



## Data in Brief

# RNA-Seq analysis and whole genome DNA-binding profile of the *Vibrio cholerae* histone-like nucleoid structuring protein (H-NS)



Julio C. Ayala<sup>a,b</sup>, Hongxia Wang<sup>b,1</sup>, Jorge A. Benitez<sup>b,\*</sup>, Anisia J. Silva<sup>b,\*</sup>

<sup>a</sup> University of Alabama at Birmingham, Department of Microbiology, Birmingham, AL, United States

<sup>b</sup> Morehouse School of Medicine, Department of Microbiology, Biochemistry and Immunology, Atlanta, GA, United States

## ARTICLE INFO

## Article history:

Received 19 May 2015

Received in revised form 26 May 2015

Accepted 28 May 2015

Available online 10 June 2015

## Keywords:

*Vibrio cholerae*

Cholera

H-NS

Nucleoid associated protein

Histone-like nucleoid structuring protein

## ABSTRACT

The data described in this article pertain to the genome-wide transcription profiling of a *Vibrio cholerae* mutant lacking the histone-like nucleoid structuring protein (H-NS) and the mapping of the H-NS chromosome binding sites (Wang et al., 2015 [1]; Ayala et al., 2015 [2]). H-NS is a nucleoid-associated protein with two interrelated functions: organization of the bacterial nucleoid and transcriptional silencing (Dorman, 2013 [3]). Both functions require DNA binding and protein oligomerization (Spurio et al., 1997[4]; Dame et al., 2001 [5]). H-NS commonly silences the expression of virulence factors acquired by lateral gene transfer (Navarre et al., 2006 [6]). The highly pleiotropic nature of *hns* mutants in *V. cholerae* indicates that H-NS impacts a broad range of cellular processes such as virulence, stress response, surface attachment, biofilm development, motility and chemotaxis. We used a *V. cholerae* strain harboring a deletion of *hns* and a strain expressing H-NS tagged at the C-terminus with the FLAG epitope to generate datasets representing the *hns* transcriptome and DNA binding profile under laboratory conditions (LB medium, 37 °C). The datasets are publicly available at the Gene Expression Omnibus (GEO) repository (<http://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE62785 and GSE64249.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications

Organism/cell line/tissue	<i>Vibrio cholerae</i> serogroup O1 Ogawa, El Tor biotype
Sex	Not applicable
Sequencer or array type	Illumina
Data format	Raw data: bam files; analyzed data: xls files.
Experimental factors	Cultures of the wild type bacterium and its isogenic mutant containing a deletion of the gene encoding the histone-like nucleoid structuring protein (H-NS); cultures of the wild type bacterium expressing H-NS tagged with the FLAG epitope.
Experimental features	Differential RNA-Seq: DNA-free total RNA was extracted from wild type and <i>hns</i> mutant; the samples were depleted of ribosomal RNA and the remaining mRNA was subsequently converted to cDNA. Libraries were prepared using TruSeq technology and standard paired-end sequencing reactions were performed. Chromatin Immunoprecipitation (ChIP)-Seq: DNA from cultures of <i>V. cholerae</i> expressing H-NS-FLAG was immunoprecipitated with the anti-FLAG M2 monoclonal antibody. Libraries for Next

(continued)

Specifications	Generation Sequencing were prepared using the Illumina TruSeq ChIP kit. Short 50 bp paired-end reads were obtained using the Illumina HiSeq 2000 system.
Consent	Not applicable
Sample source location	Morehouse School of Medicine, Atlanta, GA, United States of America.

## 1. Direct link to deposited data

The raw and processed RNA-Seq datasets have been deposited in the Gene Expression Omnibus (GEO) repository (<http://www.ncbi.nlm.nih.gov/geo/>) and assigned accession number GSE62785. The ChIP-Seq raw and processed data files are accessible through the GEO series accession number GSE64249.

## 2. Value of the data

- Provides a compilation of genes positively and negatively regulated by H-NS in *V. cholerae* serogroup O1 of the El Tor biotype in mid-exponential and early stationary growth phases [1,2].

