PNAS

Supporting Information for:

"Expanding the genetic toolkit helps dissect a global stress response in the early-branching species *Fusobacterium nucleatum*"

Falk Ponath, Yan Zhu, Valentina Cosi, Jörg Vogel

Corresponding author: Jörg Vogel
Email: joerg.vogel@uni-wuerzburg.de

This PDF file includes:

        Figures S1 to S8
        Material and Methods
        Dataset S1 to S5 (separate files)
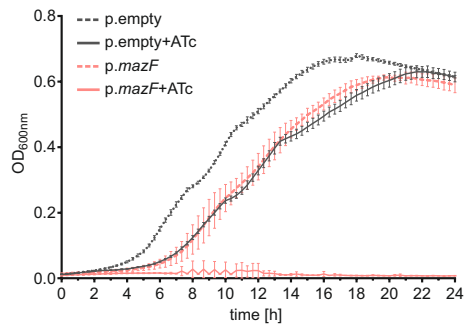   SI references

Figure S1

Figure S1: Testing of *mazF* toxicity during growth in the presence of selection pressure.

Growth curve for *F. nucleatum* carrying either p. empty or p.*mazF* in the presence or absence of 100 ng ml^{-1} ATc. Thiamphenicol was added to select for plasmid maintenance. Shown is the optical density (OD$_{600nm}$) over time (h, hours) for three biological replicates and the standard deviation.
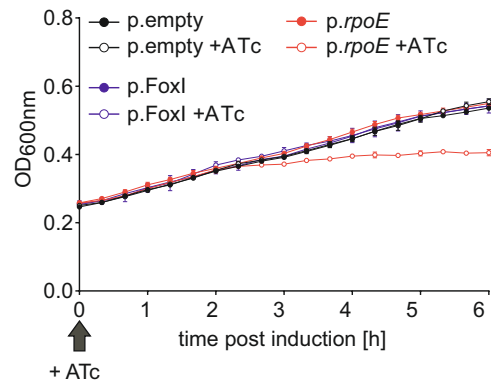
Figure S2

Figure S2: Impact of *rpoE* or FoxI inducible expression on growth.

Growth curve for *F. nucleatum* carrying either p. empty, p.*rpoE* or p.FoxI in the presence or absence of 100 ng ml^{-1} ATc. The time point of ATc addition is indicated. Shown is the optical density (OD$_{600nm}$) over time (h, hours) for three biological replicates and the standard deviation.
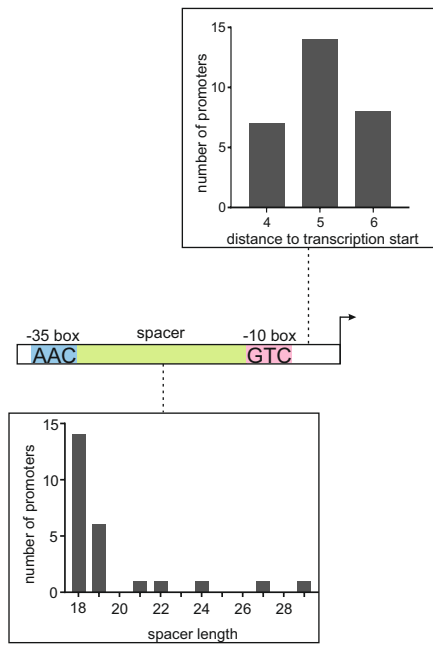
Figure S3

Figure S3: Analysis of spacing for the *rpoE* promoter motif.

The bar charts display the distance between the -10 and -35 boxes (bottom) as well as from the -10 box to the TSS (top).

Numbers of associated promotors are plotted on the Y-axis.
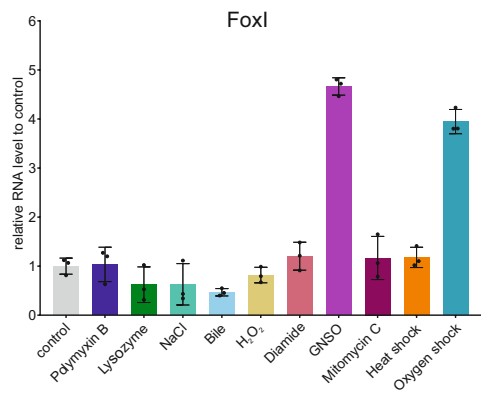
Figure S4

Figure S4: Quantification of FoxI levels upon exposure to different stress conditions.

