The timing of transcription of RpoS-dependent genes varies across multiple stresses

Josephine Adams*
Department of Biology, Harvey Mudd College, Claremont CA
Current email: joadams@g.hmc.edu

Johnson Hoang*
Department of Biology, Harvey Mudd College, Claremont CA; Current affiliation: Doheny Eye Institute, Pasadena, CA
Current email: jhoang@g.hmc.edu

Emily Petroni*
Department of Biology, Harvey Mudd College, Claremont CA
Current affiliation: Division of Molecular and Cellular Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD
Current email: epetroni@g.hmc.edu

Ethan Ashby
Department of Mathematics and Statistics, Pomona College, Claremont CA
Current affiliation: University of Washington, Department of Biostatistics, Seattle, WA
Current email: eashby@uw.edu

Johanna Hardin
Department of Mathematics and Statistics, Pomona College, Claremont CA
Current email: jo.hardin@pomona.edu

Daniel M. Stoebel
Department of Biology, Harvey Mudd College, Claremont CA
Current email: stoebel@g.hmc.edu

*These authors contributed equally

Abstract

The alternative sigma factor RpoS regulates transcription of over 1000 genes in *Escherichia coli* in response to many different stresses. RpoS levels rise continuously after exposure to stress, and the consequences of changing levels of RpoS for the temporal patterns of expression of RpoS-regulated genes have not been described. We measured RpoS levels at various times during the entry to stationary phase, or in response to high osmolarity or low temperature, and found that the time required to reach maximum levels varied by several hours. We quantified the transcriptome across these stresses using RNA-seq. The number of differentially expressed genes differed among stresses, with 1379 DE genes identified in in stationary phase, 633 in high osmolarity, and 302 in cold shock. To quantify the timing of gene expression, we fit sigmoid or double sigmoid models to differentially expressed genes in each stress. During the entry into stationary phase, genes whose expression rose earlier tended to be those that had been found to respond most strongly to low levels of RpoS. The timing of individual gene's expression was not correlated across stresses. Taken together, our results demonstrate *E. coli* activates RpoS with different timing in response to different stresses, which in turn generates a unique pattern of timing of the transcription response to each stress.

Introduction

Many bacteria live in changing environments and face fluctuating stresses. Changing patterns of transcription of their genes is a primary way to adapt to these stresses. In *E. coli* and related enteric bacteria, the alternative sigma factor RpoS is a key global regulator of this stress response. This general stress response is induced by many different physical stresses: stationary phase; starvation for carbon, phosphate, and magnesium; elevated osmolarity; low temperature; and DNA damage (Battesti *et al.*, 2011; Gottesman, 2019). In response to these and other stressors, RpoS protein accumulates in cells.

RpoS levels in the cell are regulated at many levels, including transcription, translation, and protein degradation (Lange and Hengge-Aronis, 1994). The activity of the RpoS protein can also be regulated by other molecules, including the proteins Crl and Rsd and the 6S RNA (Battesti *et al.*, 2011; Gottesman, 2019).

RpoS recruits RNA polymerase to specific promoters, regulating the transcription of a large number of genes. In stationary phase in *E. coli*, nearly ¼ of all genes in the genome are directly or indirectly regulated by RpoS (Wong *et al.*, 2017). While many different stresses all lead to the accumulation of RpoS in the cell, the transcriptional response differs across these stresses. For example, Weber et al. (2005) found that of 481 genes differentially expressed across the entry into stationary phase, osmotic upshift, and acid stress, only 140 (29%) were differentially expressed in all three conditions. The existence of different RpoS regulons in response to different stresses suggests that other components of the RpoS response could differ as well.

RpoS levels rise continuously following onset of stress. It is unknown how the levels of RpoS affect the timing of genes in its regulon. Perhaps promoters that can be strongly transcribed with low levels of a regulator will be transcribed before those that require higher levels of that regulator. This is the case for Spo0A in *B. subtilis* (Fujita *et al.*, 2005), a master regulator of sporulation in this organism. Genes that only require

low levels of Spo0A to be transcribed are involved in physiological responses that occur early in response to starvation. Genes that require high levels of Spo0A are transcribed after more prolonged starvation, when cells begin a sporulation program.

The case of PhoB and the response to phosphate starvation in *E. coli* shows similar patterns to Spo0A in *B. subtilis* (Gao and Stock, 2015). In this system, the timing of expression of PhoB regulated genes in response to starvation is correlated to the *in vitro* measured affinity of PhoB for the promoters of those genes. In this system, the genes that are expressed first in response to phosphate starvation are genes that have high affinity for PhoB, and hence require lower amounts of PhoB for expression, compared to those expressed later that have lower affinity for PhoB and so require higher levels of PhoB in the cell for expression. Similar to the Spo0A case, there are distinct physiological differences between early and late genes. Genes involved in high-affinity phosphate uptake are transcribed early from promoters with high affinity PhoB sites, while genes involved in phosphate scavenging are transcribed later from promoters with low affinity sites.

To study the relationship between RpoS levels and transcription, in earlier work we controlled the level of expression of RpoS in stationary phase and measured the transcriptional response using RNA-seq (Wong *et al.*, 2017). We classified genes that achieved high levels of transcription in response to low amounts of RpoS as sensitive, and those genes that required high levels of RpoS for high levels of transcription as insensitive to RpoS. Using a comparison to an existing RNA-seq timecourse dataset (Conway *et al.*, 2014), we found that sensitive genes were transcribed earlier in the transition to stationary phase than insensitive genes were (Wong *et al.*, 2017). However, the low number of timepoints and lack of replication of individual timepoints in the earlier dataset made it impossible to quantify the magnitude of any difference in timing of transcription. In addition, RpoS is unusual in that it is produced in response to multiple distinct stresses. Accordingly, it was unclear whether the tentative relationship found in the stress response to stationary phase would hold across other stresses.

In this study, we hypothesized that the timing of RpoS dependent genes could be predicted based on the sensitivity of the genes to RpoS, and that the difference in timing would be large enough to be physiologically meaningful. In addition, we wished to explore if there was a similar timing difference during the response to other stresses. To address these questions, we used RNA-seq to measure the levels of expression of the RpoS regulon during the transition to stationary phase, and in response to high osmolarity and to low temperature.

Materials & Methods

Strains & Growth Conditions

The wild-type and *ΔrpoS E. coli* strains used for these experiments (Table 1) have been used for previous experiments in our laboratory (Wong *et al.*, 2017).

Table 1: Strains used in RNA-seq and protein isolation experiments

Published Designation	Genotype	Reference
DMS2537	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ(araH-araF)570(::FRT), ΔaraEp-532::FRT, ϕPcp13araE534, Δ(rhaD-rhaB)568, hsdR514	(Wong *et al.*, 2017)
DMS2545	ΔrpoS mutant (F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, Δ(araH-araF)570(::FRT), ΔaraEp-532::FRT, ϕPcp13araE534, Δ(rhaD-rhaB)568, hsdR514, ΔrpoS746::kan)	(Wong *et al.*, 2017)

For protein and RNA isolation, strains were inoculated from -80°C frozen cultures and grown aerobically overnight in 5 mL of Luria-bertani (LB) broth (1% tryptone, 0.5% yeast extract, 1% NaCl) at 37˚C with shaking at 225 rpm. The next morning, 250μL of each culture was diluted in 25mL of LB and grown at 37˚C with shaking at 225 rpm until the OD_{600} of the culture reached approximately 0.3. At this point, samples were taken for RNA and protein isolation and then the culture was exposed to the stressor of interest: for low temperature, they were moved to a 15°C shaking water bath; for high osmolarity, 3mL of 5M NaCl was added to the cultures; or they were left to continue growing to reach stationary phase. For all three stressors, OD_{600} measurements, protein, and RNA samples were taken every half hour for 150 minutes. Longer experiments were also performed for stationary phase and low temperature. For the entry to stationary phase, measurements and samples were taken every hour. For low temperature, samples were taken at 90, 120, 180, 240, and 300 minutes.

