Single-cell based

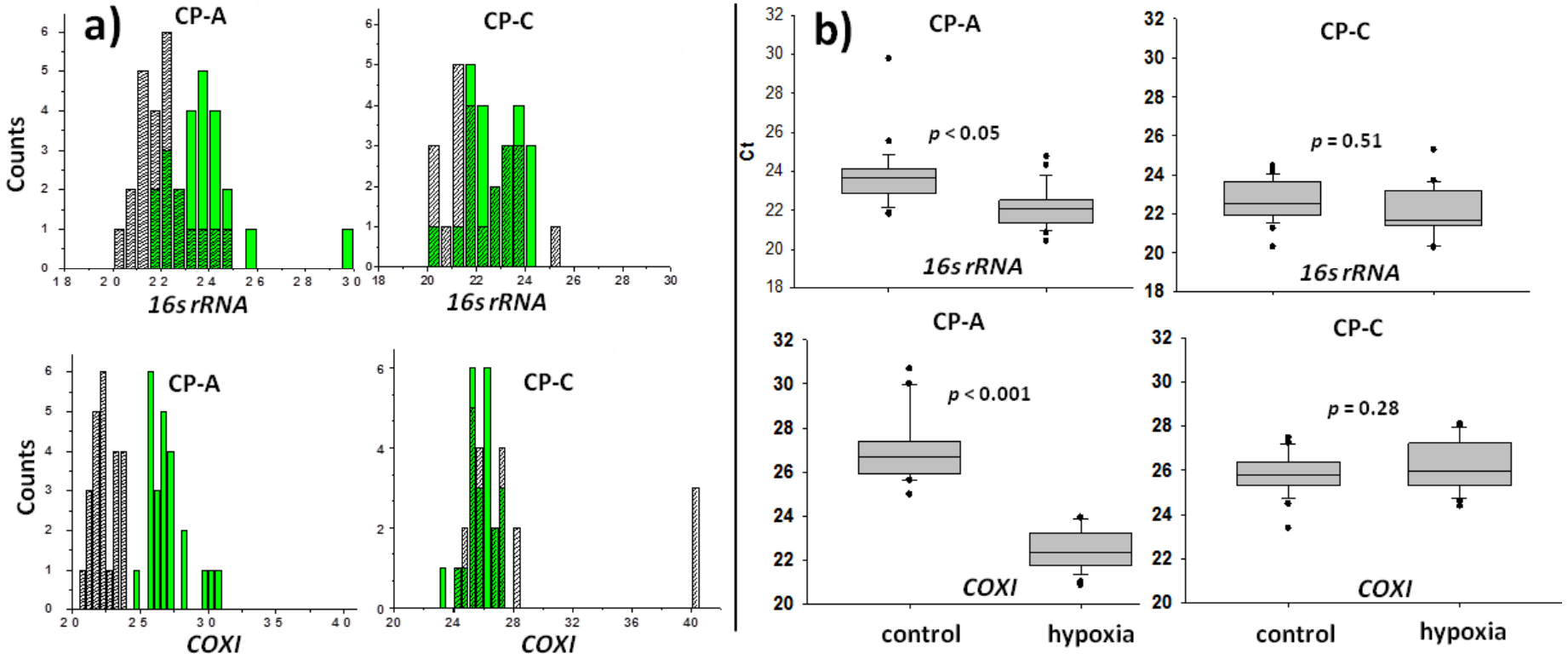




Simultaneous measurement of multiple genes encoded by chr and mt DNA in single cells



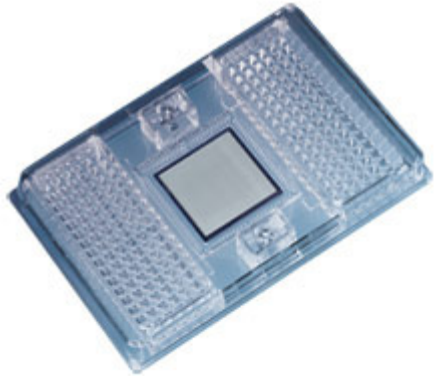
We proposed that mitochondria may be one of the key factors in the early cancer progression



Wang *et al.*, 2013, PloS One. 8:e75365

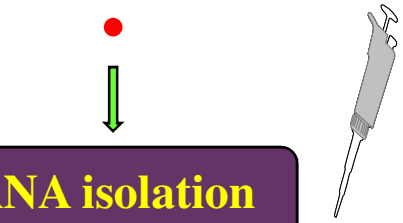
Why transcriptomics for single bacterial cell?

