# Transcriptomics of microbial cultures, visualization, and GitHub Iuke.thompson@noaa.gov

thompson@noaa.gov @lukethompson

## <u>Microbial Transcriptomics Example 1</u>

## Gene expression patterns during light and dark infection of Prochlorococcus by cyanophage



## Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism

AS

0

Luke R. Thompson<sup>a,1</sup>, Qinglu Zeng<sup>b,1</sup>, Libusha Kelly<sup>b</sup>, Katherine H. Huang<sup>b</sup>, Alexander U. Singer<sup>c</sup>, JoAnne Stubbe<sup>a,d,2</sup>, and Sallie W. Chisholm<sup>a,b,2</sup>

Departments of <sup>a</sup>Biology and <sup>b</sup>Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; <sup>c</sup>Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario Canada M5G 1L6; and <sup>d</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by James M. Tiedje, Michigan State University, East Lansing, MI, and approved July 29, 2011 (received for review February 15, 2011)

)S	ONE
----	-----

**RESEARCH ARTICLE** 

## Gene Expression Patterns during Light and Dark Infection of *Prochlorococcus* by Cyanophage

Luke R. Thompson<sup>1©¤a</sup>\*, Qinglu Zeng<sup>2©¤b</sup>, Sallie W. Chisholm<sup>1,2</sup>\*

1 Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 2 Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America





## Methods

## Mapping and counting RNA-Seq reads

\* Map to reference genomes using the <u>BWA</u> (Burrows–Wheeler Aligner), generating SAM alignment files.

\* Calculate the number of reads perfectly aligning to the sense and antisense strands of ORFs, rRNAs, tRNAs, and intergenic regions using <u>SAMtools</u> and <u>pysam</u>.

\* Paired reads were mapped separately; paired reads mapping to the same ORF were counted as one transcript for that ORF, whereas paired reads mapping to separate adjacent ORFs were counted once for each ORF.

#### Normalization of phage and host transcript abundance

\* Counts were normalized per sample using the <u>RPKM</u> method (reads per kbp gene length per million reads).

\* Phage transcript counts were normalized to the total of phage plus host transcript counts. Host transcript counts were normalized to the total of host counts only.

#### **Clustering of genes by expression pattern**

\* Phage genes were clustered by transcript relative abundance patterns using two independent approaches: <u>partitioning</u> around medoids (PAM) and hierarchical clustering.

### **Detection of differential gene expression**

\* Differentially expressed host transcripts (messenger RNAs) and antisense RNAs) were identified using the R packages DESeq2 and NOISeq.

\* Transcript abundances were analyzed at each timepoint separately (no proper t=0 control).

\* Lists of differentially expressed genes (DEGs) were the intersection of lists derived from from DESeq2 and NOISeq.

#### Terminology of differential gene expression

\* We avoided the terms "up-regulation" and "down-regulation" because we lack evidence of regulatory mechanisms and our data reflect relative but not absolute abundances. We instead favor the terms "enriched" and "diminished" in reference to relative transcript abundance.

\* In some cases we used the common term "differentially expressed genes" (DEGs); we emphasize that "expression" in this sense refers to transcript abundance only and reflects the net result of transcription minus transcript degradation.







