Systematic Discovery of Archaeal Transcription Factor Functions in Regulatory Networks through Quantitative Phenotyping Analysis

Cynthia L. Darnell,[a] Peter D. Tonner,[a,b] Jordan G. Gulli,[a*] Scott C. Schmidler,[b,c,d] Amy K. Schmid[a,b]

Department of Biology, Duke University, Durham, North Carolina, USA[a]; Program in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina, USA[b]; Department of Statistical Science, Duke University, Durham, North Carolina, USA[c]; Department of Computer Science, Duke University Durham, North Carolina, USA[d]

ABSTRACT Gene regulatory networks (GRNs) are critical for dynamic transcriptional responses to environmental stress. However, the mechanisms by which GRN regulation adjusts physiology to enable stress survival remain unclear. Here we investigate the functions of transcription factors (TFs) within the global GRN of the stress-tolerant archaeal microorganism *Halobacterium salinarum*. We measured growth phenotypes of a panel of TF deletion mutants in high temporal resolution under heat shock, oxidative stress, and low-salinity conditions. To quantitate the noncanonical functional forms of the growth trajectories observed for these mutants, we developed a novel modeling framework based on Gaussian process regression and functional analysis of variance (FANOVA). We employ unique statistical tests to determine the significance of differential growth relative to the growth of the control strain. This analysis recapitulated known TF functions, revealed novel functions, and identified surprising secondary functions for characterized TFs. Strikingly, we observed that the majority of the TFs studied were required for growth under multiple stress conditions, pinpointing regulatory connections between the conditions tested. Correlations between quantitative phenotype trajectories of mutants are predictive of TF-TF connections within the GRN. These phenotypes are strongly concordant with predictions from statistical GRN models inferred from gene expression data alone. With genome-wide and targeted data sets, we provide detailed functional validation of novel TFs required for extreme oxidative stress and heat shock survival. Together, results presented in this study suggest that many TFs function under multiple conditions, thereby revealing high interconnectivity within the GRN and identifying the specific TFs required for communication between networks responding to disparate stressors.

IMPORTANCE To ensure survival in the face of stress, microorganisms employ inducible damage repair pathways regulated by extensive and complex gene networks. Many archaea, microorganisms of the third domain of life, persist under extremes of temperature, salinity, and pH and under other conditions. In order to understand the cause-effect relationships between the dynamic function of the stress network and ultimate physiological consequences, this study characterized the physiological role of nearly one-third of all regulatory proteins known as transcription factors (TFs) in an archaeal organism. Using a unique quantitative phenotyping approach, we discovered functions for many novel TFs and revealed important secondary functions for known TFs. Surprisingly, many TFs are required for resisting multiple stressors, suggesting cross-regulation of stress responses. Through extensive validation experiments, we map the physiological roles of these novel TFs in stress

Address correspondence to Amy K. Schmid, amy.schmid@duke.edu.

* Present address: Jordan G. Gulli, Department of Biology, Georgia Institute of Technology, Atlanta, Georgia, USA.

C.L.D. and P.D.T. contributed equally to this work.

High throughput phenotyping reveals novel transcription factors and regulatory networks in archaea.

response back to their position in the regulatory network wiring. This study advances understanding of the mechanisms underlying how microorganisms resist extreme stress. Given the generality of the methods employed, we expect that this study will enable future studies on how regulatory networks adjust cellular physiology in a diversity of organisms.

KEYWORDS *Archaea*, functional ANOVA, phenomics, transcription factors

Free-living cells experience frequent stress from the extracellular environment. Transcription factors (TFs) and their regulons of target genes comprise a gene regulatory network (GRN), which functions to alter gene expression dynamically in response to stressful and changing environments. Many environmental conditions are chemically and/or physically inextricable (e.g., oxygen levels and salinity) (1), and different types of stresses can cause similar types of cellular damage (2, 3). For example, both exposure to excess metals and radiation can result in redox imbalance (4). Disparate stressors also elicit similar gene expression programs (5). These observations have led to the hypothesis that GRNs responding to each stressor are highly interconnected (6–8). However, the specific mechanisms that underlie these connected responses remain unclear.

Microorganisms known as extremophiles thrive in environments at the limits of life, representing model systems well suited for understanding how GRNs enable physiological adjustment to strong environmental forces. One group of extremophiles, the hypersaline-adapted archaea, colonize salt lakes where salt concentrations can reach saturation. Fluctuations in temperature and oxygen level, intense radiation, and desiccation/rehydration cycles pose a constant challenge to macromolecular and cellular integrity (9). To respond, archaea use a hybrid system of bacterial-like and eukaryotic-like proteins to regulate transcription. The basal transcriptional machinery resembles that of eukaryotes, including RNA polymerase (RNAP), TATA-binding proteins (TBPs), and transcription factor IIB (TFIIB) homologs (10–12). In contrast, the regulatory proteins are homologous to those found in bacteria, such as TFs containing a helix-turn-helix (HTH) DNA binding domain (13). Many archaeal TFs directly sense environmental changes, binding ligands or changing redox status to alter TF conformation and TF-DNA binding (14).

In the genetically tractable hypersaline-adapted species *Halobacterium salinarum*, GRN inference (7, 15) and subsequent validation experiments suggested that the GRN is required for dynamic adjustment of gene expression in response to extreme and interconnected stress regimes (16–19). Specifically, the integration of transcriptome data in response to environmental and genetic perturbations (1, 7, 15, 17, 19–25), gene functions (26, 27), and *cis*-regulatory motif predictions in the context of statistical inference algorithms (28, 29) resulted in a genome-wide Environment and Gene Regulatory Influence Network (EGRIN) model that predicted regulatory connections for more than 70 TFs and their target genes (7, 15). More recently, similar GRN models have been constructed for other species of archaea (30) and bacteria (31). These models are highly predictive of gene expression in response to stress and enable generation of novel hypotheses regarding the roles of TFs in stress response.

However, for organisms across the domains of life, it remains a central challenge to decipher whether and how genetic and environmental perturbation to the GRN directly impacts cellular phenotype and survival in ecologically relevant contexts (3). Systematic "phenomics" approaches hold promise for understanding the roles of TFs and other regulators in GRNs and how this role impacts cellular physiology (32–34); however, relative to other systems biology methods such as transcriptomics and proteomics, phenomics remains an underrepresented data source.

In response to these challenges, a library of 27 TF deletion mutants was generated in *H. salinarum*. These mutants were assessed for growth under a variety of stresses endemic to the salt flat environment. A novel nonparametric model was developed using a Gaussian process framework to quantify these phenotypes. We used recently developed statistical tests (35) and developed new tests to identify significant differ-

ential growth between the trajectories of these TF knockouts relative to that of the control strain. The results revealed that a surprising number of TFs are required for optimal growth under multiple stress conditions, indicating a high level of interconnectivity within the GRN. Clustering analysis of phenotype trajectories revealed that TFs with related phenotypes function together, regulating each other and common sets of genes in stress-specific subnetworks. Through further analysis of TF roles in gene regulation, stress-related functions were validated for novel TFs. We detected strong concordance of newly discovered TF functions with statistical predictions of TF gene regulatory relationships from GRN models inferred from gene expression data alone.

RESULTS AND DISCUSSION

Identification of transcription factor candidates. To prioritize candidate TFs for phenotypic characterization, genes encoding TFs were first detected in the *Halobacterium salinarum* genome (27) using the Systems Biology Experimental Management System database (SBEAMS) (36), where annotations were cross-referenced with other databases (PFAM, COG [clusters of orthologous groups of proteins], and PSI-BLAST [37–39]) (Fig. 1). Previous *ab initio* protein structure prediction results (26) and subsequent matches to the Protein Data Bank (PDB) were used to identify possible DNA binding domains in proteins of unknown function. This pipeline resulted in a list of 130 putative TFs, which were included as priors in inference of the *H. salinarum* EGRIN model (7, 15). To identify candidates of interest for further analysis here, TFs were more stringently defined as those with strong homology (E value $\leq 1 \times 10^{-7}$) to a sequence-specific DNA binding domain or structural fold experimentally characterized in other bacterial or archaeal species (13, 26) (Fig. 1). This definition excluded 24 DNA-binding proteins with peripheral roles in transcription (e.g., helicases). Another 18 genes encoding previously published, well-characterized TFs, were also excluded (21, 40–42). Of the 88 remaining TFs, a final subset of 27 were selected for further study based on transcriptional changes during fluctuations in a wide array of environmental conditions (1, 7, 20, 23–25, 43) and functional predictions from EGRIN gene regulatory network models (7, 15) (Fig. 1; see Table S1 in the supplemental material). The TFs in the resultant collection are members of a variety of functional families and contain diverse structural domains (Table S1). These include DNA binding domains known from other archaea and bacteria (winged helix-turn-helix, ribbon-helix-helix) and domains unique to halophilic archaea (e.g., HalX). Some TFs contain ligand binding domains homologous to those of known function in bacteria, such as DtxR family iron-dependent repressors. Other TFs possess domains of novel function specific to halophiles, such as the RosR C-terminal domain (16) (Table S1). Together, the results of these bioinformatic analyses suggest that the panel of selected TFs is representative of the global transcription regulatory landscape of *H. salinarum*.

Quantification of significant differential growth of TF knockout strains under stress. (i) Experimental design and data. To determine the physiological function for each TF and systematically compare these functions across TFs, knockout mutant strains for each of the 27 selected TFs were grown under five conditions: standard growth, low salinity, paraquat (PQ), hydrogen peroxide, and heat shock (see Materials and Methods). The growth of these knockout mutant strains was compared to that of the isogenic parent control strain, Δ*ura3* strain, a uracil auxotroph used to generate knockout mutants (44) (see Materials and Methods). Growth conditions were chosen based on their relevance to the hypersaline habitat of *H. salinarum*. Knockout mutants for 10 of these 27 TFs were constructed in previous studies, where the role of each TF was assessed under a single stress condition (Table 1). These strains were included for the following reasons: (i) as a control to validate our methods and (ii) to test possible secondary functions for these previously studied regulators. Strains with deletion of the genes encoding the remaining 17 TFs of interest were constructed in the current study using established genetic methods for *H. salinarum* (see Materials and Methods) (44) (Table S2). The genotypes of all 27 mutants, regardless of prior publication, were also verified here (see Materials and Methods; Fig. S1 and Table S2). Growth of each
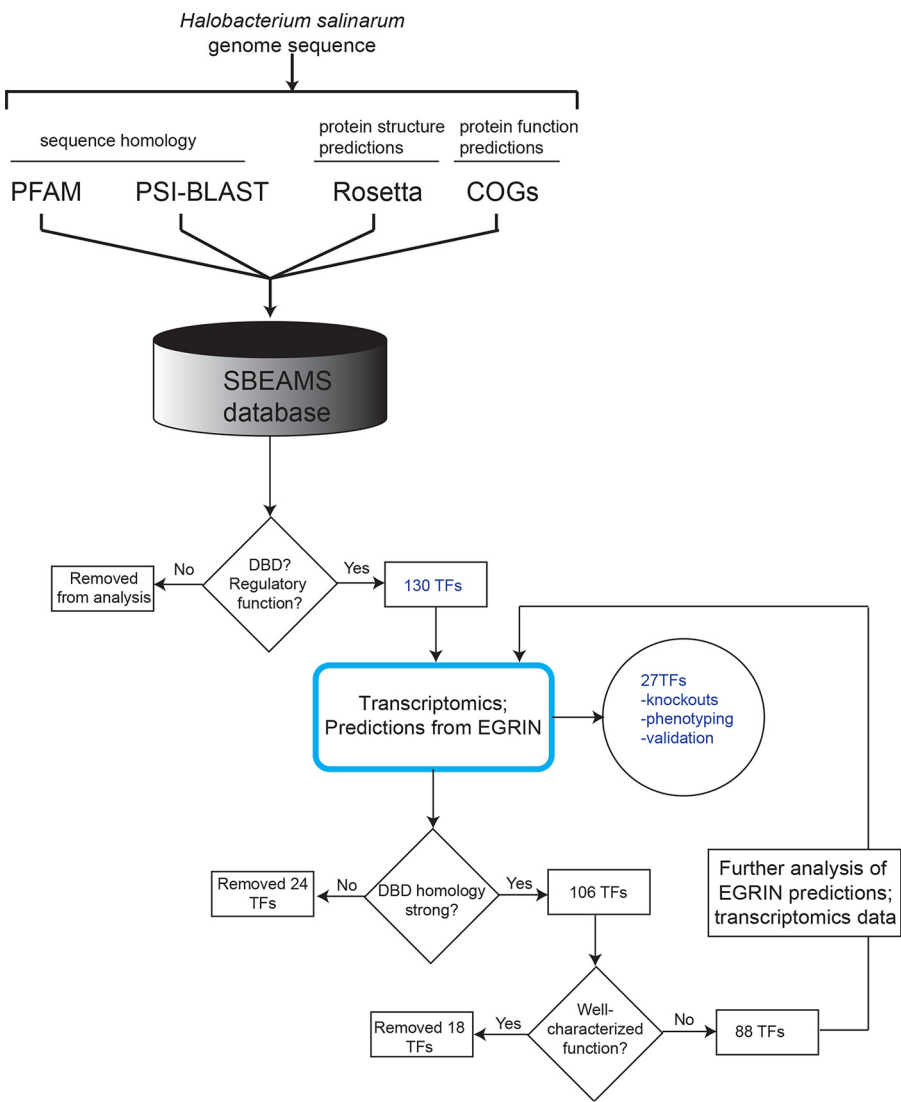
FIG 1 TF candidate selection pipeline. Genes encoding proteins with a putative DNA binding domain were annotated using sequence databases PFAM (37) and PSI-BLAST (38), structural predictions (26), and protein functions from COG (39). These annotations were stored in the Systems Biology Experiment and Analysis Management System (SBEAMS) database (36), resulting in 130 putative TFs. Transcriptome analysis across 1,495 experimental conditions (43) and GRN network inference models (7, 15) were then used to generate predictions regarding TF functions. Details of GRN predictions and criteria for selection of the final collection of 27 TFs are given in Table S1 in the supplemental material. DBD, DNA binding domain.