FoxI signal detected via northern blot was quantified for bacteria exposed to the indicated stress conditions. The average of three biological replicates relative to that of the control is displayed together with the standard deviation.
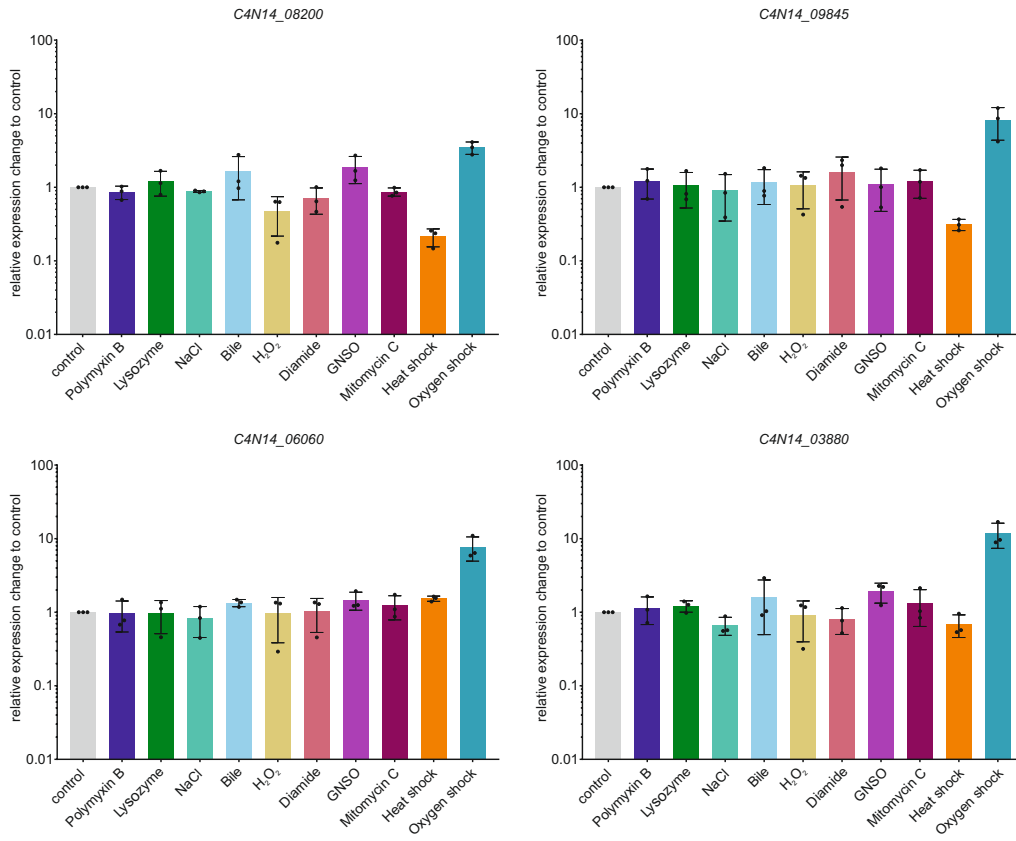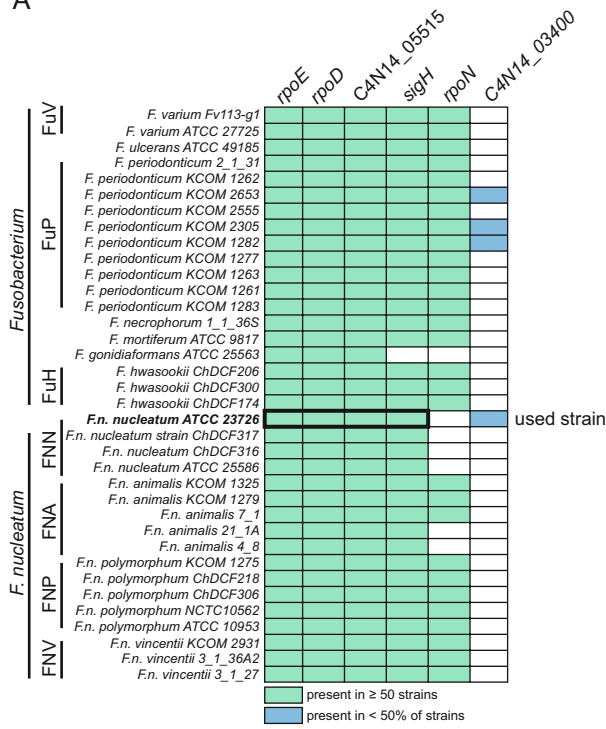
Figure S5