- qRT-PCR: 5~20 genes/cell
- Fluidigm: 96 or more genes/cell
- 1,000 ~ 10,000 (and more) genes per microorganism

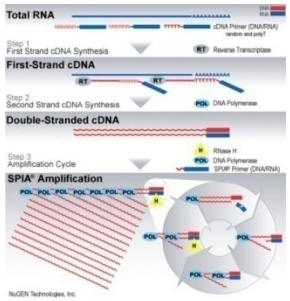


BaSiC-RNAseq: Bacterial Single Cell-RNAseq

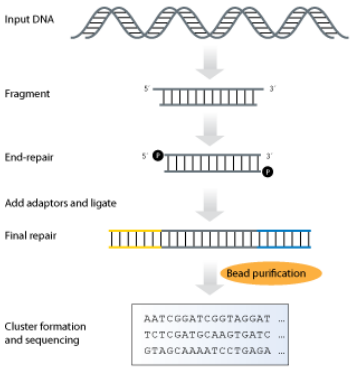
Single bacterial cells



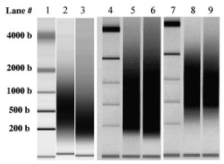
RNA isolation



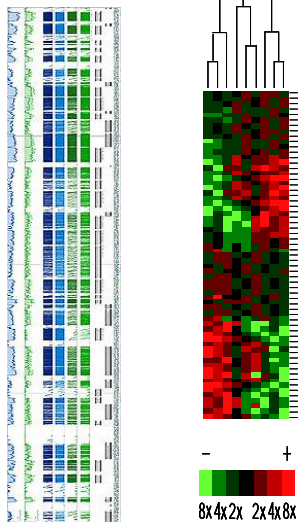
RNA amplification



QC: verification



cDNA labeling



Illumina Hi-Seq