\* Corresponding authors.

E-mail addresses: [jbenitez@msm.edu](mailto:jbenitez@msm.edu) (J.A. Benitez), [asilva-benitez@msm.edu](mailto:asilva-benitez@msm.edu) (A.J. Silva).

<sup>1</sup> Current address: State Key Laboratory for Infectious Disease Prevention and Control and National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing, China.

**Table 1**  
Summary of computational tools used to analyze RNA-Seq and ChIP-Seq data.

Software	Purpose	Location	URL
TopHat/Bowtie	RNA-Seq: alignment of spliced reads to the reference genome (NC_002505.1 and NC_002506.1).	Galaxy Tool Shed	<a href="https://galaxy.uabgrid.uab.edu">https://galaxy.uabgrid.uab.edu</a>
Cufflinks	RNA-Seq: generation of differential gene expression reports.	Galaxy Tool Shed	<a href="https://galaxy.uabgrid.uab.edu">https://galaxy.uabgrid.uab.edu</a>
CASAVA v1.8.2	ChIP-Seq: conversion of .bcl files into .fastq.gz files.	Illumina web site	<a href="http://support.illumina.com/sequencing/sequencing_software/casava.html">http://support.illumina.com/sequencing/sequencing_software/casava.html</a>
Bowtie	ChIP-Seq: alignment of short reads to the reference genome to generate .SAM files.	Galaxy Tool Shed	<a href="https://galaxy.uabgrid.uab.edu">https://galaxy.uabgrid.uab.edu</a>
SAM tools	ChIP-Seq: filtering and removal of unmapped reads and conversion of .SAM to .bam files.	Galaxy Tool Shed	<a href="https://galaxy.uabgrid.uab.edu">https://galaxy.uabgrid.uab.edu</a>
SeqMonk v0.27.0	ChIP-Seq: peak calling using the MACS algorithm [9]; quantitation using the Read Count Quantitation algorithm and annotation of H-NS binding sites.	Babraham Institute	<a href="http://www.bioinformatics.babraham.ac.uk/projects/seqmonk">http://www.bioinformatics.babraham.ac.uk/projects/seqmonk</a>

- Characterizes the clustering of H-NS occupancy along *V. cholerae* chromosomes I and II in mid-exponential growth phase [2].
- Provides a basis to investigate the role of H-NS in the regulation of multiple cellular processes [1,2].
- Provides a basis to investigate (i) the role of H-NS in the organization of the nucleoid in a two-chromosome bacterium and (ii) the link between nucleoid architecture and environmental regulation of gene expression [3–6].

### 3. Experimental design, materials and methods

#### 3.1. Objective of the experiment

Our objective was to investigate H-NS global regulation of gene expression with emphasis on cellular functions affecting bacterial virulence, stress response, motility and chemotaxis.

#### 3.2. Strains and growth conditions

Strain C7258ΔlacZ is a derivative of C7258 (clinical isolate from a 1991 Perú outbreak, serogroup O1 Ogawa, El Tor biotype) lacking endogenous β-galactosidase activity [7]. Strain AJB80 contains a kanamycin resistance cassette replacing the chromosomal *hns* gene of C7258ΔlacZ [7]. Strain C7258H-NS-FLAG is a derivative of C7258 in which an *hns-flag* allele was integrated by homologous recombination into the *hns* locus [8]. Strain AJB80 is resistant to kanamycin (50 μg/mL). Strain C7258H-NS-FLAG expresses H-NS-FLAG from native transcription and translation initiation signals and is resistant to ampicillin (100 μg/mL). Strains were grown in 50 mL of LB medium (pH 7.4) with the appropriate antibiotic at 37 °C in a New Brunswick C24 orbital shaker set to 250 rpm. For RNA-Seq, cells were collected by centrifugation at an optical density of 0.5 (mid-exponential growth phase) and 2.0 (early stationary phase) at 600 nm (OD<sub>600</sub>). For ChIP, cells were grown to OD<sub>600</sub> 0.5 and 40 mL of culture was sequentially

