Published in final edited form as:

J Mol Biol. 2008 November 14; 383(3): 713–730. doi:10.1016/j.jmb.2008.08.017.

Organization and Evolution of the Biological Response to Singlet Oxygen Stress

Yann S. Dufour^{1,2}, Robert Landick^{1,3}, and Timothy J. Donohue^{1,*}

¹Department of Bacteriology, University of Wisconsin-Madison, 1550 Linden Drive, Madison, WI 53706, USA

²BACTER Institute, University of Wisconsin-Madison, Madison, WI 53706, USA

³Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, WI 53706, USA

Abstract

The appearance of atmospheric oxygen from photosynthetic activity led to the evolution of aerobic respiration and responses to the resulting reactive oxygen species. In *Rhodobacter sphaeroides*, a photosynthetic α -proteobacterium, a transcriptional response to the reactive oxygen species singlet oxygen ($^1\text{O}_2$) is controlled by the group IV σ factor σ^E and the anti- σ factor ChrR. In this study, we integrated various large datasets to identify genes within the $^1\text{O}_2$ stress response that contain σ^E -dependent promoters both within *R. sphaeroides* and across the bacterial phylogeny. Transcript pattern clustering and a σ^E -binding sequence model were used to predict candidate promoters that respond to $^1\text{O}_2$ stress in *R. sphaeroides*. These candidate promoters were experimentally validated to nine *R. sphaeroides* σ^E -dependent promoters that control the transcription of 15 $^1\text{O}_2$ -activated genes. Knowledge of the *R. sphaeroides* response to $^1\text{O}_2$ and its regulator σ^E -ChrR was combined with large-scale phylogenetic and sequence analyses to predict the existence of a core set of approximately eight conserved σ^E -dependent genes in α -proteobacteria and γ -proteobacteria. The bacteria predicted to contain this conserved response to $^1\text{O}_2$ include photosynthetic species, as well as free-living and symbiotic/pathogenic nonphotosynthetic species. Our analysis also predicts that the response to $^1\text{O}_2$ evolved within the time frame of the accumulation of atmospheric molecular oxygen on this planet.

Keywords

singlet oxygen; reactive oxygen; species gene regulation; stress responses; sigma factors

Introduction

Genome sequences and genomewide datasets provide a way to transfer accumulated knowledge from model systems to other organisms. For example, by merging high-throughput

*Corresponding author. E-mail address: tdonohue@bact.wisc.edu.

Publisher's Disclaimer: This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues. Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit: <http://www.elsevier.com/copyright>

experimental technologies with computational tools, one can predict conservation of protein families, regulatory networks, and transcription factor binding sequences across prokaryotes and eukaryotes.^{1,2}

We analyze molecular aspects of the function, evolution, and phylogenetic conservation of prokaryotic transcriptional networks because of the large number of bacterial genome sequences and the deep understanding of these organisms. Despite these advantages, the relatively high evolution rate of prokaryotes makes this task challenging. For example, because of potentially divergent evolutionary histories, the proper functional assignment within groups of homologous genes is difficult when genes are considered in isolation.³ In addition, making inferences about prokaryotic regulatory networks or stress responses based solely on the presence of a transcription factor can be hazardous because DNA-binding proteins can represent interchangeable components in signal transduction pathways.⁴ Therefore, it is critical to consider gene function in the context of regulation, interaction of gene products with other functions, the organism's metabolic lifestyle, and phylogenetic distances when constructing bacterial transcriptional network models.

In this study, we analyze an important, yet incompletely characterized stress response to the reactive oxygen species singlet oxygen ($^1\text{O}_2$) in the facultative phototrophic bacterium *Rhodobacter sphaeroides*. This α -proteobacterium has been a model for the analysis of photosynthesis and other metabolic activities from a molecular, biochemical, and genomic perspective.⁵ Recently, a transcriptional response to $^1\text{O}_2$ was discovered in this bacterium.⁶ In *R. sphaeroides* and other phototrophs, $^1\text{O}_2$ is a toxic reactive oxygen species generated by energy transfer from excited photosynthetic pigments to oxygen.⁷ $^1\text{O}_2$ can also be generated in other ways by photosynthetic and nonphotosynthetic cells; its reactivity with biomolecules makes $^1\text{O}_2$ lethal to prokaryotes and eukaryotes.⁸ Thus, many organisms should or are known to contain a system used to sense the presence of $^1\text{O}_2$ and to activate target genes needed to mount a stress response to this reactive oxygen species.

The *R. sphaeroides* transcriptional response to $^1\text{O}_2$ is controlled by the group IV σ /anti- σ factor pair σ^E -ChrR.⁶ In this study, we used computational, experimental, and phylogenetic analyses to define most, if not all, of the direct targets for σ^E . Our results predict that the transcriptional response to $^1\text{O}_2$, the master regulators of this response, and various members of the σ^E -ChrR regulon are found in many other photosynthetic and nonphotosynthetic prokaryotes, suggesting that free-living or pathogenic/symbiotic species also encounter $^1\text{O}_2$ in nature. Our findings provide new perspectives on the evolution, natural history, and biological relevance of the transcriptional response to $^1\text{O}_2$.

Results

Identification of potential σ^E target genes

The *R. sphaeroides* *rpoE*ChrR operon, which encodes σ^E and ChrR, contains a σ^E -dependent promoter, so its expression increases in ΔChrR cells (due to high σ^E activity).^{6,9} To identify potential members of the σ^E -ChrR regulon, we clustered transcript levels from 67 publicly available *R. sphaeroides* global gene expression datasets (Materials and Methods). The Pearson's correlation coefficient between the abundance of *rpoE* transcripts and that of *chrR* transcripts across this dataset was high (0.95) as expected, since these two genes are cotranscribed.^{6,9} In addition, the correlation coefficient between transcript levels for *rpoE* and another known σ^E target gene (*rpoH*)⁶ was 0.88, suggesting that clustering genes with expression patterns that correlated with that of *rpoE* could predict other potential members of the σ^E -ChrR regulon.