knockout and parent control strain was measured under each of the five conditions every 30 min for 48 h, resulting in 210,180 data points (raw data given in Table S3).

(ii) Development of a FANOVA model and test for differential growth of knockout mutants under stress. Typically, microbial population growth is modeled using parametric models, such as Gompertz regression (45). The effects of stress and

TABLE 1 Strains used in this study with known phenotypes and functions, and types of evidence previously generated for each strain

Condition	Strain(s)	References	Knockout growth	Gene expression	TF-DNA binding	Metabolomics
Oxidative stress	ΔrosR, ΔVNG0194H, Δhrg, ΔsirR, ΔasnC	16, 35, 50, 54	✓	✓	✓	
Nutrient acquisition	ΔtrmB	17, 18, 48, 49	✓	✓	✓	✓
Manganese and iron homeostasis	Δidr1, Δidr2, ΔsirR	19, 22	✓	✓	✓	
Copper overload	ΔcopR	22, 54	✓	✓	✓	

genetics on growth are then quantified by testing for statistically significant differences in the estimated parameters for different conditions (46). However, we previously showed that the impact of genetics and stress perturbations on growth are more accurately captured with a nonparametric Gaussian process (GP) model (35). In contrast to parametric models, GPs have the advantage of learning these relations directly from the data and do not require explicit equations describing the effects of stress and genetics on growth, which are typically unknown in the case of phenotypic discovery. In this study, we again use GPs to model microbial population growth, now in the form of functional analysis of variance (FANOVA) (47). In the FANOVA setting, microbial population growth is divided into a linear combination of different experimental variables: condition, genetic background, and the interaction between the two (see Materials and Methods) (Fig. 2A). Additionally, FANOVA allows for the explicit comparison of different effects in terms of significance, even if the shapes of those effects are quite different. For these reasons, FANOVA is particularly well adapted to modeling and comparing the effects of many different stress and genetic perturbations. This extends our previous GP model (35) by including all data from multiple conditions and genetic backgrounds into a single model, which allows for direct comparison of different effects in a common framework.

Two metrics were used to assess the significance of differential growth based on our FANOVA model of growth data. The first metric, change in the optical density (OD_Δ) (35), is a function representing the difference in growth levels between the mutant strain [$f_m(x)$] and parent strain [$f_p(x)$] over the duration of the growth curve under a specific condition (Fig. 2B and Fig. S2) as follows:

$$OD_\Delta(t) = f_m(t) - f_p(t) \tag{1}$$

Where the OD_Δ curve differs significantly from zero (based on a 95% credible interval), the growth phenotype is considered significantly different from the parent strain for that section of the growth curve. The second metric, $||OD_\Delta||$, represents the overall magnitude of the functional difference between parent and mutant strains under a specific condition and is calculated based on OD_Δ (Fig. 2C and Fig. S3):

$$||OD_\Delta|| = \sqrt{\sum_{t=t_0}^{t_n} [OD_\Delta(t)]^2} \tag{2}$$

Higher values of $||OD_\Delta||$ indicate an overall larger deviation from parent strain behavior, regardless of positive or negative growth phenotype, and were used as a rank ordering of phenotype severity to prioritize TF mutants for further analysis.

(iii) FANOVA modeling of microbial population growth enables the discovery of new TF knockout phenotypes. On the basis of $||OD_\Delta||$ ranking, the Δ*trmB* strain exhibited the strongest growth impairment under standard conditions (Fig. 2A). This result was expected based on previous characterization of TrmB as a global regulator of nutrient metabolism (17, 18, 35, 48, 49). Also as expected from prior work, the Δ*rosR* mutant with deletion of a key oxidative stress response TF was among the top-ranking quartile of mutants under oxidative stress (16, 35, 50). Consistent with prior Gaussian process model results, the Δ*sirR* strain exhibited a significant growth impairment under peroxide stress (35); indeed, here the Δ*sirR* strain exhibited the top-ranking peroxide phenotype of all mutants tested (Fig. 2C). The consistency of these results with those of prior studies demonstrates the validity of the FANOVA model for recapitulating known phenotypes.

Surprisingly, however, several novel mutant phenotypes were observed. For instance, the Δ*cspD1* mutant showed the strongest growth impairment relative to the parent control strain under PQ stress, and the Δ*copR* mutant exhibited the strongest growth impairment under heat shock (Fig. 2A and B). In order to compare the number of significant mutant phenotypes under each condition, a universal cutoff of $||OD_\Delta||$ of 0.337 was chosen (see Materials and Methods). This cutoff allowed rank ordering of the stress conditions themselves in terms of the strength of perturbation to physiology, as measured by the number of significant mutant phenotypes. Peroxide treatment had
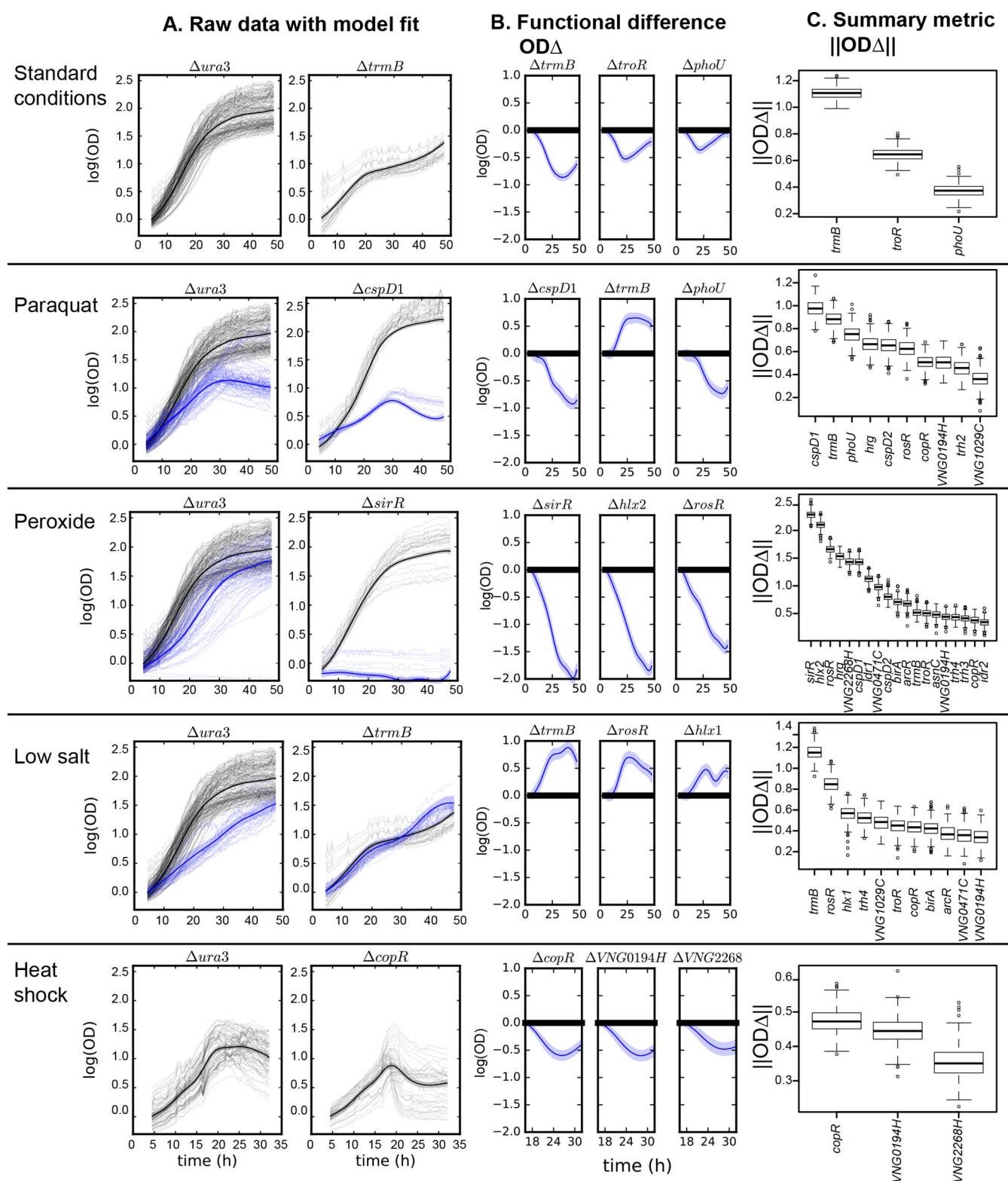
FIG 2 FANOVA modeling and statistical ranking of TF knockout mutant growth phenotypes across five environmental conditions, standard growth conditions, paraquat stress, peroxide, low salt, and heat shock. (A) Mutants with the largest difference in growth compared to the Δ*ura3* control strain under each condition. Raw data from growth trajectories for individual cultures under standard conditions (thin gray lines) were fit and compared with growth under stress conditions (thin blue lines) using FANOVA (see Materials and Methods). Solid lines indicate the mean of the fit to all replicate trajectories, and shaded regions are the 95% confidence interval of the fit. (B) Functional difference (OD_Δ) of the TF knockout strain relative to the isogenic Δ*ura3* parent strain. The mutants with the top three scoring phenotypes according to $\|OD_\Delta\|$ are shown. (C) Summary statistical metric ranking ($\|OD_\Delta\|$) for those mutants with strongly different growth trajectories compared to that of the Δ*ura3* strain ($\|OD_\Delta\| \geq 0.337$ across all conditions).
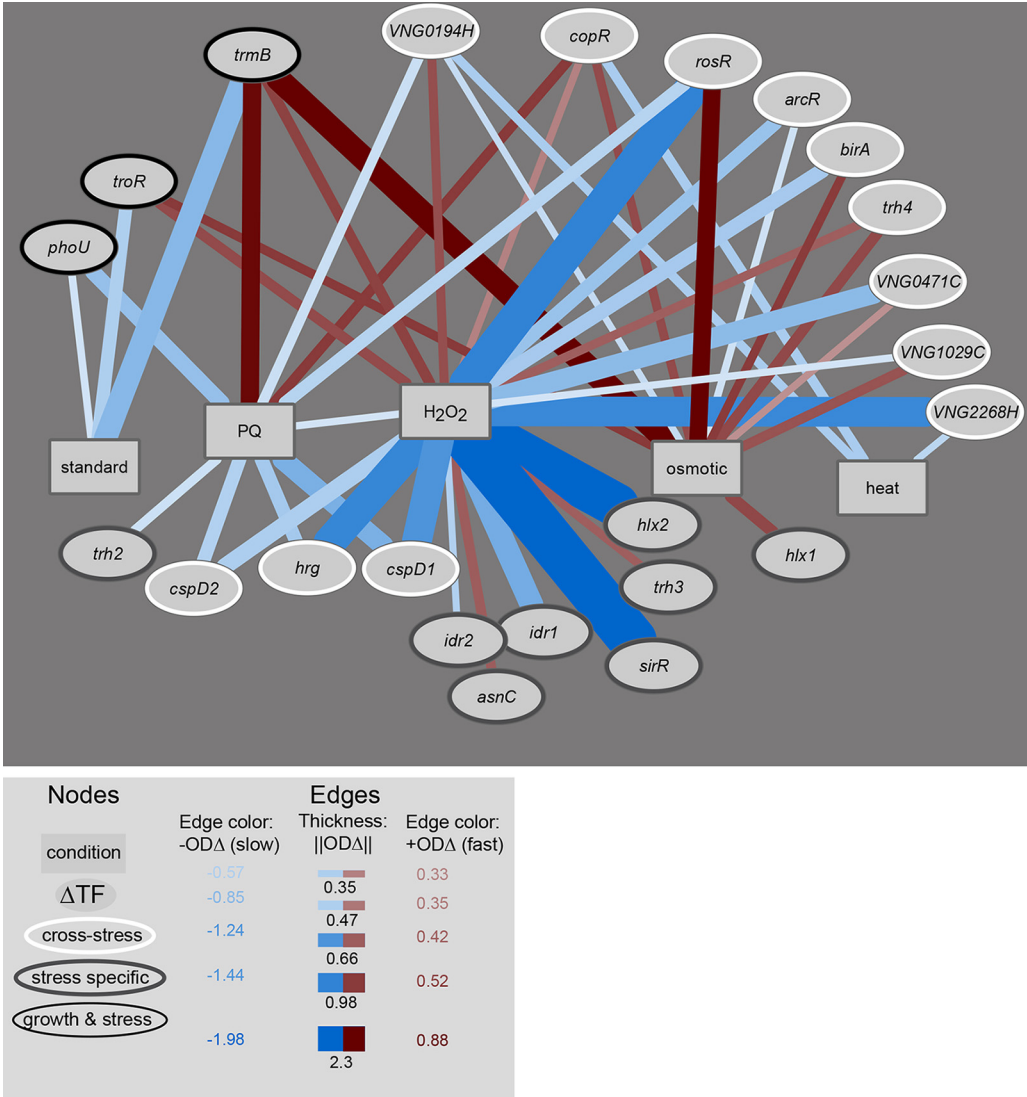
FIG 3 Phenotype network analysis reveals three major classes of TF mutants and extensive cross-regulation of stress responses by TFs. Node and edge attributes are given in the key at the bottom of the figure. OD_Δ numbers in the edge color legend refer to the maximum (fast growth) or minimum (slow growth) value of the mean of the posterior prediction from the FANOVA model across the entire growth time course. $\|OD_\Delta\|$ numbers for edge thickness refer to the median value of the distribution from $\|OD_\Delta\|$ boxplots (see Fig. 2C, Fig. S3, and Materials and Methods).