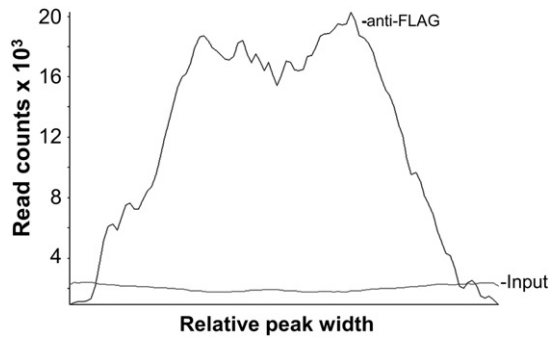
treated with rifampicin (150 μg/mL, 20 min, 37 °C), 1% formaldehyde (cross-linking, 10 min, 30 °C), and 227 mM glycine (quenching, 30 min, 4 °C). Cells were collected by centrifugation, washed twice with phosphate-buffered saline pH 7.4 (PBS) supplemented with protease inhibitor cocktail (PIC) and phenylmethylsulfonyl fluoride (PMSF; Roche Applied Science), divided into aliquots equivalent to 1/(OD<sub>600</sub> reading) mL and the cell pellets were maintained at –80 °C if not processed immediately.

#### 3.3. Total RNA extraction and removal of ribosomal RNA

Cultures of strains C7258ΔlacZ and AJB80 were divided into 5 mL aliquots and the cells were harvested by centrifugation at 4000 ×g for 10 min at room temperature. The pellets were resuspended in 5 mL of RNAlater (Invitrogen) and agitated on a rotator for 10 min at room temperature. The cells were collected by centrifugation at 4000 ×g for 10 min and resuspended in 5 mL of RNAlater. Then, cell pellets corresponding to 1 mL aliquots were collected by centrifugation for 10 min at 4000 ×g. Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. RNA integrity was determined by formaldehyde agarose gel electrophoresis and the RNA was stored at –80 °C. Contamination with DNA was further eliminated using the TURBO DNA-free kit (Invitrogen), which involves a second treatment with DNase I for 30 min at 37 °C. Reactions were terminated by addition of 0.2 volumes of the DNase inactivation reagent and RNA was purified using the Agencourt RNAClean XP kit (Beckman) following the manufacturer's instructions. Total RNA was eluted in 60 μL of RNase-free water. The absence of DNA contamination was confirmed by PCR with 16S ribosomal DNA specific primers. To this end, 2 μL of the DNase treated RNA was added to each reaction in a final reaction volume of 20 μL. Each reaction was run in parallel to *Escherichia coli* DNA (positive control) and nuclease-free water (negative control) with the following cycling conditions: 95 °C for 1 min, 30 cycles of 95 °C for 30 s, 50 °C for 30 s, and 68 °C for 1 min. Next, ribosomal RNA (rRNA) was removed using Ribo-Zero Magnetic Kit (Epicentre) following the manufacturer's instructions. Briefly, 6 μg of total RNA sample and

**Table 2**  
TopHat statistics of the quality control of .fastq files used for alignment.

Analysis	Sample (GEO accession #)	Forward sequencing reads (overall alignment rate)	Reverse sequencing reads (overall alignment rate)
RNA-Seq OD <sub>600</sub> 0.5	Wild type sample 1 (GSM1533384)	67,772,927 (81.4%)	67,729,235 (78.7%)
	Δhns sample 1 (GSM1533385)	45,204,039 (81.6%)	45,181,086 (79.6%)
	Wild type sample 2 (GSM1533386)	39,404,653 (79.7%)	39,388,107 (78.1%)
	Δhns sample 2 (GSM1533387)	40,998,335 (80.2%)	40,981,501 (78.4%)
RNA-Seq OD <sub>600</sub> 2.0	Wild type sample 1 (GSM1533388)	35,001,271 (76.6%)	34,973,987 (75.6%)
	Δhns sample 1 (GSM1533389)	43,629,892 (75.9%)	43,605,111 (74.9%)
	Wild type sample 2 (GSM1533390)	36,404,561 (72.6%)	36,375,868 (71.8%)
	Δhns sample 2 (GSM1533391)	38,798,568 (74.4%)	38,776,510 (73.3%)
ChIP-Seq OD <sub>600</sub> 0.5	Anti-FLAG (GSM1567051)	57,236,395 (63.0%)	57,236,395 (63.0%)
	Input control (GSM1567052)	58,286,380 (84.2%)	58,286,380 (84.2%)