Protein Isolation & Western Blotting

Samples with cell density equivalent to 1mL of OD_{600} = 0.3 were pelleted at > 20,000 x g for 30 s, and resuspended in 100μL of 2x Laemmli sample buffer (LiCor), heated at 95°C for 5 minutes, and then stored at -20°C until they were used.

A quantitative western blot was used to measure the levels of RpoS in the cell at each time point, in a procedure slightly modified from Wong et al. (2017). 10μL of each protein sample was electrophoresed on a 4 - 20% gradient polyacrylamide gel (Bio-Rad) in Tris-glycine running buffer (25 mM Tris base, 250 mM glycine, 0.5% SDS) at 200V for 35 minutes at room temperature. Electrophoresis was used to transfer proteins to a Immobilon-FL PVDF membrane in transfer buffer (48 mM Tris base, 39 mM glycine, 0.0375% SDS, 20% methanol) at 100V for 45 minutes at 4°C. In order to quantify the total amount of protein in the samples, the membranes were stained with 5 mL of REVERT Total Protein Stain solution and imaged using the LiCor Odyssey CLx imager. The membranes were blocked with 10mL of Odyssey Blocking Buffer for 1 hour at room temperature.

RpoS was detected using an anti-RpoS monoclonal antibody (clone 1RS1, BioLegend). The blocked membranes were labeled with primary antibody (0.4 μg/mL mouse anti-RpoS, 10mL Odyssey Blocking Buffer, 0.2% Tween 20) at 4°C with shaking at 55 rpm overnight. Once labeled, the membrane was washed four times with 10 - 15mL of 1x Tris-buffered saline with Tween 20 for 5 minutes each wash. The membranes were probed with a fluorescent secondary antibody (IRDye 800CW goat anti-mouse, LiCor) that was diluted in 10mL of Odyssey Blocking Buffer with 0.2% Tween 20 and 0.01% SDS. The membranes were incubated with the diluted secondary antibody for 1 hour at room temperature and were washed as described previously, with an additional wash at the end with 1x Tris-buffered saline without Tween 20. Membranes were dried for 2-3 hours between 2 sheets of Whatmann 3MM blotting paper before imaging on a Li-Cor Odyssey CLx imager.

RpoS levels were quantified by analyzing the amount of fluorescence of each band with Image Studio v 2.1 software. Data analysis was performed in R v4.1.2 (R Core Team, 2022). RpoS levels were normalized by dividing them by the total amount of protein at each time point, which was found in the quantification of the total protein stain. Loess regression was applied to the normalized levels to examine the relationship between RpoS levels and time. The loess model provided the maximal RpoS levels and the first time corresponding to half the maximal level.

RNA Isolation

200 µL of culture was pelleted and resuspended in 500 µL of Trizol heated to 65°C and then stored at -80°C until use. Samples were purified on a column using the Zymo Direct-Zol RNA Miniprep kit. They then went through two DNase I treatments using Turbo DNA-free (Invitrogen) for 30 minutes each at 37˚C to ensure that all DNA was degraded. The samples were further purified on a column using the Zymo RNA Clean and Concentrator Kit and again stored at -80˚C until they were ready to be sent to get sequenced. The concentration of the samples was checked using a NanoDrop spectrophotometer and degradation was checked for by electrophoresing samples on an ethidium bromide agarose gel. For osmotic shock, 3 replicates were prepared for each strain. For stationary phase, 3 replicates were prepared for each strain from both the short time-course experiment (150 minutes) and the long time-course experiment (300 minutes), resulting in 6 replicates total ranging from either 0 to 150 or 300 minutes. For cold shock, 3 replicates were made for each sample in the short time-course experiment (150 minutes) and 4 were prepared from the long time-course experiment (300 minutes) for a total of 7 replicates per strain.

Generation of RNA-Seq data

Illumina cDNA libraries were generated using a modified version of the RNAtag-seq protocol (Shishkin *et al.*, 2015). Briefly, 500 ng to 1 µg of total RNA was

fragmented, depleted of genomic DNA, dephosphorylated, and ligated to DNA adapters carrying 5'-AN_8-3' barcodes of known sequence with a 5' phosphate and a 3' blocking group. Barcoded RNAs were pooled and depleted of rRNA using the RiboZero rRNA depletion kit (Epicentre). Pools of barcoded RNAs were converted to Illumina cDNA libraries in 2 main steps: (i) reverse transcription of the RNA using a primer designed to the constant region of the barcoded adaptor with addition of an adapter to the 3' end of the cDNA by template switching using SMARTScribe (Clontech) as described (Zhu *et al.*, 2001); (ii) PCR amplification using primers whose 5' ends target the constant regions of the 3' or 5' adaptors and whose 3' ends contain the full Illumina P5 or P7 sequences. cDNA libraries were sequenced on the Illumina [NextSeq 500] platform to generate paired end reads.

RNA-seq Data Analysis

Sequencing reads from each sample in a pool were demultiplexed based on their associated barcode sequence using custom scripts. Up to 1 mismatch in the barcode was allowed provided it did not make assignment of the read to a different barcode possible. Barcode sequences were removed from the first read as were terminal G's from the second read that may have been added by SMARTScribe during template switching.

Reads were aligned to the NCBI *E. coli* K-12 MG1655 genome (RefSeq assembly GCF_000005855.2) using BWA v0.7.17 (Li, 2013). Reads were counted with htseq-count v2.0.1 (Anders *et al.*, 2015). Further analysis was done in R v4.1.2 (R Core Team, 2022). Differential expression analysis was performed using the package DESeq2 v1.34 (Love *et al.*, 2014). Testing for genes that changed with time in a RpoS dependent manner was done with a likelihood ratio test, comparing a full model with genotype, time, and genotype x time interaction with a reduced model with only genotype and time effects. All p-values were FDR adjusted using the method of Benjamini-Hochberg.

Onset times of RNA-seq data were estimated with the R package Sicegar v0.2.4 (Caglar *et al.*, 2018). All analyses used the default parameters, except that the threshold intensity ratio was set to 0.65 and the threshold AIC was set to 5. The stationary phase and cold data sets combined two different experiments and sequencing batches. Batch effects were removed on the variance stabilized transformed data using the removeBatchEffects() from the package limma v3.50.3 (Ritchie *et al.*, 2015). Data were reverse transformed back to the original scale by applying an exponential function with base 2. The package infer v1.0.0 was used for permutation based statistical tests (Couch *et al.*, 2021).

GO analysis was done using the PANTHER Classification System website (Mi *et al.*, 2019). A list of genes was uploaded and compared to the entire *E.coli* genome. A statistical overrepresentation test was run using the GO biological process complete annotation set.

The RNA-seq FASTQ files are deposited in the Gene Expression Omnibus with ascension number GSE224463.

All other data data and scripts used for data analysis are at doi:10.5281/zenodo.7627190.

Results

RpoS expression across stresses

To understand how patterns of RpoS-dependent transcription change in response to the onset of a stress, we chose three stressors that had been previously characterized as inducing RpoS: entry to stationary phase, high osmolarity, and low temperature (Lange and Hengge-Aronis, 1994; Sledjeski *et al.*, 1996). Starting with *E. coli* K12 in exponential phase in rich media at 37°C, we induced stress by allowing cells to grow into stationary phase, by raising the NaCl concentration by 600 mM, or by shifting the culture to 15°C.