the strongest effect, with 19 mutants exhibiting significant differences in growth compared to the Δ*ura3* strain (12 mutants grew more slowly than the Δ*ura3* strain, and 7 mutants grew faster than the Δ*ura3* strain [Fig. S2 and S3]). The next strongest effect was low salt (11 mutants, 9 faster and 2 slower than the Δ*ura3* strain), followed by PQ (10 mutants, 2 faster and 8 slower than the Δ*ura3* strain), heat shock and standard conditions (3 mutants in each condition, all with significantly impaired growth phenotypes [Fig. 2C]). We conclude that FANOVA analysis recapitulates known roles of TFs but also suggests novel contributions of TFs to cell physiology.

Phenotype network analysis reveals extensive cross-regulation of stress responses by TFs. Many mutants exhibited significant differential growth phenotypes under multiple conditions tested (Fig. 2). At the rank order $\|OD_\Delta\|$ cutoff (Fig. 2C), 23 of the 27 mutants studied exhibited significant differential growth under at least one condition (Fig. 3). Network analysis of these phenotypes across conditions enabled the classification of the mutants into three categories. (i) Significant differential growth was detected for 12 mutants relative to the Δ*ura3* strain under two or more stress condi-

tions. These TFs were considered to have "cross-stress" functions (Fig. 3). (ii) Significant differential growth was detected for three mutants (Δ*trmB*, Δ*phoU*, and Δ*troR* mutants) under both standard conditions and one or more stress conditions. These TFs were considered to have "growth & stress" functions. (iii) Eight TFs were required for normal growth under only one of the conditions tested here and therefore considered "stress-specific" TFs. However, three of the stress-specific mutants also exhibited growth defects under other stress conditions tested *ad hoc* in previous studies but not included in the systematic phenomic testing here (e.g., Δ*sirR*, Δ*idr1*, and Δ*idr2* mutants are required for metal homeostasis [19, 22]), suggesting that some of these "stress-specific" TFs might also be required for protection across multiple stressors.

Several mutants with strong growth impairments relative to the parent control strain under one condition exhibited growth improvement under others. For example, although the Δ*trmB* mutant grew poorly under standard conditions, it showed significantly improved growth under PQ and osmotic stress conditions (maximum OD$_\Delta$ of 0.5 and 1.0, respectively; Fig. 2A and 3). Similarly, the Δ*rosR* mutant exhibited substantially improved growth under low osmolarity (maximum OD$_\Delta$ of 0.8) but strong growth defects under oxidative stress induced by PQ and peroxide. Deletion of the gene encoding a third TF, CspD1, led to increased growth relative to the control strain under standard conditions but impaired growth under oxidative stress (Fig. 2A and 3 and Fig. S2). These observations suggest novel functions for these previously characterized TFs. These opposing phenotypic patterns for individual TFs could result from direct regulation of genes required for growth and/or stress resistance under these conditions. For example, ribosome levels are directly related to growth rate across the tree of life (51, 52), including archaea (25), and *H. salinarum* RosR directly regulates ribosome biosynthesis genes (50). Alternatively, alteration in growth rate *per se* could change cellular stress resistance properties; for example, slow growth in wild-type yeast cells is associated with heat shock resistance (53). Together, the classes of TF mutants identified here suggest that a surprisingly high number of TFs are required for growth homeostasis during exposure to multiple stressors, suggesting extensive network interconnectivity.

The structure of the gene regulatory network corresponds strongly with TF physiological functions. To gain insight into possible regulatory mechanisms underlying such phenotypes, we determined the relationship between growth trajectories by clustering phenotypes according to the OD$_\Delta$ metric (Fig. 4) and asked how correlated phenotypes mapped to GRN topology. In particular, the Δ*rosR* mutant was previously shown to regulate 20 other TFs, 7 of which were included in the strain collection evaluated here (Δ*arcR*, Δ*cspD2*, Δ*hlx2*, Δ*hrg*, Δ*trh4*, Δ*VNG0039H*, and Δ*VNG0194H* strains). Hierarchical clustering of OD$_\Delta$ phenotypes across the five conditions revealed that these mutants clustered closely together under oxidative stress, including PQ (Fig. 4A) and peroxide conditions (Fig. 4B). Under PQ stress, correlation of the OD$_\Delta$ trajectories for these mutants with that of the Δ*rosR* mutant were significantly enriched for strong positive correlations ($\hat{\rho} \geq 0.4$) relative to all other pairwise correlations between the 27 mutants ($P \leq 6.90 \times 10^{-3}$ by the hypergeometric distribution test [Fig. 4C]). Under peroxide stress, according to $\|OD_\Delta\|$, knockouts in genes encoding TFs regulated by RosR were also significantly enriched for impaired growth relative to all other mutants ($P \leq 0.034$ by the hypergeometric distribution test). Such significant phenotype correlations were not observed under other conditions tested (Fig. 4C and Fig. S4), suggesting that the interconnection between these TFs is coordinated specifically under oxidative stress (50). Together, these results are consistent with the hypothesis that TFs that regulate each other in GRN subnetworks controlling the cellular response to a particular stress are predictive of TF knockout phenotypes and vice versa.

Previous predictions of TF functions from computational GRN models specific to oxidative stress for *H. salinarum* (15) were compared with phenotypic results obtained here. Of the 27 knockouts studied, 15 were predicted by network analysis to play a role
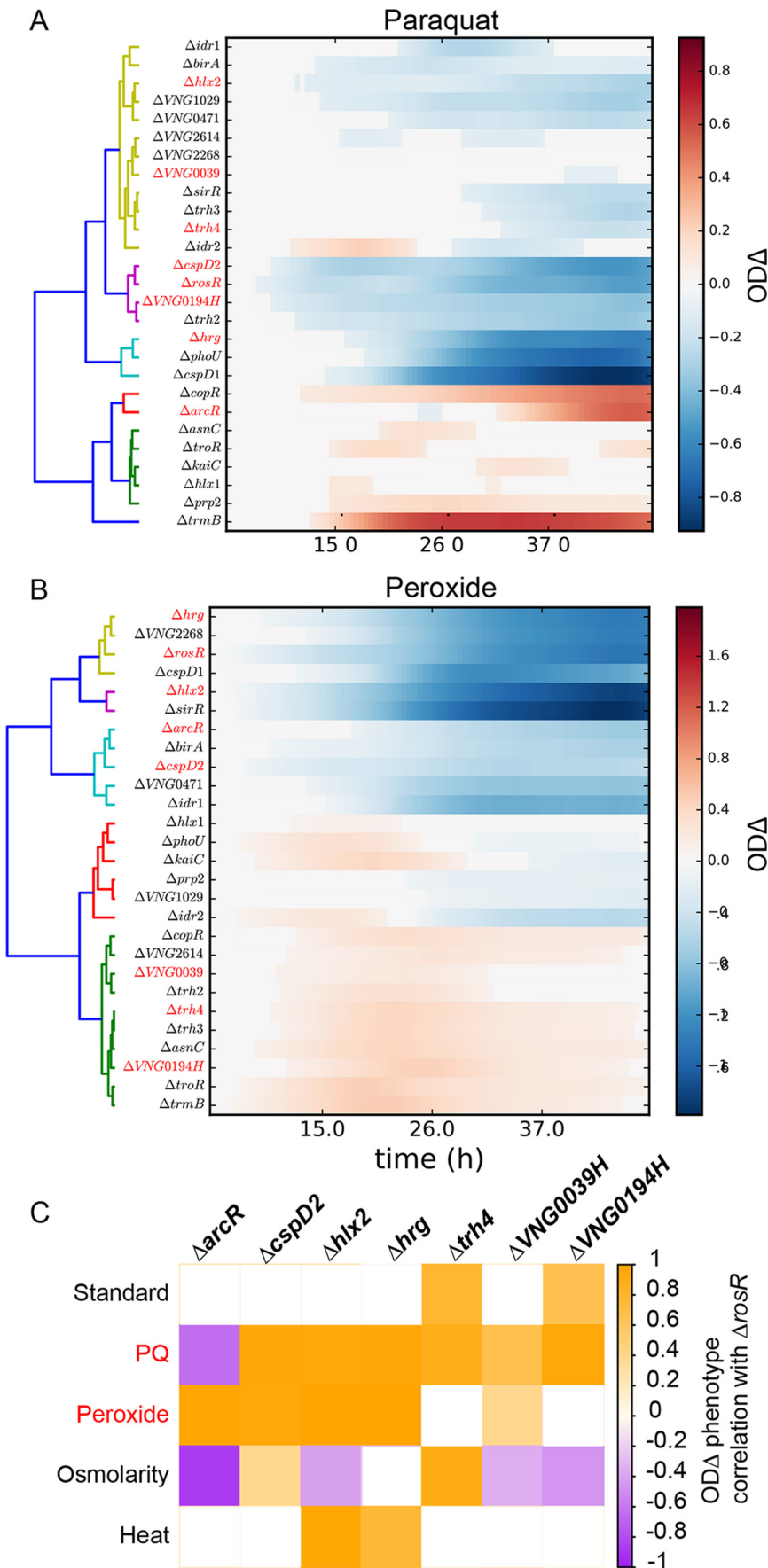
in gene regulation during oxidative stress (15). Of these 15, 14 TF knockouts showed significantly altered growth relative to the Δura3 control in oxidative stress induced by PQ, peroxide, or both (Fig. 2C, Fig. S3, and Table S1). This suggests that the GRN computationally predicted from gene expression data has strong predictive power for the roles of TFs in cell physiology.

