



Modeling the *Pseudomonas* Sulfur Regulome by Quantifying the Storage and Communication of Information

Peter E. Larsen,^{a,b} Sarah Zerbs,^a Philip D. Laible,^a Frank R. Collart,^a Peter Korajczyk,^a Yang Dai,^b Philippe Noirots^a

^aBiosciences Division, Argonne National Laboratory, Lemont, Illinois, USA

^bDepartment of Bioengineering, University of Illinois at Chicago, Chicago, Illinois, USA

ABSTRACT Bacteria are not simply passive consumers of nutrients or merely steady-state systems. Rather, bacteria are active participants in their environments, collecting information from their surroundings and processing and using that information to adapt their behavior and optimize survival. The bacterial regulome is the set of physical interactions that link environmental information to the expression of genes by way of networks of sensors, transporters, signal cascades, and transcription factors. As bacteria cannot have one dedicated sensor and regulatory response system for every possible condition that they may encounter, the sensor systems must respond to a variety of overlapping stimuli and collate multiple forms of information to make “decisions” about the most appropriate response to a specific set of environmental conditions. Here, we analyze *Pseudomonas fluorescens* transcriptional responses to multiple sulfur nutrient sources to generate a predictive, computational model of the sulfur regulome. To model the regulome, we utilize a transmitter-channel-receiver scheme of information transfer and utilize principles from information theory to portray *P. fluorescens* as an informatics system. This approach enables us to exploit the well-established metrics associated with information theory to model the sulfur regulome. Our computational modeling analysis results in the accurate prediction of gene expression patterns in response to the specific sulfur nutrient environments and provides insights into the molecular mechanisms of *Pseudomonas* sensory capabilities and gene regulatory networks. In addition, modeling the bacterial regulome using the tools of information theory is a powerful and generalizable approach that will have multiple future applications to other bacterial regulomes.

IMPORTANCE Bacteria sense and respond to their environments using a sophisticated array of sensors and regulatory networks to optimize their fitness and survival in a constantly changing environment. Understanding how these regulatory and sensory networks work will provide the capacity to predict bacterial behaviors and, potentially, to manipulate their interactions with an environment or host. Leveraging the information theory provides useful quantitative metrics for modeling the information processing capacity of bacterial regulatory networks. As our model accurately predicted gene expression profiles in a bacterial model system, we posit that the information theory-based approaches will be important to enhance our understanding of a wide variety of bacterial regulomes and our ability to engineer bacterial sensory and regulatory networks.

KEYWORDS *Pseudomonas fluorescens*, regulome, systems modeling, transcriptomics

Bacteria may be simple organisms, no more than a few microns in length and comprised from the interactions of a few thousand types of proteins, but they are active, dynamic participants in virtually all ecosystems. Predatory bacterial species can

Received 20 November 2017 Accepted 3


June 2018 Published 19 June 2018

Citation Larsen PE, Zerbs S, Laible PD, Collart FR, Korajczyk P, Dai Y, Noirots P. 2018. Modeling the *Pseudomonas* sulfur regulome by quantifying the storage and communication of information. mSystems 3:e00189-17. <https://doi.org/10.1128/mSystems.00189-17>.

Editor Sergio Baranzini, University of California, San Francisco

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Peter E. Larsen, plarsen@anl.gov.

 Modeling the *Pseudomonas* Sulfur Regulome by Quantifying the Storage and Communication of Information

hunt their prey in coordinated packs (1, 2), communicate with one another across biofilms using electrical signals like a primitive nervous system (3, 4), and select molecular compounds from an array of secondary metabolism biosynthetic pathways to stun prey, escape predators, or manipulate eukaryotic organisms (5–12). These activities highlight the abilities of bacteria to collect data from their surroundings, to store and process that information, and to use it to adapt its behavior to maximize fitness in their environment (13). This capacity for information processing is fundamental to understanding how bacteria survive in complex environments and respond to stimuli.

Our appreciation of bacteria as cognitive entities has continued to grow as a consequence of our increased understanding of the complex regulatory networks that drive a bacterium's interaction with its environment and with other organisms (13–17). While a significant number of regulatory circuits have been identified in a range of living organisms (18–20), the understanding of how these regulatory circuits form an architecture that supports bacterial information processing and decision-making remains elusive. As our view of bacteria evolves with respect to their role as information computing systems, new opportunities to model bacterial networks in terms of the collection, storage, and application of data become available (15, 17, 21, 22). Using the powerful and well-developed metrics and methods from information theory allows us to consider bacteria as possessors of data as well as metabolizers of nutrients. Here, we apply common tools and metrics for modeling and quantifying the flow of information in a bacterial regulome. The regulome is the set of interacting components of a cell that links information sensing to gene and protein function regulation and may include networks of genes, genomic regulatory elements, proteins, and RNA molecules (23, 24).

The sulfur regulome is defined here as the set of sensors, transcription factors (TFs), and regulated genes responding to various levels of sulfur nutrient availability and sulfur starvation. *Pseudomonas fluorescens* is a useful laboratory model for the investigation of regulome networks as it is a genetically tractable organism, enabling direct interrogation of the regulatory circuits controlling responses to environmental sulfur sources. We propose that the *P. fluorescens* sulfur regulome can be modeled using results from laboratory manipulation of the *P. fluorescens* sulfur nutrient environment and regulatory circuits to build and test our models. We consider the ability of the soil bacterium *P. fluorescens* to detect and adapt its transcriptome in response to the presence of a variety of sulfur compounds as a consequence of the flow of information through a transmitter-channel-receiver data transmission scheme for the sulfur regulome. By utilizing the tools and metrics of information theory—specifically, those of Shannon's entropy, Hamming distances, data compression, and a transmitter-channel-receiver model of information transfer—we gain the capacity to quantify the flow of information in biological regulatory systems and to use those metrics to gain novel insights into biological regulatory systems.

In this investigation, *P. fluorescens* SBW25 grown in rich medium was shifted into minimal medium containing a variety of compounds used as sole sulfur sources and relative growth levels were measured as well as transcriptional responses at one selected time point during this adaptation. A model of the sulfur regulome was generated from transcriptomic data that predict transcriptomic expression patterns in response to chemoinformatic attributes of sulfur nutrients, by means of analysis of the expression profiles of 14 sulfur regulome-associated TFs. The relevance of selected TFs to the sulfur regulome was validated using gene-knockout mutants. Analysis of our generated model indicates that metrics and concepts drawn from information theory can accurately predict biological observations and provide insights into the predicted molecular mechanisms of environmental sensing and response in bacteria.

RESULTS AND DISCUSSION

***P. fluorescens* growth depends on the sulfur nutrient.** The following nine sulfur sources were selected to represent a wide variety of molecular classifications: sodium

sulfate (a sulfur containing ion), 2-aminoethyl hydrogen sulfate (a linear sulfate), taurine (a linear sulfonate), L-methionine and L-cysteine (amino acids), α -keto- γ (methylthio)butyric acid (thioester), potassium 4-nitrophenyl sulfate (aromatic sulfate), L-methionine sulfone (an organosulfur compound with a modified S-group), and glutathione (a complex sulfur-containing molecule). An additional “no-sulfur” condition was also considered.

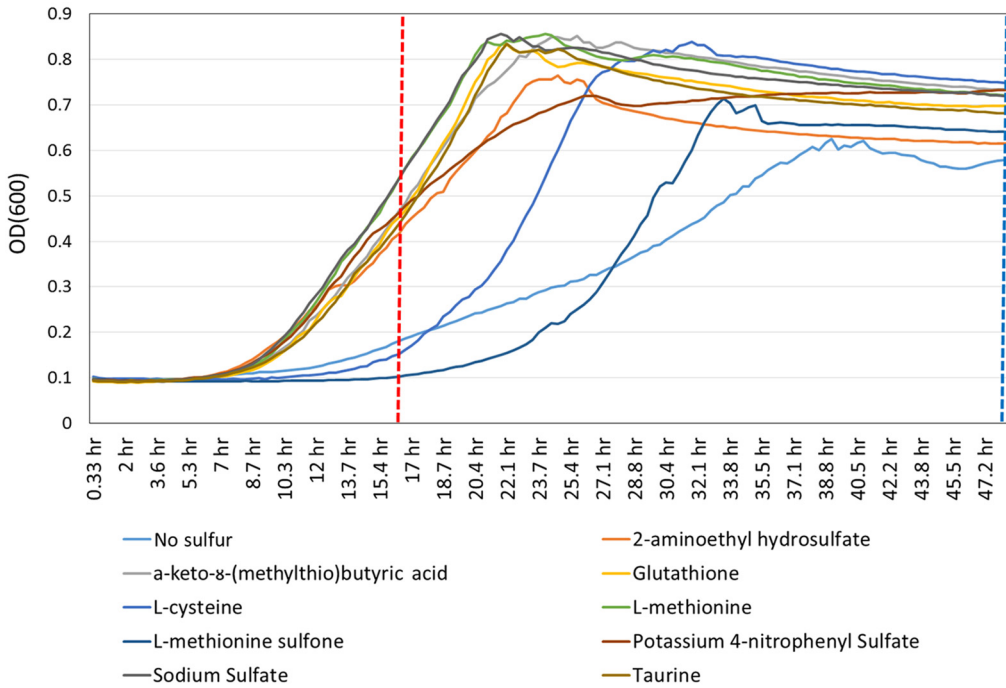
Pseudomonas minimal media (PMM) was modified to lack sulfur and was supplemented with an excess of each of the nine compounds as the sole sulfur source. Cultures of *P. fluorescens* SBW25 were grown in rich Luria broth plates (LB; 10-g/liter tryptone, 5-g/liter yeast extract, 5 g/liter NaCl), and cells were washed, diluted, and inoculated into minimal media containing a single sulfur source. To resume growth after this shift, cells must adapt to the minimal medium condition and utilize the only sulfur source available. We monitored growth (optical density at 600 nm [OD₆₀₀]) over time, measuring the lag phase before growth resumed, the growth phase, and the OD₆₀₀ after 48 h (Fig. 1). In the absence of a sulfur source, an extended phase of slow growth was observed, presumably corresponding to a sulfur-sparing response in which sulfur that had accumulated during growth in rich medium was reallocated to sulfur-containing amino acids to support cell growth. In contrast, the presence of a sulfur source triggered a distinct phase of accelerated growth after lag phases of various durations (Fig. 1A and C), reflecting an adaptation of the cell metabolism to utilize the available sulfur source. The coefficient of variation of OD₆₀₀ at 16 h after the shift was 43.4%, indicating a wide diversity of growth phenotypes at this time point. The cultures supplemented with methionine sulfone exited a long lag phase and initiated growth, while cultures supplemented with other sulfur sources had already transitioned to rapid growth, including for L-cysteine. After 48 h, growth under all medium conditions had reached a stable plateau. The coefficient of variation of growth at 48 h was 8.4%, indicating that all sulfur medium conditions had eventually reached similar ODs, the control without added sulfur reproducibly showing the lowest level.