To prevent spurious correlation, only genes that showed a statistically significant differential expression between Δ ChrR and wild-type cells (which have low σ^E activity) were included in clustering analysis. RNA abundance levels from the resulting 110 genes were clustered (Supplemental Fig. 1) in a hierarchical tree based on the Pearson's correlation coefficient of their respective transcript patterns. In this tree, a cluster of transcript patterns containing *rpoE* and some other known σ^E target genes was identified (Fig. 1),⁶ indicating that this method successfully predicts other members of the σ^E -ChrR regulon. Likewise, genes that are likely to belong to the same transcription unit, based on computational predictions of *R. sphaeroides* operons,^{10,11} were present in the *rpoE*-containing transcript cluster, providing further confidence in the predictive nature of the output. This *rpoE*-containing cluster contains one gene that had not previously been identified as a potential σ^E target (RSP1852), suggesting the existence of additional yet-to-be-identified members of this regulon.

If these operons were all directly transcribed by σ^E , each one should contain a conserved promoter motif. To test whether DNA sequences upstream of these clustered operons contain a common sequence element, we queried them for a conserved motif using BioProspector.¹² As predicted, a conserved motif was found upstream of all the potential operons contained within the *rpoE*-containing cluster. This motif contains two high-information-content regions separated by a variable spacer sequence, which is typical of the -35 and -10 regions recognized by group IV bacterial σ factors (Fig. 2a). The motif also contains elements of an *R. sphaeroides* σ^E promoter derived from past mutational analysis.⁹ All but one of the known or predicted σ^E promoters have 13 bp between their putative -10 and -35 elements; the exception (RSP1409) has a 14-bp spacer in its putative promoter.

We recognize that clustering of global gene expression patterns would likely not identify operons containing multiple promoters if σ^E only makes a small contribution to the total transcription, as is true for the *cycA* gene.⁹ To search for additional candidate members of the σ^E -ChrR regulon, a library of putative *R. sphaeroides* promoters (Materials and Methods) was queried for sequences related to the conserved sequence motif. To perform this query, the conserved sequence motif was converted into a position-specific weighted matrix (PSWM) and used to score the promoter library based on the content information of the best match in each sequence. To relax the stringency of the query, spacer lengths of 13 bp or 14 bp were allowed without penalty, even though a spacing of 13 bp between the -35 region and the -10 region was more frequent among operons with gene expression patterns that clustered with the *rpoE*-*chrR* operon (see above). Among the best 15 matches (Table 1), six were known σ^E target operons and one was a candidate σ^E promoter upstream of a gene within the *rpoE*-containing cluster (RSP1852). On the other hand, none of the promoter regions for any of the other 110 differentially expressed genes analyzed above (Supplemental Fig. 1) contained a match to the motif with a score $\geq 75\%$ of the maximum. Thus, it appears that the other differentially expressed operons (Supplemental Fig. 1) are not likely to contain σ^E promoters.

Experimental analysis of candidate σ^E target genes

To test these candidates (Table 1) for the presence of a functional σ^E promoter, the regions spanning from -200 bp to +80 bp relative to the predicted transcription initiation site were analyzed for function *in vitro* and *in vivo*. Control experiments with known σ^E target genes allowed us to assess the relative strength of these candidate promoters and to analyze the ability of ChrR to inhibit their function.

When multiple-round transcription reactions were performed using *R. sphaeroides* core RNA polymerase reconstituted with recombinant σ^E , a product of the predicted size was detected from the candidate RSP1852 promoter (Fig. 3a) at a level similar to that obtained from the strong σ^E -dependent promoter (RSP1092) upstream of *rpoE*.⁶ As expected, addition of the anti- σ factor ChrR prevented the accumulation of the σ^E transcript from the known RSP1092

promoter and the candidate RSP1852 promoter (Fig. 3a). In contrast, the levels of transcript obtained from the candidate RSP3336 and RSP6222 promoters were lower, and detection of these products required a σ^E concentration that was 5- to 10-fold (250–500 nM recombinant protein) in excess over *R. sphaeroides* core RNA polymerase (50 nM; Fig. 3b). Control experiments indicate that similar increases in the concentration of σ^E do not generate a higher amount of product from strong promoters such as RSP1092 in multiple-round transcription assays under these conditions.¹³ Thus, DNA binding by σ^E holoenzyme may be a limiting kinetic step for transcription at the RSP3336 and RSP6222 promoters. With RSP3336 and RSP6222, we were unable to show inhibition of transcript formation by ChrR because the anti- σ factor could not be obtained at concentrations sufficiently high to inhibit all the σ^E in the assay.

All other candidate promoters shown in Table 1 failed to produce a detectable transcript *in vitro*, regardless of the levels of σ^E added to core RNA polymerase (data not shown). Upon examination of the promoters tested, those active with σ^E *in vitro* contain a GTA motif in their –10 region, whereas those that were inactive in this assay have one or more substitutions in this motif. The RSP1207 promoter, which was also not transcribed by σ^E *in vitro*, contains a GTA motif in its putative –10 region but contains a substitution in the first T of the –35 region that it is conserved in all the active promoters. Thus, it appears that specific bases at these positions of the –10 and –35 elements are critical for σ^E -dependent transcription.

We also fused these and other candidate promoter sequences to a promoterless β -galactosidase gene on a low-copy plasmid to test for activity *in vivo*. The activity of these reporter genes was tested in wild-type *R. sphaeroides* cells (low σ^E activity), Δ ChrR cells (high σ^E activity), and cells lacking the *rpoEchrR* operon (no σ^E activity). Consistent with the results of *in vitro* transcription assays, we observed high reporter gene activity in Δ ChrR cells from known σ^E target genes (RSP601, RSP1091, RSP1092, RSP1409, and RSP2143) and RSP1852, while activities from both RSP6222 and RSP3666 were lower (Fig. 4). All detectable activities of each of these promoters were lost in $\Delta\sigma^E$ cells, suggesting that function under these conditions required a functional *rpoE* gene. For unknown reasons, the activity of the RSP6222 promoter is reproducibly higher than that of RSP3666 *in vivo* even though they generated comparable levels of transcript under the conditions used *in vitro*. When we tested each of the remaining candidate genes in Table 1 for activity *in vivo* or *in vitro* (RSP6076, RSP4129, RSP3284, RSP1207, RSP1521, and RSP0357), there was no evidence for the presence of a σ^E -dependent promoter in either assay (data not shown).