For all three stresses, RpoS levels changed gradually with time (Figure 1; Figure S1), but expression kinetics varied across stressors. During the entry to stationary phase, RpoS levels increased for nearly the entire experiment, peaking at 265 minutes with a normalized RpoS level of 0.047. In the presence of high osmolarity, RpoS levels rose rapidly and peaked at 50 minutes with a normalized RpoS level of 0.050. In low temperature, RpoS levels peaked at 240 minutes with a maximum normalized RpoS level of 0.024. Overall, the timing of maximal RpoS levels varies more than the amount of RpoS at the peak.

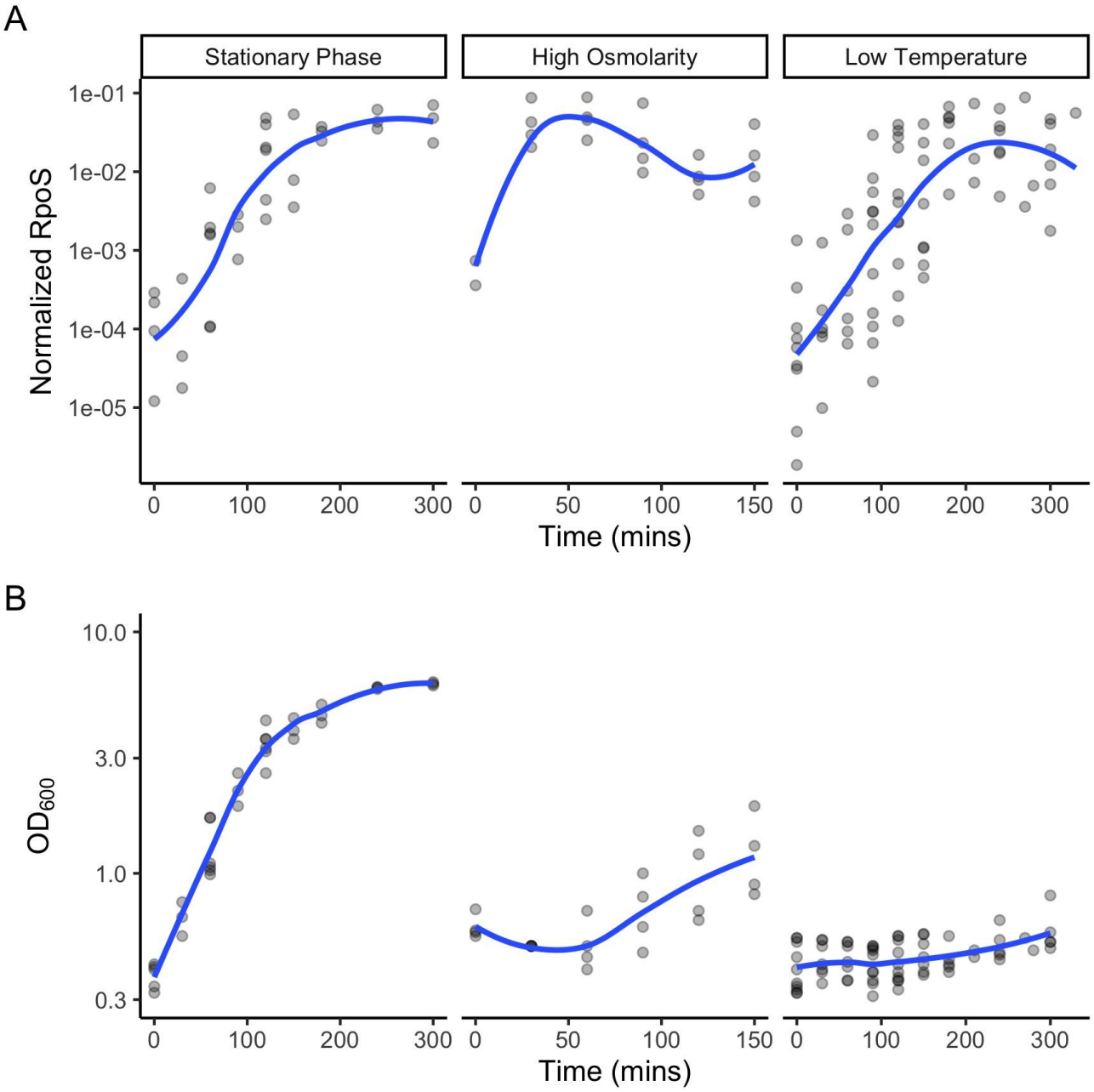

Figure 1. RpoS levels in *E.coli* in different stressors over time. Levels of (A) RpoS and (B) cell density over time for the three stress conditions in this study, the entry into stationary phase, high osmolarity, and low temperature. RpoS levels were measured by western blot. Cell density was measured by spectrophotometry. Gray dots represent individual measurements and blue curves are loess regressions.

Transcriptome changes during three stresses.

To study the role of RpoS in the transcriptional response to the three stresses, we used RNA-seq to measure transcript abundance in wild-type and Δ*rpoS* strains in each condition. In order to study the timing of RpoS-dependent gene expression in these stresses, we first characterized the regulon more generally. We modeled the counts for each gene with a linear model in DESeq2 (Love *et al.*, 2014) with terms for time + genotype + time x genotype, and considered as significant those genes with a significant time x genotype interaction term, as judged by a likelihood ratio test (FDR adjusted p-value < 0.01). Less technically, this means that we considered as significant genes where the counts changed across time in a different manner in wild-type and Δ*rpoS* strains. The number of differentially expressed (DE) genes differed over four-fold across the three different stressors, with stationary phase having 1379 DE genes, high osmolarity 633 DE genes, and low temperature 302 DE genes (Figure 2, supplementary table S1).

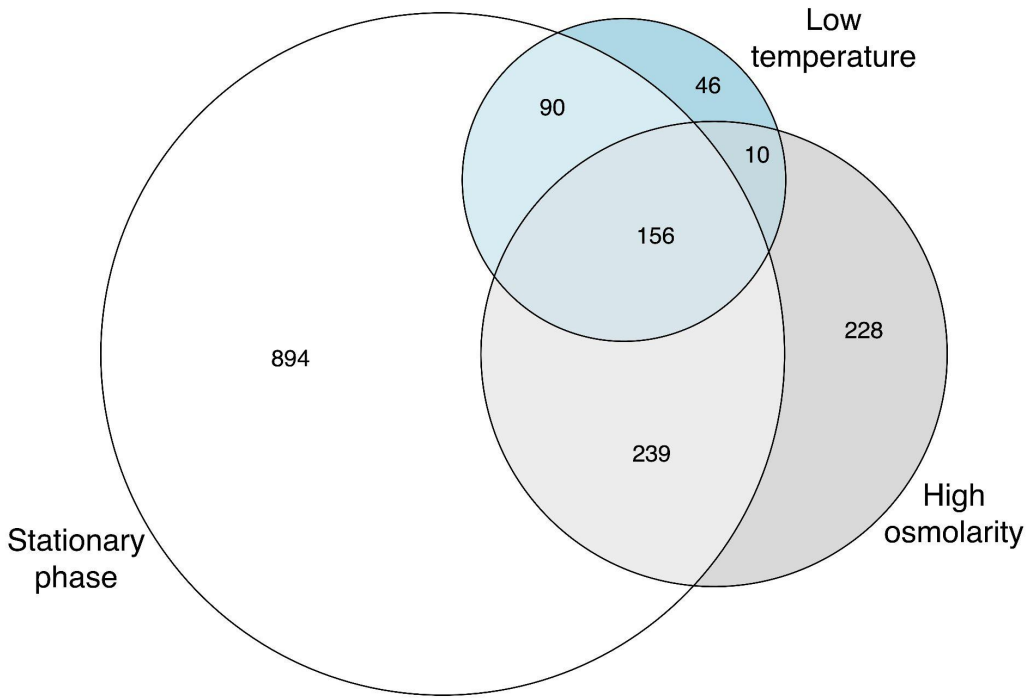

Figure 2. Venn diagram of DE genes during three different stressors as determined by DESeq2 with a LRT. White represents entry to stationary phase, gray represents high osmolarity, and blue represents low temperature. The overlapping regions indicate how many genes were differentially expressed in the overlapping stressors. In total, 1379 DE genes were identified in in stationary phase, 633 in high osmolarity, and 302 in cold shock.