Figure S5: RT-qPCR analysis for additional genes upon exposure to different stress conditions.

RT-qPCR analysis for mRNA of the indicated genes after exposing *F. nucleatum* to indicated stress conditions for 60 min. Data are normalized to the control and plotted as the average of three biological replicates with the standard deviation.

Figure S6

A

legend:
- present in ≥ 50 strains
- present in < 50% of strains

B

Figure S6: Response of sigma factors to the different stress conditions.

(A) Schematic representation of the presence or absence of annotated sigma factors in species of *F. nucleatum* and members of the *Fusobacterium* genus. White fields mark the absence of the sigma factor. The black box indicates all conserved sigma factors investigated in this study.(B) qRT-PCR analysis for mRNA of the indicated genes after exposing *F. nucleatum* to indicated stress conditions for 60 min. Data are normalized to the control and plotted as the average of three biological replicates with the standard deviation.
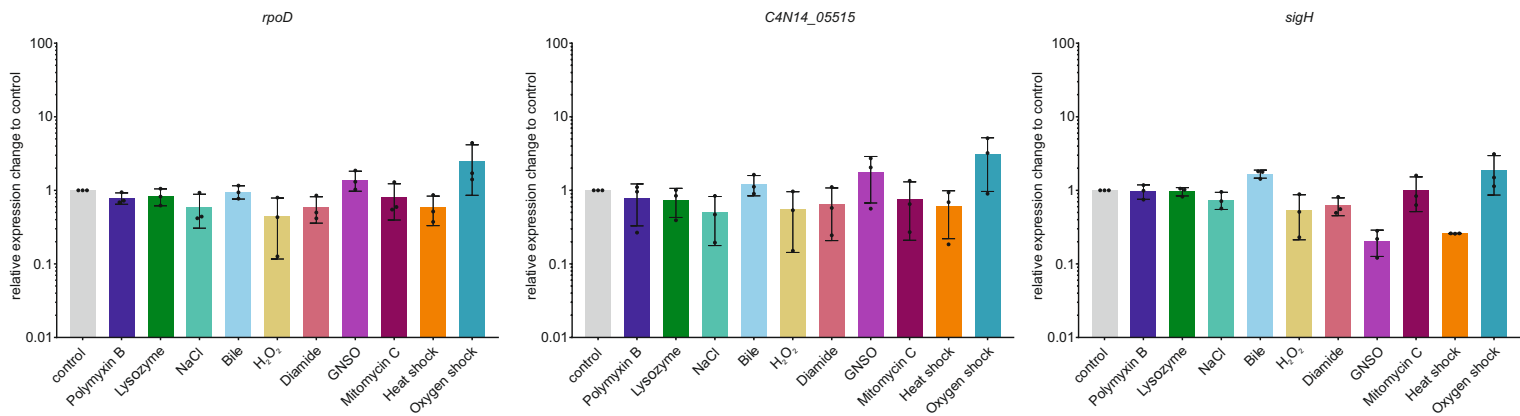
Figure S7
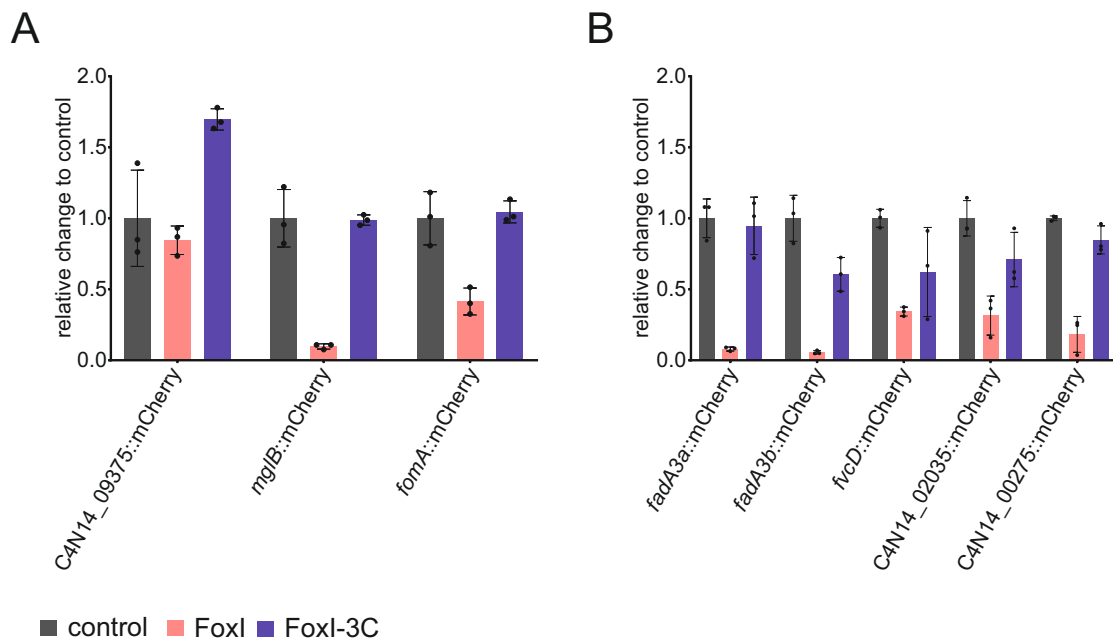
Figure S7: Western blot quantification for different translational reporters.