In sum, the combination of clustering analysis, promoter motif predictions, and functional assays allowed us to identify additional members of the σ^E –ChrR regulon. If these operons constitute the entire regulon (see below), then the number of genes directly transcribed by σ^E is relatively small (9 operons, 15 genes), and its promoter motif has a high information content.

Genomewide mapping of σ^E interaction with the genome

While these experiments identified new members of the σ^E –ChrR regulon, we performed another experiment to assess what fraction of the σ^E promoters has been captured. To do this, we used a chromatin immunoprecipitation and microarray (ChIP-chip) assay to score genomewide interactions of σ^E or the β' subunit of RNA polymerase with DNA (Materials and Methods). For immunoprecipitation, we used specific polyclonal antibodies to enrich DNA bound by each protein in Δ ChrR cells (high σ^E activity). Enriched genomic regions with $p \leq 0.01$ were detected using TAMALPAIS¹⁴ and ranked by the average of the $\log_2$ ratios of the 10 best consecutive probes in each selected region (Table 2).

The data show that regions containing strong σ^E -dependent promoters (RSP2143–2144, RSP0601, RSP1091–1087, RSP1092–1093, RSP1852, and RSP1409) were significantly enriched by immunoprecipitation with antibody against σ^E (Table 2). In addition, each of these operons showed enrichment for the β' subunit of RNA polymerase (Fig. 5; Supplemental Fig. 2), indicating that they were actively transcribed in Δ ChrR cells (high σ^E activity). Furthermore, if the center of the enrichment peak for σ^E is indicative of DNA binding, the maximal occupancy by σ^E coincides with the position of the conserved motif (Fig. 2a). In contrast, no evidence for σ^E or RNA polymerase interaction was detected at promoters that had very low activity *in vitro* or when using a *lacZ* reporter gene (RSP0296, RSP3336, and RSP6222; Fig. 4). Therefore, these results cannot rule out the existence of other weak σ^E -dependent promoters.

This experiment identified 13 other regions enriched for σ^E that were not among candidates from the clustering or motif-scanning datasets. However, most of these σ^E occupancy sites are within protein coding sequences or large intergenic regions (Table 2; Supplemental Fig. 2), which were not part of the previous searches. When we used the conserved motif (Fig. 2a) to predict the location and orientation of putative σ^E -dependent promoters in these regions (Table 1), several have a relatively good match, but most are oriented in the opposite direction of the nearby open reading frame (Supplemental Fig. 2). In addition, with the exception of RSP1612 (Fig. 5), RSP3832, and RSP2401 (Supplemental Fig. 2), there is no evidence for RNA polymerase occupancy (active transcription) beyond these putative σ^E binding sites. Finally, the global gene expression dataset indicates that none of the open reading frames neighboring these 13 other regions of σ^E occupancy showed increased expression levels when σ^E activity was high. Thus, at this point, we lack evidence supporting the hypothesis that these other σ^E occupancy sites are functional σ^E -dependent promoters *in vivo*. However, we cannot exclude the existence of additional σ^E -dependent promoters directing the transcription of small nonannotated open reading frames or small RNA. Alternatively, those apparently nontranscribed regions that are enriched for σ^E may represent poised RNA polymerase complexes in promoters that require different environmental conditions or additional proteins to be actively transcribed. If any of these regions contains functional σ^E promoters, further efforts will be needed to assay their activity or to elucidate the function of any product. In sum, when the results of ChIP-chip assays are combined with other experimental and computational approaches, we conclude that we have identified the vast majority of the protein coding genes directly and actively transcribed by *R. sphaeroides* σ^E .

Phylogenetic analysis of σ^E -ChrR across the bacterial kingdom

Based on what we learned about the *R. sphaeroides* σ^E -ChrR regulon and its promoters, we tested for the presence of this response across the bacterial kingdom. To do this, we analyzed the genomes of 84 bacteria that contain a gene encoding a group IV alternative σ factor adjacent to one coding for a ChrR homolog.¹⁵ We limited our analysis to 73 bacteria with almost complete genome sequences, since we wanted to probe for both σ^E -ChrR (regulators of the $^1\text{O}_2$ stress response) and members of the σ^E -ChrR regulon. These 73 σ^E -ChrR homologs were distributed almost evenly between α -proteobacteria and γ -proteobacteria, with only one encoded by a β -proteobacterium (*Acidovorax anenae* subsp. *citrulli*) or a δ -proteobacterium (*Myxococcus xanthus*).

To test for lateral transfer of the σ^E -ChrR structural genes, we constructed a species tree based on the amino acid sequence of RuvB, RpoD, and GyrB. These proteins were selected to construct the species phylogeny because their genes are found in single copy in the selected bacteria and they are thought to be subject to lateral gene transfer only rarely. The phylogenetic trees constructed with each protein were in agreement, supporting the assumption that no lateral gene transfer occurred. Consequently, we concatenated the RuvB, RpoD, and GyrB protein sequences to build a species phylogenetic tree. Individual phylogenetic trees constructed for

σ^E and ChrR also suggested their coinheritance, so their amino acid sequences were concatenated to build a σ^E -ChrR phylogenetic tree.

By comparing the two phylogenetic trees (Fig. 6), it appears that the σ^E -ChrR tree generally mirrors the species tree. The only exceptions are as follows: *Pseudomonas syringae* species and *Oceanospirillum* sp. MED 92, members of the γ -proteobacteria, contain σ^E -ChrR proteins that appear more related to those found in α -proteobacteria. In addition, the σ^E -ChrR proteins of δ -proteobacteria and β -proteobacteria cluster with those found in the α -proteobacteria, suggesting lateral transfer of these genes into these species. Taken together, this analysis suggests that the σ^E -ChrR pair evolved before the branching of the α -proteobacteria and γ -proteobacteria, a point that will be revisited in the Discussion.

Characterizing the putative σ^E -ChrR regulon across species

To predict common or divergent aspects of this network across these bacteria, we also analyzed these 73 genomes for possible orthologs of genes within the *R. sphaeroides* σ^E -ChrR regulon and genes predicted to contain σ^E promoters. Some of these 73 genomes were not completely assembled, so not all potential σ^E target genes might be captured in other bacteria. We used a two-step process to predict members of this regulon in these bacteria based on the *R. sphaeroides* σ^E promoter motif and the known target genes.