One hundred fifty-six genes were DE in all three conditions, but many genes were DE in only a single condition, confirming that there is both a core RpoS-dependent transcriptional response and extensive stress-dependent responses, as first reported by Weber *et al*. (2005). The core 156 genes found in this study include a number of genes that have been well studied as classic RpoS-dependent genes, including *bolA, dps, gadB, gadC, gadX, hdeA, hdeB, hdeD, osmE, osmF, osmY, otsA, otsB, otsA, and rssB* (Bohannon *et al.*, 1991; Hengge-Aronis *et al.*, 1991; Lange *et al.*, 1993; Yim *et al.*, 1994; Altuvia *et al.*, 1994; Arnqvist *et al.*, 1994; Conter *et al.*, 1997; Castanie-Cornet and Foster, 2001; Ruiz *et al.*, 2001; Tramonti *et al.*, 2002; Checroun and Gutierrez, 2004). Weber *et al.* (2005) found 140 DE genes across stationary phase, high osmolarity, and acidic environment, and 81 genes are shared between our data set and that of Weber *et al*. (2005) In addition, 48 of the 156 genes were identified as

directly bound by RpoS in the ChIP-seq analysis of Wong *et al.* (2017). It is not clear if the other genes that are DE in all three conditions are not directly bound by RpoS, or if they were missed in the ChIP-seq experiment.

To better understand the functions of the RpoS-dependent genes under different conditions, we used the PANTHER Classification System to examine which gene functions, as described by the Gene Ontology (GO) vocabulary, are more abundant in certain groups of DE genes. In general, metabolic functions were enriched in all three stresses, although the specific metabolic functions were different. During the entry into stationary phase, 595 genes are annotated as having a cellular metabolic function (Table S1). There are twice as many genes annotated with peptidoglycan metabolic processes in this set of DE genes than the null expectation (FDR adjusted p-value = 0.04) but those genes represent only 41 of the 1379 DE genes over this transition. In high osmolarity, enriched metabolic functions include those for colanic acid biosynthesis, ATP metabolism, glycolysis, polyol metabolism, and response to heat (FDR adjusted p-value < 0.05; Table S2), all though each of these categories accounted for only 2 to 3% of all of the DE genes in this condition. In low temperature, genes associated with amino acid catabolism are enriched, but they represent less than 5% of all DE genes in this condition (FDR adjusted p-value = 0.02; Table S3). GO enrichment reveals a few gene functions more abundant than expected in the DE genes but those enriched functions are only associated with a few of the genes.

The timing of RpoS-dependent gene expression

RpoS comes on at different times in different stresses. What does this mean for the timing of RpoS dependent gene expression? To quantify the time it takes for a gene to change its expression pattern from ‘on’ to ‘off’ (or vice versa), we attempted to fit sigmoid and double sigmoid models to the expression data using Sicegar (Caglar *et al.*, 2018). A sigmoid curve is S shaped, starting at one level, then changing to another (Fig 3A). A double sigmoid curve starts at one level, changes in a S-shaped curve to a

second level, and then returns back toward the initial level in a second S-shaped curve. (Fig 3B.) The Sicegar algorithm tests if the model satisfies criteria related to minimal and maximal expression, expression at the start of the time course, and the model's Akaike Information Criterion (Caglar *et al.*, 2018). If one of the two models (sigmoid or double sigmoid) fit well, we extracted the "onset time," or time at which half maximal transcription is reached. If neither model fit well, the expression pattern for that gene in that condition was labeled as ambiguous (Fig 3C), and no onset time could be estimated. The data used for the model fitting were the normalized counts of each sample of the wild-type strain for a given condition. We only attempted to fit models for a given gene in a given condition if the gene had earlier been determined to be significantly differentially expressed.

Sigmoid or double sigmoid curves fit well to about half of the DE genes in each of the conditions. More precisely, 537 of 1142 (47%) of DE genes during the entry to stationary phase were well-fit by one of the models, as were 278 of 574 (48%) of DE genes during exposure to high osmolarity, and 177 of 284 (62%) of DE genes during exposure to high osmolarity. Visual inspection of plots of genes for which a sigmoid or double sigmoid model fit poorly showed that most such genes were noisy and generally had low levels of expression. In some cases, the model appeared not to fit well because gene expression was still rising at the last time point, rather than having started to or completely plateaued.