From these observations, we can conclude that *P. fluorescens* SBW25 is capable of utilizing all selected sulfur sources, albeit with different efficiencies. Most sulfur sources were detectably utilized 8 to 9 h after the shift and promoted faster growth and higher maximal OD than were seen with the control without sulfur. Interestingly, L-methionine sulfone and, to a lesser extent, L-cysteine appear to extend the lag phase by inhibiting growth relative to the no-sulfur control before being detectably utilized and promoting faster growth. These responses are consistent with cells sensing the sulfur source, inhibiting the response observed in the control (i.e., the sulfur-sparing response), and triggering an adaptation of the cellular metabolism to utilize it. Thus, the diversity in growth phenotypes likely reflects adaptive responses that can be associated with sulfur source-specific patterns of gene regulation. To study these adaptive responses, we have selected 16 h after the shift as the time point for sampling the bacterial transcriptome and capturing the specific gene expression patterns associated with adaptation to each sulfur source.

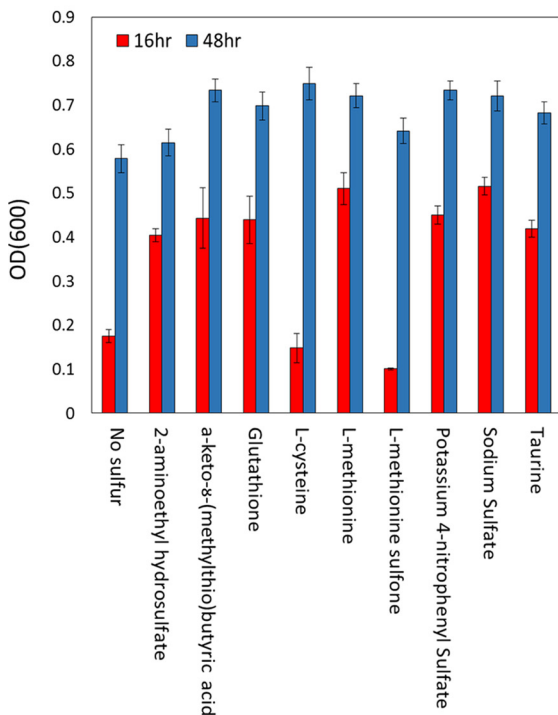
Specific transcriptomic responses to sulfur nutrients. The transcriptomes were collected from *P. fluorescens* cells cultured using the sulfur supplement conditions as described above, albeit they were collected from larger (25-ml)-volume cultures. A total of 327 genes were identified by analysis of their statistically significant differential expression (DE) (false-discovery-rate [FDR]-adjusted analysis of variance [ANOVA] *P* value of <0.05). The clusters of orthologous groups (COG) annotation categories (25) identified as significantly enriched in the set of DE genes relative to the annotated genome (*P* value of <0.05, calculated as a hypergeometric distribution) were “amino acid transport,” “posttranslational modification,” “energy production,” “lipid transport,” and “secondary metabolism.” Categories of COG annotations significantly depleted in the set of DE genes were “signal transduction” and “cell motility.”

Of the 327 DE genes, 14 are annotated as TFs (Table 1). These TFs belong to the following TF protein families: PFLU2455, AraC family; PFLU1958, PFLU3460, and PFLU4596, GntR family; PFLU0548, PFLU3260, PFLU4291, and PFLU5186, LysR family;

(A) Growth (OD₆₀₀) of *P. fluorescens* on multiple sole sulfur sources after transfer from LB media



(B) Growth (OD₆₀₀) at 16 and 48 hrs



(C) Lag and growth phase times (hrs) of *P. fluorescens* after transfer from LB media

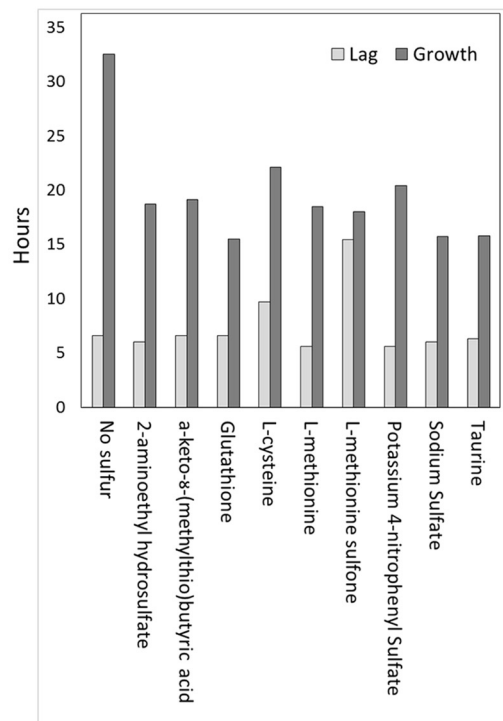


FIG 1 Relative *P. fluorescens* SBW25 growth after a shift from rich to minimal medium with a single sulfur source. (A) Each growth curve represents averages of data from 4 independent experiments. OD₆₀₀ data were measured from the time of the shift in PMM with a single sulfur source (T0). (B) OD₆₀₀ of cultures at 16 and 48 h for the different sulfur sources. These time points are indicated in panel A by dashed red and blue lines. (C) Duration of the lag phase after the shift.

TABLE 1 Sulfur regulome-associated transcription factors^a

Transcription Factor	2-aminoethyl hydrogen sulfate	L-Cysteine	Glutathione	Potassium 4-nitrophenyl sulfate	L-Methionine	L-Methionine Sulfone	α-keto-γ(methylthio) butyric acid	Sodium Sulfate	Taurine	# Co-regulated	Shannon Entropy
PFLU2455 AraC	D	D	D	D	D	N	D	D	I	56	3.21
PFLU1958 GntR	N	N	N	D	D	N	D	N	N	38	0.75
PFLU3460 GntR	N	D	D	D	N	D	N	N	I	62	1.70
PFLU4596 GntR	D	D	N	D	N	N	N	N	N	50	0.75
PFLU0548 LysR	N	N	N	N	N	N	N	N	N	23	0.00
PFLU3260 LysR	N	D	N	D	D	N	N	N	N	49	0.75
PFLU4291 LysR	N	N	N	N	N	D	N	N	N	17	0.14
PFLU5186 LysR	N	D	N	D	N	N	N	N	N	28	0.39
PFLU3257 ArsR	D	D	D	D	D	D	D	D	D	63	3.54
PFLU4114 AsnC	N	N	N	N	N	N	N	N	N	42	0.00
PFLU2053 redox-sensitive	D	D	N	D	N	N	N	D	N	57	1.20
PFLU3284 TetR	N	I	I	I	I	I	I	I	D	65	3.21
PFLU4781 TetR	N	I	N	I	N	I	I	I	N	64	1.70
PFLU5852 TetR	N	N	N	N	N	N	N	N	N	13	0.00

^aThe 14 transcription factors identified as being part of the sulfur regulome are listed together with the transcription factor family to which they belong. For each transcription factor gene (PFLU identifier number [ID] and gene family), a profile of differential expression across sulfur nutrients is shown, with significant differential expression (two-tailed t test [compared to “no-sulfur” growth conditions]) marked as “D” (decreased expression), “I” (increased expression), or “N” (no change in expression) (see Materials and Methods). “# Co-regulated” indicates the number of genes identified as potentially regulated by transcription factor. Data in the “Shannon Entropy” column were calculated as the amount of information, defined as the number of possible sulfur nutrients, that is provided by a significant change in transcription factor expression. Transcription factors in bold were selected for deletion.