Quantification of western blots probed for mCherry expressed from the translational reporter plasmids in the presence of FoxI or FoxI-3C analyzed in Figure 5 for (A) and Figure 6 for (B). 5'UTRs of the indicated genes were analyzed. The average of three biological replicates relative to that of the control is displayed together with the standard deviation.

Figure S8


```
mglB
        52                              84
        |                               |
5'-AUA...UUAU   GA        A   G  AAAGU   G      GUUA...UUA-3'
             UUAG  GGGUAUUU UAU AA       UC  GUAU
             ||||  |||||||| ||| ||       ||  ||||
             AAUC  CCCAUAAA AUA UU       AG  CAUA
3'-UUU...UAAC    A              G   AAAAU  A     -5'
             |   |                              |
            31   |                              1
FoxI-C4A         |
               C4A mutation
```

Figure S8: IntaRNA target prediction between *mglB* mRNA and FoxI-C4A.

Schematic representation of the IntaRNA base-pairing prediction between the *mglB* mRNA and the seed region mutant FoxI-C4A.

The site of mutation on the sRNA (grey) and the start codon of the mRNA (red) are indicated.

Material and Methods

Electroporation of F. nucleatum

Preparation of electro-competent cells as well as delivery of plasmid DNA was achieved as described previously (1). In short, cells were harvested from the mid-exponential phase, pelleted and washed five times with ice-cold pre-reduced 10 % (V:V) glycerol solution. 5 OD cells per transformation were used to transform 100 ng (replicative plasmid) or 10 μg (suicide vectors; dialyzed) DNA via electroporation (2.0 kV, 1-mm gap). Bacteria were recovered for 2 h prior to plating on BHI-C plates containing 5 μg ml^{-1} of thiamphenicol enabling selection of successful transformants.