#### Α

Ratio Infected/Uninfected

Light				Dark					
0.5	1.5	2.5	4.5	8.5	0.5	1.5	2.5	4.5	8.5



psbA, Photosystem II protein D1 amt1, Ammonium transporter family *psbK*, Photosystem II protein PsbK Possible ABC transporter, ATP-binding component tRNA-lle1, tRNA-lle-GAT tRNA-Ala1, tRNA-Ala-TGC 5S RNA Plastoquinol terminal oxidase \* Cyanobacterial protein slr0575 Possible reverse transcriptase, RNA-dependent AbrB family transciptional regulator (1) AbrB family transciptional regulator (2) *clpS*, ATP-dependent Clp protease adaptor protein ClpS *hli22*, High light inducbile protein (1) *hli22*, High light inducbile protein (2) Tryptophan-rich protein DUF2389, Ssr2843 homolog Possible major surface glycoprotein *petN*, Cytochrome b6f complex subunit VIII \* *pcyA*, Phycocyanobilin:ferredoxin oxidoreductase \* Bacterial regulatory proteins, Crp family Cytochrome oxidase C subunit VIb-like rpsR, SSU ribosomal protein S18p *rpmG*, LSU ribosomal protein L33p Possible cAMP phosphodiesterases class-II precursor rpsO, SSU ribosomal protein S15p (S13e) COG1939: Ribonuclease III family protein rpsU, SSU ribosomal protein S21p ftsK, Cell division protein FtsK tRNA-Lys1, tRNA-Lys-TTT som, Possible porin *som*, Possible porin *apt*, Adenine phosphoribosyltransferase hli14, High light inducbile protein \* Possible 7kD DNA-binding domain \* Macrophage migration inhibitory factor family isiB, Flavodoxin 1 *tyrA*, Chorismate mutase II *rpmH*, LSU ribosomal protein L34p ftsZ, Cell division protein FtsZ \* rpIU, LSU ribosomal protein L21p *hli11*, High light inducbile protein **\*** hli7, High light inducbile protein \* hli6, High light inducbile protein (1) \* Possible hemagglutinin-neuraminidase Ferredoxin-NADP(+) reductase hli5, High light inducbile protein \* Possible chorismate binding enzyme groES, Heat shock protein 60 family co-chaperone GroES ccdA, Cytochrome c-type biogenesis protein CcdA Possible NDP-sugar pyrophosphorylase tRNA-Ser1, tRNA-Ser-GCT Hypothetical membrane protein \* tRNA-Arg2, tRNA-Arg-CCG \* tRNA-Cys1, tRNA-Cys-GCA Possible conserved carboxylase domain Cvanobacteria-specific RpoD-like sigma factor, typ tig, Cell division trigger factor Photosystem I assembly related protein Ycf37 tRNA-Val2, tRNA-Val-TAC







TPR-repeat protein, specific for cyanobacteria rpaA, Two-component system response regulator *thyX*, Thymidylate synthase *dcd*, Deoxycytidine triphosphate deaminase cob(I)alamin adenosyltransferase glyA, Serine hydroxymethyltransferase Possible LysM domain Plastoquinol terminal oxidase \* Possible spectrin repeat Proline iminopeptidase aroD, 3-dehydroquinate dehydratase II ctaE, Cytochrome c oxidase polypeptide III Single-stranded DNA-binding protein talB, Transaldolase COG3339 Cyanobacteria-specific RpoD-like sigma factor, type-14 Possible RNA recognition motif nrdJ, Ribonucleotide reductase class II (B12-dependent) Protein family PM-1 petN, Cytochrome b6f complex subunit VIII \* pcyA, Phycocyanobilin:ferredoxin oxidoreductase \* Phosphoribosylformylglycinamidine synthase, PurS hli19, High light inducbile protein hli17, High light inducbile protein (1) hli17, High light inducbile protein (2) glnA, Glutamine synthetase type I urtA, Urea ABC transporter, substrate binding protein UrtA *petG*, Cytochrome b6-f complex subunit V Ribosomal RNA small subunit methyltransferase D hli14, High light inducbile protein \* *pntB*, NAD(P) transhydrogenase subunit beta pntA-2, NAD(P) transhydrogenase alpha subunit Possible 7kD DNA-binding domain \* Thioredoxin reductase Possible helix-turn-helix protein, CopG family Permease of drug/metabolite transporter (DMT) superfamily Cyanobacteria-specific RpoD-like sigma factor, type-12 tRNA-Leu3, tRNA-Leu-CAA ftsZ, Cell division protein FtsZ \* Bacterial histone-like DNA-binding protein HAD-superfamily hydrolase, subfamily IA, variant 3 hli11, High light inducbile protein \* hli7, High light inducbile protein \* hli6, High light inducbile protein (1) \* *hli6*, High light inducbile protein (2) Possible hemagglutinin-neuraminidase \* hli5, High light inducbile protein \* adk, Adenylate kinase Hypothetical membrane protein \* tRNA-Arg2, tRNA-Arg-CCG \* thiC, Thiamin biosynthesis protein ThiC *clpX*, Clp protease ATP-binding subunit ClpX Cyanobacteria-specific RpoD-like sigma factor, type-13 sps, Sucrose phosphate synthase

pe-6	bold – photosynthetic electron	(A) Uninf.	-1.6	0.0	+1.6	Infected
	transport or nucleotide					
	biosynthesis	(B) Light	-3.5	0.0	+3.5	Dark
	★ – found in both (A) and (B)	()		log2(fold change)		