PFLU3284, PFLU4781, and PFLU5852, TetR family; PFLU3257, putative ArsR family; PFLU4114, putative AsnC family; and PFLU2053, a predicted redox-sensitive transcriptional activator. A TF can regulate gene expression through direct interaction with a DNA motif near or within regulated genes and can also influence the expression of additional genes through indirect regulatory mechanisms, such as regulation of post-translational modification, in the cell. As both types of regulation can be biologically relevant, we considered here that a gene is “regulated” by a transcription factor if the patterns of expression are strongly correlated across all conditions tested (calculated as described in Materials and Methods). The 14 TFs are predicted to regulate the remaining 313 DE genes. A breakdown of the numbers of regulated genes is shown in Table 1.

The sulfur content of proteins encoded by the expressed genes is proportionate to bacterial growth. The proportion of the transcriptome that codes for sulfur-containing amino acids can be estimated from transcriptomic data and the predicted protein sequence of transcribed genes (as described in Materials and Methods). Differentially expressed genes coded for proteins that have average sulfur content (3.57%) similar to that of proteins coded by genes not differentially expressed (3.36%). There was a positive correlation (Pearson correlation coefficient [PCC] value of 0.60 [P value less than 0.05; calculated as 10,000× bootstrap]) between the total sulfur content of predicted expressed proteome and the relative growth of bacterial culture. While protein abundance is not necessarily proportionate to the level of gene expression, this observation suggests that a lower level of assimilation of the sulfur source, indicated by reduced growth, may be associated with sulfur-sparing responses in which bacterial cells downregulate genes for proportionately sulfur-rich proteins. Such a sulfur-sparing

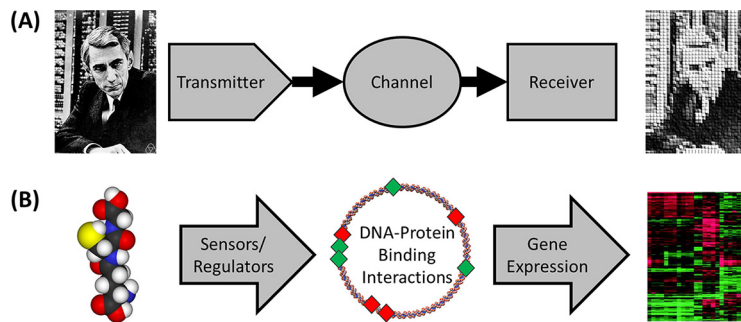


FIG 2 Modeling the bacterial regulome transmitter-channel-receiver scheme. (A) Transmitter-channel-receiver scheme for information transfer. (B) Scheme used to describe information flow in biological networks with specific molecular mechanisms that fulfill each role in the transmitter-channel-receiver indicated.

response has been well characterized in yeast (26). This observed link between sulfur assimilation and global regulation of the sulfur content of the bacterium's proteome is a strong indication of the broad regulatory capacity of the sulfur regulome.

Model *P. fluorescens* SBW25 regulome as transmitter-channel-receiver. Information transfer can be modeled as being comprised of three components (27): information is detected and collected by a transmitter and then encoded into a more compact form and passed along via a (potentially noisy) channel; the information from the channel is collected by a receiver; and the original message is reconstructed (Fig. 2). This transfer of information can be lossless, if the recovered data are identical to the original data, or lossy, if the data cannot be exactly recovered and if some information is lost. In using the model of information transfer to describe a bacterial regulome, specific biological mechanisms are proposed to fulfill the functions of transmitter, channel, and receiver. Here, the information being conveyed is the composition of nutrients in the extracellular environment. It is unlikely that bacteria possess a specific sensor for every possible nutrient that they may encounter. Therefore, we propose that bacteria have a “transmitter” comprised of multiple membrane-bound sensors with overlapping activities that, by acting in coordination, accurately discern far greater numbers of environmental conditions than they have sensor proteins. In the model, we represent this capacity by considering compounds in the environment to be vectors of chemoinformatic attributes, allowing a potentially great number of possible molecules to be described by relatively few features. The role of the “channel” in bacterial systems involves protein-DNA interactions, as TFs bind to their cognate regulatory elements in the genome. In this fashion, information about the extracellular environment can be symbolically encoded and stored through protein-DNA binding interactions. The “receiver” in this system is the gene expression output, brought about by the binding/release of transcription factors that modulate expression patterns for genes, ultimately optimizing fitness for the nutrient environment.

The methods of construction of the transmitter, channel, and receiver components of the regulome model are described separately below. In the last section, the individual elements are combined in a predictive, system-scale model of the regulome.

The transmitter: expression of the sulfur nutrient environment as a vector of chemoinformatic attributes. Our model of the sulfur regulome presumes that *P. fluorescens* collects information from its environment, not as the presence or absence of specific sulfur compounds but rather as assemblages of key chemical features present in the extracellular environment. The sulfur nutrients used in this experiment can be described as vectors of chemoinformatic attributes that can be grouped into atoms, bonds, functional groups, and molecular characteristics. This approach provides the model with powerful extrapolative abilities. By defining a nutrient as a vector of attributes rather than as a distinct chemical entity, new nutrients that were not used in model training sets can be considered by describing new nutrients as recombinations

TABLE 2 Chemoinformatic attributes for sulfur nutrients^a

		2-aminoethyl hydrogen sulfate	L-Cysteine	Glutathione	Potassium 4-nitrophenyl sulfate	L-Methionine	L-Methionine Sulfone	α -keto- γ (methylthio) butyric acid	Sodium Sulfate	Taurine
Atoms	C	2	3	10	6	5	5	5	0	2
	N	1	1	3	1	1	1	0	0	1
	O	3	2	6	6	2	4	3	4	3
	S	1	1	1	1	1	1	1	1	1
	H	7	5	15	4	11	11	8	0	7
Bonds	C-C	1	4	7	3	3	3	4	0	1
	C=C	0	0	0	3	0	0	0	0	0
	C-H	4	3	8	4	7	8	7	0	4
	C-O	1	1	2	0	1	1	0	0	0
	C=O	0	1	3	0	1	1	2	0	0
	C-N	1	0	5	0	1	1	0	0	1
	C-S	0	1	1	0	2	2	2	0	1
	S-O	2	0	0	2	0	0	1	2	1
	S=O	2	0	0	2	0	2	2	2	2
	O-H	1	1	2	0	0	0	1	0	1
	N-H	2	2	4	0	2	2	0	0	2
	S-H	0	1	1	0	0	0	0	0	0
	N-O	0	0	0	2	0	0	0	0	0
Functional Groups	NO ₂	0	0	0	1	0	0	0	0	0
	CNH ₂	1	1	1	0	1	1	0	0	1
	Phenol	0	0	0	1	0	0	0	0	0
	COOH	0	1	2	0	1	1	1	0	0
Molecular Characteristics	H-donors	2	2	5	1	2	2	1	0	2
	H-acceptors	2	3	9	8	3	5	3	4	4
	Rotatable bonds	3	2	9	3	4	4	4	0	2

^aChemoinformatic attributes are grouped into number of atoms, number of chemical bonds, number of functional groups, and number of specific molecular characteristics. "H-donors" and "H-acceptors" data indicate the number of hydrogen bond donors and acceptors in the molecule (at pH 7.0). "Rotatable bonds" data represent the number of bonds which allow free rotation around themselves (a measure of molecule's flexibility). For each attribute (row), values are highlighted in colors that range from lowest (red) to highest (green) values.

of the attributes used in the training set. Twenty-five chemoinformatic attributes were selected to represent the 9 sulfur nutrients used in our experiment as follows: 5 atoms, 13 molecular bonds, 4 functional groups, and 3 molecular characteristics (Table 2).

The channel: environmental conditions encoded as TF expression profiles. The channel in our model is described as using DNA-protein binding interactions of TFs to encode information about cell environmental conditions. Here, we consider the expression level of a gene encoding a TF and presume that increased expression of a TF will result in a proportionately greater frequency of binding of the TF to the chromosome. Although this assumption represents a simplification of the complexity of

biological regulatory circuits, it allows us to use the measurable level of TF expression as a proxy for DNA-protein regulatory interactions in a context where the regulatory interactions taking place in the cell remain poorly characterized.

(i) A unique TF expression pattern “code” indicates the identity of a sulfur nutrient. Our proposition that multiple transcription factors encode information regarding extracellular environmental conditions implies that there must be a unique profile of transcription factor expression that corresponds to each sulfur nutrient. Indeed, the patterns of significant differential expression (DE) of transcription factors (Table 1) can be viewed as bar codes that are unique for each sulfur nutrient, thereby allowing association of specific sulfur nutrients with patterns of transcription factor expression.

(ii) Gene knockout experiments indicate that identified TFs are active players in the sulfur regulome. While 14 TFs were identified as differentially expressed in response to sulfur nutrient conditions, all those TFs may not be specifically controlling a response to a sulfur nutrient. For example, some transcription factors may have more general roles associated with different growth rates. To validate that the identified TFs play a role in the *Pseudomonas* sulfur regulome, we generated knockout mutants for half of them (Table 1). Of the 14 TFs, 7 were selected to represent a broad range of transcription factor families, medium-specific gene expression patterns, and Shannon’s entropy levels, namely, PFLU2053, PFLU2455, PFLU3460, PFLU4782, PFLU5187, PFLU5853, and PFLU4597. Gene deletions were generated by homologous gene replacement and verified. TF-knockout mutants were grown on the set of 9 sulfur sources, and growth profiles were monitored.