Construction of improved backbone pVoPo-00

Our previously generated plasmid pEcoFus (1) was optimized and reduced in size as follows: The ColE1 replication of origin for *E. coli* as well as the *catP* resistance cassette for chloramphenicol/thiamphenicol were amplified from pEcoFus using primer pairs JVO-18069/ JVO-18070 and JVO-18071/ JVO-18072, respectively. The first primer pair includes a multiple cloning site and both fragments were assembled using NEBuilder Hifi Assembly Cloning kit (New England Biolabs) together with the promoter and 5'-UTR of the constitutive expressed flavodoxin C4N14_09865 amplified from genomic DNA of *F. nucleatum* using JVO-18073/ JVO-18074. This yielded pFP76, which contains the *catP* gene constitutively, expressed from the inserted fusobacterial promoter. pFP76 was digested with PvuI and NotI and ligated with a similarly digested origin of replication for *F. nucleatum* (ORI$_{FN}$) excised from pEcoFus. This resulted in the improved *F. nucleatum* – *E. coli* shuttle vector pVoPo-00.

Construction of transcriptional reporter system pVoPo-01

The here generated pVoPo-00 was opened via inverted PCR (JVO-18273/ JVO-18274) and used for an assembly reaction together with mCherry amplified from pDSW1728

(2) (JVO-18339/ JVO-18340) as well as a single strand oligonucleotide (ssOligo) JVO-18338 forming the 5'-UTR of *acpP* in *F. nucleatum*. The resulting vector was opened once more (JVO-18341/ JVO-18342) to insert a 50 bp promoter region of the constitutively expressed *accD* (C4N14_10115; ssOligo JVO-18344). This resulted in pVoPo-01, which was used to introduce the different promoter regions tested in this work. In all cases, pVoPo-03 was opened by inverted PCR (JVO-18341/ JVO-18342) and assembled with the different ssOligo containing the different promoter sites (Dataset S5).

Construction of translation reporter system pVoPo-02

The vector pVoPo-01 was opened by inverted PCR (JVO-19090/ JVO-19091) and assembled with an ssOligo (JVO-20214) to include a ScaI restriction site at transcriptional start and an in-frame XhoI site with the coding sequence of mCherry. This yielded pVoPo-02, which was used to generate translational fusions (see below).

Construction of inducible system pVoPo-03

To construct an inducible expression plasmid, the *tetR*-GUS cassette from pRPF185 (3) was amplified (JVO-18371/JVO-18372) and assembled into the SpeI site of pVoPo-00. The resulting vector was opened via inverse PCR to remove the GUS gene and insert an XhoI restriction site (JVO-17537/ JVO-17538). The vector can be opened at transcriptional start site of the TetR-dependent promoter via the BamHI digestion.

Construction of gene deletion system pVoPo-04

pVoPo-03 was digested with XhoI and BamHI. The *mazF* gene was amplified to include an XhoI and BamHI restriction site as well as its natural 5'-UTR. Both fragments were ligated. To generate the fusobacterial suicide vector pVoPo-04, the *tetR*-*mazF* expression cassette was transferred into pFP76 via the SpeI site (JVO-18075/ JVO-18076).

Construction of translational fusions for studying the post-transcriptional regulation mediated by FoxI

pVoPo-02 was further digested with EcoRI to insert the FoxI or FoxI-3C overexpression cassette from pFP10 and pFP70, respectively, via isothermal assembly reaction to ensure the desired directionality. Each vector was then digested with ScaI and XhoI and ligated with similarly digested PCR products and regions of interest containing the 5' UTR and the first 30 nucleotides of the target genes, generating the target vectors (see Dataset S5 for primers used).