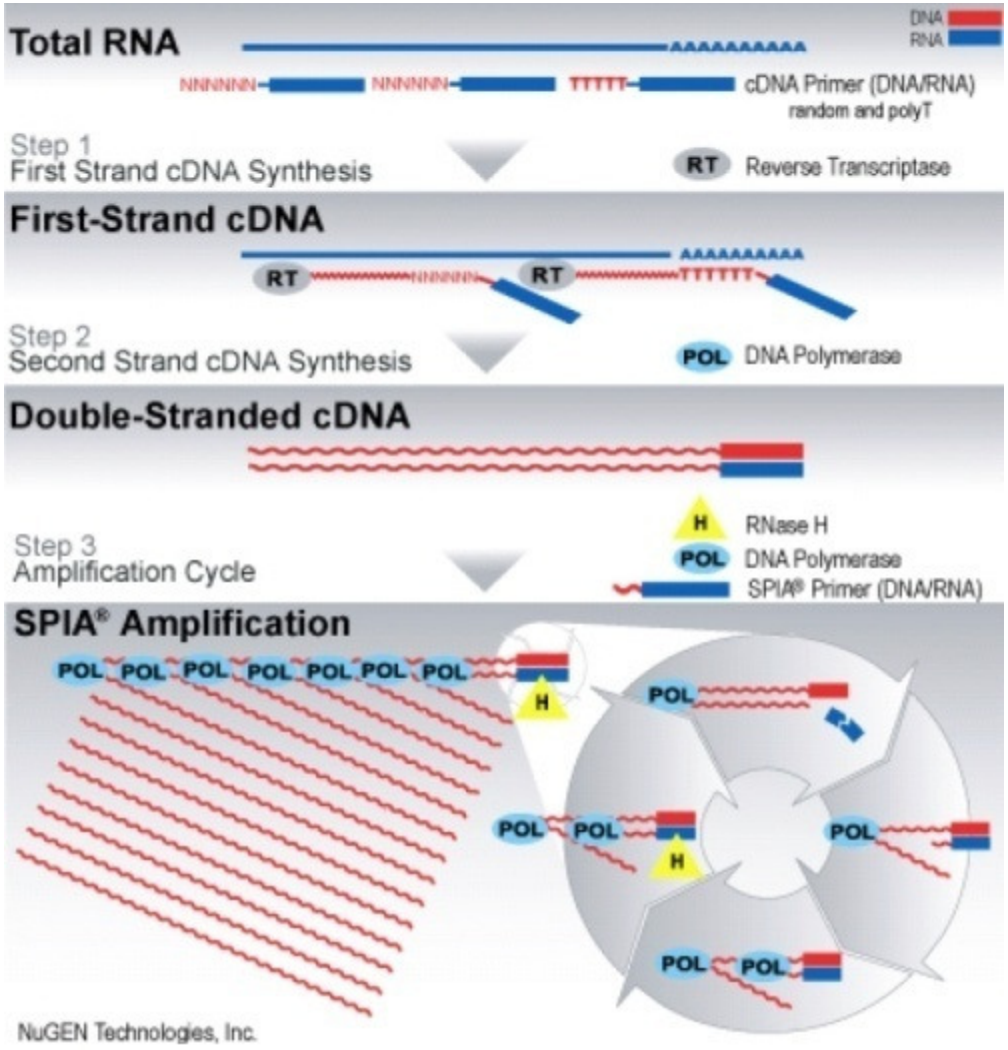
BaSiC-RNAseq
RNA Amplification

NuGen RNA Amplification Kit

Unique at:

- **primers: random/polyT**
- **Poly DNA polymerase**
- **RNase H**
- **SPIA DNA/RNA primer**

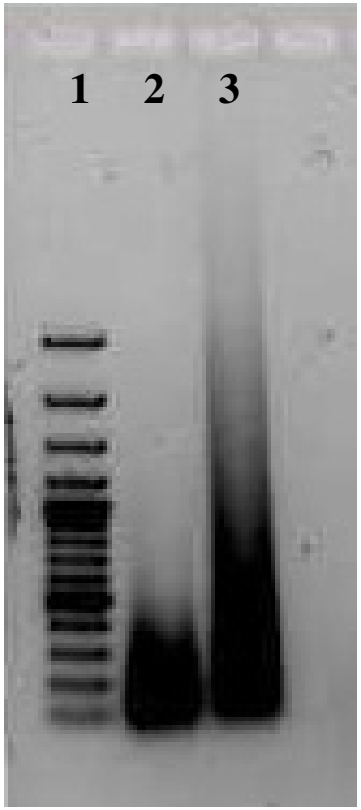
1 bacterial cell generated
7~19 µg cDNA



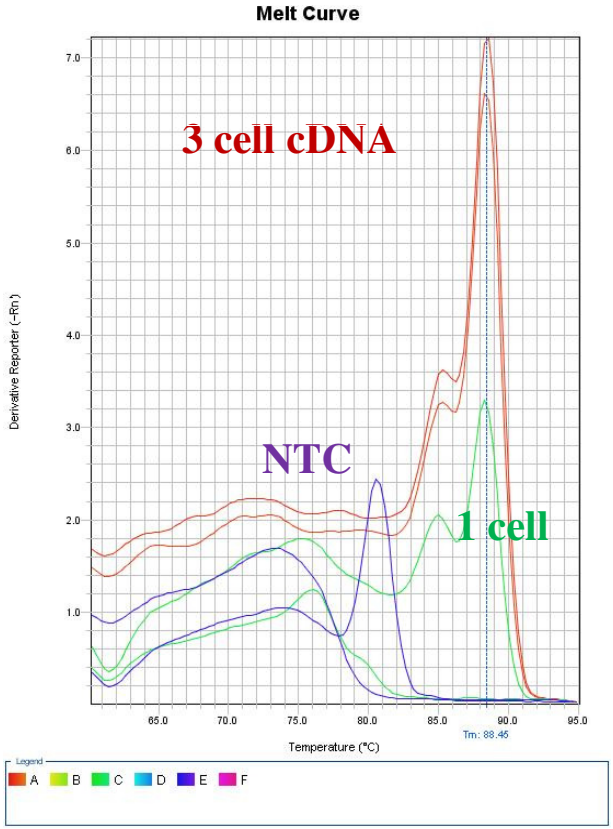
BaSiC-RNAseq: Quality Control

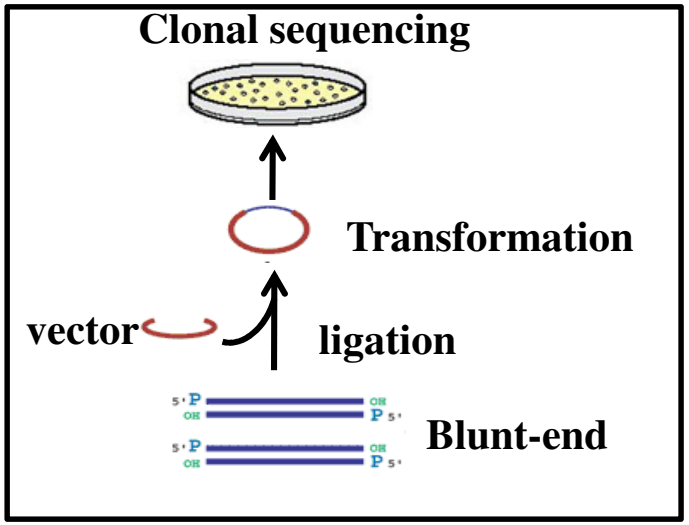
Cyanobacterial Synechocystis sp. PCC 6803