Validation and characterization of novel TF functions. (i) CopR functions as a regulator of both heat shock and copper overload responses. To support the stress-specific novel functions for individual TFs discovered here, we performed validation experiments. First we tested the observation of secondary functions for the previously characterized TF, CopR (Fig. 3). We detected significantly impaired growth of the Δ*copR* mutant relative to the Δ*ura3* parent strain during heat shock (Fig. 2A). The heat shock phenotype of the Δ*copR* mutant was the top-ranking mutant under this condition (Fig. 2C). These observations were surprising given that CopR (previously called VNG1179C [22, 54]) was previously characterized as a repressor of P1-type ATPases that export copper during overload (22). To conduct further functional validation of the role for CopR in response to elevated temperature, a wild-type copy of the *copR* gene was expressed in *trans* on a plasmid in the Δ*copR* deletion background, which returned the growth of this strain to levels indistinguishable from that of the Δ*ura3* parent control (Fig. 5A). This complementation result indicates that the Δ*copR* heat shock growth defect was caused by loss of the *copR* gene alone and not by polar or off-target secondary site genetic effects.

To detect potential gene targets of CopR regulation under heat shock, existing transcriptomic and TF-DNA binding data sets were reanalyzed. In the wild-type *H. salinarum* strain exposed to heat shock, 247 genes were differentially expressed (55, 56). A significant fraction of these heat-responsive genes were also differentially expressed in the Δ*copR* mutant grown under copper overload conditions (22) and/or bound by CopR under optimum growth conditions (54) (63 of 247; significance by the hypergeometric distribution test, $P \leq 1.42 \times 10^{-8}$; Fig. 5B and Fig. S5). Chaperones and amino acid metabolic functional categories were significantly enriched among heat-induced and CopR-repressed genes (Fig. 5B). In contrast, energy generation, translation, and transcription functions were significantly enriched among heat-repressed and CopR-induced genes (Fig. 5B). Of these 63 heat-responsive CopR-regulated genes, 6 overlapped with the list of 10 genes whose transcripts and proteins were most strongly induced in response to heat shock (56) (hypergeometric test $P \leq 5.0 \times 10^{-5}$). These six genes encoded proteins required for cellular repair following heat shock (chaperones CctA and Hsp5) and protection from further damage (metalloprotein NirJ, ferritin DpsA, and anaerobic metabolic genes ArcAC). Together, these CopR-regulated gene functions are consistent with a cellular need to arrest growth to refold and regenerate degraded proteins under elevated temperatures (57).

CopR regulation of gene expression described above was tested under standard and copper overload conditions (22, 54). To validate whether CopR is also specifically required for regulation during elevated temperature, gene expression was measured by quantitative reverse transcription-PCR (qRT-PCR) in the Δ*copR* strain immediately before and 30 min after a shift from 42°C to 54°C. *cctA* expression in the Δ*copR* mutant cells was 3- to 4-fold higher under standard conditions than in Δ*ura3* parent control cells and remained elevated upon heat shock (Fig. 5C). This indicates that CopR is required for *cctA* repression under standard conditions and that relief of CopR repres-

FIG 4 TFs that regulate each other have similar OD$_\Delta$ phenotype trajectories. (A and B) Heat maps depict hierarchical clustering of OD$_\Delta$ trajectories for the 27 TF knockout mutants under paraquat (PQ) (A) and peroxide (B) conditions. TFs transcriptionally regulated by RosR are known (50) and indicated by red text. Colors in the dendrogram represent different clusters. The color scale indicates the mean of the posterior OD$_\Delta$ distribution for each mutant across the growth time course (x axis). (C) Each of the seven mutants regulated by RosR are statistically enriched for strongly correlated phenotypes ($\hat{\rho} \geq 0.4$) with the Δ*rosR* mutant under PQ and standard growth conditions relative to other conditions (red text). Colored squares represent significant correlations ($P \leq 0.001$), and white squares represent nonsignificant correlations. See the color scale to the right of the figure for correlation values.
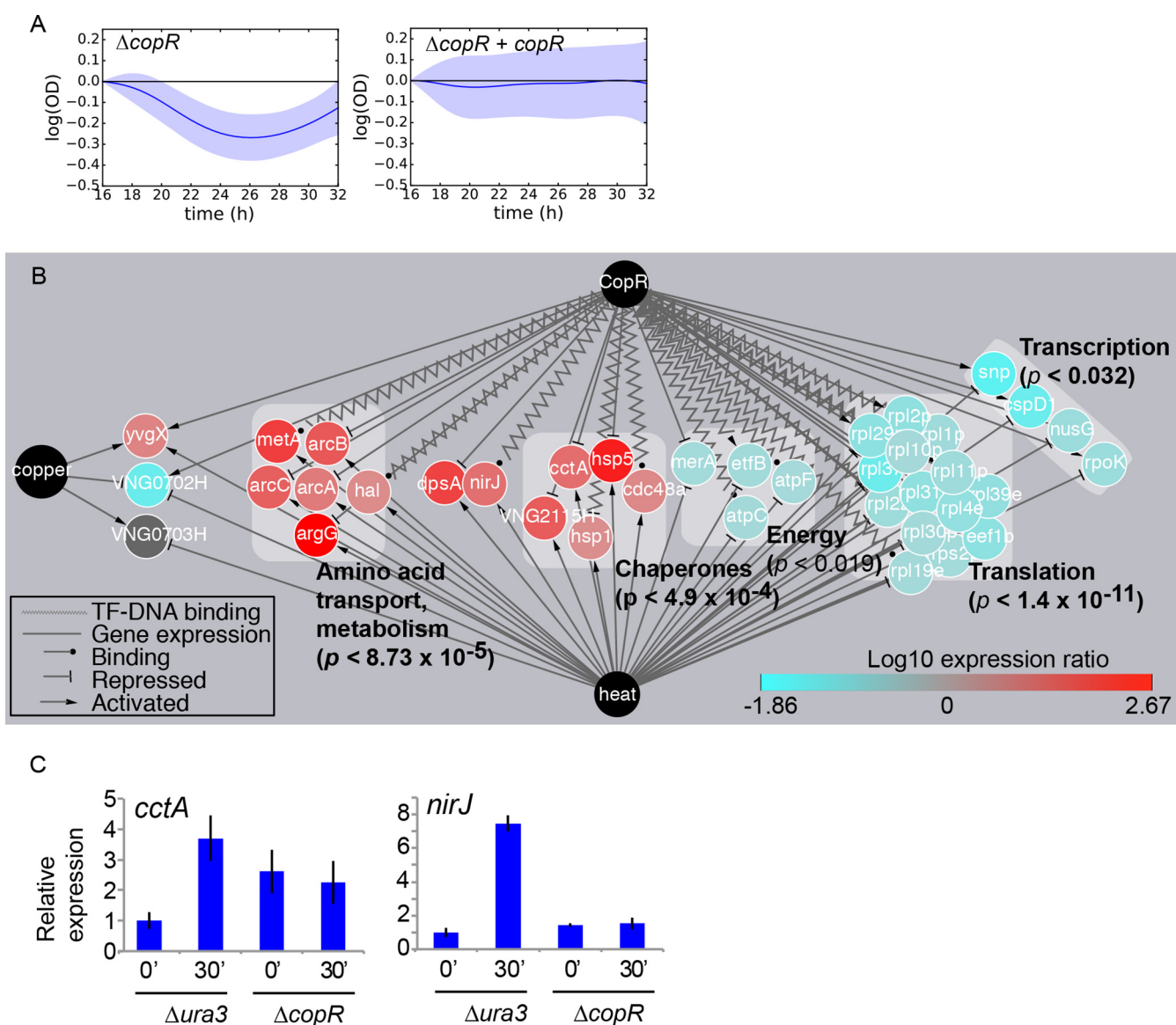
FIG 5 Phenotype validation: a novel heat shock function for the copper-responsive regulator CopR. (A) Slow growth under heat shock conditions in the Δ*copR* mutant (left graph) is complemented by expression of the *copR* gene in *trans* (right graph). (B) Cytoscape gene regulatory network depicting the significant overlap between genes regulated by CopR in response to copper overload (CopR node at the top) and in wild-type cells in response to heat shock (heat node at the bottom) (55) and copper (copper node on the left) (22). Node colors (representing expression levels in wild-type cells under heat shock) and edge line types are indicated in the keys. Gray boxes behind groups of nodes represent arCOG functional categories (61). *P* values of enrichment were calculated by the hypergeometric distribution test. Genes of unknown function are not shown here for clarity but are given in Fig. S5. (C) Quantitative RT-PCR gene expression of *cctA* and *nirJ* genes in the knockout strain compared to the Δ*ura3* control strain. The levels of expression before heat shock (0′) and 30 min after induction of heat shock (30′) are shown. Expression is normalized relative to a control gene whose expression does not change during heat shock (see Materials and Methods). Error bars represent standard errors of the means (SEM) of three biological replicate cultures, each with three technical replicate trials.

sion is necessary for *cctA* heat induction. In contrast, 7-fold induction of *nirJ* in the Δ*ura3* strain was abrogated in the Δ*copR* mutant, indicating that CopR is an activator of this gene under heat shock (Fig. 5C). Taken together, these validation analyses suggest that CopR functions both as a global regulator of gene expression under heat shock and as a specific regulator of a copper efflux transporter during copper overload (22, 58). The connection between heat shock and copper overload has not been established in archaea; however, because both heat shock and copper overload can induce the accumulation of oxidative radicals, the transcriptional responses to these common types of cellular damage may be linked (4).

(ii) A novel oxidative stress response function for CspD1, a conserved cold shock family protein. The Δ*cspD1* strain exhibited the strongest growth defect of all

27 mutants under oxidative stress induced by PQ and the sixth strongest phenotype under peroxide (Fig. 2C and 3). The CspD1 sequence showed strong similarity to the cold shock family of proteins (E value of 2.30×10^{-21}; Table S1). Cold shock domain (CSD) proteins are broadly conserved from bacteria to humans but serve diverse functions, including RNA protection, inhibition of DNA replication during oxidative stress, and transcriptional regulation under various stress conditions (59, 60). Thus, we further investigated the role of CspD1 in response to multiple stresses. We found that, although the *cspD1* gene was significantly repressed in response to high temperature (Fig. 5B) (56), the growth of the Δ*cspD1* mutant strain was indistinguishable from that of the Δ*ura3* control strain under heat and cold shock conditions (Fig. S6). In addition, *cspD1* expression was induced under copper overload conditions in a CopR-dependent manner (Fig. 5B) (22); however, the Δ*cspD1* mutant did not exhibit impaired growth under copper overload conditions (22). These data suggest either that CspD1 does not play a role in the temperature shock and copper overload responses or that CspD1 is functionally redundant under these conditions with other, as yet unknown, TFs.