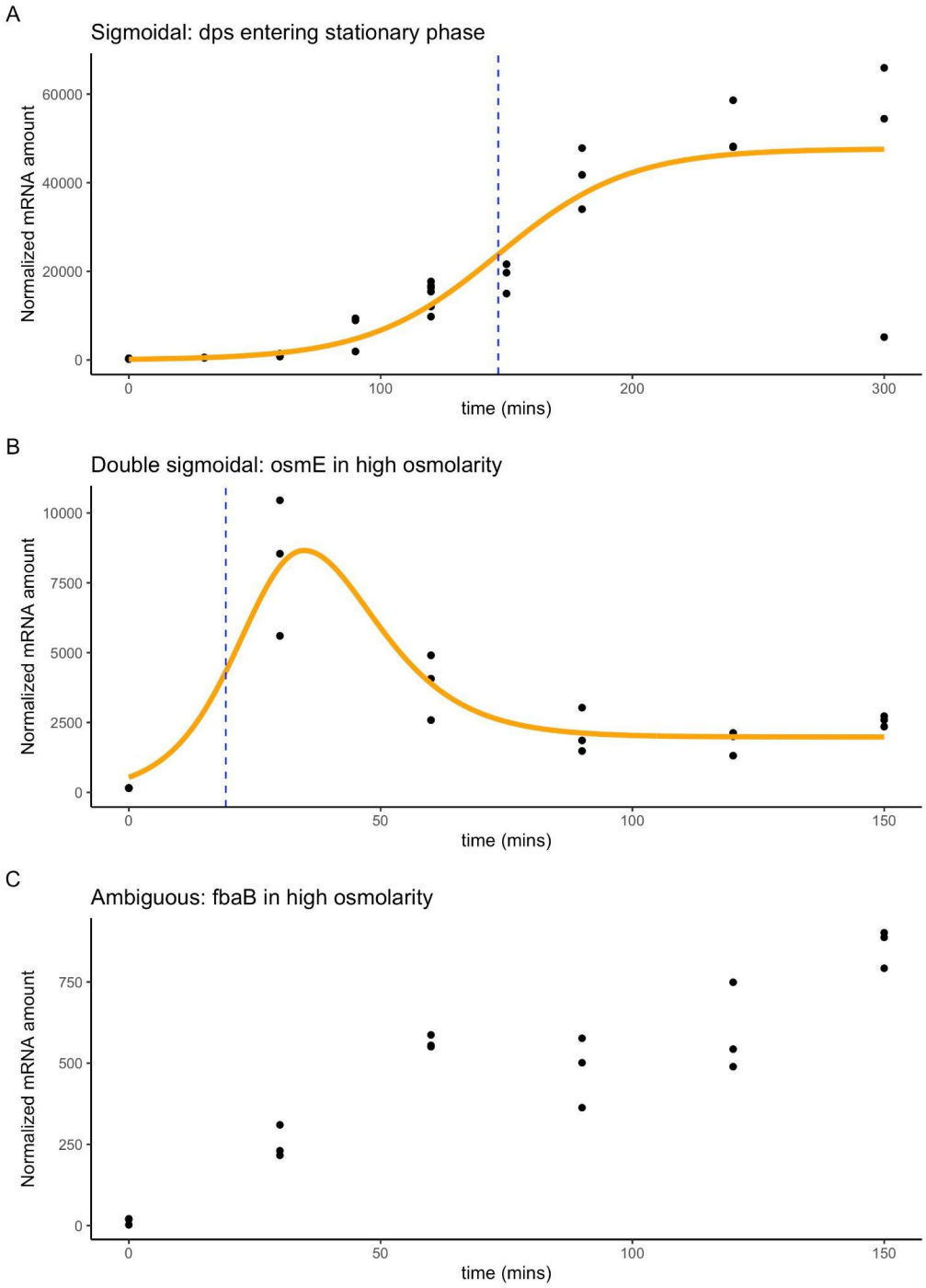

Figure 3. Examples of the three classifications of individual gene expression patterns. Expression patterns could be classified as either (A) sigmoidal, like *dps* entering stationary phase, (B) double sigmoidal, like *osmE* in high osmolarity, or (C) ambiguous, like *fbaB* in high osmolarity. A model can be fit for the first two cases. Black dots represent individual measurements, the best-fit line is shown in solid yellow, and the onset time (time of half-maximal expression) is shown as the black dashed line.

The median onset time of RpoS-regulated genes in the three stresses varied, and all three times are all significantly different from each other (Figure 4A). In general, the median onset time within each condition was similar to the time at which RpoS reached its half-maximal level (Figure 4B). The median onset time for all of the genes in stationary phase was 181 mins (95% CI: 169 - 190 mins, 1000 bootstrap replicates), 18 minutes after RpoS protein levels reached half-maximal levels (163 mins). The median onset time for all of the genes in high osmolarity was 31 mins (95% CI: 30 - 35 mins, 1000 bootstrap replicates), only 1 minute after RpoS protein levels reached their half-maximal levels (at 30 minutes). The median onset time for all of the genes in low temperature was 140 mins (95% CI: 131 - 149 mins, 1000 bootstrap replicates), which occurred *before* RpoS protein levels reached their half-maximal levels at 172 minutes. In general, the median onset time of genes is similar to the time that RpoS reached half-maximal levels in a given stress, although there is some difference between environments in the relationship between time to RpoS half-maximal levels and gene onset time.

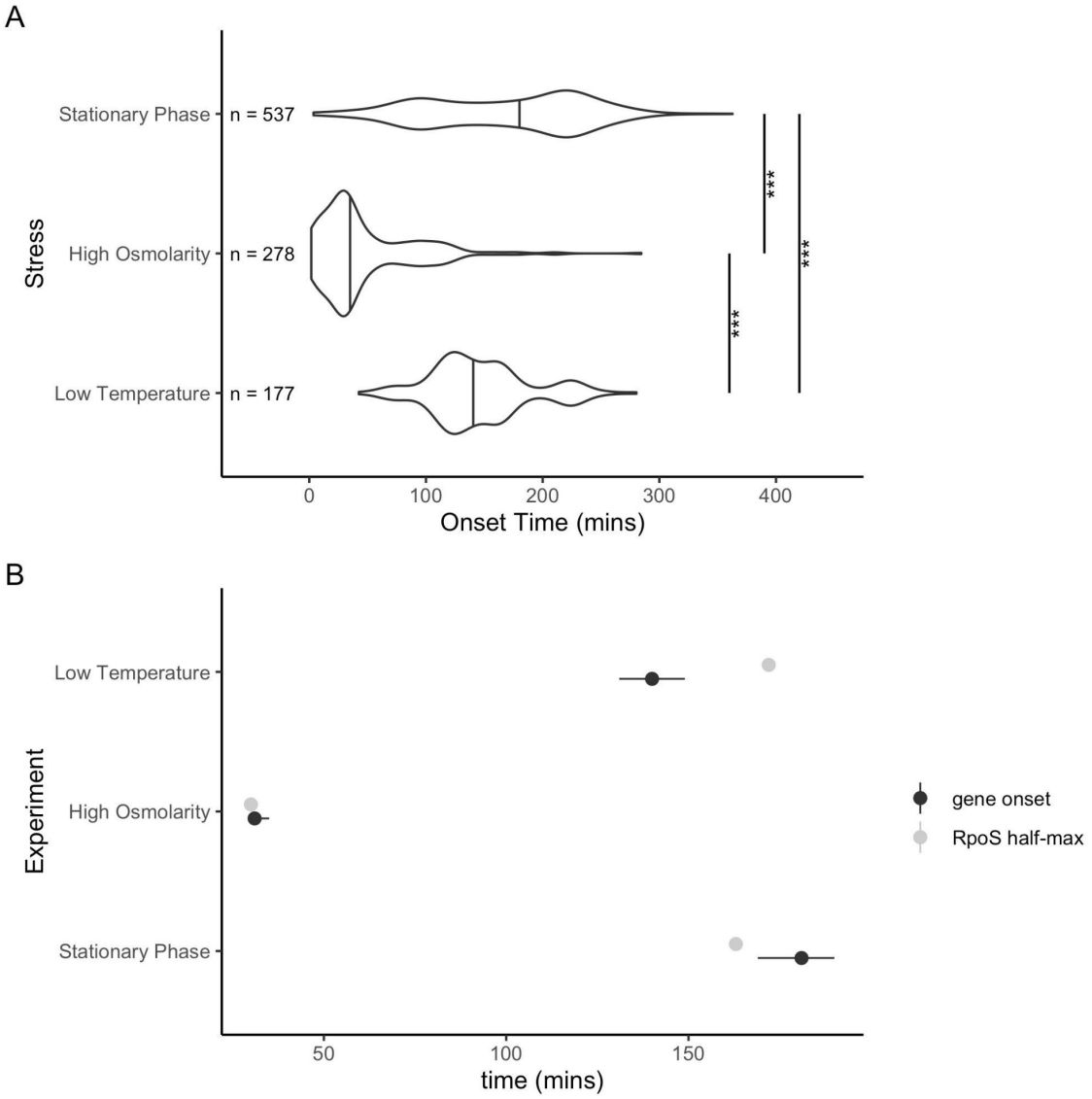

Figure 4. Onset times of DE genes that were able to be fit by Sicegar. A) Distribution of onset times of all DE genes in three conditions. Horizontal lines in the middle of the violin plot are the 50th percentile of the density estimate. Median onset time was 181 mins stationary phase, 31 minutes for high osmolarity, and 140 minutes for low temperature. All three median onset times are significantly different ($p < 0.001$, randomization test of difference in medians). B) Comparison of gene onset times and the time when RpoS reached half maximal levels in that condition. The black circle is the median gene onset time, the black line is the 95% CI on the median (1000 bootstrap replicates), and the gray dot is the time when RpoS reached half its maximal level.

In earlier work, we categorized RpoS-dependent genes as sensitive, linear, or insensitive based on how they responded to various levels of RpoS in stationary phase (Wong *et al.*, 2017). We predicted that sensitive genes should turn on before linear genes, which should turn on before insensitive genes. This prediction was correct in stationary phase (Figure 5). In stationary phase, sensitive genes had a median onset time of 102 mins, which was significantly quicker than the value for both linear genes (148 mins) and insensitive genes (180 mins) (FDR adjusted p-value < 0.01, permutation test for difference in medians, 3000 permutations). The difference in median onset time between linear and insensitive genes was not significant (FDR adjusted p-value > 0.7).