Construction of protein and sRNA inducible expression vectors using pVoPo-03

To evaluate the effect of σ^E expression, we amplified *rpoE* from the genomic DNA of *F. nucleatum* and placed it in the BamHI site in pVoPo-03. To enable higher translation, we added a short synthetic 5' UTR (4) by opening the vector via inverse PCR (JVO-19601/ JVO-18355) and assembling it with the ssOligo (JVO-19605). This generated pVoPo-03-σ^E (p.*rpoE*). To achieve inducible expression of mCherry, we amplified the *mCherry* gene from pVoPo-01 (JVO-19865/JVO-19866) and replaced *rpoE* in p.*rpoE* through inverse PCR (JVO-18355/JVO-18346). To insert FoxI, FoxI-3C or FoxI-C4A under control of TetR, we amplified *foxI* from gDNA of *F. nucleatum*, FoxI-3C from pFP70 and FoxI-C4A from pFP181 and placed in the BamHI site of pVoPo-03 via isothermal assembly to have the sRNA start at the transcriptional site of the system.

Construction of pVoPo-FP for the expression of different fluorescence proteins

Codon-optimized sequences of superfolder GFP (sfGFP), mNeonGreen and mScarlet-I were synthesized by Eurofins. The sequences can be found in Dataset S5. First, p.*rpoE* used as a template to insert mScarlet-I downstream of the 5' UTR by opening the vector with JVO-19859 / JVO-19860. This product was used together with mScarlet-I amplified via JVO-19869 / JVO-19870 for an assembly reaction. The subsequent vector was opened with JVO-20852 / JVO-20853 and used with the ssOligo JVO-20859 in an assembly reaction to insert the promoter of the fusobacterial

acpP. This resulted in pVoPo-mSc. To insert sfGFP and mCherry, the individual gene products were amplified with JVO-21080 / JVO-21081 and JVO-21082 / JVO-21083, respectively, and assembled with pVoPo-mSc opened via JVO-21079 / JVO-19859. This resulted in pVoPo-GFP and pVoPo-mCh. As the same strategy was unsuccessful for the generating a vector expressing mNeonGreen, we placed mNeonGreen in the backbone of pVoPo-02. For this, pVoPo-02 was opened with JVO-19275 / JVO-19276 and assembled with mNeonGreen amplified using JVO-19279 / JVO-19280. This yielded pVoPo-mNG. Of note, due to the difference in codon usage between *E. coli* and *F. nucleatum*, high expression of any gene, codon-optimized for or native to *F. nucleatum*, can result in spontaneous mutation or otherwise inactivation of the gene product in *E. coli*.

Northern blot

Detection of RNA via northern blot was carried out as described before (1). In short, 3 µg of DNaseI treated total RNA was separated on a 6% polyacrylamide gel (7 M urea). Afterwards, the RNA was transferred to Hybond-XL membranes and hybridized overnight at 42 °C with $[\gamma^{32}]$-ATP end-labelled deoxyribonucleotide probes (Dataset S5). The signal was visualized using a Typhoon FLA 7000 phosphoimager (GE Healthcare).

Western blot

Detection of proteins using western blot was carried as described before (1). Briefly, 0.2 $OD_{600\ nm}$ units were loaded on denaturing SDS-polyacrylamide gel for SDS PAGE analysis. Proteins were transferred to polyvinylidene fluoride (PVDF) membranes. After transfer, equal loading was verified by staining with Ponceau S solution. mCherry was detected using rabbit anti-mCherry polyclonal antibody (Life Technology PA534974) as a primary antibody in combination with an anti-rabbit secondary antibody (Thermo Fisher Scientific, catalogue no. 31460).

cDNA library preparation for RNA-seq

The cDNA library preparation was carried out by Vertis (Munich, Germany). The first step consisted in rRNA removal. The rRNA depleted RNA was then fragmented via ultrasound (1 pulse of 30 seconds; 4 °C). The fragmented RNA was used to ligate an adapter to the 3' end of the molecules. For first-strand cDNA synthesis, the M-MLV reverse transcriptase was used in conjunction with the introduced 3'-adapter serving as a primer. After purification, 5' Illumina TruSeq sequencing adapters were ligated to the cDNA. The resulting cDNA was used as input for PCR amplification (10 – 20 ng µl-1). The amplified cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics) and evaluated via capillary electrophoresis. The cDNA was pooled and further purified to include only cDNA from 200 – 600 bp using a preparative agarose gel. The finished pooled libraries were sequenced by the Core unit SysMed (University of Würzburg) using an Illumina NextSeq 500 system and 75 bp read length.