To further test the role of CspD1 in oxidative stress, we expressed the *cspD1* gene in *trans* on a plasmid in the Δ*cspD1* knockout background. The oxidative stress phenotype was complemented in this strain, confirming that deletion of *cspD1* was responsible for the observed phenotypes (Fig. 6A). In previously published transcriptomic data, the expression of the *cspD1* gene was strongly correlated with fluctuations in oxygen levels (1), induced 30 min after an increase in oxygen (cross-correlation $\hat{\rho} = 0.730$; Fig. 6B). Such expression follows a pattern similar to that observed for *rosR*, which encodes a known global regulator of the oxidative stress response (16, 50). In the Δ*cspD1* strain exposed to fluctuating oxygen levels over time (see Materials and Methods), the transcription of 132 of the 660 known oxygen-responsive genes in *H. salinarum* (1) exhibited significantly altered expression (Fig. 6C and Table S4). Of these 132 genes, 89 are induced in a CspD1-dependent manner during the transition from anaerobic to aerobic conditions. According to arCOG (clusters of orthologous genes for archaea) ontology (61), these aerobic genes are significantly enriched for functions crucial for cell growth (e.g., translation; $P \leq 5.68 \times 10^{-20}$ by the hypergeometric distribution test; Table 2). In contrast, CspD1 is required to repress 43 genes under anaerobic conditions (Fig. 6C and Table 2). Of the 132 genes requiring CspD1 for appropriate expression under oxygen fluctuation, 106 are also differentially expressed under oxidative stress induced by PQ (Fig. 6D) (15).

The EGRIN model based on gene expression under a wide array of conditions accurately predicted a significant fraction of the 132 CspD1-dependent genes (42%; $P \leq 6.33 \times 10^{-32}$ by the hypergeometric distribution test) (7) (Fig. 6E). These EGRIN-predicted genes were also significantly enriched for functions involved in ribosome biogenesis ($P \leq 1.8 \times 10^{-26}$). Similarly, EGRIN predictions based solely on gene expression under oxidative stress also predicted that CspD1 regulates a significant fraction of the 132 CspD1-dependent genes ($P \leq 3.83 \times 10^{-15}$) (15) (Fig. 6F). Together, these results provide validation of GRN functional predictions and phenotype analysis, implicating CspD1 in the regulation of functions critical to growth under oxidative stress and fluctuating oxygen conditions.

Conclusions and perspectives. Data and analyses presented here enabled the discovery of physiological roles for 17 previously uncharacterized TFs in the archaeal species *H. salinarum*. New physiological roles for previously characterized TFs were also revealed (e.g., CspD1 and CopR). This demonstrates the power of our combined high-throughput growth analysis and quantitative growth modeling for discovering unknown gene functions. The functions of a large fraction of genes and pathways remain unknown even in well-characterized model microorganisms. Recently, other high-throughput, genome-wide forward and reverse genetic approaches have made great strides in gene functional discovery, such as population genomics in *Saccharomyces cerevisiae* (62), clustered regularly interspaced short palindromic repeat interference (CRISPRi) in *Bacillus subtilis* (63), and genome-wide knockout collections in *Esch-*
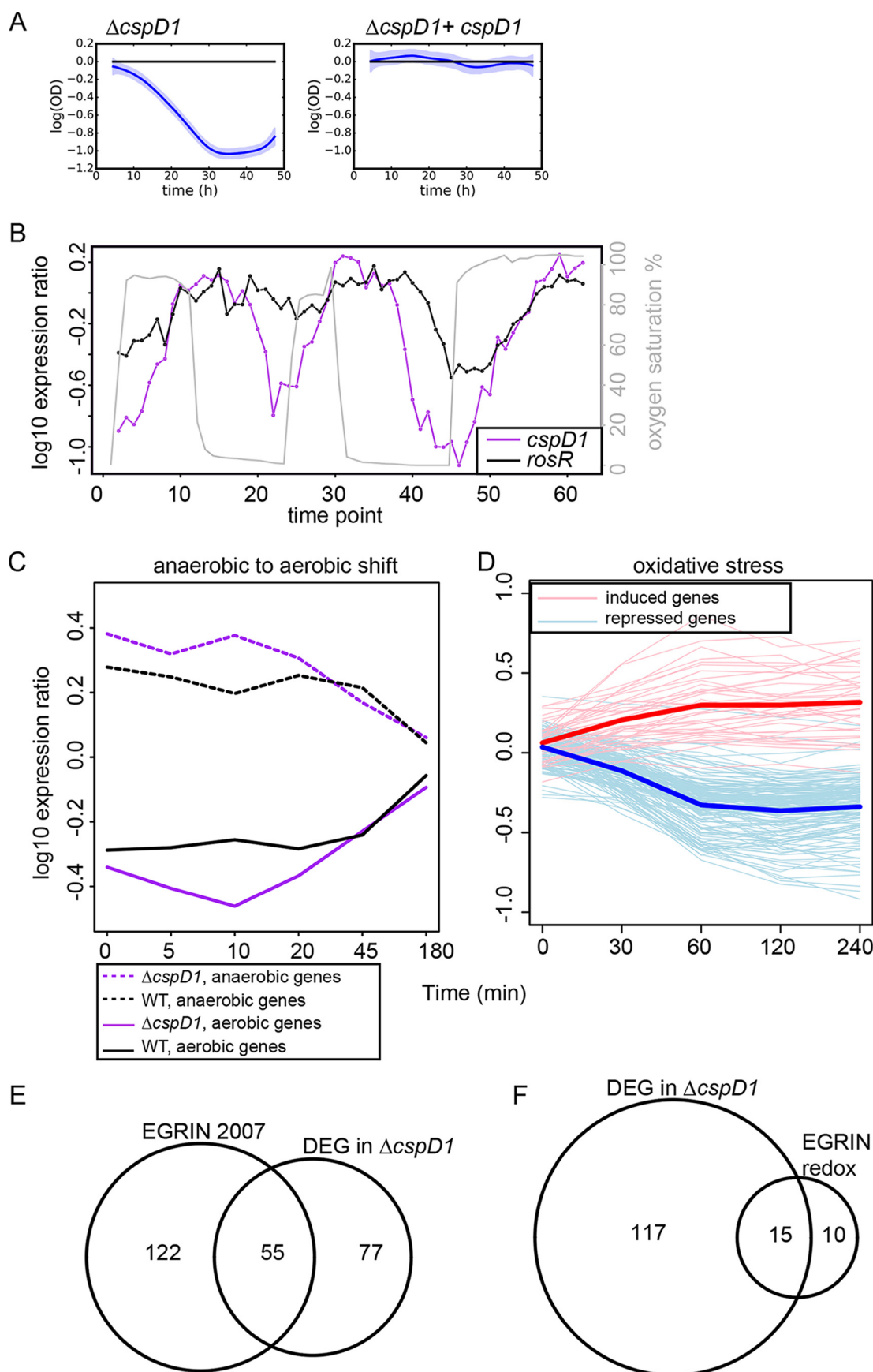
TABLE 2 Functions significantly enriched among CspD1 target genes under fluctuating oxygen conditions

Condition	Category	P value[a]
Aerobic	Coenzyme transport and metabolism	2.48×10^{-4}
	Energy production and conversion	6.33×10^{-3}
	Translation; ribosomal structure and biogenesis	5.68×10^{-20}
Anaerobic	Amino acid transport and metabolism	1.81×10^{-3}

[a]P values result from the hypergeometric distribution test in arCOG categories.

erichia coli (33, 64), as well as nontraditional model organisms such as methanogenic archaea (65). Previously we showed that Gaussian process (GP) methods for phenotype discovery are generally applicable across diverse species (35). Additionally, GPs have been shown to be useful in other domains of biological research, such as modeling genome-wide expression data (66–69). Here we extend the use of GP methods in the context of functional ANOVA to compare the growth of a large collection of strains (Fig. 2), demonstrating the promise of these methods for genome-wide functional discovery across a variety of species.

Here we directly test the predictions of computationally inferred, global GRN models such as EGRIN (7, 15). Because these and other statistical GRN inference models were inferred directly from gene expression data, the direct impact of the GRN activity on cellular physiology remained unclear. Here we show that models such as EGRIN predict not only gene expression but also the phenotypic impact of such expression (Fig. 2 and 6). In previous work, we also demonstrated the accuracy of EGRIN in predicting TF functions. For example, hypotheses generated from EGRIN predictions enabled the discovery of RosR, an archaeon-specific, novel master regulator of oxidative stress response (7, 15, 16, 50). Similarly, EGRIN model predictions regarding the cross-regulation of phosphate metabolism and methanogenesis pathways were validated by gene knockout studies in methanogens (30). Here we also observed extensive cross-regulation: each TF was important for resistance of multiple stressors, and multiple TFs played a role in surviving each stressor (Fig. 3). Such cross-regulation of gene expression has also been observed in bacteria as a means to integrate environmental cues that cooccur (e.g., heat and singlet oxygen [8, 70]) or that induce functionally related response pathways (e.g., metal homeostasis and oxidative stress [71]). Together with these previous studies, the broader investigation of 27 TF knockout phenotypes reported here demonstrates that GRN models such as EGRIN are effective tools for generating accurate hypotheses regarding TF functions. The combination of GRN modeling and phenomic validation reveals the direct impact of a complex web of regulatory interactions on cell physiology (Fig. 4 and 5). This work supports an emerging general principle that cross-regulation between TFs within the GRN enables a coordinated response to a variety of environmental stimuli.

MATERIALS AND METHODS

Culturing and construction of transcription factor mutants. *Halobacterium salinarum* NRC-1 (ATCC 700922) was used as the wild-type strain background. Constructed mutants are derivatives of the Δ*ura3* strain (44), and the Δ*ura3* strain was used as the isogenic parent strain as a control in all assays. *H. salinarum* was grown routinely in complex medium (CM) (250 g/liter NaCl, 20 g/liter MgSO$_4$ · 7H$_2$O,

FIG 6 Phenotype validation: a novel oxidative stress function for the cold shock family protein CspD1. (A) Slow growth under oxidative stress conditions in the Δ*cspD1* mutant (left graph) is complemented by expression of the *cspD1* gene in *trans* (right graph). (B) Line plot depicting the expression of the *cspD1* gene (left axis) during fluctuations in oxygen concentrations (gray line, right axis). *cspD1* expression is compared to that of the gene encoding known oxidative stress regulator RosR (left axis). (C) Line plot depicting the expression of genes requiring CspD1 for appropriate dynamic expression in response to oxygen. Each line represents the mean expression value of the groups of genes indicated in the legend. Expression data and annotations for individual genes are given in Table S4. WT, wild type. (D) Expression profiles for 106 of the 132 CspD1-dependent genes differentially expressed under PQ conditions in wild-type cells (15). Thin lines represent the expression of individual genes. Thick lines show the mean of induced or repressed genes. (E) Overlap between EGRIN predictions for CspD1 target gene regulatory influences (7) and differentially expressed genes (DEG) in the Δ*cspD1* background. (F) Similar to panel E except that predictions under oxidative stress conditions are shown (15).

3 g/liter sodium citrate, 2 g/liter KCl, 10 g/liter peptone) supplemented with 50 μg/ml uracil to complement the uracil auxotrophy of the $\Delta ura3$ parent background. *E. coli* strain DH5α used for routine cloning was grown in LB containing carbenicillin (50 μg/ml) to maintain plasmids. Mutants were constructed as previously reported using the standard double-crossover counterselection method (44). Briefly, approximately 500 bp of flanking regions up- and downstream of the gene of interest were integrated into the StuI restriction site of plasmid pNBKO7 by blunt-end ligation (details of all plasmid constructs listed in Table S2 in the supplemental material). The resultant constructs were transformed into the $\Delta ura3$ strain and selected on CM plates containing mevinolin (10 μg/ml). The resulting merodiploid strains were then plated on CM plates containing 5-fluoroorotic acid (250 μg/ml) and uracil to remove the integrated plasmid, yielding unmarked TF deletion strains. Complementation strains were constructed using the pMTF-cmyc vector backbone (72) by isothermal Gibson assembly (73) and routinely maintained in liquid culture in CM supplemented with mevinolin (1 μg/ml). All strains were verified as described in reference 19. Genotype results are given in Fig. S1, and primer, strain, and plasmid details are given in Table S2.