**Fig. 1.** Average shape of H-NS peaks across the genome. The average peak height and width corresponding to H-NS occupied sites (anti-FLAG) and the ChIP-Seq negative control sample (input) was computed using the SeqMonk Probe Trend Plot function. The magnitude of the plots is given in sequence read counts along the y axis. The diagram shows the relative enrichment in read counts of the immunoprecipitated DNA sample compared to the DNA shearing control.

20  $\mu$ L of Ribo-Zero rRNA Removal Solution were combined in a final reaction volume of 80  $\mu$ L. Samples in Ribo-Zero rRNA removal solution were incubated at 68  $^{\circ}$ C for 10 min followed by a 5 min incubation at room temperature. To remove the rRNA molecules from the mRNA, reaction mixtures were incubated with the magnetic beads provided in the kit, mixed and placed at room temperature for 5 min and then at 50  $^{\circ}$ C for 5 min. The rRNA bound to the beads was then removed by magnetic separation. Finally, mRNA was purified using Agencourt RNAClean XP kit and eluted in 15  $\mu$ L of RNase-free water.

#### 3.4. Differential RNA-Seq and data analysis

RNA-Seq was performed on the Illumina HiSeq 2000 system using the latest versions of sequencing reagents and flow cells providing up to 300 GB of sequence information per cell. The quality of the RNA was assessed using the Agilent 2100 Bioanalyzer and samples were subsequently converted to cDNA. Libraries were constructed using the TruSeq library generation kits as per the manufacturer's instructions (Illumina, San Diego, CA). The cDNA libraries were quantitated using qPCR in a Roche LightCycler 480 with the Kapa Biosystems kit for library quantitation (Kapa Biosystems, Woburn, MA) prior to cluster generation. Clusters were generated to yield approximately 725 K–825 K clusters/ $\text{mm}^2$ . Paired-end 50 bp sequencing runs were conducted to align the cDNA sequences to the reference genome. The TopHat software and the short read aligner Bowtie were used to align the raw RNA-Seq

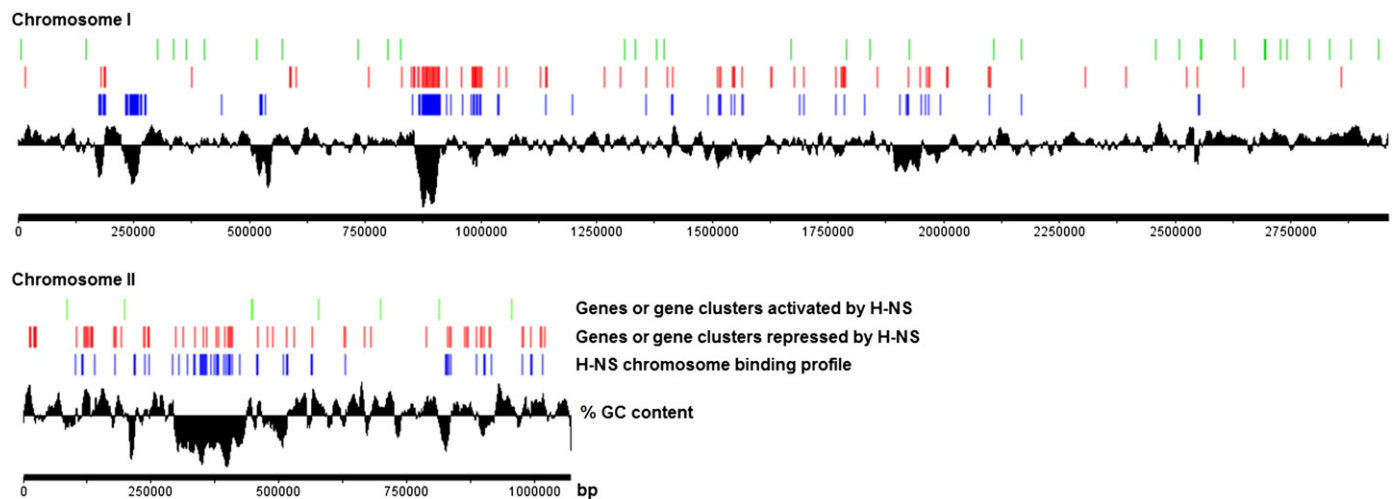
fastq reads to the reference genome (Tables 1 and 2). Transcripts assembly, abundance and evaluation of differential expression were accomplished using the Cufflinks software. Genes exhibiting a fold change  $\geq \pm 2.0$  and q-value  $< 0.05$  were considered differentially expressed in the *hns* mutant.