The first step was to identify candidate promoters in these 73 genomes. This prediction was based on amino acid conservation across the 73 *R. sphaeroides* σ^E homologs. Because residues of *Escherichia coli* σ^E that make contact with the -35 promoter element are known¹⁶, we can predict residues of the 73 *R. sphaeroides* σ^E homologs that are involved in the -35 promoter element recognition. Based on the conservation of these residues among the 73 *R. sphaeroides* σ^E homologs, we propose that the sequence of the -35 promoter element recognized by these proteins is conserved. While information about interactions between domains 2.3–2.4 and the -10 promoter element is not available for group IV σ factors, alignments of the 73 homologs reveal a high degree of amino acid sequence conservation in this region (Fig. 7). These facts supported a hypothesis stating that the DNA motif recognized by these σ factors would be sufficiently conserved so that the *R. sphaeroides* σ^E promoter motif could be used to query for potential target genes in the 73 bacterial genomes that contain σ^E -ChrR homologs.

To test this hypothesis, we performed a *de novo* search for the *R. sphaeroides* σ^E promoter motif upstream of *rpoEchrR* in other bacteria. This query identified a sequence upstream of the putative σ^E structural gene in 57 of these 73 microbes that is almost identical with that of the *R. sphaeroides* σ^E motif (Fig. 2b). This finding is not surprising, since genes for group IV σ factors are often positively autoregulated.¹⁷ When we constructed a new PSWM using the putative promoter sequences upstream of these 57 *rpoEchrR* operons, this phylogenetically determined PSWM was almost identical with that of the *R. sphaeroides* σ^E motif. We then used the phylogenetically determined σ^E PSWM to score all upstream regions in these 73 genomes to identify candidate members of the σ^E -ChrR regulons. The threshold used to score a positive match was intentionally set relatively low to increase the sensitivity of the analysis, resulting in an increased number of possible false positives. However, these presumed false-positive matches are not conserved across the phylogeny.

Subsequently, we asked whether groups of orthologous genes (constructed using OrthoMCL; Materials and Methods) were consistently associated with a predicted σ^E promoter. When orthologs were identified downstream of a candidate σ^E promoter, they were scored to reflect their potential membership in the σ^E -ChrR regulon (Materials and Methods). To account for the existence of σ^E -dependent operons, genes likely to be cotranscribed with one containing a match to the σ^E promoter motif were considered positive. After each protein coding sequence

had been scored, ortholog group-specific values were calculated and ranked by summing individual values within each group.

After the groups of orthologous proteins had been ranked, the dataset was split to include proteins found in α -proteobacteria or γ -proteobacteria (Fig. 8a and b; data derived from one δ -proteobacterium and one β -proteobacterium were removed, since they did not exhibit any informative pattern). This analysis showed that the σ^E (protein ID 21307) and ChrR (protein ID 20362) homologs ranked highest in both α -proteobacteria and γ -proteobacteria as expected, since their expression should be positively autoregulated (see above). The remaining highly ranked groups from both α -proteobacteria and γ -proteobacteria contain orthologs of known members of the *R. sphaeroides* σ^E -ChrR regulon (RSP1087; protein ID no. 20348; RSP1088, protein ID no. 21508; RSP1090, protein ID no. 20013; RSP1091, protein ID no. 21520; RSP2143, protein ID no. 21876; RSP2144, protein ID no. 254; Table 3). The presence of a σ^E promoter motif upstream of these genes supports their assignment to a core set of genes for this regulon and the hypothesis that some part of the biological response controlled by σ^E and ChrR is conserved in α -proteobacteria and γ -proteobacteria. Unfortunately, the biological functions of most of those conserved proteins are unknown. Only RSP2143 and RSP2144 have been annotated (DNA photolyase and cyclopropane fatty acid synthase, respectively) in *R. sphaeroides*,^{18–20} providing some clues as to the response of these bacteria to the signal that increases σ^E activity (see the Discussion).

Other ortholog groups appear to be specific to the α -proteobacteria or the γ -proteobacteria. For example, RSP0601 (*rpoH_{II}*, protein ID no. 20947), which encodes a second member of the σ^{32} family, is part of the *R. sphaeroides* σ^E -ChrR regulon,^{6,21} so it is not surprising to find *rpoH_{II}* orthologs that appear to contain a σ^E promoter in α -proteobacteria such as *Roseovarius* sp. 217, *Hyphomonas neptunium* ATCC 15444, *Oceanicaulis alexandrii* HTCC2633, *Jannaschia* sp. CCS1, *Loktanella vestfoldensis* SKA53, *Maricaulis maris* MCS10, and *Silicibacter pomeroyi* DSS3. However, other α -proteobacteria contain an *rpoH_{II}* gene that lacks elements related to the conserved σ^E promoter motif even though this sequence is present upstream of other predicted core members of this regulon. In addition, the predicted σ^E -ChrR regulons in these α -proteobacteria do not appear to be extended to compensate for the loss of any RpoH_{II} target gene, raising questions about the existence of a second tier of ¹O₂-responsive genes in these species and the exact role of RpoH_{II} in bacteria containing this gene product (see the Discussion).

In the γ -proteobacteria that contain *rpoEchrR* operons, no *rpoH_{II}* orthologs were found. On the other hand, the motif search revealed that a putative polyketide cyclase (protein ID no. 22231 in Fig. 8b) specific for the γ -proteobacteria often possesses a σ^E binding motif in its promoter region. In addition, this gene is generally found in the genomic neighborhood of orthologs of the *R. sphaeroides* σ^E -regulated genes in these γ -proteobacteria[†].

In addition, several species of α -proteobacteria and γ -proteobacteria do not have the conserved σ^E binding motif associated with any of the predicted conserved regulon members. To test the hypothesis that the σ^E promoter motif evolved into a different sequence, we queried the promoter regions of possible regulon members in each species individually for a conserved sequence that related to a σ factor binding site. This approach failed to yield any conserved motif with high information content, so it is possible that σ^E is not functional or that both the σ^E binding motif and the regulon composition have diverged in these bacteria. However, no evidence for such divergence was evident in sequence alignments of their σ^E homologs in regions 2.3–2.4 and 4.2.