Growth curve assays. Cultures were pregrown in standard conditions, which were defined as growth in CM containing uracil at 42°C with shaking (225 rpm) under ambient light until early stationary phase, measured by an optical density at 600 nm (OD_{600}) $\approx$ 2.0 (74). Each strain was then subcultured to an $OD_{600} \approx 0.05$ in 200 μl CM containing uracil under continuous shaking at 42°C in a Bioscreen C analysis system (Growth Curves USA, Piscataway, NJ) set to measure OD_{600} every 30 min for the duration of the 48-h experiment. Each strain was tested in at least biological quadruplicate samples, each with three technical replicates. For heat stress experiments, the temperature was shifted to 54°C at 16 h, and the elevated temperature was maintained for the remainder of the experiment. For oxidative stress experiments, hydrogen peroxide (5 mM) or paraquat (PQ) (0.333 mM) was added at the beginning of growth. For low-salinity experiments, strains were grown in CM medium containing 2.9 M NaCl. For cold growth curves (Fig. S6), cultures were pregrown in standard conditions until stationary phase, then subcultured to a starting OD_{600} of 0.1 into 5 ml of CM containing uracil and incubated at 15°C with shaking (225 rpm). Sample aliquots were taken every 24 h for 5 days to measure OD_{600}.

FANOVA growth curve model framework. Growth data were then modeled using functional analysis of variance (FANOVA) (75), using a Bayesian approach (47). FANOVA models data as a linear combination of functional effects, where the number of effects is determined by the experiment. For example, in the case of two experimental perturbations, observations at time point t for effects i and j are modeled as:

$$y_{i,j}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + (\alpha\beta)_{i,j}(t) + \varepsilon_{i,j}(t) \tag{3}$$

where $\mu_{(t)}$ is a mean function, $\alpha_i(t)$ and $\beta_j(t)$ are the effect functions, $(\alpha\beta)_{i,j}$ is the interaction between them, and $\varepsilon_{i,j}(t)$ is observation noise. Functional effects and interactions can be added and removed from the model as needed for different experimental designs. In order to make the latent effect functions identifiable, they are constrained to sum to zero:

$$\sum_{i=1}^{n_\alpha} \alpha_i(t) = \sum_{j=1}^{n_\beta} \beta_j(t) = \sum_{i=1}^{n_\alpha} (\alpha\beta)_{i,j}(t) = \sum_{j=1}^{n_\beta} (\alpha\beta)_{i,j}(t) = 0 \ \forall \ t \tag{4}$$

The mean function $\mu(t)$ is given a GP prior directly:

$$\mu(t) \sim \mathrm{GP}(0, \kappa_\mu(t_1, t_2)) \tag{5}$$

In order to satisfy the identifiability constraints (equation 4), effect functions are parameterized using a set of contrast functions. For example, $\alpha_i(t)$ is defined as a linear combination of the contrast functions:

$$\alpha_i(t) = \sum_{k=1}^{n_\alpha - 1} c_{ik} \times \alpha_k^*(t) = c_i^T \alpha^*(t) \tag{6}$$

where

$$\alpha^*(t) = \left\{ \alpha_1(t), \alpha_2(t), \ldots, \alpha_{n_\alpha - 1}(t) \right\} \tag{7}$$

Gaussian process (GP) priors were assigned to the latent functions as described in reference 47. GPs place a distribution on a continuous function, any finite number of observations of which are distributed as multivariate normal (76). Each GP prior is parameterized by a mean function $m(x)$ and a covariance function, $\kappa(x_1, x_2)$. All prior mean functions in this analysis were set identically to zero, as is standard. Covariance functions were modeled using radial basis functions (RBFs):

$$\kappa(x_1, x_2) = \sigma^2 \times \exp\left(\frac{-\| x_1 - x_2 \|^2}{\ell} \right) \tag{8}$$

where σ^2 and ℓ are hyperparameters defining the variance and length scale of the GP, respectively. The variance σ^2 determines the magnitude of variability for a given GP prior distribution, with higher variances leading to more-variable functions. The length scale parameter controls the rate of decay of covariance between two time points, and larger length scales place higher probability on smoother (slower changing) functions.

All contrast functions for a given effect are defined by a shared GP prior:

$$\alpha_i^*(t) \sim GP(0, \kappa_\alpha) \tag{9}$$

where $1 \leq i \leq n_\alpha - 1$. Kernel hyperparameters were given noninformative priors. Posterior quantities were obtained by Markov chain Monte Carlo (MCMC) simulation. Sampling of the contrast functions was accomplished using Gibbs sample updates (47). Kernel hyperparameters were sampled via slice sampling (77).

Determination of significant growth phenotypes. All raw growth data (Table S3) were first normalized to the log$_2$ scale. The first 4 h of growth (less than one generation) were removed because technical and instrument variability is often observed during this time frame. Growth data were then grouped by strain and experimental design (standard growth, 0.333 mM PQ, 2.9 M NaCl, or heat shock), referred to as the condition. Growth data corresponding to the same condition were scaled by a fixed value so the mean of the condition at the earliest time point was equal to zero. The growth data for *H. salinarum* TF mutants under various stress conditions was modeled with GP FANOVA as described below.

(i) Standard, oxidative stress, and osmotic stress conditions. Growth data were modeled with two effects corresponding to strain and experimental design. The strain effect, α_i, varied from $i = 1$ to $i = 28$, where $i = 1$ corresponded to the Δ*ura3* parent strain and $i > 1$ corresponded to one of the 27 mutant strains. The experimental design effect, β_j ($1 \leq j \leq 4$), represented one of four experimental designs: standard growth ($j = 1$), low-osmolarity stress ($j = 2$), PQ ($j = 3$), or peroxide stress ($j = 4$). Interactions between the two effects were also modeled to determine the strain-specific responses to each of the stress effects. The full GP FANOVA model for the analysis was then the same as in equation 3 and was estimated using the GP FANOVA model as described above. Variation between phenotypes arising from separate batches of experiments were controlled by adding a specific batch function, $\gamma(t)$.

(ii) Heat shock. Growth data for heat shock conditions was modeled using a GP FANOVA modeling the individual effect of strain, α_i. This model has the form

$$y_i(t) = \mu(t) + \alpha_i(t) + \varepsilon_i(t) \tag{10}$$

All metrics for the heat shock condition (OD$_\Delta$ and $\|$OD$_\Delta\|$, described below) were computed starting at the 16-h time point, the beginning of the shock. Additionally, the difference between the Δ*ura3* parent and mutant strain was subtracted from all metrics, which removes any confounding differential growth that occurred between the two strains prior to the shock initiation.

(iii) Δ*copR* complementation. Complementation was modeled as the combined effect of strain (α_i, Δ*ura3* or Δ*copR*), empty vector (β), and presence or absence of the *copR* complementation on the plasmid (γ). The FANOVA model for this condition is then

$$y_{i,j,k}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + \gamma_k(t) + \varepsilon_{i,j,k}(t) \tag{11}$$

There are two states for each effect, and we are interested in estimating the fixed effect of the *copR* complementation only under the heat stress condition starting at 16 h, so no interactions between condition and strain were needed in this model.

(iv) Δ*cspD1* complementation. Complementation of Δ*cspD1* in H$_2$O$_2$ was modeled with an extension of the model for Δ*copR* in heat shock (equation 11). Functions for strain (α), condition (β), and their interaction (α,β) are included, as in equation 3. In addition, a function is included to model the presence or absence of the empty vector γ, as well as the presence or absence of Δ*cspD1* on the plasmid (δ). In this case, we are specifically interested in the complementation provided by the plasmid-expressed *cspD1* to the Δ*cspD1* strain in the H$_2$O$_2$ condition, so we modeled the interaction of this strain with the condition (β,δ):

$$y_{i,j,k,l} = \mu + \alpha_i + \beta_j + \gamma_k + \delta_l + (\alpha\beta)_{i,j} + (\beta\delta)_{j,l} \tag{12}$$

(v) Significance test. Two metrics were used to assess the significance of the difference between the growth of each mutant versus the parent strain under each condition.

The first metric, OD$_\Delta$, was first computed for each strain under standard conditions as

$$\mathrm{OD}_\Delta(t) = y_{i,1}(t) - y_{1,1}(t) \tag{13}$$

$$\mathrm{OD}_\Delta(t) = \alpha_i(t) - \alpha_1(t) + (\alpha\beta)_{i,1}(t) - (\alpha\beta)_{1,1}(t) \tag{14}$$

where *i* represents the strain of interest. This represents the difference between the parent strain and TF mutant strain under standard conditions. When comparing mutant and parent strain under nonstandard conditions, the function described in equation 14 is used as the control of the difference between the parent and mutant strains. This leads to the formulation of OD$_\Delta$ under stress conditions as

$$\mathrm{OD}_\Delta(t) = [y_{i,j}(t) - y_{1,j}(t)] - [y_{i,1}(t) - y_{1,1}(t)] \tag{15}$$

$$\mathrm{OD}_\Delta(t) = (\alpha\beta)_{i,j}(t) - (\alpha\beta)_{1,j}(t) - (\alpha\beta)_{i,1}(t) + (\alpha\beta)_{1,1}(t) \tag{16}$$

which is the difference between strain *i* and the parent strain under condition *j*, normalized by the difference between the two strains under standard conditions. In this formulation, if the difference between the parent strain and mutant strain is the same under both standard conditions and one of the stress conditions, OD$_\Delta$ will be close to zero under the stress condition for that strain. Both formulations of OD$_\Delta$ in equation 14 and equation 16 use effect functions estimated by the GP FANOVA model and can therefore be calculated from the model posterior samples. The posterior distribution of OD$_\Delta$ for each strain under a given condition was used to determine where a mutant strain and the Δ*ura3* parent were significantly different as a function of time. Ninety-five percent credible intervals for OD$_\Delta$ were constructed using the posterior samples of the GP FANOVA model, and any time point where zero was not included in the interval was considered significantly different from the parent strain.

To obtain an overall test to rank order the significance of the phenotype under a specific condition, a second metric was derived from OD$_\Delta$ representing the overall magnitude of growth difference between the parent and mutant strains. This metric, $\|$OD$_\Delta\|$, was calculated as in equation 2 which represents the magnitude of OD$_\Delta$ over the entire growth curve. Larger values of $\|$OD$_\Delta\|$ indicate growth phenotypes

that deviate most from that of the parent strain. This metric can be used to directly compare different mutant strains for the overall significance of their growth phenotype.

Phenotype networks (Fig. 3) were constructed and visualized using Cytoscape (78). The $\|OD_\Delta\|$ cutoff of 0.337 was used to enable comparison across mutants and across conditions. This cutoff was chosen to (i) exclude mutants whose maximum or minimum OD_Δ value was inside the 95% confidence interval and (ii) exclude mutants outside the top 10% of $\|OD_\Delta\|$ values under standard conditions (Fig. 2).

Transcript quantitation with qRT-PCR. Samples were harvested at mid-log phase ($OD_{600} \approx 0.4$) and 30 min after the shift to 54°C. RNA was purified using an Absolutely RNA Miniprep kit (Agilent Technologies, Santa Clara, CA), and 500 ng was tested for DNA contamination by PCR with *Taq* DNA polymerase (Thermo Scientific, Grand Island, NY) and for integrity with a 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA). Quantitative reverse transcription-PCR (qRT-PCR) primers (Integrated DNA Technologies, Coralville, IA) are listed in Table S2. The iTaq universal SYBR green one-step kit (Bio-Rad, Hercules, CA) was used in 20-μl reaction mixtures according to the manufacturer's instructions. Quantitative analysis was performed in a Bio-Rad real-time thermal cycler (Bio-Rad, Hercules, CA). Expression relative to the *VNG1756G* reference locus (18) was calculated using the $\Delta\Delta C_T$ method (18, 79).