The sensitivity of a gene (as measured in stationary phase in LB media) can predict the onset time of that gene as it transitions to stationary phase in LB media. Can that same sensitivity classification predict the timing of expression in response to other stresses, like high osmolarity or low temperature? Unlike stationary phase, we found little variation in median onset time of the different classes in high osmolarity or low temperature (Figure 5). The median onset time in high osmolarity ranged from 29 mins in sensitive genes to 36 mins for insensitive genes, a nonsignificant difference (FDR adjusted p-value > 0.3, permutation test for difference in medians, 3000 replicates). The median onset time in low temperature ranged from 139 mins for linear genes to 162 mins for insensitive genes, a nonsignificant difference (FDR adjusted p-value > 0.3, permutation test for difference in medians, 3000 replicates). These results suggest that the classification of sensitivity in stationary phase stress response cannot predict timing of gene expression in other stresses.

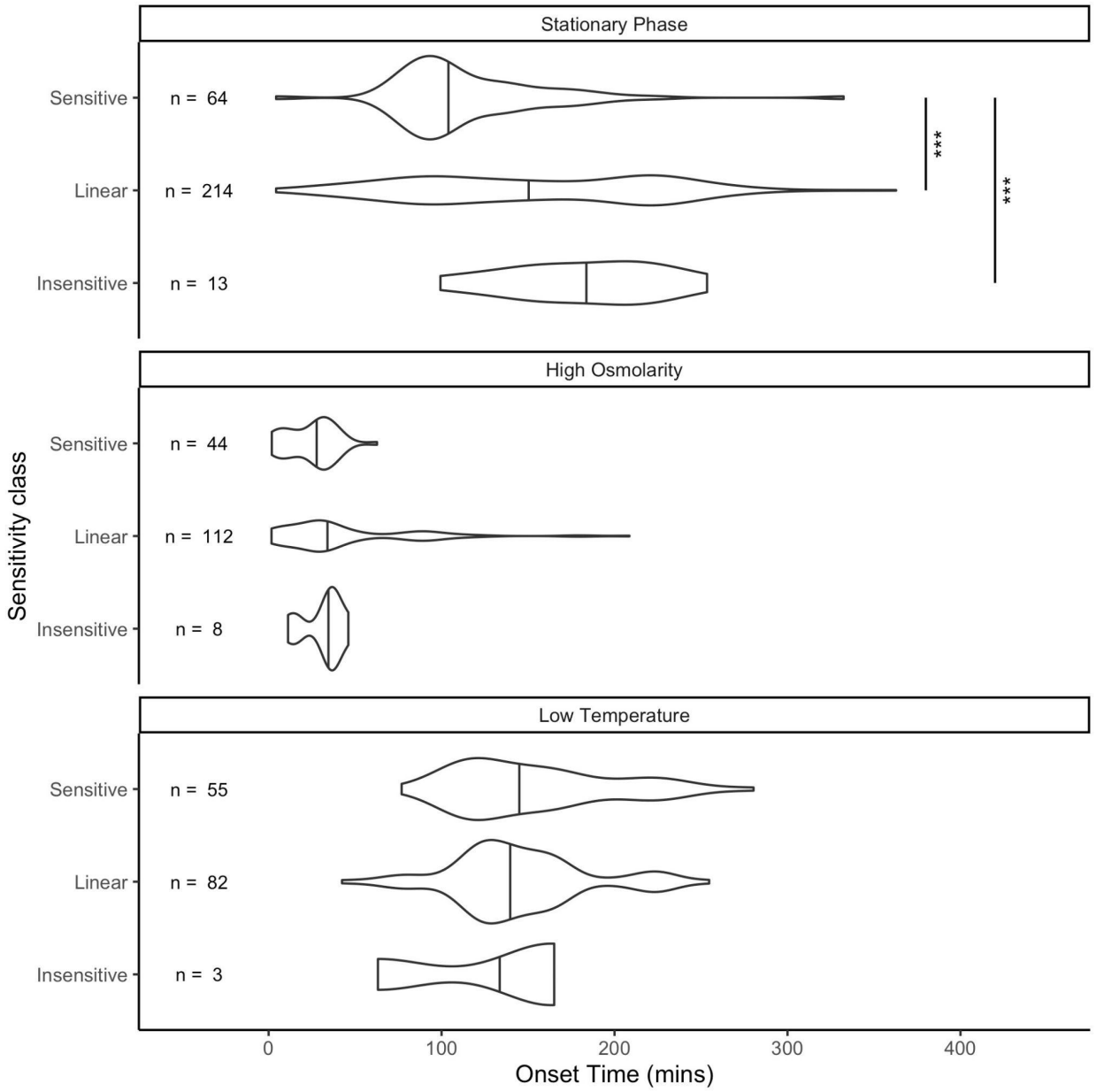

Figure 5. Distribution of onset times of genes classified as Sensitive, Insensitive, and Linear by Wong *et al.* (2017) in each of the three stresses. Vertical lines in the middle of the violin plot are the 50th percentile of the density estimate, which can deviate from the empirical median for small sample sizes. In stationary phase, sensitive genes had a median onset time of 102 mins, significantly faster than for linear (148 mins) or insensitive (180 mins) ($p < 0.001$, randomization test of difference in medians). There was no significant difference in medians between pairs of sensitivity classes for high osmolarity of low temperature.

Since sensitivity measured in stationary phase cannot predict timing in other stresses, it is worth asking if there is any correlation at all between onset times of a gene between two stressors. We examined the onset time for all genes with a successfully fit onset time in each pair of stresses (Figure 6). There is no observable correlation between onset times in any pair of stresses (absolute value of $r < 0.1$, $p > 0.3$ for all three pairs of stresses.) This lack of correlation is consistent with our finding that sensitivity classification in stationary phase can predict timing in that stress, but not in others (figure 5).

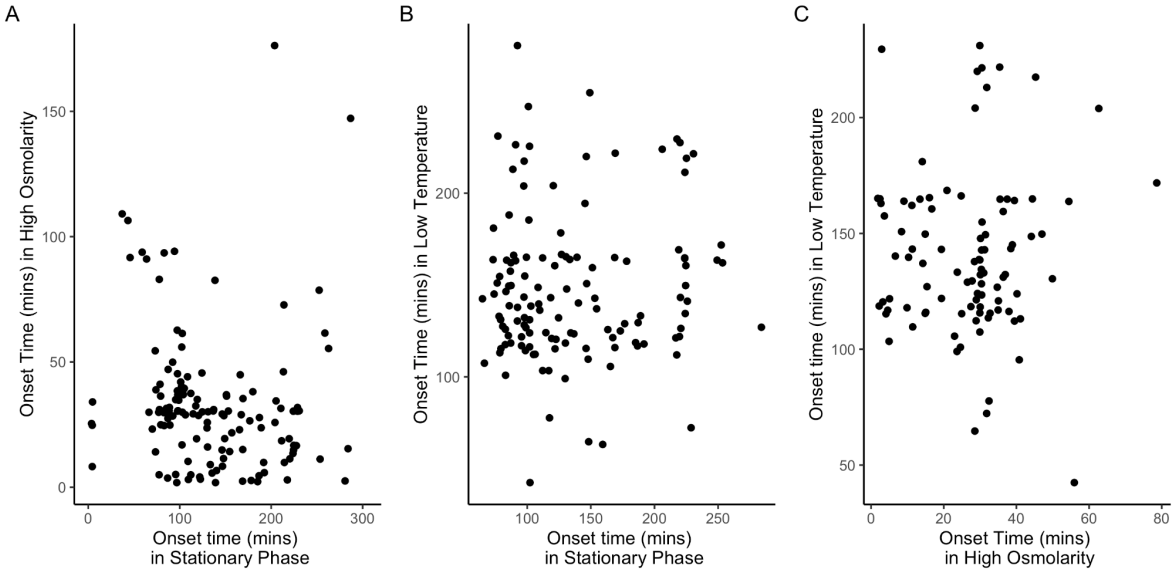

Figure 6. Onset time is uncorrelated across three stress conditions. A) Onset times in stationary phase and high osmolarity for n = 133 genes with a well-fit onset time fit in both conditions. r = - 0.09, p = 0.30. B) Onset times in stationary phase and low temperature for n = 129 genes with an onset time fit in the two conditions. r = 0.07, p = 0.44. C) Onset times in high osmolarity phase and low temperature for n = 95 genes with an onset time fit in the two conditions. r = 0.02, p = 0.82.