[†]<http://img.jgi.doe.gov/>

In summary, this analysis uncovered a set of genes that potentially constitute the core of a σ^E -dependent biological response to $^1\text{O}_2$ that could be conserved across distantly related bacterial species. The implications of this finding are presented in the Discussion.

Discussion

As oxygen appeared in the biosphere, cells evolved through aerobic respiration and signal transduction pathways to respond to the toxic byproducts of reactive oxygen species. In *R. sphaeroides*, the reactive oxygen species $^1\text{O}_2$ increases the activity of the alternative σ factor σ^E by somehow inactivating its anti- σ factor ChrR.^{6,13,15} Although $^1\text{O}_2$ can kill cells,^{8,22} the process by which they respond to and repair the damage from this compound is unknown. To understand how bacteria react to $^1\text{O}_2$, we sought to define the entire *R. sphaeroides* σ^E -ChrR regulon and to investigate the conservation of target genes in other cells containing these proteins. Our results predict that *R. sphaeroides* σ^E directly transcribes nine operons (15 genes), a small regulon size compared to other characterized group IV alternative σ factors (Table 4).^{2,23–25} It also appears that a core set of putative σ^E target genes is conserved in many α -proteobacteria and γ -proteobacteria. Finally, the ecology of the organisms predicted to mount a σ^E -dependent response to $^1\text{O}_2$ is diverse, reinforcing the notion that there are different sources of this reactive oxygen species in α -proteobacteria and γ -proteobacteria.

Predicted σ^E binding site

We identified *R. sphaeroides* σ^E -dependent promoters using a combination of transcription profiling, bioinformatic analysis, and ChIP-chip assays. Our results suggest that we have found most, if not all, proteins members of the *R. sphaeroides* σ^E -ChrR regulon. However, we cannot exclude the existence of other promoters with activities or features similar to those of RSP0296, RSP6222, and RSP3336, and the function of the putative promoters that are not associated with open reading frames still needs to be investigated. Experimental testing of several candidate promoters that contained a sequence deviating only by one base from the predicted *R. sphaeroides* σ^E consensus failed to give activity *in vitro* and *in vivo*, or to show occupancy by $\text{E}\sigma^E$ in ChIP-chip analysis.

The other genes that show significant correlation with known σ^E targets by expression clustering do not contain sequences related to the σ^E promoter motif. We know that $^1\text{O}_2$ activates a gene expression cascade; a known σ^E target gene is one encoding the alternative σ factor RpoH_{II}.⁶ We propose that some or all of the genes in this other cluster that do not contain a σ^E promoter are targets for RpoH_{II} (Green *et al.*, unpublished data). If this is the case, there would be a large number of additional operons activated in the presence of $^1\text{O}_2$.

The motif defined by nine confirmed *R. sphaeroides* σ^E promoters (TGATCC-[15,16]-CCGTA) agrees with the results of past mutational analysis⁹ and with the prediction made from a smaller dataset derived from analyzing the effect of blue light on global transcript levels.²⁶ The previous mutational study found some single-base substitutions in the nonconserved region downstream of the -10 region that significantly increased its activity. The test promoter for this mutational study contains a suboptimal -10 sequence, so it is possible that sequences in this region may facilitate promoter melting by RNAP holoenzyme to initiate transcription.

The high sequence conservation of the *R. sphaeroides* σ^E promoter (21.5 bit) is consistent with other characterized group IV σ factor motifs,^{2,16,27} since their information content can be higher than those of other members of the σ^{70} superfamily.^{28,29} Group IV σ factors have not been reported to function in conjunction with DNA-binding activators, so their higher information content may reflect a large dependence on DNA sequence elements to ensure binding specificity. Our analysis of *R. sphaeroides* σ^E , together with other analyses of group

IV σ factors, supports the value of using sequence motif models to predict promoter elements. 2,23–25

By analyzing how region 4 of the group IV *E. coli* σ factor σ^E binds the –35 region, it has been predicted that other group IV σ factors contain a critical AAC motif in this promoter element.¹⁶ However, the predicted *R. sphaeroides* σ^E –35 region contains an ATC motif, suggesting that the DNA geometry created by the AA/TT tract in the *E. coli* σ^E –DNA complex might not be present in the cognate *R. sphaeroides* σ^E promoters. If we infer the residues that contact the ATC motif in the *R. sphaeroides* σ^E –35 motif by amino acid sequence alignment with *E. coli* σ^E , they differ from those in *E. coli* σ^E but are conserved in the *R. sphaeroides* σ^E orthologs we considered (Fig. 7). This finding supports the hypothesis that region 4.2 side chains of *R. sphaeroides* σ^E make specific DNA contacts and explains why the sequence of its –35 region is different from that of *E. coli* σ^E promoters but is conserved in the ones we predict to be transcribed by *R. sphaeroides* σ^E orthologs.

Phylogenetic conservation of σ^E promoter elements

When we searched for the σ^E binding motif in other genomes, 57 out of 73 species that contain *rpoE*/*chrR* homologs possess an identical or nearly identical sequence upstream of the gene for the *R. sphaeroides* σ^E ortholog. We also frequently found this promoter motif upstream of genes encoding homologs of known *R. sphaeroides* σ^E targets. The fact that both the promoter sequence and the composition of the regulon show significant conservation is remarkable, considering the phylogenetic distance represented by the groups of bacterial species considered.

Only a few of the bacterial genomes analyzed that encode σ^E –ChrR homologs do not contain the conserved σ^E promoter motif upstream of a gene that is known to be part of the *R. sphaeroides* regulon. If σ^E and ChrR are functional in these bacteria, their regulons might have a different composition or function. It is also possible, but less likely, that the binding motif for σ^E has diverged even though the amino acid sequence alignments show a similar degree of identity to bacteria that contain a conserved promoter upstream of homologs of *R. sphaeroides* σ^E target genes.

Composition of the core σ^E –ChrR regulon

Our analysis predicts that eight of the known *R. sphaeroides* σ^E target genes are conserved across 45 of the 73 species analyzed (Table 3), suggesting that they define a core set of functions needed for the response to $^1\text{O}_2$. Few of these genes have been experimentally characterized to date, but some predictions can be made about the biological response to $^1\text{O}_2$ stress.