There are three anticipated outcomes of a transcription factor knockout: (i) no effect on bacterial growth, suggesting that the transcription factor is not relevant to the sulfur regulome; (ii) negative effects on bacterial growth that are independent of the sulfur source, suggesting that while the transcription factor is generally important to growth or metabolism, it is not necessarily associated with any of the environmental conditions that we tested; and (iii) changes in mutant growth relative to the wild type that are specific to the combination of sulfur source and transcription factor knockout, indicating that those TFs are part of the sulfur regulome.

The results for mutant growth on sulfur media are summarized in Fig. 3 and are presented in full in Fig. S1 in the supplemental material. The changes in OD_{600} at 16 h and in lag time duration for each sulfur media were compared for each mutant relative to the wild type under the same conditions. Interestingly, growth without sulfur at 16 h was significantly reduced in 6 of the 7 knockout mutants, suggesting that TFs are principally responsible for cell adaptation to the shift from rich to minimal media with a single sulfur source. There was a unique, medium-specific effect on growth at 16 h and on lag times for each sulfur regulome-associated TF analyzed. A TF knockout’s effect on growth at 16 h is not necessarily correlated with a change in lag time. No knockout mutant had a significant effect on growth under L-methionine sulfone conditions. However, the deletion of PFLU2455 caused a significant decrease in lag times for cultures in L-methionine sulfone and L-cysteine. There was no significant change in either growth at 16 h or lag time for growth in sodium sulfate media for any knockout mutants, indicating that the regulatory circuits perturbed by this experiment are mainly relevant for the organosulfur regulome. From this, we can conclude that for each knockout mutant, medium-specific changes in bacterial growth and lag times were observed, supporting anticipated outcome iii above and suggesting that all TFs selected for validation were actively contributing to the sulfur regulome. Interestingly, some knockout mutations resulted in increased growth relative to the wild-type strain, suggesting that the deregulated genes in the knockout mutant affected directly or indirectly the adaptation to minimal media and utilization of the sulfur source, which was nonlimiting under our conditions. An indirect effect may be that a TF activated genes that compete with the utilization of the sulfur source for coping with a particular stress (e.g., redox stress). In such a case, the sulfur source not only provides a nutrient that fuels the metabolism of the wild-type bacterium but also provides information

	Change in OD at 16 hrs							Change in Lag Time						
	PFLU_2058	PFLU_2455	PFLU_3460	PFLU_4781	PFLU_4596	PFLU_5852	PFLU_5186	PFLU_2058	PFLU_2455	PFLU_3460	PFLU_4781	PFLU_4596	PFLU_5852	PFLU_5186
No sulfate	-0.43	-0.21	-0.47	-0.58	-0.30	-0.47	-0.37	0.08	0.33	0.08	0.08	0.28	0.28	-0.07
2-aminoethyl hydrosulfate	-0.06	0.00	-0.07	-0.25	-0.42	-0.03	0.02	-0.10	0.00	-0.18	0.22	0.34	-0.18	-0.18
α -keto- β -(methylthio)butyric acid	-0.11	0.03	0.15	-0.21	-0.24	0.24	0.29	-0.14	0.00	-0.24	0.20	0.40	-0.52	-0.32
Glutathione	0.19	0.29	0.26	-0.06	-0.08	0.35	0.18	-0.32	-0.14	-0.32	0.15	0.33	-0.24	-0.14
L-cysteine	-0.71	-0.21	-0.60	-0.75	0.15	-0.53	-0.63	1.48	-3.28	0.78	1.46	0.31	0.78	1.09
L-methionine	-0.13	-0.06	-0.04	-0.41	-0.35	0.15	0.06	0.00	0.32	0.10	0.38	0.57	-0.08	-0.16
L-methionine sulfone	-0.01	0.23	0.12	-0.09	-0.01	0.17	0.19	-0.04	-0.89	-0.32	0.12	0.12	-0.07	-0.11
Potassium 4-nitrophenyl Sulfate	0.03	0.06	-0.03	-0.37	-0.48	0.05	0.05	-0.08	0.00	0.00	0.44	0.64	0.10	0.00
Sodium Sulfate	0.01	-0.06	0.00	-0.25	-0.12	0.11	0.04	-0.18	0.00	-0.26	0.14	0.22	-0.26	-0.18
Taurine	0.09	0.09	0.08	-0.11	-0.43	0.30	0.19	-0.25	0.00	-0.17	0.15	0.40	-0.45	-0.55

FIG 3 Relative growth of transcription factor knockout mutants on different sulfur sources. Changes in OD₆₀₀ are indicated as the log₂ of the ratio between the culture OD₆₀₀ at 16 h for a knockout (KO) mutant and the wild type on the same sulfur nutrient media. Changes in lag time are indicated as log₂ of the ratio between the lag times in the KO mutant and wild-type cultures with the same sulfur nutrient media. Cells are highlighted using a color gradient from the lowest values (blue) to the highest values (red). Values that are statistically significantly different from wild-type values (*P* value of <0.05) are highlighted in bold.

about the environment that is used by the cell to maximize its fitness, which does not necessarily imply maximizing its growth rate.

(iii) Vector of chemoinformatic features predicts TF expression patterns. The expression level of each TF can be described as a mathematical function of the chemoinformatic attributes of the available sulfur source. A leave-one-out cross-validation (LOO-CV) approach was used to train the models of TF expression, and only the validation results are presented here. The overall correlation between predicted and observed TF expression profiles across all medium types was significantly high (PCC = 0.82, LOO-CV *P* value less than 0.05 in 10,000× bootstrap analyses). Considering the results for individual sulfur sources, correlations between predicted and observed patterns of TF expression were also significant for all sulfur medium types except glutathione (Fig. 4).

There are many possible reasons for the comparatively poor prediction of TF profile for glutathione. Glutathione is known to have multiple roles in redox signaling, protection from various stresses, and posttranslational modification of proteins in *Proteobacteria* (28, 29), including *Pseudomonas*. These roles may be indirectly related to the utilization by cells of glutathione as a sulfur nutrient and may not be apparent in the data collected from our simple experimental design, which did not include redox stress. Additionally, glutathione is an outlier for 17 of the 24 chemoinformatic features (Table 2), which might make predictions for glutathione more difficult in utilizing a LOO-CV scheme.

The receiver: environmental condition information decoded as gene expression patterns. The receiver element of our model of the sulfur regulome translates the

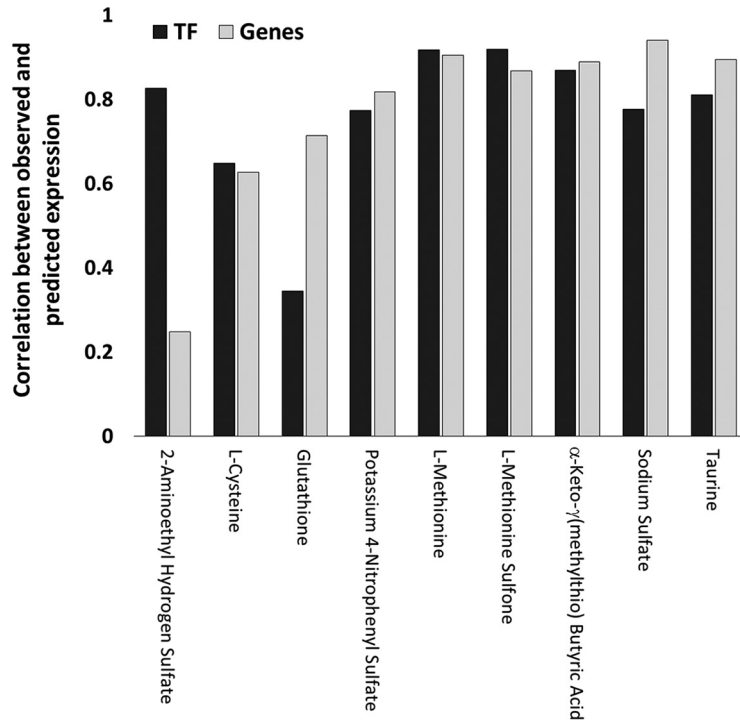


FIG 4 Correlations between computationally predicted and observed gene expression patterns. The correlations between observed and predicted gene expression patterns are shown for 14 sulfur-related TFs (black bars) and 313 SDE genes in response to sulfur source (gray bars).

information that is encoded as TF expression profiles into the transcriptome expression patterns specific to a cell's environmental conditions.

The correlations between the observed gene expression levels and the gene expression levels predicted as a function of TF expression profile were significant for every sulfur source (LOO-CV P value of ≤ 0.05 in 10,000 bootstrap analysis) and had an average PCC value of 0.77 (Fig. 4). Correlations between the predicted and observed gene expression patterns were lowest for 2-aminoethyl hydrogen sulfate and highest for sodium sulfate. Interestingly, the predicted gene expression patterns for the 313 significantly differentially expressed (SDE) genes with 2-aminoethyl hydrogen sulfate were poorly accurate relative to the predicted expression pattern of the 14 sulfur-related TFs. This result suggests either that TFs important for the adaptive response to 2-aminoethyl hydrogen sulfate are not present in the 14 sulfur regulome-related TFs used in this model or that there are posttranscriptional regulatory mechanisms involved in response to 2-aminoethyl hydrogen sulfate.

What is gained from modeling the sulfur regulome using the transmitter-channel-receiver scheme? Predictions of TF expression profiles as functions of chemoinformatic attributes and gene regulation patterns were found significantly correlated with biological observations. However, we must now ask the following question. Does the incorporation of the transmitter-channel-receiver concept in the model of the regulome lead to greater predictive power or biological insight than a simpler approach that does not use such a scheme?