1, 100 bp ladder
2, NTC (H_2O as input)
3, single bacterial cell



Agarose gel 1%

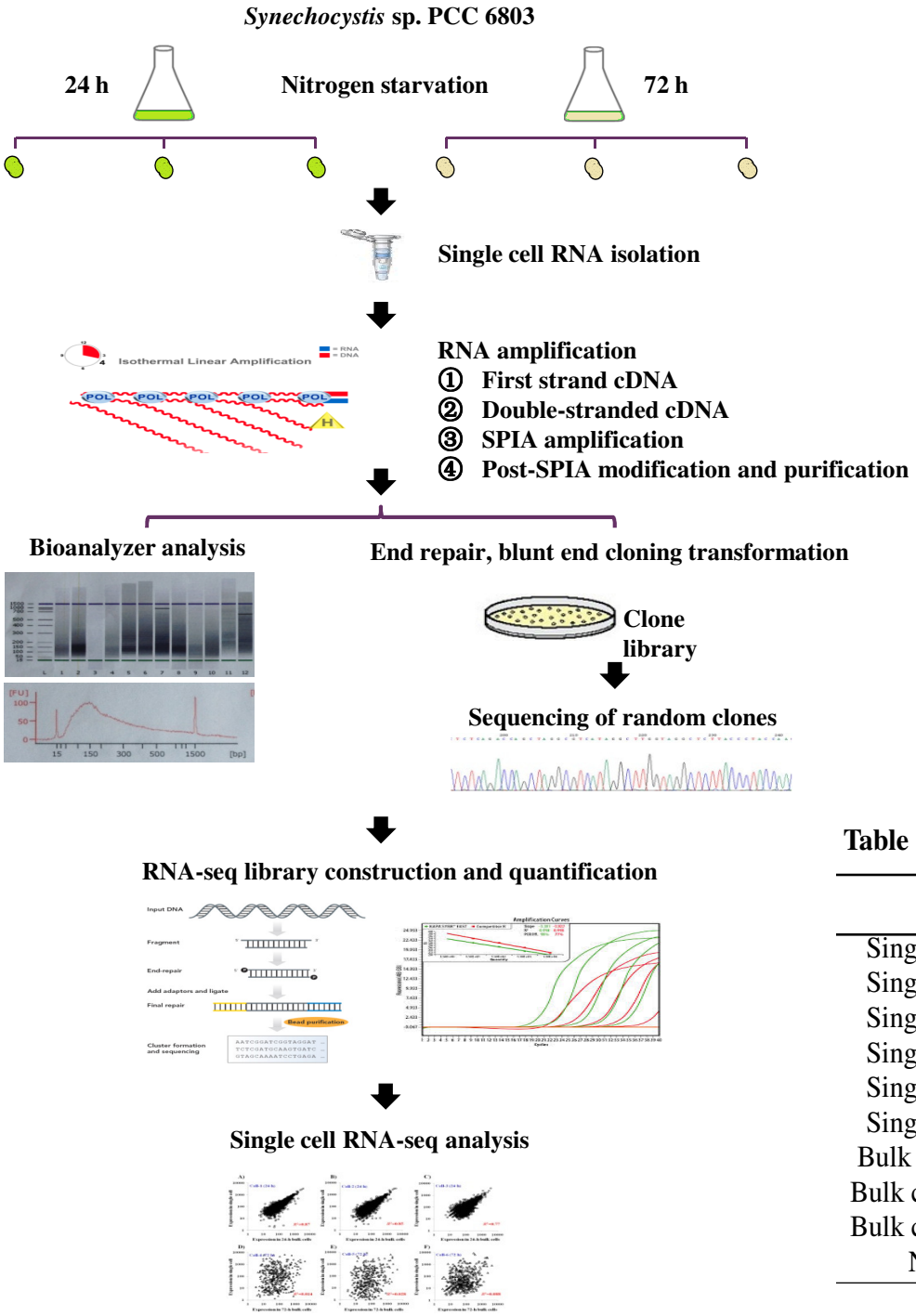




Sequencing of clone library:
All 30 clones are from cyanobacteria

BlastN against GenBank

All *Synechocystis* sp. PCC 6803 genes!