#### 3.5. Chromatin immunoprecipitation

Cells were lysed by suspending the frozen pellets in 500  $\mu$ L of 10 mM Tris-HCl, pH 8.0, 50 mM NaCl containing 20 ng/ $\mu$ L of RNase A, and  $10^5$  kU of Ready-Lyse lysozyme (Epicentre Biotechnologies), followed by a 30 min incubation at 37  $^{\circ}$ C. One volume of double-strength immunoprecipitation (IP) buffer (200 mM Tris-HCl, pH 7.5, 600 mM NaCl, 4% Triton X-100) containing PIC and PMSF was added to each lysate, and DNA was broken down to a range of 150 to 1000 bp by sonication. The cell debris was removed by centrifugation, and 100  $\mu$ L of cleared lysate was diluted 10-fold in IP buffer. Protein-DNA complexes were immunoprecipitated by overnight incubation at 4  $^{\circ}$ C with 8  $\mu$ g of anti-FLAG M2 monoclonal antibody (Sigma-Aldrich). The antibody-protein-DNA complexes were pulled down with salmon sperm DNA-treated protein A/G agarose beads (Imgenex, San Diego, CA) for 1 h at 4  $^{\circ}$ C. The beads were washed twice with 100 mM Tris-HCl (pH 7.5), 250 mM LiCl, 2% Triton X-100, collected in Spin-X centrifuge tube filters (Costar), and washed three times with each of the following buffers: IP buffer containing 600 mM NaCl, IP buffer, and TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The immunoprecipitated complexes were eluted from the beads by incubation at 65  $^{\circ}$ C for 30 min in TE buffer containing 1% SDS. After reversal of cross-linking (4 h, 65  $^{\circ}$ C), proteins were removed by incubation with 20  $\mu$ g of proteinase K (1 h, 45  $^{\circ}$ C). The DNAs immunoprecipitated from 9 reactions were combined, purified using a MinElute PCR purification kit (Qiagen) and eluted in 30  $\mu$ L of DNase-free water. The quality of the immunoprecipitated DNA was assessed by quantifying the enrichment of a DNA known to contain an H-NS binding site [8] over a mock sample obtained by using the IgG1 isotype control G3A1 mAb (Cell Signaling Technology) for IP. This was done using the iTaq<sup>TM</sup> Universal SYBR<sup>®</sup> Green Supermix kit and a CFX96 Touch<sup>TM</sup> Real-Time PCR Detection System (Bio-Rad) and specific primers for the *tcpA* (positive control) and VC1922 (negative control) promoters.

#### 3.6. ChIP-Seq and data analysis

The DNA concentration of the immunoprecipitated (7.2 ng/ $\mu$ L) and input (a 10  $\mu$ L sample taken prior to IP) (23.0 ng/ $\mu$ L) samples were



**Fig. 2.** Graphic representation of RNA-Seq (GSE62785) and ChIP-Seq (GSE64249) datasets. Lineal representation of *V. cholerae* chromosomes showing the H-NS-regulated genes at OD<sub>600</sub> 0.5 in LB medium determined by RNA-Seq and H-NS binding regions determined by ChIP-Seq. The numbering indicates the chromosome positions in base pairs (bp). The diagram was prepared using the software DNAPlotter for genome visualization [10].

