GadEWX regulons DESCRIPTION DATA FILE

1. GENERAL INFORMATION.

Title: GadEWX regulons.

Description of the dataset:

GadEWX dataset from ChIP-exo and RNA-seq analysis.

Confidence:

ChIP analysis and statistical validation of TFBSs (CHIP-SV) and Mapping of signal intensities by RNA-seq (MSI).

Reference:

Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in *Escherichia coli*. Seo SW, Kim D, O'Brien EJ, Szubin R, Palsson BO. Nat Commun. 2015 Aug 10;6:7970. doi: 10.1038/ncomms8970. PMID: 26258987

Citation:

Dataset provided and maintained by RegulonDB (PUBMED: #18158297) from the original source published in*:* (PUBMED: #26258987)

Contact person for this dataset:

Questions concerning the content of the data set that are raised by users of RegulonDB will be forwarded to this person. We would appreciate receiving a copy of the response to the user, so we can keep track of taking care of user requests.
Person: RegulonDB staff
Email address: regulondb@ccg.unam.mx

Licensing:

See the license of RegulonDB in:

http://regulondb.ccg.unam.mx/LicenseRegulonDBWithoutSing.jsp

2. DATASET DESCRIPTION.

Summary:

GadEWX regulons consist of 45 target genes in 31 TUs (GadE: 20 genes in 15 TUs; GadW: 6 genes in 4 TUs; GadX: 29 genes in 19 TUs, several genes are co-regulated) and 28 of these genes were associated with RpoS-binding sites. ChIP-exo was also used to determine the *in vivo* binding profile of RpoS under acid stress.

To determine the causal relationships between the binding of GadEWX and changes in RNA transcript levels of genes in the GadEWX regulons, transcript levels between the wild-type strain and that of each deletion mutant (*gadE*, *gadW* and *gadX*) grown under acidic stress conditions were compared.

Experiment:

ChIP-exo.

Based on Rhee, H. S. & Pugh, B. F. ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Curr. Protoc. Mol. Biol. Chapter 21, 21–24 (2012), with following modifications as shown in the previous study: Seo, S. W. et al. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. Nat. Commun. 5, 4910 (2014). MACE program (https://code.google.com/p/ chip-exo/) was used to define peak candidates from biological duplicates with sequence depth normalization Wang, L. et al. MACE: model based analysis of ChIP-exo. Nucleic Acids Res. 42, e156 (2014). To reduce false-positive peaks Mock-IP was used: were removed as in the previous study: Seo, S. W. et al. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. Nat. Commun. 5, 4910 (2014).

RNA-seq expression profiling.

Based on: Levin, J. Z. et al. Comprehensive comparative analysis of strand-specific RNA sequencing methods. Nat. Methods 7, 709–715 (2010), with following modifications as shown in the previous study: Seo, S. W. et al. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in *Escherichia coli*. Nat. Commun. 5, 4910 (2014).

Motif search and analysis.

The GadEWX-binding motif analyses were completed using the MEME tool from the MEME software suite with default settings: Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208 (2009). The sequence of each binding site by 10–30 bp at each end to allow for adjacent sequences to be included in the analysis was extended.

COG functional enrichment.

The GadEWX regulons were categorized according to their annotated COG category.

Acid resistance assays.

AR assays followed the procedures of previous study with modification of usage of growth media: Mates, A. K., Sayed, A. K. & Foster, J. W. Products of the *Escherichia coli* acid fitness island attenuate metabolite stress at extremely low pH and mediate a cell density-dependent acid resistance. J. Bacteriol. 189, 2759–2768 (2007).

Conservation analysis of GadEWX regulons.

Gene annotation of strains and species: (http://theseed.org) and ortholog calculation: (Rapid Annotation using Subsystem Technology). Conservation level of gadEWX and genes in GadEWX regulons were calculated from orthologs retained from RAST output.

More details for each Methods, see: Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in *Escherichia coli*. Seo SW, Kim D, O'Brien EJ, Szubin R, Palsson BO. Nat Commun. 2015 Aug 10;6:7970. doi: 10.1038/ncomms8970. PMID: 26258987

Methods:

Version of programs:

Not apply.

Version of datasets:

Not apply.

Protocol or algorithm

See section Experiment.

Specificity and sensitivity:

Not apply.

3. COLUMN FORMAT OF DATA FILE BY

Version of *E.coli´s* Genome: *E. coli* K-12 MG1655 and its derivatives. The version used is not mentioned.

Sequence Identifier:

Suggested column format of the data file:

1) Transcription Factor Name

2) Gene or operon regulated by the TF (regulated gene)

3) Regulatory effect of the TF on the regulated gene (+ activator, - repressor, +- dual, nd not determined)

4) ChIP-exo Start

5) ChIP-exo End

6) Distance to TSS (Transcription Start Site) or C.P. Central position of the TF binding-site

7) Growth condition of the experimental procedures.

8) Evidence that supports the existence of the regulatory interaction

9) PMID Reference(s)

10) RpoS-dependent. RpoS-depedent. Red characters indicate novel RpoS-binding determined by ChIP-exo.

11) S/N ratio value. The signal-to-noise (S/N) ratios of the OxyR-, SoxR-, and SoxS-binding peaks and used them as a proxy of the in vivo binding intensity of each binding site.

12) TFBSs in RegulonDB