(i) Incorporation of the transmitter-channel-receiver structure into the regulome model improves predictions of gene expression patterns. To validate the use of the transmitter-channel-receiver scheme, we have constructed a gene regulatory model that does not use this structure. The gene expression pattern was calculated directly as a function of chemoinformatic attributes, without considering the intermediate level of the TF expression profile. Models were trained using a LOO-CV approach identical to that used for the prediction of gene expression patterns as a function of TF

expression. As with the transmitter-channel-receiver scheme model, only validation data are considered to represent a metric of model prediction accuracy. The overall value corresponding to the PCC between predicted and observed gene expression patterns was 0.44, which is lower than the overall PCC value of 0.77 for predicting gene expression patterns using the model incorporating the transmitter-channel-receiver scheme. Incorporation of the transmitter-channel-receiver structure into the model provides relevant biological information to the model and generates better predictions of the observed gene expression patterns than a model that disregards this proposed biological structure.

(ii) The information content of TF expression is proportionate to the number of genes that it regulates. The Shannon's entropy value represents quantification of the expected value of the information contained in a message, measured as the reduction of uncertainty. In this case, the "message" is defined as an observed, significant change in TF expression and a change in expression of a TF reduces the uncertainty regarding the bacterium's nutrient environment. The set of calculated Shannon's entropy values for each TF can be found in Table 1. Using data from Table 1, a significant positive correlation between the Shannon's entropy value for a TF and the number of genes that it regulates is observed (PCC = 0.78, P value ≤ 0.05 , calculated as $10,000 \times$ bootstraps). This result suggests that TFs that encode more information about the extracellular environment tend to regulate (directly and/or indirectly) a greater number of genes, which may be a general characteristic of information processing in regulatory networks.

(iii) A robust method of encoding environmental information. If the biological networks can be modeled as the flow of information, then we might expect that the method of coding environmental conditions as patterns of TF-DNA binding interactions should be robust against channel noise. Considering the patterns of TF profiles in Table 1, the average value for the Hamming distance (30) between TF expression patterns is 4.7. This Hamming distance result means that, on average, about 5 transcription factors (36% of all of the sulfur regulome-associated TFs) would have to be altered with respect to their regulation before one sulfur nutrient could be mistaken for another. This indeed exhibits an encoded signal of TF expression patterns that is robust against channel noise.

(iv) Drawing biological inferences from a visualization of the model of the sulfur regulome. The transmitter-channel-receiver scheme for depicting the regulome increases the accuracy of gene expression profile predictions and enables the application of metrics from information theory (i.e., Hamming distance, Shannon's entropy) to the model for the quantification of information flow in the regulome. However, can this model be used to make biological inferences with respect to the molecular mechanisms of the regulome? To engage a biological analysis of the model, we have generated a visualization of the regulome suitable for direct interpretation, as described below.

The three components of the *Pseudomonas* sulfur regulome, i.e., the transmitter, channel, and receiver, can be combined to form a single, system-scale model of the sulfur regulome. The interactions between the vector of sulfur source chemoinformatics features and the TF expression profile were generated as a set of evolutionary algorithm-derived equations. A network visualization was generated such that the parent nodes of transcription factors were those chemoinformatic attributes used in the model equations. Those equations were used to generate a network in which every node in the network is a child of the specific features (i.e., chemoinformatic attributes of the transcription factor expression level) in the function that describes its relationship to its parent nodes. The visualization of the network comprising all the links between TFs and the 313 genes whose expression patterns most closely correlate with them results in a network too dense for easy visual inspection. Therefore, we used a different approach to visualize interactions between transcription factor expression profiles and regulated genes. We calculated the Pearson's correlation coefficient (PCC) values corresponding to the gene expression patterns of the 14 transcription factors

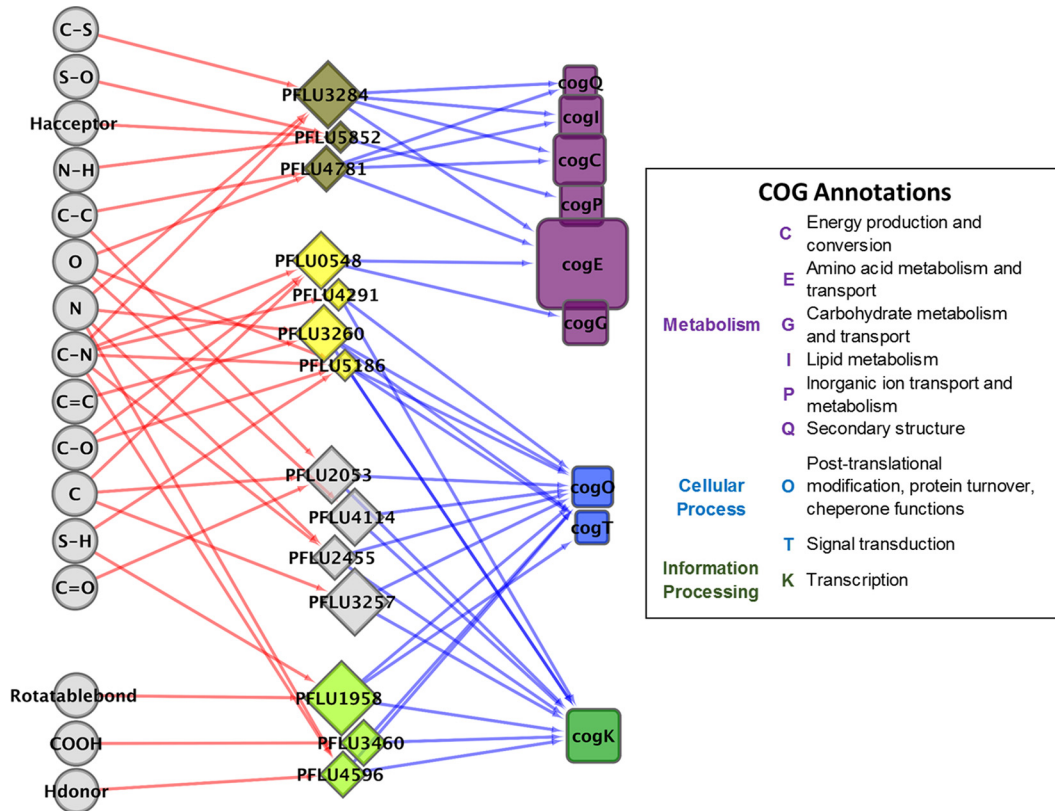


FIG 5 The sulfur regulome of *P. fluorescens* SBW25. Circles represent chemoinformatic features of nutrients. Diamonds represent transcription factors, and colors indicate transcription factor families as follows: TetR family, brown; LysR family, yellow; GntR family, light green; other transcription factor families, gray. Diamond size is proportionate to the Shannon's entropy value for the transcription factor. Rounded rectangles represent groups of genes predicted to be regulated by transcription factors. Rectangle color indicates COG annotation category, as indicated in the inset. Rectangle size is proportionate to the number of regulated genes with the indicated COG annotation. Edges between nodes indicate information-driven interactions between chemoinformatic features and transcription factors (red arrows) and transcription factors and group of regulated genes (blue arrows).

and the remaining 313 significantly differentially regulated genes. Genes were grouped into sets that were coregulated with the sulfur regulome-associated TFs. A visualization of the *Pseudomonas* sulfur regulome network is shown in Fig. 5.

Three TetR family TFs (PFLU3284, PFLU4781, and PFLU5852) exclusively regulate genes annotated as "metabolism" related, with the largest subgroup within metabolism being "amino acid transport and metabolism." TetR family TFs are also the only transcription factors predicted to be regulated by the chemoinformatic attributes of C-S and C-O bonds, which are predicted to play an important role in the sulfur regulome. The members of the TetR family of transcriptional regulators are one-component signal transduction systems, in which a ligand binds directly to the transcription factor to regulate transcription factor activity. TetR family members are known to bind to a wide range of ligands and to regulate a variety of biological functions, including antibiotic resistance, metabolism, and quorum sensing. From the results of this model, we hypothesize that transcription factors PFLU3284 and PFLU5852 directly bind sulfur-containing nutrients or amino acids. Note that at the time of writing, there was no available molecular characterization of these proteins to support our hypothesis.

The network can be examined to identify the portions of the regulome that are predicted to respond specifically to sulfur. In a subnetwork that is poorly connected to the rest of the network (Fig. 5), the bond between a sulfur atom and an oxygen atom uniquely drives the expression of TetR family TF PFLU5852 and regulates genes annotated as "inorganic ion transport" genes. This subnetwork suggests that a portion of the regulome is devoted to detection of and response to sulfates (i.e., 2-aminoethyl

hydrogen sulfate, potassium 4-nitrophenyl sulfate, and sodium sulfate). The number of atomic bonds between sulfur and hydrogen atoms is found to drive the expression of the members of GntR family PLU1958 and LysR family PFLU5186 TFs.

The nonsulfur components of the selected nutrient molecules also have a predicted effect on the regulome. In fact, sulfur itself is not the most significant factor that drives gene expression patterns in this regulome, indicating that the “sulfur regulome” in fact incorporates interactions with a wider array of biological functions than the incorporation of sulfur into metabolism. The chemoinformatic attributes that are the largest drivers of the complete regulome, identified as the number of child nodes in the network, are the numbers of C-N bonds and the counts of atomic nitrogen in the nutrient. The genes associated with “carbohydrate transport and metabolism” appear exclusively regulated by the chemoinformatic attributes consisting of C-O and C-N bonds and C atoms through the action of members of LysR family TF PFLU0548. This suggests that while sulfur may influence a broad range of biological functions, carbon and nitrogen present in the media primarily affect metabolism.