CspD1 validation experiments. (i) Transcriptomic analysis of the Δ*cspD1* mutant exposed to different oxygen levels. Although these data were published previously as part of a larger gene regulatory network study (43), here we report the details of the experiment and specific analysis of the effect of the Δ*cspD1* deletion on gene expression. Briefly, the Δ*cspD1* mutant and Δ*ura3* strain were grown to mid-logarithmic phase in batch mode in a New Brunswick BioFlo100 modular benchtop fermentor (New Brunswick Scientific) in CM medium as described in reference 1. At mid-log phase, oxygen sparging and agitation were stopped to induce anoxia. The cultures were incubated anaerobically overnight, then oxygen was sparged, and RNA was collected at time points immediately prior to the addition of oxygen and 5, 10, 20, 45, and 180 min afterwards. RNA extraction, microarray hybridization, scanning, and preprocessing were conducted as described in reference 1. Data were analyzed using a modified ANOVA in the Statistical Analysis of Microarrays (SAM) package in the TM4 freeware program (80) to determine the final list of genes differentially expressed in response to oxygen and *cspD1* deletion (Table S4).

(ii) Analysis of PQ transcriptomic data. Wild-type *H. salinarum* NRC-1 cultures in mid-logarithmic phase were exposed to 4 mM PQ, and transcriptomics by microarrays were monitored immediately prior to PQ exposure and 30, 60, 120, and 240 min afterward. These data were previously described in reference 15. In this study, we performed hierarchical clustering of these data in the TM4 program to determine which genes were induced and which genes were repressed in response to PQ. These PQ-responsive gene sets were then filtered to include only the 132 genes differentially expressed in the Δ*cspD1* mutant (Fig. 6C) and plotted in the R environment base package (Fig. 6D) (81).

(iii) Comparison of CspD1 target genes from EGRIN predictions with transcriptomic validation data. EGRIN predictions of CspD1-regulated genes were filtered in Cytoscape (78). For predictions from reference 7, cluster residuals of <0.4 and CspD1 target gene edge weights of ≥0.2 were considered significant; for predictions from reference 15, residuals of <0.4 and edge weight of >0.1 were considered significant. These criteria are consistent with those used in the original EGRIN publications. The significance of the overlap between genes in resultant clusters and genes differentially expressed in the Δ*cspD1* mutant (Table S4) was determined using the hypergeometric distribution.

General statistical methods used for analysis of validation experiments. OD_Δ phenotype correlations and their significance described in the legend to Fig. 4 were calculated in R using the rcorr() function in the Hmisc package (82) and visualized using the corrplot package (83). Significance of overlap between disparate gene lists was calculated by the hypergeometric distribution test. For all gene lists, enrichment in arCOG functional categories (61) relative to the genomic background was calculated by the hypergeometric distribution test as described previously (84, 85).

Data availability. Raw and normalized microarray data and metadata from the Δ*cspD1* mutant exposed to oxygen are freely accessible at the NCBI Gene Expression Omnibus (GEO) database (86) (https://www.ncbi.nlm.nih.gov/geo/) under accession numbers GSE97933 and GPL22925. Data for PQ gene expression are accessible in the GEO database via accession GSE17515. Phenotyping data generated in this study are available in the supplemental material (Table S3). The code repository for the FANOVA model is freely available at https://github.com/ptonner/hsalinarum_tf_phenotype. Code for determining enrichment in arCOG functional groups is freely available at https://github.com/amyschmid/histone_arCOG (85).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00032-17.

FIG S1, PDF file, 0.4 MB.

FIG S2, JPG file, 0.9 MB.

FIG S3, PDF file, 1 MB.

FIG S4, PDF file, 0.7 MB.

FIG S5, PDF file, 1 MB.

FIG S6, PDF file, 0.2 MB.

TABLE S1, XLSX file, 0.04 MB.

TABLE S2, XLSX file, 0.02 MB.
TABLE S3, XLSX file, 3.5 MB.
TABLE S4, XLSX file, 0.3 MB.

REFERENCES

1. Schmid AK, Reiss DJ, Kaur A, Pan M, King N, Van PT, Hohmann L, Martin DB, Baliga NS. 2007. The anatomy of microbial cell state transitions in response to oxygen. Genome Res 17:1399–1413. https://doi.org/10.1101/gr.6728007.
2. Kültz D. 2005. Molecular and evolutionary basis of the cellular stress response. Annu Rev Physiol 67:225–257. https://doi.org/10.1146/annurev.physiol.67.040403.103635.
3. Brooks AN, Turkarslan S, Beer KD, Lo FY, Baliga NS. 2011. Adaptation of cells to new environments. Wiley Interdiscip Rev Syst Biol Med 3:544–561. https://doi.org/10.1002/wsbm.136.
4. Imlay JA. 2003. Pathways of oxidative damage. Annu Rev Microbiol 57:395–418. https://doi.org/10.1146/annurev.micro.57.030502.090938.
5. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000. Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11:4241–4257. https://doi.org/10.1091/mbc.11.12.4241.
6. Beer KD, Wurtmann EJ, Pinel N, Baliga NS. 2014. Model organisms retain an "ecological memory" of complex ecologically relevant environmental variation. Appl Environ Microbiol 80:1821–1831. https://doi.org/10.1128/AEM.03280-13.
7. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, Longabaugh W, Vuthoori M, Whitehead K, Madar A, Suzuki L, Mori T, Chang DE, Diruggiero J, Johnson CH, Hood L, Baliga NS. 2007. A predictive model for transcriptional control of physiology in a free living cell. Cell 131:1354–1365. https://doi.org/10.1016/j.cell.2007.10.053.
8. Dufour YS, Donohue TJ. 2012. Signal correlations in ecological niches can shape the organization and evolution of bacterial gene regulatory networks. Adv Microb Physiol 61:1–36. https://doi.org/10.1016/B978-0-12-394423-8.00001-9.
9. Oren A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. Saline Systems 4:2. https://doi.org/10.1186/1746-1448-4-2.
10. Decker KB, Hinton DM. 2013. Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. Annu Rev Microbiol 67:113–139. https://doi.org/10.1146/annurev-micro-092412-155756.
11. Werner F, Grohmann D. 2011. Evolution of multisubunit RNA polymerases in the three domains of life. Nat Rev Microbiol 9:85–98. https://doi.org/10.1038/nrmicro2507.
12. Grohmann D, Werner F. 2011. Recent advances in the understanding of archaeal transcription. Curr Opin Microbiol 14:328–334. https://doi.org/10.1016/j.mib.2011.04.012.
13. Pérez-Rueda E, Janga SC. 2010. Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin. Mol Biol Evol 27:1449–1459. https://doi.org/10.1093/molbev/msq033.
14. Peeters E, Peixeiro N, Sezonov G. 2013. Cis-regulatory logic in archaeal transcription. Biochem Soc Trans 41:326–331. https://doi.org/10.1042/BST20120312.
15. Kaur A, Van PT, Busch CR, Robinson CK, Pan M, Pang WL, Reiss DJ, DiRuggiero J, Baliga NS. 2010. Coordination of frontline defense mechanisms under severe oxidative stress. Mol Syst Biol 6:393. https://doi.org/10.1038/msb.2010.50.
16. Sharma K, Gillum N, Boyd JL, Schmid A. 2012. The RosR transcription factor is required for gene expression dynamics in response to extreme oxidative stress in a hypersaline-adapted archaeon. BMC Genomics 13:351. https://doi.org/10.1186/1471-2164-13-351.
17. Schmid AK, Reiss DJ, Pan M, Koide T, Baliga NS. 2009. A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability. Mol Syst Biol 5:282. https://doi.org/10.1038/msb.2009.40.
18. Todor H, Sharma K, Pittman AM, Schmid AK. 2013. Protein-DNA binding dynamics predict transcriptional response to nutrients in archaea. Nucleic Acids Res 41:8546–8558. https://doi.org/10.1093/nar/gkt659.
19. Schmid AK, Pan M, Sharma K, Baliga NS. 2011. Two transcription factors are necessary for iron homeostasis in a salt-dwelling archaeon. Nucleic Acids Res 39:2519–2533. https://doi.org/10.1093/nar/gkq1211.
20. Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, Kottemann MC, Hood L, DiRuggiero J. 2004. Systems level insights into the stress response to UV radiation in the halophilic archaeon Halobacterium NRC-1. Genome Res 14:1025–1035. https://doi.org/10.1101/gr.1993504.
21. Baliga NS, Pan M, Goo YA, Yi EC, Goodlett DR, Dimitrov K, Shannon P, Aebersold R, Ng WV, Hood L. 2002. Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. Proc Natl Acad Sci U S A 99:14913–14918. https://doi.org/10.1073/pnas.192558999.
22. Kaur A, Pan M, Meislin M, Facciotti MT, El-Gewely R, Baliga NS. 2006. A systems view of haloarchaeal strategies to withstand stress from transition metals. Genome Res 16:841–854. https://doi.org/10.1101/gr.5189606.
23. Whitehead K, Kish A, Pan M, Kaur A, Reiss DJ, King N, Hohmann L, DiRuggiero J, Baliga NS. 2006. An integrated systems approach for understanding cellular responses to gamma radiation. Mol Syst Biol 2:47. https://doi.org/10.1038/msb4100091.
24. Whitehead K, Pan M, Masumura K, Bonneau R, Baliga NS. 2009. Diurnally entrained anticipatory behavior in archaea. PLoS One 4:e5485. https://doi.org/10.1371/journal.pone.0005485.
25. Facciotti MT, Pang WL, Lo FY, Whitehead K, Koide T, Masumura K, Pan M, Kaur A, Larsen DJ, Reiss DJ, Hoang L, Kalisiak E, Northen T, Trauger SA, Siuzdak G, Baliga NS. 2010. Large scale physiological readjustment during growth enables rapid, comprehensive and inexpensive systems analysis. BMC Syst Biol 4:64. https://doi.org/10.1186/1752-0509-4-64.
26. Bonneau R, Baliga NS, Deutsch EW, Shannon P, Hood L. 2004. Comprehensive de novo structure prediction in a systems-biology context for the archaea Halobacterium sp. NRC-1. Genome Biol 5:R52. https://doi.org/10.1186/gb-2004-5-8-r52.
27. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H, Isenbarger TA, Peck RF, Pohlschroder M, Spudich JL, Jung KW, Alam M, Freitas T, Hou S, Daniels CJ, Dennis PP, Omer AD, Ebhardt H, Lowe TM, Liang P, Riley M, Hood L, DasSarma S. 2000. Genome sequence of Halobacterium species NRC-1. Proc Natl Acad Sci U S A 97:12176–12181. https://doi.org/10.1073/pnas.190337797.

28. Reiss DJ, Plaisier CL, Wu WJ, Baliga NS. 2015. cMonkey2: automated, systematic, integrated detection of co-regulated gene modules for any organism. Nucleic Acids Res 43:e87. https://doi.org/10.1093/nar/gkv300.

29. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V. 2006. The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. Genome Biol 7:R36. https://doi.org/10.1186/gb-2006-7-5-r36.

30. Yoon SH, Turkarslan S, Reiss DJ, Pan M, Burn JA, Costa KC, Lie TJ, Slagel J, Moritz RL, Hackett M, Leigh JA, Baliga NS. 2013. A systems level predictive model for global gene regulation of methanogenesis in a hydrogenotrophic methanogen. Genome Res 23:1839–1851. https://doi.org/10.1101/gr.153916.112.

31. Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, Shuster B, Barry SN, Gallitto M, Liu B, Kacmarczyk T, Santoriello F, Chen J, Rodrigues CD, Sato T, Rudner DZ, Driks A, Bonneau R, Eichenberger P. 2015. An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. Mol Syst Biol 11:839. https://doi.org/10.15252/msb.20156236.