B. Host differentially expressed genes – Dark infection





## <u>Microbial Transcriptomics Example 2</u>

## Transcriptional characterization of Vibrio fischeri during colonization of juvenile Euprymna scolopes

Environmental Microbiology (2017) 00(00), 00-00

## Transcriptional characterization of Vibrio fischeri during colonization of juvenile *Euprymna scolopes*

Luke R. Thompson,<sup>1†</sup> Kiel Nikolakakis,<sup>2†</sup> Shu Pan,<sup>2</sup> Introduction Jennifer Reed,<sup>2</sup> Rob Knight<sup>1</sup> and Edward G. Ruby<sup>2,3\*</sup> Bacterial colonization of host organisms has been studied <sup>1</sup>Department of Pediatrics, University of California, in many different model systems, and in the context of San Diego, CA, USA. both pathogenesis and beneficial symbiosis (Bry et al., <sup>2</sup>Department of Chemical and Biological Engineering, 1996; Dedeine et al., 2001; Russell and Rychlik, 2001; University of Wisconsin, Madison, WI, USA. Hongoh, 2010; Gilbert et al., 2012; Nyholm and Graf, <sup>3</sup>Pacific Biosciences Research Center, University of 2012; Bulgarelli et al., 2013; McFall-Ngai et al., 2013; Hawaii, Manoa, HI, USA. Almagro-Moreno et al., 2015; Uzal et al., 2015; Kao et al.,



doi:10.1111/1462-2920.13684

## Methods

### Sequence read processing and mapping

\* Map to reference genomes using the <u>BWA</u> (Burrows–Wheeler Aligner), generating SAM alignment files, then processed with SAMtools.

\* The numbers of reads mapping to protein-coding (CDS) or rRNA genes were calculated using the htseq-count command of <u>HTSeq</u>.

### **Detection of differential gene expression**

\* Differentially expressed host transcripts (messenger RNAs and antisense RNAs) were identified using the R packages DESeq2 and NOISeq.

\* NOISeq was used to filter low counts.

\* DESeq2 was used to detect differential expression. The three conditions (squid-associated, planktonic, and cultured) were contrasted pairwise for all genes, and results exported as Benjamini–Hochberg adjusted p-values and log2(fold change). Differentially expressed genes were identified using tiered cutoffs of these values, with the most stringent cutoff being an adjusted p-value < 0.001 and abs(log2(fold change)) > 3.0(three replicates per condition).

### **RNA-Seq from ribo-depleted, low-biomass samples**

\* We determined the lower limit of ribo-depleted RNA that would produce robust results when following the standard protocol of the TruSeq RNA Sample Preparation Kit (Illumina). \* Using a single sample of V. fischeri total RNA, we made TruSeq libraries from three amounts of non-ribo-depleted total RNA (1000, 500, and 100 ng), and nine amounts of ribodepleted RNA (1000, 500, 100, 50, 25, 10, 5, 2.5, and 1 ng). \* Ribo-depletion using the Ribo-Zero Gold Epidemiology Kit before library prep reduced the percentage of rRNA in the sample from  $\sim 90\%$  to  $\sim 1\%$ .

\* Ribo-depletion did not appreciably affect the relative abundance of individual mRNAs detected. DESeq2 did not identify any genes significantly differentially abundant between ribo-depleted and non-ribo-depleted samples (FDR-adjusted p-value < 0.05).

\* Genes with the lowest relative mRNA abundance (across all samples) were those most likely to be undetected in the lowinput RNA samples.

\* Input total RNA could be reduced to 50 ng without loss of coverage, reduced further to 10 ng without loss of mRNA relative abundance fidelity, reduced further to 2.5 to 5 ng with ~10% reduction in coverage and mRNA relative abundance fidelity.







A









Thompson, L. R., Zeng, Q. & Chisholm, S. W. Gene expression patterns during light and dark infection of *Prochlorococcus* by cyanophage. *PLOS ONE* 11, e0165375 (2016). doi:10.1371/ journal.pone.0165375

Thompson, L. R., Nikolakakis, K., Pan, S., Reed, J., Knight, R. & Ruby, E. G. Transcriptional characterization of Vibrio fischeri during colonization of juvenile Euprymna scolopes. *Environ Microbiol* 19, 1845–1856 (2017). doi:10.1111/1462-2920.13684



## @lukethompson