Summary. We have utilized a transmitter-channel-receiver scheme to model the *P. fluorescens* sulfur regulome. The input to this model is a vector of chemoinformatic attributes that can be used to potentially describe a wide range of organosulfur compounds. While this analysis does not provide evidence that the chemoinformatic features chosen for the model are related to the features that *P. fluorescens* actually utilizes to recognize environmental nutrients, our results support the general hypothesis initially proposed: the bacterial regulome responds to a complex environment through a set of overlapping sensor functions that integrate environmental data to drive specific patterns of gene expression. The unique expression profiles of 14 TFs can be linked to 1 of 10 possible sulfur nutrient environments and to predict the expression patterns of hundreds of other genes. The prediction of gene expression patterns is more accurate using a model that considers a transmitter-channel-receiver scheme than one that attempts to predict gene expression directly from extracellular chemoinformatic features, implying that there is indeed both utility and biological relevance in the structure of the computational model.

Our model allows one to formulate some specific hypotheses about the environmental attributes used by *P. fluorescens*, and these could be tested experimentally in the future. For example, we have previously described a combination of biophysical and biochemical assays to identify the ligand binding specificity of proteins (31–33) that could be directly applied to characterize the ligands of our selected transcription factors. Further validation of the sulfur regulome model could be achieved by collecting transcriptomic data from our transcription factor knockout mutants across sulfur sources. In addition, the model allows one to understand how to introduce additional biochemical strategies (“knock-in” of function) that would allow the utilization of a new panel of nutrients. A means of understanding how a bacterium parameterizes and regulates the utilization of common environmental nutrients, such as is provided by our modeling approach, is needed to enable engineering approaches to utilize advanced strains for conversion of exotic feedstocks in biomanufacturing processes.

These observations have general significance with respect to our understanding of and ability to model complex bacterial regulomes. We propose that bacteria undergo continuous evolutionary pressure to maximize error detection/correction across potentially noisy channels and to maximize the information content of information-containing interactions while minimizing the number of discrete biological elements (e.g., proteins, genes, DNA binding motifs) required for the collection, storage, and manipulation of information. We additionally propose that maximizing data compression also represents an evolutionary pressure that shapes bacterial regulomes. While it could be trivially calculated that, in the computational model, 14 TFs can effectively encode 25 chemoinformatic features (1.8-fold data compression) or that 14 transcription factors encode the expression features of 313 genes (22-fold data compression), it is not a metric that is likely to provide meaningful biological insights into the sulfur

regulome. Nonetheless, it is likely that efficient data compression plays a role in the regulome. For example, considering only three possible states (“upregulated,” “down-regulated,” and “no change” in expression) per transcription factor, the total possible number of nutrients that could be encoded by 14 TFs is ($3^{14} =$) 4,782,969. Extrapolating from this, it is easy to see how a bacterium could potentially store information on high numbers of potential environmental conditions utilizing a relatively small number of TFs. Application of information theory and implementation of quantifiable metrics with respect to the design and optimization of proposed biological regulatory networks will provide vital tools for the understanding, computational modeling, and rational engineering of bacterial regulomes. This understanding is required for optimization of any strain that is planned to be used for conversion of complex feedstocks in biomanufacturing strategies.

MATERIALS AND METHODS

***Pseudomonas* strain and annotated genome.** *Pseudomonas fluorescens* SBW25 was a gift from Gail Preston, Department of Plant Sciences, Oxford University, United Kingdom. The *P. fluorescens* SBW25 annotated genome, predicted gene sequences, and functional annotations of gene products were collected from “The *Pseudomonas* Genome DB” (<http://www.pseudomonas.com/>), which ascribes 6,106 genes to the SBW25 genome (34, 35).

***Pseudomonas* utilization of organosulfur sources.** A modified *Pseudomonas* minimal medium (PMM) (36) was used to study gene expression patterns in response to a shift from rich media to a single organosulfur source. The medium was modified to use $MgCl_2$ instead of $MgSO_4$ as a source of magnesium (0.2 g/liter KCl, 1 g/liter $NH_4H_2PO_4$, 2.3 g/liter $NaH_2PO_4 \cdot H_2O$, 4.96 g/liter Na_2HPO_4 , 0.4 g/liter $MgCl_2$, 0.054 mg/liter $FeCl_3$, 0.2% [wt/vol] glucose). The following sulfur sources were added to PMM (final concentration, 1.62 mM) (31): sodium sulfate (Na_2SO_4), L-cysteine, L-methionine, α -keto- γ (methylthio) butyric acid, taurine, 2-aminoethyl hydrogen sulfate, potassium 4-nitrophenyl sulfate, and reduced glutathione. L-Methionine sulfone is poorly soluble and was added at its maximum concentration of 0.4 mM. A modified PMM with no added sulfur was also utilized.

To analyze growth after a shift from a rich media to a minimal media with a single organosulfur source, cells were grown overnight on Luria broth plates (LB; 10-g/liter tryptone, 5-g/liter yeast extract, 5 g/liter NaCl) and inoculated into 5-ml precultures grown with shaking at 225 rpm at 28°C for 4 h. Cells were pelleted by centrifugation, washed four times in phosphate-buffered saline (PBS), and resuspended in 1 ml PBS, and the cell suspension was adjusted to an OD_{600} of 1.0. In a 96-well microtiter plate, 200 μ l of modified PMM supplemented with a single sulfur source was added per well and inoculated with the cell suspension to an OD_{600} of 0.1. Cultures were incubated for 48 h in a Hidex Sense plate reader (Hidex, Turku, Finland), and OD_{600} readings were taken at 20-min intervals. The lag time value was calculated as the first time point at which the OD_{600} was consistently greater than the average value plus 2 standard deviations of the OD_{600} for the first 2 h of growth. The growth phase value was calculated as the difference between the lag time, calculated as described above, and the time point of maximum OD_{600} . All growth data can be found in Table S1 in the supplemental material. Wild-type growth data can be seen in Fig. 1, and all mutant growth patterns can be seen in Fig. S1 in the supplemental material.

RNA extraction for transcriptomic analysis. For transcriptomic analysis, *P. fluorescens* SBW25 was precultured in LB overnight at 30°C, and cells were washed, resuspended in 1 vol of PBS (pH 7.4), and used to inoculate 25-ml cultures of modified PMM supplemented with a unique sulfur source. Cultures were grown in 125-ml Erlenmeyer flasks at 30°C with shaking at 225 rpm and harvested 16 h after inoculation into modified PMM. Approximately 3×10^9 cells were treated with 2 volumes of Qiagen RNeasy Protect Bacteria reagent (Qiagen, Hilden, Germany) for 5 min at room temperature, and cell pellets were frozen at $-80^\circ C$. For each condition, three independent biological replicates were collected.

Total bacterial RNA was extracted using a Qiagen RNeasy minikit and the method described for enzymatic lysis of bacteria. A combination of RNase-free Rlysozyme (EMD-Millipore, Darmstadt, Germany) at 100 kU/ml and 3 freeze-thaw cycles was used to lyse cells. An on-column DNase I digestion was performed as described in the kit instructions, after which total RNA was eluted in nuclease-free water and stored at $-80^\circ C$. To remove residual genomic DNA, all RNA samples were treated with one unit of Baseline-Zero RNase-free DNase I (Illumina/Epicentre, San Diego, CA) per 5 μ g of total bacterial RNA for 15 min at 37°C. Digestion was stopped by immediate cleanup with RNA Clean and Concentrator-5 (Zymo Research, Irvine, CA) spin columns. RNA quality was examined by using a Bioanalyzer and an RNA 6000 Nano Chip (Agilent, Santa Clara, CA). All RNA samples used for transcriptome sequencing (RNA-seq) library preparation had an RNA integrity number higher than 9.4. To deplete ribosomal RNAs, samples were concentrated and treated with Ribo-Zero (Bacteria) probes (Illumina/Epicentre, San Diego, CA) according to the manufacturer’s instructions. Depleted RNA samples were eluted into RNase-free water and characterized using Agilent RNA Pico 6000 chips (Agilent, Santa Clara, CA) to confirm removal of 16S and 23S rRNA subunits. All samples still exhibited peaks at sizes consistent with the presence of tRNAs and 5S rRNAs.

Sequencing libraries were produced from mRNA amounts ranging from 12.6 to 50.1 ng using a Script-Seq version 2 kit (Illumina/Epicentre, San Diego, CA) and following the manufacturer’s instructions. Next-generation sequencing services were provided by the High-throughput Genome Analysis Core of the Institute for Genomics and Systems Biology (IGSB) at the University of Chicago.