Functions within the proposed core σ^E –ChrR regulon include activities for the repair or prevention of damage caused by $^1\text{O}_2$. RSP2143 (*phrA*) is predicted to be a DNA photolyase,¹⁹ so this protein could repair pyrimidine dimers caused by $^1\text{O}_2$. RSP2144 (*cfaS*) is predicted to encode a cyclopropane–fatty acyl–phospholipid synthetase that could maintain cell membrane integrity by adding a methylene group across the double bond of unsaturated fatty acids,³⁰ which are particularly subject to peroxidation by $^1\text{O}_2$.³¹ RSP1087 and RSP1091 each have homology to proteins with oxidation–reduction activity (short-chain dehydrogenases/reductases and flavin-containing oxidoreductases, respectively) so they could detoxify peroxidation products of fatty acids or replenish an unknown molecule(s) that is critical for survival in the presence of $^1\text{O}_2$.³² The product of a σ^E target gene that has so far been only found in γ -proteobacterial genomes (protein ID no. 22231) is related to the small polyketide cyclase family and steroid isomerase. It is possible that this protein works in conjunction with *cfaS* for positioning double bonds in unsaturated fatty acids before or after their modification by *cfaS*. Given the known ability of carotenoids to quench $^1\text{O}_2$ ³³ and the presence of genes

encoding enzymes for synthesis of these pigments in several organisms that contain σ^E -ChrR orthologs, it is surprising that no gene coding for proteins predicted to be involved in the synthesis of these compounds appears to be part of the σ^E -ChrR regulon in any of the bacteria analyzed.

The extended σ^E -ChrR regulon

Other members of the σ^E -ChrR regulon are not widely distributed across species (Table 3). For example, *R. sphaeroides* RSP1088 (putative glycosyl hydrolase) and RSP1089 (sugar/cation symporter) are part of an operon with members of the conserved core regulon in *R. sphaeroides* species and close relatives. There was also little conservation across species for RSP1852, RSP1409, RSP3336, and RSP6222—proteins with unknown biological functions.

RSP0601, which encodes RpoH_{II} (a second σ^{32} family member), is only present in a fraction of the α -proteobacteria, and most of them appear to have a σ^E -dependent promoter upstream of this gene. We know that RpoH_{II} is part of the *R. sphaeroides* $^1\text{O}_2$ stress response⁶ (Green *et al.*, unpublished data) and that some RpoH_{II} target genes are also transcribed by RpoH_I.²¹ Therefore, in *R. sphaeroides* and potentially other α -proteobacteria, $^1\text{O}_2$ activates a transcriptional cascade composed of a small σ^E -ChrR regulon and a second possibly larger suite of RpoH_{II}-dependent target genes. We failed to find evidence for another conserved transcription factor that contains a σ^E promoter in cells lacking an RpoH_{II} gene, so additional analysis of those α -proteobacteria and γ -proteobacteria lacking this protein is needed to understand whether and how cells respond to $^1\text{O}_2$ in the absence of this second tier of a gene expression cascade.

RSP3077 and, to a lesser extent, RSP2389 contain promoters with a σ^E binding motif in many γ -proteobacteria. RSP2389 is a known RpoH_{II} target in *R. sphaeroides*, and RSP3077 is also a likely candidate for transcription by this alternative σ factor (Green *et al.*, unpublished data). RSP3077 encodes another putative DNA photolyase,¹⁹ while RSP2389 encodes a putative glutathione peroxidase that can reduce peroxidation products caused by $^1\text{O}_2$.³² Expression of a glutathione peroxidase homolog is activated by $^1\text{O}_2$ formation in *Chlamydomonas reinhardtii*, so this aspect of the response appears to be conserved across biological kingdoms.³⁴

Most additional members of an extended σ^E -ChrR regulon in a *Shewanella* group and a *Vibrio* group lack experimentally supported annotation (Table 3). In the pathogenic *Vibrio* group, several annotated cytochrome *c* biogenesis proteins (protein ID no. 22038, cytochrome *c* biogenesis protein E; protein ID no. 127, cytochrome *c* maturation protein F; protein ID no. 20054, ABC heme exporter inner membrane subunit) are predicted to be part of the σ^E -ChrR regulon. Tetrapyrroles are a potential source of $^1\text{O}_2$, and heme-containing proteins are damaged by this reactive oxygen species, so placement of cytochrome *c* biogenesis operons in this stress response might help replenish the pool of these proteins in some species. In this regard, we note that transcription profiling experiments in *Vibrio fischeri* indicate that the transcript levels for our predicted σ^E targets correlate with the *rpoEchrR* homolog transcripts (Y. Wang and E. G. Ruby, personal communication, 2008).

Sources of $^1\text{O}_2$ in various bacteria

Our analysis predicts that the core σ^E -ChrR regulon is present in diverse bacteria found in various environments. One major source of $^1\text{O}_2$ for *R. sphaeroides* is via light energy transfer from photosynthetic pigments to molecular oxygen.³⁵ Thus, it is not surprising to find σ^E -ChrR, members of the *R. sphaeroides* σ^E regulon, and conserved promoters in other photosynthetic bacteria. However, among photosynthetic α -proteobacteria,

Rhodopseudomonas palustris and *Rhodospirillum rubrum* contain σ^E -ChrR, but we could not define their regulon because we could not detect the conserved σ^E promoter sequence.

Only a few bacteria containing σ^E -ChrR are photosynthetic, so nonphotosynthetic bacteria are likely to encounter this reactive oxygen species in nature. Of those containing the σ^E -ChrR regulon, the bacteria most similar to *R. sphaeroides* include the phylogenetically close members of the *Roseobacter* clade (*Jannaschia*, *Silicibacter*, *Roseobacter*, *Loktanella*, *Oceanicola*, and *Roseovarius*). Most *Roseobacter* species synthesize photosynthetic pigments in the presence of oxygen,³⁶ so there is relevance of a $^1\text{O}_2$ stress response to these bacteria. On the other hand, *Silicibacter* is not photosynthetic, but its close association with photosynthetic dinoflagellates³⁷ may expose it to $^1\text{O}_2$ in nature.