Research hypotheses?

1) Heterogeneity could vary upon stress in isogenic bacterial population?

2) The change as a driver for adaption and evolution of the population?

Table 1. Summary of RNA amplification from single bacterial cells.

	Concentration（ug/uL）	Total (ug)	A260/280
Single cell-1	0.36	9.00	1.82
Single cell-2	0.30	7.40	1.82
Single cell-3	0.57	14.18	1.84
Single cell-4	0.63	15.78	1.90
Single cell-5	0.67	16.78	1.88
Single cell-6	0.38	9.45	1.82
Bulk cells-0 h	0.34	8.60	1.81
Bulk cells-24 h	0.27	6.85	1.92
Bulk cells-72 h	0.28	6.90	1.84
NTC	0.11	2.80	1.81

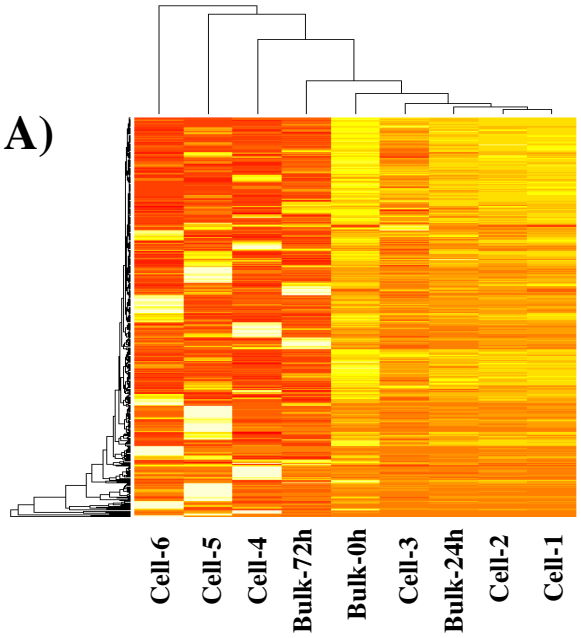
RNA-seq coverage of transcripts

	O h	24 h				72 h				Transcriptome detected in all samples
	Bulk cells	Bulk cells	Single cell-1	Single cell-2	Single cell-3	Bulk cells	Single cell-4	Single cell-5	Single cell-6	
Total transcripts identified	**3117**	**2937**	**3117**	**3102**	**2615**	**1961**	**1132**	**1521**	**982**	
Percentage of the genome (%)	0.98	0.92	0.98	0.98	0.82	0.62	0.36	0.48	0.31	
Amion acid metabolism	75	73	75	75	64	57	25	38	18	6
Biosynthesis of cofactors, prosthetic groups, and carrie	82	81	82	81	71	47	23	34	21	6
Cell envelope	69	69	69	69	63	46	31	36	23	12
Cellular process	91	84	91	90	79	64	28	48	31	10
Central intermediary metabolism	57	53	57	57	48	35	20	33	14	6
DNA metabolism	50	49	50	50	44	35	23	38	21	10
Energy metabolism	269	246	269	267	217	159	90	118	80	26
Fatty acid and phospholipid metabolism	35	33	35	35	32	26	15	18	8	1
Hypothetical protein	1275	1193	1275	1272	1027	753	437	582	387	180
Mobile and extrachromosomal element functions	81	70	81	77	65	62	54	51	42	28
Protein fate	62	61	62	62	57	46	29	30	21	12
Protein synthesis	98	84	98	96	67	43	20	34	16	5
Purines, pyrimidines, nucleosides, and nucleotides	42	42	42	41	38	26	13	24	15	6
Regulatory functions	140	135	140	140	126	102	62	82	55	21
Transcription	24	22	24	24	15	11	4	11	8	4
Transport and binding proteins	166	164	166	166	158	118	70	98	66	16
Unclassified	501	478	501	500	444	331	188	246	156	87