Read mapping and differential gene expression analysis

Reads from the RNA-seq experiments were trimmed and filtered using the FASTX toolkit (v.0.10.1; http://hannonlab.cshl.edu/fastx_toolkit). Mapping was performed using READemption (v.1.01) (5) against the genome sequence for *F. nucleatum* subsp. *nucleatum* ATCC 23726 (NZ_CP028109.1) downloaded from the National Center for Biotechnology Information (NCBI). Differential gene expression analysis was performed using DEseq2 (v.1.18.01) (6). The data from biological triplicates was used for the analysis. For the experiment concerning the σ^E expression, exposure to oxygen or polymyxin B, the genes had to show a $\log_2$ fold change $\leq$ -1 or $\geq$ 1 to be marked as significantly regulated. For the experiment concerning the pulse expression of FoxI or seed region mutants of the sRNA, the fold change of genes to be considered significantly changed was set to either $\leq$ -0.5 or $\geq$ 0.5. In both cases, the differentially expressed genes were only considered if their false-discovery-rate (FDR) was $\leq$ 0.05.

Generation of phylogenetic tree

We obtained the data of the phylogenetic tree from *Coleman et al.* (2021) (7) and used it as input to visualize the phylogenetic tree via ggtree (8) to generate Figure 1A.

In silico target prediction

Prediction of sRNA targets was carried using IntaRNA (v. 2.0.4) (9). As input we used all genes significantly downregulated with σ^E expression (see Fig. 3). For this, we extracted the nucleotide sequences for the entire CDS as well as the 5'-UTR. In case we were not able to determine the 5'-UTR, we took additional 50 nt upstream of the start codon. The prediction was carried out using standard settings with the exception of running it in the heuristic mode and allowing the seed region to consist out of up to three base pairs.

Datasets (separate files)

Dataset S1 (separate file): Overview of differential gene expression analysis for the pulse expression of sigma E comparing WT or ΔfoxI strain to the respective empty vector control.

Dataset S2 (separate file): Overview of differential gene expression analysis for *F. nucleatum* exposed to polymyxin B or oxygen for 20 min comapred to the untreated control.

Dataset S3 (separate file): Overview of DEseq2 analysis for gene expression analysis for pulse expression of FoxI and two seed-region mutants.

Dataset S4 (separate file): IntaRNA predicition for FoxI and genes downregulated upon sigma E induction.

Dataset S5 (separate file): Overview of strains, plasmids, oligonucleotides and gene products used in this study.

SI References

1. Ponath F, *et al.* (2021) RNA landscape of the emerging cancer-associated microbe Fusobacterium nucleatum. *Nature microbiology* 6(8):1007-1020.
2. Ransom EM, Ellermeier CD, & Weiss DS (2015) Use of mCherry Red fluorescent protein for studies of protein localization and gene expression in Clostridium difficile. *Applied and environmental microbiology* 81(5):1652-1660.
3. Fagan RP & Fairweather NF (2011) Clostridium difficile has two parallel and essential Sec secretion systems. *The Journal of biological chemistry* 286(31):27483-27493.
4. Xiao J, *et al.* (2020) Facilitating Protein Expression with Portable 5'-UTR Secondary Structures in Bacillus licheniformis. *ACS synthetic biology* 9(5):1051-1058.
5. Förstner KU, Vogel J, & Sharma CM (2014) READemption-a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinformatics (Oxford, England)* 30(23):3421-3423.
6. Love MI, Huber W, & Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15(12):550.
7. Coleman GA, *et al.* (2021) A rooted phylogeny resolves early bacterial evolution. *Science (New York, N.Y.)* 372(6542).
8. Yu G, Lam TT, Zhu H, & Guan Y (2018) Two Methods for Mapping and Visualizing Associated Data on Phylogeny Using Ggtree. *Molecular biology and evolution* 35(12):3041-3043.
9. Mann M, Wright PR, & Backofen R (2017) IntaRNA 2.0: enhanced and customizable prediction of RNA-RNA interactions. *Nucleic acids research* 45(W1):W435-w439.