The symbiotic nitrogen fixers *Rhizobium* and *Bradyrhizobium* possess *rpoEchrR*, but only *Rhizobium leguminosarum* bv. *viciae* 3841 appears to contain the conserved core σ^E -ChrR regulon. Thus, it is not clear whether these organisms are exposed to $^1\text{O}_2$ or whether they sense a different signal such as another reactive oxygen or nitrogen species. In this regard, nitric oxide and reactive oxygen species are an important part of the plant defense system,^{38,39} so symbiotic bacteria may mount a transcriptional response to those signals. Additional support for a role of σ^E -ChrR in a symbiotic/pathogenic lifestyle is indicated by the existence of a core σ^E -ChrR regulon among the Vibrionaceae clade. These γ -proteobacteria are nonphotosynthetic and not reported to live in proximity to known sources of $^1\text{O}_2$. On the other hand, most of them are either pathogenic or symbiotic,⁴⁰ so their ability to sense or respond to the host immune system may be essential for their lifestyle.

The *Shewanella* clade also shows a strong conservation of the σ^E -ChrR regulon. In nature, these bacteria synthesize a large amount of cytochromes to reduce a variety of metals.⁴¹ The hemes of these cytochromes are a potential source of $^1\text{O}_2$, so this might explain why they contain the regulators and core members of the *R. sphaeroides* σ^E -ChrR regulon. However, the *Shewanella* ChrR homologs possess an additional cysteine residue in the N-terminal region that coordinates one of the zinc ions in the *R. sphaeroides* anti- σ factor (Supplemental Fig. 3). It has been proposed that $^1\text{O}_2$ reacts with the N-terminal zinc site to initiate a conformational change that releases σ^E so it can bind RNA polymerase and activate gene expression.¹⁵ Thus, amino acid changes in the ChrR proteins of the *Shewanella* clade might allow it to sense a different signal. If this is the case, the signal might still produce similar damage to biomolecules, since their predicted core regulon members are very similar to those in other organisms we considered.

Evolution of σ^E -ChrR and the appearance of oxygen in the biosphere

Phylogenetic analysis suggests that the origin of the σ^E -ChrR proteins cannot be placed much later than the evolutionary divergence of the α -proteobacteria and γ -proteobacteria, which is estimated to be between 2.51 and 2.36 billion years (Ba) ago.^{42,43} For comparison, the divergence of the cyanobacteria group, which is proposed to be the first microbe to perform oxygenic photosynthesis, is estimated to be between 2.65 Ba and 2.55 Ba,^{42,43} at about the time when signs of molecular oxygen appeared in the Earth's atmosphere (between 2.6 Ba and 2.3 Ba).^{44,45} For anoxygenic photosynthetic organisms, such as the ancestor of *R. sphaeroides*, solar energy capture in the presence of newly formed atmospheric oxygen is likely to have generated $^1\text{O}_2$ for the first time.³⁵ Thus, our phylogenetic predictions trace the history of σ^E -ChrR to a time coinciding with the appearance of oxygen in the biosphere. Indeed, if σ^E -ChrR and the response to $^1\text{O}_2$ did not appear until after the evolution of oxygenic photosynthesis, it explains why orthologs of these proteins are not found in cyanobacteria containing other group IV σ factors.⁴⁶ Thus, it would appear that cyanobacteria, other proteobacteria, and eukaryotes, which all respond to the toxic effects of $^1\text{O}_2$, use similar

proteins to protect themselves and repair damage from this reactive oxygen species, but have evolved different network architectures to mount a transcriptional response to this signal.

In summary, our results demonstrate the utility of using genomic, computational, and experimental tools to dissect the evolutionary history of important signal transduction pathways and to predict promoter architecture and target genes for members of the group IV family of σ factors across widely divergent bacterial species. We have identified most—or possibly all—protein members of the *R. sphaeroides* σ^E -ChrR regulon, providing information needed to begin a comprehensive and mechanistic view of how cells respond to $^1\text{O}_2$. However, the possible existence and potential function of small functional RNAs or nonannotated peptides among genes directly transcribed by σ^E remain to be investigated. Our analysis also predicts that σ^E -ChrR evolved at or near the time when oxygen appeared on the planet and that the response to $^1\text{O}_2$ was conserved across many bacterial species that were present before this change in the atmosphere altered life on Earth. However, the incomplete distribution of σ^E -ChrR among α -proteobacteria and γ -proteobacteria and their absence in δ -proteobacteria β -proteobacteria, and ϵ -proteobacteria that have been sequenced to date leave open questions regarding the modalities of its acquisition or loss through the species lineages. Future analysis of the conservation and functions of the σ^E -ChrR regulon may help form hypotheses about the evolutionary history of this system. For example, other bacteria that deal with $^1\text{O}_2$ stress may have placed members of the *R. sphaeroides* σ^E -ChrR regulon under the control of a different regulator. Finally, our findings support the hypothesis that the role of σ^E -ChrR as master regulator of $^1\text{O}_2$ stress is relevant to photosynthetic bacteria, as well as to nonphotosynthetic free-living or symbiotic/pathogenic microbes.

Materials and Methods

Media, strains, plasmids, and molecular manipulation

E. coli strains were grown in Luria–Bertani medium⁴⁷ at 37 °C and *R. sphaeroides* strains in Sistrof's succinate-based medium⁴⁸ at 30 °C. *E. coli* DH5 α was used as a plasmid carrier, and *E. coli* S17-1 was used as a donor for conjugation of plasmids into *R. sphaeroides*. The media were supplemented with ampicillin (100 $\mu\text{g}/\text{ml}$) or kanamycin (25 $\mu\text{g}/\text{ml}$) when required. Unless noted below, all reagents were used in accordance with the manufacturer's specifications.

Plasmids used are listed in Supplemental Table 1 (Supplementary Materials). Plasmids not reported previously^{6,9} were constructed by amplifying the relevant regions from *R. sphaeroides* 2.4.1 genomic DNA using oligonucleotide primers containing an upstream KpnI restriction site and a downstream StuI restriction site. The amplified DNA fragments were digested with both KpnI and StuI before being ligated in KpnI-StuI-digested pRKK96,⁴⁹ to obtain the *in vitro* transcription template, and in KpnI-StuI-digested pRKK200,⁵⁰ to create a promoter fusion to a *lacZ* reporter gene.