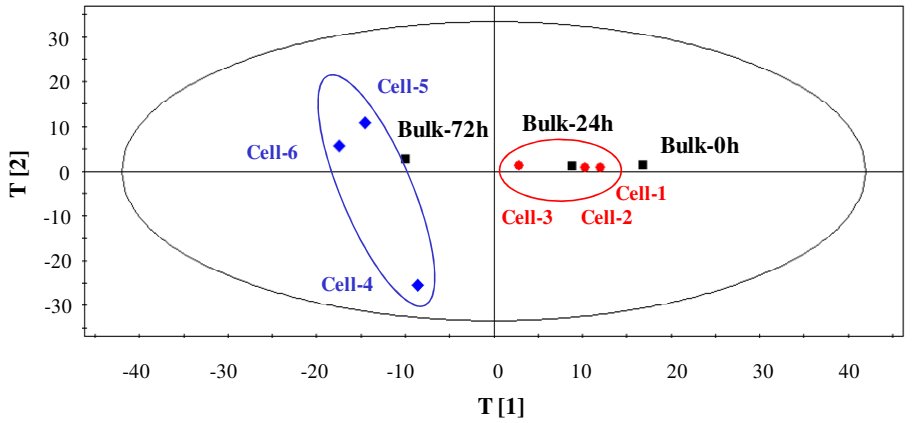
Clustering analysis

PCA analysis

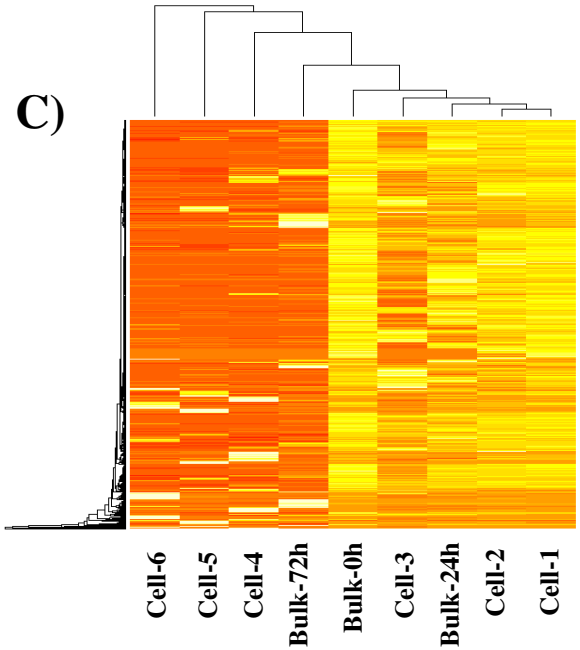
A)



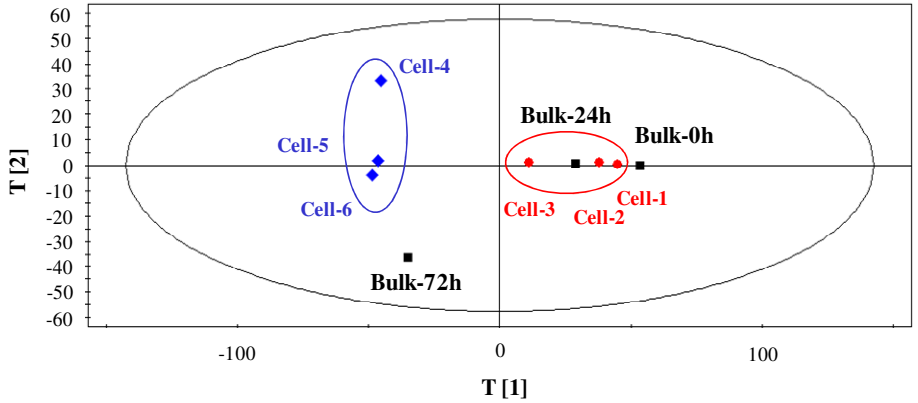
B)