32. Bochner BR. 2009. Global phenotypic characterization of bacteria. FEMS Microbiol Rev 33:191–205. https://doi.org/10.1111/j.1574-6976.2008.00149.x.

33. Nichols RJ, Sen S, Choo YJ, Beltrao P, Zietek M, Chaba R, Lee S, Kazmierczak KM, Lee KJ, Wong A, Shales M, Lovett S, Winkler ME, Krogan NJ, Typas A, Gross CA. 2011. Phenotypic landscape of a bacterial cell. Cell 144:143–156. https://doi.org/10.1016/j.cell.2010.11.052.

34. Gunsalus KC, Ge H, Schetter AJ, Goldberg DS, Han JD, Hao T, Berriz GF, Bertin N, Huang J, Chuang LS, Li N, Mani R, Hyman AA, Sönnichsen B, Echeverri CJ, Roth FP, Vidal M, Piano F. 2005. Predictive models of molecular machines involved in Caenorhabditis elegans early embryogenesis. Nature 436:861–865. https://doi.org/10.1038/nature03876.

35. Tonner PD, Darnell CL, Engelhardt BE, Schmid AK. 2017. Detecting differential growth of microbial populations with Gaussian process regression. Genome Res 27:320–333. https://doi.org/10.1101/gr.210286.116.

36. Marzolf B, Deutsch EW, Moss P, Campbell D, Johnson MH, Galitski T. 2006. SBEAMS-microarray: database software supporting genomic expression analyses for systems biology. BMC Bioinformatics 7:286. https://doi.org/10.1186/1471-2105-7-286.

37. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, Bateman A. 2006. Pfam: clans, web tools and services. Nucleic Acids Res 34:D247–D251. https://doi.org/10.1093/nar/gkj149.

38. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.

39. 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.

40. Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, Bonneau R, Shannon P, Srivastava A, Donohoe SM, Hood LE, Baliga NS. 2007. General transcription factor specified global gene regulation in archaea. Proc Natl Acad Sci U S A 104:4630–4635. https://doi.org/10.1073/pnas.0611663104.

41. Müller JA, DasSarma S. 2005. Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine *N*-oxide as terminal electron acceptors. J Bacteriol 187:1659–1667. https://doi.org/10.1128/JB.187.5.1659-1667.2005.

42. Hofacker A, Schmitz KM, Cichonczyk A, Sartorius-Neef S, Pfeifer F. 2004. GvpE- and GvpD-mediated transcription regulation of the *p-gvp* genes encoding gas vesicles in *Halobacterium salinarum*. Microbiology 150:1829–1838. https://doi.org/10.1099/mic.0.27078-0.

43. Brooks AN, Reiss DJ, Allard A, Wu WJ, Salvanha DM, Plaisier CL, Chandrasekaran S, Pan M, Kaur A, Baliga NS. 2014. A system-level model for the microbial regulatory genome. Mol Syst Biol 10:740. https://doi.org/10.15252/msb.20145160.

44. Peck RF, DasSarma S, Krebs MP. 2000. Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. Mol Microbiol 35:667–676. https://doi.org/10.1046/j.1365-2958.2000.01739.x.

45. Zwietering MH, Jongenburger I, Rombouts FM, van't Riet K. 1990. Modeling of the bacterial growth curve. Appl Environ Microbiol 56:1875–1881.

46. Banks AP, Lawless C, Lydall DA. 2012. A quantitative fitness analysis workflow. J Vis Exp 66:e4018. https://doi.org/10.3791/4018.

47. Kaufman C, Sain S. 2009. Bayesian functional ANOVA modeling using Gaussian process prior distributions. Bayesian Anal 5:123–149. https://doi.org/10.1214/10-BA505.

48. Todor H, Gooding J, Ilkayeva OR, Schmid AK. 2015. Dynamic metabolite profiling in an archaeon connects transcriptional regulation to metabolic consequences. PLoS One 10:e0135693. https://doi.org/10.1371/journal.pone.0135693.

49. Todor H, Dulmage K, Gillum N, Bain JR, Muehlbauer MJ, Schmid AK. 2014. A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon *Halobacterium salinarum*. Mol Microbiol 93:1172–1182. https://doi.org/10.1111/mmi.12726.

50. Tonner PD, Pittman AM, Gulli JG, Sharma K, Schmid AK. 2015. A regulatory hierarchy controls the dynamic transcriptional response to extreme oxidative stress in archaea. PLoS Genet 11:e1004912. https://doi.org/10.1371/journal.pgen.1004912.

51. Warner JR. 1999. The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24:437–440. https://doi.org/10.1016/S0968-0004(99)01460-7.

52. Scott M, Klumpp S, Mateescu EM, Hwa T. 2014. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol Syst Biol 10:747. https://doi.org/10.15252/msb.20145379.

53. Lu C, Brauer MJ, Botstein D. 2009. Slow growth induces heat-shock resistance in normal and respiratory-deficient yeast. Mol Biol Cell 20:891–903. https://doi.org/10.1091/mbc.E08-08-0852.

54. Plaisier CL, Lo FY, Ashworth J, Brooks AN, Beer KD, Kaur A, Pan M, Reiss DJ, Facciotti MT, Baliga NS. 2014. Evolution of context dependent regulation by expansion of feast/famine regulatory proteins. BMC Syst Biol 8:122. https://doi.org/10.1186/s12918-014-0122-2.

55. Coker JA, DasSarma P, Kumar J, Müller JA, DasSarma S. 2007. Transcriptional profiling of the model archaeon *Halobacterium* sp. NRC-1: responses to changes in salinity and temperature. Saline Systems 3:6. https://doi.org/10.1186/1746-1448-3-6.

56. Weng RR, Shu HW, Chin SW, Kao Y, Chen TW, Liao CC, Tsay YG, Ng WV. 2014. OMICS in ecology: systems level analyses of *Halobacterium salinarum* reveal large-scale temperature-mediated changes and a requirement of CctA for thermotolerance. OMICS 18:65–80. https://doi.org/10.1089/omi.2012.0117.

57. Lee C, Wigren E, Lünsdorf H, Römling U. 2016. Protein homeostasis-more than resisting a hot bath. Curr Opin Microbiol 30:147–154. https://doi.org/10.1016/j.mib.2016.02.006.

58. Pang WL, Kaur A, Ratushny AV, Cvetkovic A, Kumar S, Pan M, Arkin AP, Aitchison JD, Adams MW, Baliga NS. 2013. Metallochaperones regulate intracellular copper levels. PLoS Comput Biol 9:e1002880. https://doi.org/10.1371/journal.pcbi.1002880.

59. Thieringer HA, Singh K, Trivedi H, Inouye M. 1997. Identification and developmental characterization of a novel Y-box protein from *Drosophila melanogaster*. Nucleic Acids Res 25:4764–4770. https://doi.org/10.1093/nar/25.23.4764.

60. Lyabin DN, Eliseeva IA, Ovchinnikov LP. 2014. YB-1 protein: functions and regulation. Wiley Interdiscip Rev RNA 5:95–110. https://doi.org/10.1002/wrna.1200.

61. Wolf YI, Makarova KS, Yutin N, Koonin EV. 2012. Updated clusters of orthologous genes for archaea: a complex ancestor of the archaea and the byways of horizontal gene transfer. Biol Direct 7:46. https://doi.org/10.1186/1745-6150-7-46.

62. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, Tsai IJ, Bergman CM, Bensasson D, O'Kelly MJT, van Oudenaarden A, Barton DBH, Bailes E, Nguyen AN, Jones M, Quail MA, Goodhead I, Sims S, Smith F, Blomberg A, Durbin R, Louis EJ. 2009. Population genomics of domestic and wild yeasts. Nature 458:337–341. https://doi.org/10.1038/nature07743.

63. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, Wong S, Hawkins JS, Lu CHS, Koo BM, Marta E, Shiver AL, Whitehead EH, Weissman JS, Brown ED, Qi LS, Huang KC, Gross CA. 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. Cell 165:1493–1506. https://doi.org/10.1016/j.cell.2016.05.003.

64. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2:2006.0008. https://doi.org/10.1038/msb4100050.

65. Sarmiento F, Mrázek J, Whitman WB. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon Metha-

nococcus maripaludis. Proc Natl Acad Sci U S A 110:4726–4731. https://doi.org/10.1073/pnas.1220225110.

66. Reid JE, Wernisch L. 2016. Pseudotime estimation: deconfounding single cell time series. Bioinformatics 32:2973–2980. https://doi.org/10.1093/bioinformatics/btw372.

67. Äijö T, Butty V, Chen Z, Salo V, Tripathi S, Burge CB, Lahesmaa R, Lähdesmäki H. 2014. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. Bioinformatics 30:i113–i120. https://doi.org/10.1093/bioinformatics/btu274.

68. Hensman J, Lawrence ND, Rattray M. 2013. Hierarchical Bayesian modelling of gene expression time series across irregularly sampled replicates and clusters. BMC Bioinformatics 14:252. https://doi.org/10.1186/1471-2105-14-252.

69. Fusi N, Listgarten J. 2016. Flexible modelling of genetic effects on function-valued traits, p 95–110. In Singh M (ed), Research in computational molecular biology. Lecture Notes in Computer Science, vol 9649. Lecture Notes in Bioinformatics. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-319-31957-5_7.

70. Dufour YS, Imam S, Koo BM, Green HA, Donohue TJ. 2012. Convergence of the transcriptional responses to heat shock and singlet oxygen stresses. PLoS Genet 8:e1002929. https://doi.org/10.1371/journal.pgen.1002929.

71. Seo SW, Kim D, Latif H, O'Brien EJ, Szubin R, Palsson BO. 2014. Deciphering Fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. Nat Commun 5:4910. https://doi.org/10.1038/ncomms5910.

72. Wilbanks EG, Larsen DJ, Neches RY, Yao AI, Wu CY, Kjolby RA, Facciotti MT. 2012. A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. Nucleic Acids Res 40:e74. https://doi.org/10.1093/nar/gks063.

73. Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349–361. https://doi.org/10.1016/B978-0-12-385120-8.00015-2.

74. Yao AI, Facciotti MT. 2011. Regulatory multidimensionality of gas vesicle biogenesis in Halobacterium salinarum NRC-1. Archaea 2011:716456. https://doi.org/10.1155/2011/716456.

75. Ramsay JO, Silverman BW. 2005. Modelling functional responses with multivariate covariates, p 223–245. In Functional data analysis, 2nd ed. Springer Series in Statistics. Springer, New York, NY. http://link.springer.com/chapter/10.1007/0-387-22751-2_13.

76. Rasmussen CE, Williams CKI. 2006. Gaussian processes for machine learning. In Dietterich T (ed), Adaptive computation and machine learning. MIT Press, Cambridge, MA.

77. Neal RM. 2003. Slice sampling. Ann Statist 31:705–767. https://doi.org/10.1214/aos/1056562461.

78. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang PL, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. 2007. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc 2:2366–2382. https://doi.org/10.1038/nprot.2007.324.

79. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45.

80. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J. 2006. TM4 microarray software suite. Methods Enzymol 411:134–193. https://doi.org/10.1016/S0076-6879(06)11009-5.

81. R Core Team. 2015. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

82. Harrell FE, Jr. 2016. Hmisc: Harrell Miscellaneous. R package version 4.0-0. https://CRAN.R-project.org/package=Hmisc.

83. Wei T, Simko V. 2016. corrplot: visualization of a correlation matrix. R package version 0.77. https://CRAN.R-project.org/package=corrplot.

84. Darnell CL, Schmid AK. 2015. Systems biology approaches to defining transcription regulatory networks in halophilic archaea. Methods 86:102–114. https://doi.org/10.1016/j.ymeth.2015.04.034.

85. Dulmage KA, Todor H, Schmid AK. 2015. Growth-phase-specific modulation of cell morphology and gene expression by an archaeal histone protein. mBio 6:e00649-15. https://doi.org/10.1128/mBio.00649-15.

86. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 41:D991–D995. https://doi.org/10.1093/nar/gks1193.