Discussion

The alternative sigma factor RpoS coordinates the general stress response in *E. coli* by changing patterns of transcription. The timing of RpoS induction differs across stresses. The transcription response to change in RpoS levels also differs across stresses, both in terms of which genes change transcription patterns and the timing of the change. During the entry into stationary phase, we can explain this timing using a simple conceptual model: sensitive genes come on first, insensitive genes come on last. Predictions based on gene sensitivity in stationary phase cannot predict onset timing in other stresses, as onset times in one stress are not correlated with onset times in other stresses.

E. coli has previously been reported to accumulate RpoS in response to a wide variety of stresses (Battesti *et al.*, 2011), including the three studied here: entry into stationary phase (Lange and Hengge-Aronis, 1994), high osmolarity (Hengge-Aronis *et al.*, 1991), and low temperature (Sledjeski *et al.*, 1996). Our work builds on those previous observations by measuring the rates of change of RpoS levels in response to stresses, allowing for quantitative comparisons. A notable finding is that while the peak level of RpoS varied about 2-fold across these stresses, the time it took to reach the maximal level varied nearly 5-fold. Earlier work examining how starvation for different nutrients affected RpoS levels found nearly 10-fold variation between nitrogen starvation and carbon starvation (Mandel and Silhavy, 2005). In both nitrogen and carbon starvation, maximal RpoS levels appear to have been reached within 1 hour of stress onset, more similar to elevated osmolarity than to low temperature in our study. While molecular mechanisms of RpoS upregulation in response to these stresses have been reported (Sledjeski *et al.*, 1996; Peterson *et al.*, 2012; Zafar *et al.*, 2014; Yuan *et al.*, 2017), converting that molecular understanding to understanding of the kinetics remains to be done.

The response of RpoS-dependent genes to stresses has also been reported by a number of previous studies (Patten *et al.*, 2004; Weber *et al.*, 2005; Stoebel *et al.*, 2009; Franchini *et al.*, 2015; Wong *et al.*, 2017), and our results are broadly consistent with previous findings. For example, we found that there was substantial difference in the genes that are RpoS regulated in different stresses, with only 156 genes DE across all three stresses, similar to the findings of Weber *et al.* (2005) who studied some of the same set of stresses. This work builds on those previous studies by measuring the kinetics of the response. We found substantial variation in the timing of gene transcriptional responses within individual stresses as well as across stresses. In Wong *et al.* (2017), we provided an initial analysis to suggest that differences in sensitivity of genes resulted in a difference in timing of expression of those genes, but limitations in the underlying data made it impossible to quantify the difference in timing. The results in this paper make it clear that there is substantial variation in the onset timing of RpoS-dependent genes during the entry to stationary phase, and that variation can be explained with a simple model: genes that respond strongly to limited levels of RpoS are activated before genes that require higher levels of RpoS to be activated. These results therefore suggest that differences in sensitivity of RpoS-dependent promoters coordinate the transcriptional response of cells entering stationary phase.

Is the variation in onset time among genes physiologically meaningful? In stationary phase, the median difference in onset time between sensitive genes and insensitive genes is 78 minutes. 78 minutes is long enough that *E. coli* that does not mount a response to stress can be severely harmed. For example, ΔrpoS *E. coli* exposed to high heat for 25 minutes have nearly 99% reduction in viability, while the viability of wild-type cells is reduced by about 20% in this time (Lange and Hengge-Aronis, 1991). Similarly, ΔrpoS *E. coli* exposed to hydrogen peroxide have viability reduced by about 99% within 10 minutes, while wild-type cells have viability reduced by 20% (Lange and Hengge-Aronis, 1991). With stresses affecting viability on

this time scale, a difference in onset time of 78 minutes could be a life-or-death difference for *E. coli* in response to a variety of stresses. The difference in onset timing between sensitive, linear, and insensitive genes is on a physiologically relevant time scale.

The sensitivity of genes to stationary phase RpoS levels does not predict the timing of gene induction to other stresses. We found no evidence for a difference in median onset time of the same for sensitive, linear, and insensitive genes in both osmotic shock and low temperature. In addition, onset times are not correlated across the three conditions, implying that genes that come on early in stationary phase are not more or less likely to come on early in high osmolarity or low temperature. These results are easiest to understand if the sensitivity of a gene is not an intrinsic property of its promoter, but is instead due to the interactions between the promoter and regulatory proteins that bind it. Consistent with this idea, earlier work found that regions where transcription factors other than RpoS bind to the promoter determine its sensitivity to RpoS (Wong *et al.*, 2017; Beeler *et al.*, 2019). The abundance of the transcription factors that bind these regions likely varies across environments, so we expect the sensitivity of a promoter also varies across environments. This means that our classification of promoters as sensitive, insensitive, or linear is condition dependent, and so we should not expect this classification to predict timing across stresses beyond the one in which it was measured. Biologically, this implies that the RpoS-dependent timing of expression can evolve separately across multiple stresses, consistent with the fact that different physiological responses are required for adaptation to different stresses.

In conclusion, we have studied the timing of RpoS-dependent gene expression in *E. coli* across three stresses. We have found that the the kinetics of RpoS accumulation in the cell, the identity of genes regulated in a RpoS-dependent manner, and the kinetics of RpoS-dependent transcription vary across the three stresses.

During entry into stationary phase, the timing of individual RpoS-dependent genes can be predicted by how those genes respond to changing levels of RpoS, consistent with a model where the gradual change in RpoS levels in the cell results in different timing of expression of individual genes. Moreover, the timing of gene expression is not correlated across stresses, implying that *E. coli* has distinct timings of RpoS-dependent gene regulation in response to different stresses.

Acknowledgements

RNA-Seq libraries were constructed and sequenced at the Broad Institute of MIT and Harvard by the Microbial ‘Omics Core and Genomics Platform, respectively. The Microbial ‘Omics Core also provided guidance on experimental design and performed the demultiplexing of RNA-Seq data. Thanks to Lenny Selligman and Danae Schulz for feedback and insights. This material is based upon work supported by the National Science Foundation under Grant No. 1716794.

References

Altuvia, S., Almiron, M., Huisman, G., Kolter, R., and Storz, G. (1994) The dps promoter is activated by OxyR during growth and by IHF and σs in stationary phase. *Molecular Microbiology* **13**: 265–272.

Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Arnqvist, A., Olsén, A., and Normark, S. (1994) Sigma S-dependent growth-phase induction of the *csgBA* promoter in *Escherichia coli* can be achieved in vivo by sigma 70 in the absence of the nucleoid-associated protein H-NS. *Mol Microbiol* **13**: 1021–1032.