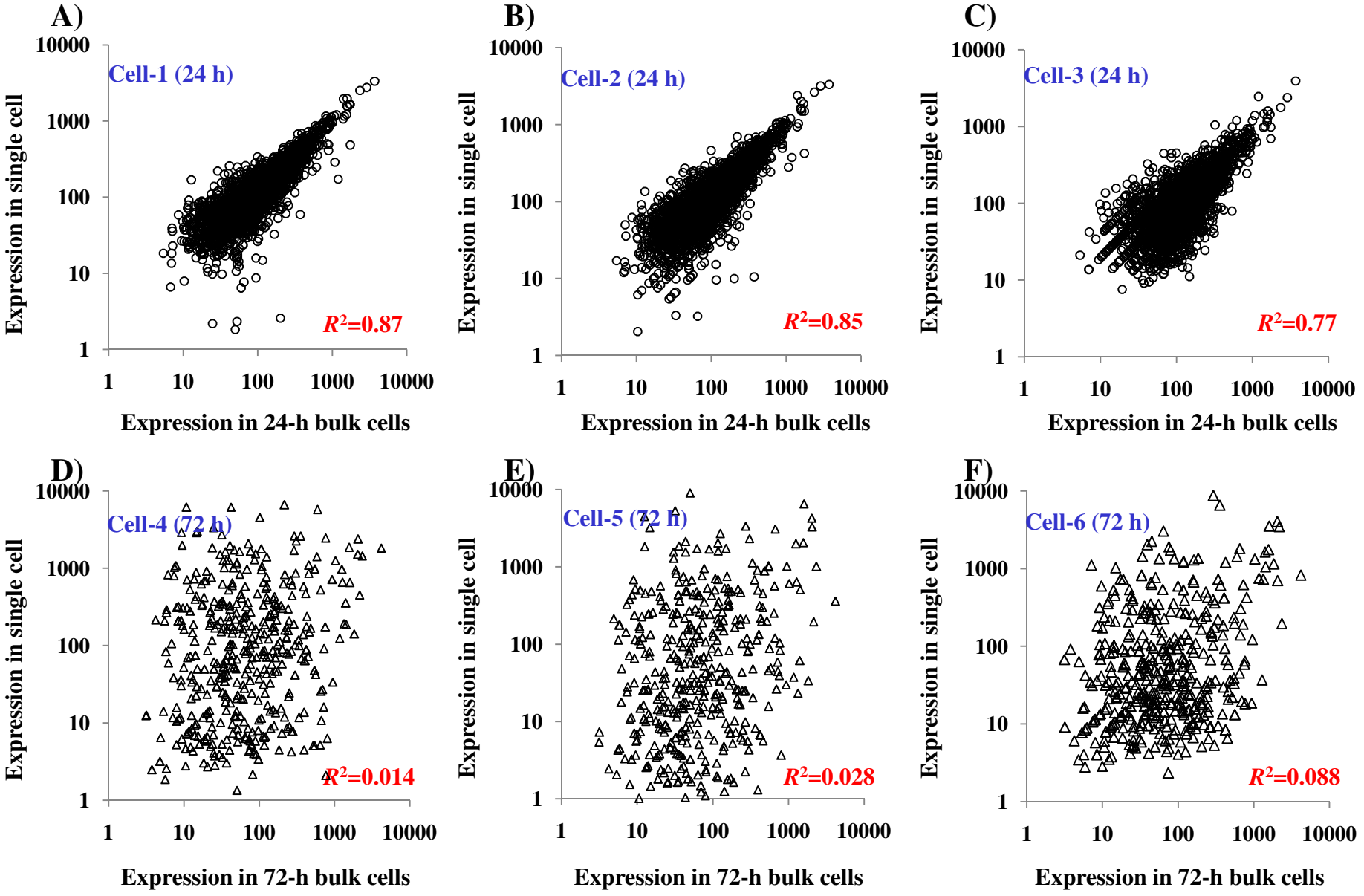
C)



D)



Heterogeneity increase as part of stress response !



Heterogeneity variation among functional categories

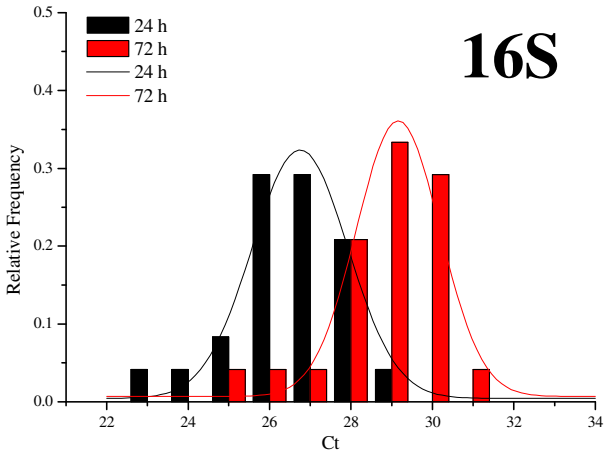
BaSiC RNA-seq	F-test Value	*P* value	Significance category**
24 h cell-1	2.41	0.0018	10*** (9, 18, 15, 7, 6)
24 h cell-2	2.8	0.0003	10 (9, 18, 15, 7, 6, 12, 14)
24 h cell-3	3.24	<0.0001	10 (17, 9, 18, 15, 3, 12, 11, 7, 14, 5, 6, 1)
72 h cell-4	0.71	0.7882	none
72 h cell-5	0.62	0.8641	none
72 h cell-6	0.9	0.575	none
24 h CV * [1]	3.57	<0.0001	10 (9, 15, 18, 17, 7, 12, 4, 6)
72 h CV * [1]	4.22	<0.0001	10 (3, 4, 5, 7, 9, 15, 17, 18)
24 h CV * [2]	6.49	<0.0001	10 (7, 9, 1, 4, 2, 5, 18, 14, 3, 8, 12, 15, 6, 17) 13 (7, 9, 18, 3, 12, 15, 6, 17)
72 h CV * [2]	4.81	<0.0001	10 (1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18)