Generation of TF knockout mutants. Selected genes encoding transcription factors were deleted from the *P. fluorescens* SBW25 genome by homologous recombination as represented in Fig. S2. Briefly, regions of ~1 kb in length flanking the targeted transcription factor coding region were PCR amplified using SBW25 genomic DNA as the template. Target-proximal primers were extended with 15-bp to 20-bp sequences complementary to a DNA cassette carrying tetracycline resistance genes (37). The two genome fragments and the cassette were joined by assembly cloning methods. Electroporation was used to transform the resulting linear DNA fragments into SBW25 cells expressing RecET-like phage recombinases from a plasmid. The expressed recombinases stimulated the homologous recombination of the targeted gene with the antibiotic cassette in a reaction similar to that previously described for *Pseudomonas syringae* (38, 39), resulting in replacement of the targeted sequence with the antibiotic resistance genes on the host chromosome. The primers used to construct the mutants are described in Table S2. A Bio-Rad Gene Pulser Xcell system (Bio-Rad, Hercules, CA) was used with settings for *P. aeruginosa* (25 μ F, 200 Ω , 2,500 V) for all transformations performed with SBW25. Transformants were selected on solid LB media containing 15 μ g/ml tetracycline, after which gene replacement was verified by colony PCR and two independent isolates were cured of the recombinase plasmid prior to further characterization. For each isolate, a 5-kb-to-6-kb region encompassing the homologous integration site was PCR amplified and sequenced. Single base pair changes were found sporadically in regions corresponding to primer sites, suggesting that mutations were most likely introduced by the use of synthesized DNA primers. In contrast, no mutations were detected in the flanking chromosomal coding regions.

Analysis of transcriptomic data. Gene expression levels were calculated from RNA-seq reads using “BowStrap” (40) and predicted gene coding sequences of SBW25 (34, 35). BowStrap performs a bootstrap analysis on the output of the short-sequence-aligning program “Bowtie” (<http://bowtie-bio.sourceforge.net/index.shtml>). In BowStrap, both unique and multiply aligned reads are considered as a means of generating a measure of gene model expression with accompanying data representing confidence interval and statistical significance of expression.

Transcriptome data are presented as \log_2 values determined for the number of aligned reads per 1,000 base pairs of gene per million aligned sequence reads (reads per kilobase per million [RPKM] values) and were normalized by quantile normalization (41). The complete set of gene expression data is available in the supplemental material.

Statistically significant differential expression (DE) of genes was determined by ANOVA in MeV4 (<http://mev.tm4.org>) with *P* values calculated from 10,000 permutations, and the data were adjusted for false-discovery rate (FDR) by the use of the Bonferroni method (42). An FDR-corrected *P* value of 0.05 was used as the threshold for significant differential gene expression. The complete set of normalized RPKM gene expression data can be found in Table S3.

Fourteen of the genes identified as showing DE by ANOVA are annotated as TFs. As ANOVA considers differential expression as a function of variance within a treatment relative to variance across all observations, ANOVA cannot provide a measure of fold change relative to a reference condition. To calculate a relative fold change value for TF expression patterns, an additional level of DE was calculated. Fold change and the significance of fold changes for the 14 ANOVA-identified TFs were calculated relative to the “no-sulfur” medium condition using the 2-tailed *t* test (*P* value < 0.05).

Annotations of clusters of orthologous groups (COG) of proteins (25) were used to determine whether subsets of SDE genes were enriched for biological functions. Enrichment for specific annotation was determined using *P* values, calculated as 1 minus the hypergeometric distribution relative to the total number of genes with that annotation in the complete SBW25 genome. A threshold of a *P* value of less than 0.05 was used for statistical significance determinations.

Prediction of the percentage of sulfur-containing amino acids in a proteome from transcriptomic data. The sulfur content of the predicted *Pseudomonas* proteome was estimated from transcriptomic data. The following formula was used for predicting proteome sulfur content:

$$\text{predicted proteome \% sulfur} = \frac{\sum_{i=1}^m \text{Gene}_i * \text{SulfurousAA}_i}{\sum_{i=1}^m \text{Gene}_i * \text{TotalAA}_i} \quad (1)$$

where *m* is the total number of genes in the *P. fluorescens* genomes, Gene_i is the normalized bootstrapped RPKM expression of gene *i*, SulfurousAA_i is the number of sulfur-containing amino acids (i.e., cysteine and methionine) in the protein coded by gene *i*, and TotalAA_i is the total number of amino acids in the protein coded by gene *i*.

Identification of the genes controlled by sulfur regulome TFs. To identify the genes potentially regulated directly by a change in expression of a TF, we calculated the PCC values of gene expression using comparisons between the set of 14 TFs and the remaining 313 DE genes. We considered a gene to be potentially directly regulated by a TF if the PCC value of the pair was greater than the average plus 1 standard deviation of all PCC values. The coregulation of a TF and a gene does not necessarily require that the TF directly controls the expression of the gene.

Calculation of Shannon’s entropy associated with each TF. Shannon’s entropy is a quantification of the expected value of the information contained in a message, measured as the reduction of uncertainty. In this case, the “message” is defined as an observed, significant change in TF expression. Differential expression of a TF (by ANOVA) reduces the uncertainty regarding the bacterium’s nutrient environment. A change in TF expression is defined as a statistically significant result (2-tailed *t* test *P* value less than 0.05) relative to expression in sodium sulfate growth condition. For example, a significant change in expression in TF PFLU4596 in this experiment indicates that the nutrient present in the environment is 2-aminoethyl hydrogen sulfate or cysteine or potassium 4-nitrophenyl sulfate,

reducing the uncertainty regarding the environment from nine possible messages describing environmental conditions to three. For this experiment, the Shannon's entropy H value for a TF is calculated as follows:

$$H = - \sum_{i=1}^n \frac{i}{n} \log_2 \left(\frac{i}{n} \right) \quad (2)$$

where n is the number of possible sulfur nutrients associated with a significant change in TF expression.

Generation of a model of the sulfur regulome. There are two main components of the sulfur regulome model: (i) modeling the TF profile as a function of sulfur nutrient chemoinformatic attributes and (ii) modeling gene expression as a function of the TF profile. For modeling, all sulfur nutrient chemoinformatic attributes and gene expression levels were normalized to arbitrary values between 1 and 100. All models were calculated using leave-one-out cross-validation (LOO-CV), a special case of a K-fold cross-validation. Only the results from the validation sets are presented here.

(i) TF expression as a function of sulfur nutrient chemoinformatic attributes. In the first part of the proposed model of the sulfur regulome, environmental information detected by the receiver is encoded into a TF expression profile in the channel. The relationship can be defined as follows:

$$TF_j = f(\text{Chem}_1, \dots, \text{Chem}_{25}) \quad (3)$$

where TF_j is the expression level of TF j and Chem_{1-25} is the vector of the 25 chemoinformatic attributes for a sulfur nutrient condition. The program "Eureqa" (Nuttonian, Boston, MA) was used to generate equations that best fit the observed data. "Eureqa" is an artificial intelligence (AI) modeling engine that uses an evolutionary algorithm approach to finding optimized equations to fit experimental data using a user-selected set of allowed mathematical operations. The operators constant, addition, subtraction, multiplication, and division were used, and the equation fitting was allowed to continue until the values corresponding to equation "stability" and "maturity" each exceeded 90%. The set of equations describing TF expression profile as a function of environmental chemoinformatic attributes can be found in Table S4.

(ii) Gene expression as a function of the TF expression profile. The set of genes regulated by the sulfur regulome in the receiver can be described as a function of TF profile of the channel using the following equation:

$$G_i = c_i + \sum_{j=1}^{TF_{\max}} TF_j \times w_{i,j} \quad (4)$$

where G_i is the expression level of gene i , c_i is a constant associated with gene i , TF_j is the expression of TF j in the set of TF_{\max} number of TFs, and $w_{i,j}$ is the weight of the effect of TF j on gene i . The set of edge weights describing gene expression as a function of TF expression profile can be found in Table S5.

As a control method, gene expression patterns were also described directly as a function of chemoinformatic attributes as follows:

$$G_i = c_i + \sum_{k=1}^{\text{Chem}_{\max}} \text{Chem}_k \times w_{i,k} \quad (5)$$

where Chem_{\max} is the number of chemoinformatic attributes and $w_{i,k}$ is the weight of the effect of chemoinformatic feature k on gene i .

Equations were solved as a set of underdetermined linear equations using QR decomposition (where Q represents an orthogonal matrix and R represents an upper triangular matrix) for solving linear least-square equations in "R." The set of edge weights describing gene expression as a function of chemoinformatic attributes can be found in Table S6.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/mSystems.00189-17>.

FIG S1, TIF file, 0.3 MB.

FIG S2, TIF file, 0.04 MB.

TABLE S1, XLSX file, 0.2 MB.

TABLE S2, TXT file, 0 MB.

TABLE S3, XLSX file, 2.5 MB.

TABLE S4, XLSX file, 0.02 MB.

TABLE S5, XLSX file, 0.4 MB.

TABLE S6, XLSX file, 0.5 MB.

ACKNOWLEDGMENTS

We thank Sara Forrester for excellent technical help and Danielle M. Larsen for a critical read of the manuscript.