Battesti, A., Majdalani, N., and Gottesman, S. (2011) The RpoS-mediated general stress response in *Escherichia coli*. *Annual review of microbiology* **65**: 189–213.

Beeler, S.M., Doering, C.R., Tran, S., and Stoebel, D.M. (2019) The transcriptional response of genes to RpoS concentration in Escherichia coli is not determined by core promoter sequences. *bioRxiv* 796656.

Bohannon, D.E., Connell, N., Keener, J., Tormo, a, Espinosa-Urgel, M., Zambrano, M.M., and Kolter, R. (1991) Stationary-phase-inducible “gearbox” promoters: differential effects of katF mutations and role of sigma 70. *Journal of bacteriology* **173**: 4482–92.

Caglar, M.U., Teufel, A.I., and Wilke, C.O. (2018) Sicegar: R package for sigmoidal and double-sigmoidal curve fitting. *PeerJ* **6**: e4251.

Castanie-Cornet, M.-P., and Foster, J.W. (2001) *Escherichia coli* acid resistance: cAMP receptor protein and a 20 bp *cis*-acting sequence control pH and stationary phase expression of the *gadA* and *gadBC* glutamate decarboxylase genes. *Microbiology* **147**: 709–715.

Checroun, C., and Gutierrez, C. (2004) Sigma(s)-dependent regulation of yehZYXW, which encodes a putative osmoprotectant ABC transporter of Escherichia coli. *FEMS Microbiol Lett* **236**: 221–226.

Conter, A., Menchon, C., and Gutierrez, C. (1997) Role of DNA supercoiling and RpoS sigma factor in the osmotic and growth phase-dependent induction of the Gene osmE

of Escherichia coli K12. *Journal of Molecular Biology* **273**: 75–83.

Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., *et al.* (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* **5**: e01442-01414.

Couch, S.P., Bray, A.P., Ismay, C., Chasnovski, E., Baumer, B.S., and Çetinkaya-Rundel, M. (2021) infer: An R package for tidyverse-friendly statistical inference. *Journal of Open Source Software* **6**: 3661.

Franchini, A.G., Ihssen, J., and Egli, T. (2015) Effect of Global Regulators RpoS and Cyclic-AMP/CRP on the Catabolome and Transcriptome of Escherichia coli K12 during Carbon- and Energy-Limited Growth. *PLoS One* **10**: e0133793.

Fujita, M., González-Pastor, J.E., and Losick, R. (2005) High- and Low-Threshold Genes in the Spo0A Regulon of Bacillus subtilis. *J Bacteriol* **187**: 1357–1368.

Gao, R., and Stock, A.M. (2015) Temporal Hierarchy of Gene Expression Mediated by Transcription Factor Binding Affinity and Activation Dynamics. *mBio* **6**: e00686-15.

Gottesman, S. (2019) Trouble is coming: Signaling pathways that regulate general stress responses in bacteria. *J Biol Chem* **294**: 11685–11700.

Hengge-Aronis, R., Klein, W., Lange, R., Rimmele, M., and Boos, W. (1991) Trehalose synthesis genes are controlled by the putative sigma factor encoded by rpoS and are involved in stationary-phase thermotolerance in Escherichia coli. *Journal of bacteriology* **173**: 7918–24.

Lange, R., Barth, M., and Hengge-Aronis, R. (1993) Complex transcriptional control of the sigma s-dependent stationary-phase-induced and osmotically regulated osmY (csi-5) gene suggests novel roles for Lrp, cyclic AMP (cAMP) receptor protein-cAMP complex, and integration host factor in the stationary-phase response of Escherichia coli. *J Bacteriol* **175**: 7910–7917.

Lange, R., and Hengge-Aronis, R. (1991) Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Molecular Microbiology* **5**: 49–59.

Lange, R., and Hengge-Aronis, R. (1994) The cellular concentration of the sigma S subunit of RNA polymerase in *Escherichia coli* is controlled at the levels of transcription, translation, and protein stability. *Genes & Development* **8**: 1600–1612.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with

BWA-MEM. *arXiv:13033997 [q-bio]* http://arxiv.org/abs/1303.3997. Accessed August 3, 2020.

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**: 550.

Mandel, M.J., and Silhavy, T.J. (2005) Starvation for different nutrients in *Escherichia coli* results in differential modulation of RpoS levels and stability. *J Bacteriol* **187**: 434–442.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., and Thomas, P.D. (2019) Protocol Update for Large-scale genome and gene function analysis with PANTHER Classification System (v.14.0). *Nat Protoc* **14**: 703–721.

Patten, C.L., Kirchhof, M.G., Schertzberg, M.R., Morton, R. a, and Schellhorn, H.E. (2004) Microarray analysis of RpoS-mediated gene expression in *Escherichia coli* K-12. *Molecular genetics and genomics : MGG* **272**: 580–91.

Peterson, C.N., Levchenko, I., Rabinowitz, J.D., Baker, T.A., and Silhavy, T.J. (2012) RpoS proteolysis is controlled directly by ATP levels in Escherichia coli. *Genes Dev* **26**: 548–553.

R Core Team (2022) R: A language and environment for statistical computing. https://www.R-project.org/.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**: e47.

Ruiz, N., Peterson, C.N., and Silhavy, T.J. (2001) RpoS-Dependent Transcriptional Control ofsprE: Regulatory Feedback Loop. *J Bacteriol* **183**: 5974–5981.

Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., *et al.* (2015) Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat Methods* **12**: 323–325.

Sledjeski, D.D., Gupta, A., and Gottesman, S. (1996) The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *EMBO J* **15**: 3993–4000.

Stoebel, D.M., Hokamp, K., Last, M.S., and Dorman, C.J. (2009) Compensatory evolution of gene regulation in response to stress by *Escherichia coli* lacking RpoS.

PLoS genetics **5**: e1000671.

Tramonti, A., Visca, P., De Canio, M., Falconi, M., and De Biase, D. (2002) Functional characterization and regulation of gadX, a gene encoding an AraC/XylS-like transcriptional activator of the Escherichia coli glutamic acid decarboxylase system. *J Bacteriol* **184**: 2603–2613.

Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., and Hengge, R. (2005) Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: σS-Dependent Genes, Promoters, and Sigma Factor Selectivity. *J Bacteriol* **187**: 1591–1603.

Wong, G.T., Bonocora, R.P., Schep, A.N., Beeler, S.M., Lee Fong, A.J., Shull, L.M., *et al.* (2017) Genome-Wide Transcriptional Response to Varying RpoS Levels in Escherichia coli K-12. *J Bacteriol* **199**: e00755-16.

Yim, H.H., Brems, R.L., and Villarejo, M. (1994) Molecular characterization of the promoter of osmY, an rpoS-dependent gene. *J Bacteriol* **176**: 100–107.

Yuan, J., Jin, F., Glatter, T., and Sourjik, V. (2017) Osmosensing by the bacterial PhoQ/PhoP two-component system. *PNAS* 201717272.

Zafar, M.A., Carabetta, V.J., Mandel, M.J., and Silhavy, T.J. (2014) Transcriptional occlusion caused by overlapping promoters. *PNAS* **111**: 1557–1561.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* **30**: 892–897.

Supplementary Material

Table S1: Values of statistical tests for differential change across time for all genes in wildtype and Δ*rpoS* strains across entry stationary phase, high osmolarity, and low temperature, and the estimated onset times when it could be estimated.

Table S2: Biological processes enriched in RpoS dependent genes in stationary phase

Table S3: Biological processes enriched in RpoS dependent genes in high osmolarity

Table S4: Biological processes enriched in RpoS dependent genes in low temperature

Figure S1: Examples of western blots of RpoS levels across the three stresses.