* CV is equal to standard deviation/average from cell-1,2,3 and cell-4,5,6 for 24 h and 72 h, respectively, "1" indiacted that we used the data of 424 transcripts detected in all single cells; "2" indicated that we used the data of 3117 transcripts detected in at least one single cell.

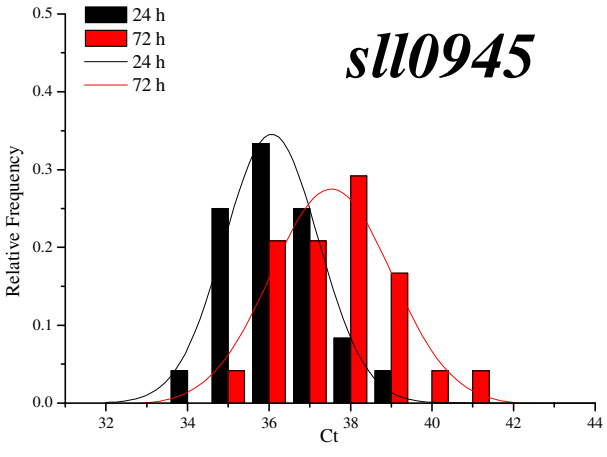
** The functional category shows significantly different from someother categories (in parentheses)

*** Function categories according to KEGG: 1-Amino acid biosynthesis; 2-Biosynthesis of cofactors, prosthetic groups, and carriers; 3-Cell envelope; 4-Cellular processes; 5-Central intermediary metabolism; 6-DNA metabolism; 7-Energy metabolism; 8-Fatty acid and phospholipid metabolism; 9-Hypothetical protein; 10-Mobile and extrachromosomal element functions; 11-No Data; 12-Protein fate; 13-Protein synthesis; 14-Purines, pyrimidines, nucleosides, and nucleotides; 15-Regulatory functions; 16-Transcription; 17-Transport and binding proteins; 18-Unclassified.

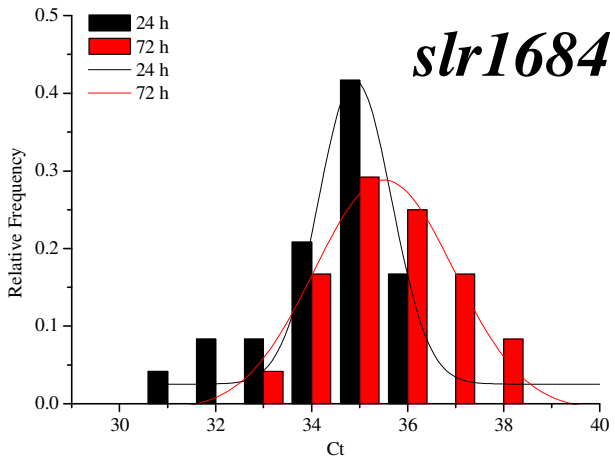
qRT-PCR verification



	Adj. R-Square	sigma
24 h	0.95977	1.17723
72 h	0.95493	1.02529



	Adj. R-Square	sigma
24 h	0.98108	1.15355
72 h	0.92073	1.48065



	Adj. R-Square	sigma
24 h	0.92063	0.76276
72 h	0.97386	1.45617

Heterogeneity increase in "Mobile elements" could be a important driver for cell adaption and evolution!

Wang *et al*., 2014, Genome Research., under review

Summary

- Microbial cell-cell heterogeneity increasingly recognized.
- Two-step qRT-PCR protocol established for analyzing gene expression in single bacterial cells.
- Transcriptomics protocol established for single bacterial cells
- Single-cell transcriptomics reveals increasing heterogeneity upon stress in isogenic cyanobacterial population.

Acknowledgments

Laboratory of Synthetic Microbiology
Tianjin University

Tianjin University "985" Program"

National "973 Program" and "863 program"

National Science Foundation of China