Research under the "Environment Sensing and Response" scientific focus area (SFA) at Argonne National Laboratory was supported by the Genomic Science Program, Office of Biological and Environmental Research (BER), Office of Science of

the U.S. Department of Energy (DOE), operated by UChicago Argonne, LLC, under contract DE-AC02-06CH11357. This work also was supported in part by the Agile BioFoundry (<http://agilebiofoundry.org>), supported by the U.S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, through contract DE-AC02-05CH11231. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

REFERENCES

- Kaiser D. 1979. Social gliding is correlated with the presence of pili in *Myxococcus xanthus*. *Proc Natl Acad Sci U S A* 76:5952–5956. <https://doi.org/10.1073/pnas.76.11.5952>.
- Dworkin M. 1983. Tactic behavior of *Myxococcus xanthus*. *J Bacteriol* 154:452–459.
- Prindle A, Liu J, Asally M, Ly S, Garcia-Ojalvo J, Süel GM. 2015. Ion channels enable electrical communication in bacterial communities. *Nature* 527:59–63. <https://doi.org/10.1038/nature15709>.
- Humphries J, Xiong LY, Liu JT, Prindle A, Yuan F, Arjes HA, Tsimring L, Süel GM. 2017. Species-independent attraction to biofilms through electrical signaling. *Cell* 168:200–209.e12. <https://doi.org/10.1016/j.cell.2016.12.014>.
- Pérez J, Moraleda-Muñoz A, Marcos-Torres FJ, Muñoz-Dorado J. 2016. Bacterial predation: 75 years and counting! *Environ Microbiol* 18:766–779. <https://doi.org/10.1111/1462-2920.13171>.
- Tyc O, Song C, Dickschat JS, Vos M, Garbeva P. 2017. The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends Microbiol* 25:280–292. <https://doi.org/10.1016/j.tim.2016.12.002>.
- Jousset A. 2012. Ecological and evolutive implications of bacterial defences against predators. *Environ Microbiol* 14:1830–1843. <https://doi.org/10.1111/j.1462-2920.2011.02627.x>.
- Sanz Y, Moya-Perez A. 2014. Software tools and algorithms for biological. *Systems* 817:291–317.
- Borre YE, Moloney RD, Clarke G, Dinan TG, Cryan JF. 2014. Software tools and algorithms for biological. *Systems* 817:373–403.
- Foster JA, McVey Neufeld KA. 2013. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci* 36:305–312. <https://doi.org/10.1016/j.tins.2013.01.005>.
- Xiao Y, Wei X, Ebricht R, Wall D. 2011. Antibiotic production by myxobacteria plays a role in predation. *J Bacteriol* 193:4626–4633. <https://doi.org/10.1128/JB.05052-11>.
- Barnard AM, Bowden SD, Burr T, Coulthurst SJ, Monson RE, Salmond GP. 2007. Quorum sensing, virulence and secondary metabolite production in plant soft-rotting bacteria. *Philos Trans R Soc Lond B Biol Sci* 362:1165–1183. <https://doi.org/10.1098/rstb.2007.2042>.
- Lyon P. 2015. The cognitive cell: bacterial behavior reconsidered. *Front Microbiol* 6:264. <https://doi.org/10.3389/fmicb.2015.00264>.
- Baker MD, Stock JB. 2007. Signal transduction: networks and integrated circuits in bacterial cognition. *Curr Biol* 17:R1021–R1024. <https://doi.org/10.1016/j.cub.2007.10.011>.
- Lan G, Tu Y. 2016. Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Rep Prog Phys* 79:052601. <https://doi.org/10.1088/0034-4885/79/5/052601>.
- Shapiro JA. 2007. Bacteria are small but not stupid: cognition, natural genetic engineering and socio-bacteriology. *Stud Hist Philos Biol Biomed Sci* 38:807–819. <https://doi.org/10.1016/j.shpsc.2007.09.010>.
- Ben Jacob E, Shapira Y, Tauber Al. 2006. Seeking the foundations of cognition in bacteria: from Schrödinger's negative entropy to latent information. *Phys A Stat Mech Appl* 359:495–524. <https://doi.org/10.1016/j.physa.2005.05.096>.
- Straube R. 2017. Analysis of network motifs in cellular regulation: structural similarities, input-output relations and signal integration. *Biosystems* 162:215–232. <https://doi.org/10.1016/j.biosystems.2017.10.012>.
- Dendooven T, Luisi BF. 2017. RNA search engines empower the bacterial intranet. *Biochem Soc Trans* 45:987–997. <https://doi.org/10.1042/BST20160373>.
- Valentini M, Filloux A. 2016. Biofilms and cyclic di-GMP (c-di-GMP) signaling: lessons from *Pseudomonas aeruginosa* and other bacteria. *J Biol Chem* 291:12547–12555. <https://doi.org/10.1074/jbc.R115.711507>.
- Barato AC, Hartich D, Seifert U. 16 October 2014. Efficiency of cellular information processing. *New J Phys* 16. <https://doi.org/10.1088/1367-2630/16/10/103024>.
- Riera-Fernández P, Munteanu CR, Escobar M, Prado-Prado F, Martín-Romalde R, Pereira D, Villalba K, Duardo-Sánchez A, González-Díaz H. 2012. New Markov-Shannon entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway, parasite-host, neural, industry, and legal-social networks. *J Theor Biol* 293:174–188. <https://doi.org/10.1016/j.jtbi.2011.10.016>.
- Kendall SL, Movahedzadeh F, Wietzorrek A, Stoker NG. 2002. Microarray analysis of bacterial gene expression: towards the regulome. *Comp Funct Genomics* 3:352–354. <https://doi.org/10.1002/cfg.193>.
- Vicente M, Mingorance J. 2008. Microbial evolution: the genome, the regulome and beyond. *Environ Microbiol* 10:1663–1667. <https://doi.org/10.1111/j.1462-2920.2008.01635.x>.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.
- Fauchon M, Lagniel G, Aude JC, Lombardia L, Soularue P, Petat C, Marguerie G, Sentenac A, Werner M, Labarre J. 2002. Sulfur sparing in the yeast genome in response to sulfur demand. *Mol Cell* 9:713–723. [https://doi.org/10.1016/S1097-2765\(02\)00500-2](https://doi.org/10.1016/S1097-2765(02)00500-2).
- Shannon CE, Weaver W. 1949. The mathematical theory of communication, p 117. University of Illinois Press, Urbana, Illinois.
- Masip L, Veeravalli K, Georgiou G. 2006. The many faces of glutathione in bacteria. *Antioxid Redox Signal* 8:753–762. <https://doi.org/10.1089/ars.2006.8.753>.
- Sporer AJ, Kahl LJ, Price-Whelan A, Dietrich LEP. 2017. Redox-based regulation of bacterial development and behavior. *Annu Rev Biochem* 86:777–797. <https://doi.org/10.1146/annurev-biochem-061516-044453>.
- Hamming RW. 1982. Citation classic - error detecting and error correcting codes. *Curr Contents Eng Technol Appl Sci* 31:18–31.
- Zerbs S, Korajczyk PJ, Noirot PH, Collart FR. 2017. Transport capabilities of environmental *Pseudomonads* for sulfur compounds. *Protein Sci* 26:784–795.
- Michalska K, Chang C, Mack JC, Zerbs S, Joachimiak A, Collart FR. 2012. Characterization of transport proteins for aromatic compounds derived from lignin: benzoate derivative binding proteins. *J Mol Biol* 423:555–575. <https://doi.org/10.1016/j.jmb.2012.08.017>.
- Tan K, Chang C, Cuff M, Osipiuk J, Landorf E, Mack JC, Zerbs S, Joachimiak A, Collart FR. 2013. Structural and functional characterization of solute binding proteins for aromatic compounds derived from lignin: p-coumaric acid and related aromatic acids. *Proteins* 81:1709–1726. <https://doi.org/10.1002/prot.24305>.
- Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SA, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hotherhall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, Thomson NR. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* 10:R51. <https://doi.org/10.1186/gb-2009-10-5-r51>.
- Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FS. 2016. Enhanced annotations and features for comparing thousands of *Pseu-*

- domonas genomes in the Pseudomonas genome database. *Nucleic Acids Res* 44:D646–D653. <https://doi.org/10.1093/nar/gkv1227>.
36. Bolton H, Elliott LF, Gurusiddaiah S, Fredrickson JK. 1989. Characterization of a toxin produced by a rhizobacterial *Pseudomonas* sp. that inhibits wheat growth. *Plant Soil* 114:279–287. <https://doi.org/10.1007/BF02220808>.
 37. Heeb S, Itoh Y, Nishijyo T, Schnider U, Keel C, Wade J, Walsh U, O’Gara F, Haas D. 2000. Small, stable shuttle vectors based on the minimal pVS1 replicon for use in gram-negative, plant-associated bacteria. *Mol Plant Microbe Interact* 13:232–237. <https://doi.org/10.1094/MPMI.2000.13.2.232>.
 38. Swingle B, Bao ZM, Markel E, Chambers A, Cartinhour S. 2010. Recombineering using RecTE from *Pseudomonas syringae*. *Appl Environ Microbiol* 76:4960–4968. <https://doi.org/10.1128/AEM.00911-10>.
 39. Bao Z, Cartinhour S, Swingle B. 2012. Substrate and target sequence length influence RecTE(Psy) recombineering efficiency in *Pseudomonas syringae*. *PLoS One* 7:e50617. <https://doi.org/10.1371/journal.pone.0050617>.
 40. Larsen PE, Collart FR. 2012. BowStrap v1.0: assigning statistical significance to expressed genes using short-read transcriptome data. *BMC Res Notes* 5:275. <https://doi.org/10.1186/1756-0500-5-275>.
 41. Bolstad BM, Irizarry RA, Astrand M, Speed TP. 2003. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19:185–193. <https://doi.org/10.1093/bioinformatics/19.2.185>.
 42. Hochberg Y. 1988. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 75:800–802. <https://doi.org/10.1093/biomet/75.4.800>.