determined using the Qubit® Fluorometric Quantitation System (Life Technologies). The anti-FLAG immunoprecipitated DNA and input DNA were used for library preparation for Illumina sequencing using the TruSeq ChIP kit (Illumina, San Diego, CA) following the manufacturer's protocol. Briefly, 20 ng of immunoprecipitated DNA samples was end repaired-ligated to Illumina adaptors and selected for a fragment size of approximately 300 bp by gel extraction. Multiplex Illumina primers were used to amplify gel-extracted products. The amplified ChIP-Seq libraries were quantitated by qPCR and loaded to a concentration of 2.5 pM per lane in the Illumina HiSeq 2000 platform. A standard paired-end sequencing reaction was performed to generate 50 bp of sequence in each direction. The raw data was converted from .bcl file format to .fastq format for downstream analysis. This was done using CASAVA v1.8.2 software from Illumina. The raw fastq data files were aligned to the *V. cholerae* El Tor N16961 reference genome (NC\_002505.1 and NC\_002506.1) using Bowtie, an ultrafast memory-efficient short read aligner available on the Galaxy web-based platform. The generated sam data files were filtered to remove unmapped reads and converted to bam data files using the Galaxy Next Generation Sequencing (NGS): SAM Tools. SeqMonk v0.27.0 was used for peak calling, quantitation and annotation. The MACS (Model-Based Analysis for ChIP-Seq) algorithm within SeqMonk was used for peak calling with p-value  $1.0E - 05$ , a sonicated fragment size of 300 bp and input sample as control. The Read Count Quantitation algorithm within SeqMonk was used for peak quantitation. The average enrichment of sequence reads in the immunoprecipitated sample over the input sample is shown in Fig. 1. A graphic representation of the RNA-Seq and ChIP-Seq datasets is shown in Fig. 2.

#### Conflict of interest

The authors state that there are no conflicts of interest.

#### Acknowledgments

This study was supported by awards 5SC1AI104993-03 (to AJS), 5R21AI103693-03 (to JAB) and F31AI106288 (to JCA) from the National Institutes of Health. We are grateful to Michael Crowley, Mei Han and David Crossman from the University of Alabama at Birmingham Heflin Center for Genomic Sciences for their assistance in NGS methods.

#### References

- [1] H. Wang, et al., RNA-Seq analysis identifies new genes regulated by the histone-like nucleoid structuring protein (H-NS) affecting *Vibrio cholerae* virulence, stress response and chemotaxis. *PLoS ONE* 10 (2) (2015) e0118295.
- [2] J.C. Ayala, et al., Repression by H-NS of genes required for the biosynthesis of the *Vibrio cholerae* biofilm matrix is modulated by the second messenger cyclic diguanylic acid. *Mol. Microbiol.* (2015) <http://dx.doi.org/10.1111/mmi.13058>.
- [3] C.J. Dorman, Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat. Rev. Microbiol.* 11 (5) (2013) 349–355.
- [4] R. Spurio, et al., The oligomeric structure of nucleoid protein H-NS is necessary for recognition of intrinsically curved DNA and for DNA bending. *EMBO J* 16 (7) (1997) 1795–1805.
- [5] R.T. Dame, C. Wyman, N. Goosen, Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* 83 (2) (2001) 231–234.
- [6] W.W. Navarre, et al., Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* 313 (5784) (2006) 236–238.
- [7] A.J. Silva, et al., Role of the histone-like nucleoid structuring protein in the regulation of *rpoS* and *RpoS*-dependent genes in *Vibrio cholerae*. *J. Bacteriol.* 190 (22) (2008) 7335–7345.
- [8] H. Wang, et al., Interaction of the histone-like nucleoid structuring protein and the general stress response regulator *RpoS* at *Vibrio cholerae* promoters that regulate motility and hemagglutinin/protease expression. *J. Bacteriol.* 194 (5) (2012) 1205–1215.
- [9] Y. Zhang, et al., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9 (9) (2008) R137.
- [10] T. Carver, et al., DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25 (1) (2009) 119–120.