Microarray expression data

To identify genes that are coexpressed with known members of the σ^E -ChrR regulon, *R. sphaeroides* gene expression datasets were collected from the Gene Expression Omnibus database[‡] as of May 2005. The dataset contains RNA abundance measurements from 67 GeneChip Custom Express microarrays (Affymetrix, Santa Clara, CA⁵¹) obtained from wild-type or mutant strains grown in a succinate-based minimal medium under different conditions (GSM1670, GSM1671, GSM1672, GSM1673, GSM2416, GSM2417, GSM2418, GSM2419, GSM2420, GSM2421, GSM2422, GSM2423, GSM2424, GSM2425, GSM2426, GSM2427,

[‡]<http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi%3Facc%3DGPL162>

GSM2429, GSM2430, GSM2450, GSM25295, GSM25296, GSM25297, GSM25298, GSM25299, GSM25300, GSM25301, GSM25302, GSM25303, GSM26242, GSM26243, GSM26244, GSM26245, GSM26260, GSM26262, GSM26263, GSM26265, GSM26266, GSM27348, GSM27349, GSM27350, GSM27351, GSM27352, GSM27353, GSM3030, GSM3031, GSM3032, GSM3258, GSM3260, GSM3262, GSM3272, GSM3273, GSM3274, GSM38774, GSM38775, GSM38776, GSM38777, GSM38778, GSM38779, GSM38780, GSM38781, GSM38782, GSM38783, GSM38810, GSM40560, GSM8107, GSM8108, and GSM8109).^{6,26,51–53} RNA levels in all datasets were normalized using the Robust Multichip Average method⁵⁴ to log₂ scale.

Clustering genes by expression patterns and motif discovery

RNA species with abundance levels that showed a twofold or more difference and a significant difference at FDR 0.1, using the software EBarrays,⁵⁵ in expression level between wild-type cells and those that have increased σ^E activity ($\Delta chrR$ cells) were selected (~100 loci) to limit spurious correlation in the clustering steps. Hierarchical clusters were constructed with the R statistical software environment⁸ using $(1 - \text{Pearson's correlation coefficient})/2$ as the distance between expression patterns and the 'complete' method for cluster linkage. The BioProspector¹² algorithm was used to search for a conserved bipartite motif within 300 bp upstream of the predicted translation starting site in sets of candidate promoter sequences. The parameters were set to search for two blocks of 6 bp and for a gap between 13 bp and 16 bp.

Determining promoter activity

In vivo promoter activities were determined by measuring β -galactosidase activity in cells containing *lacZ* operon fusions harbored on a low-copy plasmid.⁵⁰ β -Galactosidase activity (U/ml) was calculated as follows: $(A_{420} \times 1000)/[\text{cell culture volume in assay (ml)} \times \text{culture optical density (OD}_{600}) \times \text{time (min)}]$.⁵⁶ Culture optical density and absorption at 420 nm were measured with a BioSpec 1601 spectrophotometer (Schimatzu, Columbia, MD). All data are reported from cultures performed in triplicate.

In vitro transcription assays were performed using plasmid templates in which a test promoter was cloned in pRKK96.⁴⁹ Standard assay conditions mixed recombinant His₆- σ^E (50 nM)¹³ with *R. sphaeroides* core RNA polymerase (50 nM) for 30 min at 30 °C in transcription buffer (40 mM Tris-HCl pH 7.9, 200 mM KCl, 10 mM Mg acetate, 1 mM DTT, and 62.5 mg/ml acetylated bovine serum albumin). Next, template DNA was added and incubated for 30 min before transcription was initiated by addition of nucleotide triphosphates.⁹ Samples were incubated for 20 min before RNA products were analyzed on 6% (wt/vol) denaturing PAGE, and products were R-project.org visualized on a phosphorimager (Molecular Dynamics, Sunnyvale, CA).

ChIP-chip hybridization

Our protocol was adapted from Lee *et al.*⁵⁷ *R. sphaeroides* $\Delta chrR$ cells (increased σ^E activity) were grown by bubbling 500 ml of cell culture with 1% CO₂, 30% O₂, and 69% N₂. At midexponential phase (~2×10⁸ colony-forming units/ml), formaldehyde and sodium phosphate were added to a final concentration of 1% and 10 mM, respectively. This mixture was incubated at 30 °C for 4 min before glycine was added to 100 mM, and the solution was placed on ice for 30 min with gentle agitation to quench excess formaldehyde. Cells were centrifuged at 3000g, washed twice with chilled phosphate-buffered saline, centrifuged, and flash-frozen at -80 °C. About 2×10¹⁰ cells were suspended in 0.5 ml of 100 mM Tris (pH 8.0), 300 mM NaCl, 2% Triton X-100, and 1 mM phenylmethylsulfonyl fluoride, and sonicated eight times for 20

[§]<http://wwwR-project.org>.

s with a Branson Sonifier (Branson Ultrasonics Corp., Danbury, CT) set to level 6 and 50% output using a 3-mm microtip. A mixture of micrococcal nuclease (50 U) and RNase A (0.5 μ g) in 200 μ M CaCl₂, 1.2 mM KCl, 6 mM sucrose, and 10 μ M DTT was added; the mixture was incubated for 1 h at 4 °C; and then nuclease activity was inhibited by addition of 10 mM ethylenediaminetetraacetic acid (EDTA). Cell debris was removed by centrifugation for 10 min at 12,000g, and an aliquot was removed to analyze DNA fragmentation by agarose gel electrophoresis (desired size of ~200–1000 bp with enrichment for ~500-bp molecules).

The resulting supernatant was incubated with gentle mixing with 20 μ l of *Staphylococcus aureus* protein A Sepharose beads (Sigma-Aldrich, St. Louis, MO) for 3 h at 4 °C as pretreatment to remove potential nonspecific binding to the beads. After the beads had been removed by centrifugation (5 min at 3000g), one-tenth of the sample was removed and used as non-antibody-treated control. Two microliters of the anti-*R. sphaeroides* σ^E rabbit polyclonal antibody serum was added, and the mixture was incubated overnight at 4 °C with gentle mixing before being incubated with protein A Sepharose beads (30 μ l) for 2 h at 4 °C. The beads were recovered by centrifugation; washed once (4 °C) with 250 mM LiCl, 100 mM Tris (pH 8.0), and 2% Triton X-100; washed twice with 600 mM NaCl, 100 mM Tris (pH 8.0), and 2% Triton X-100; washed twice with 300 mM NaCl, 100 mM Tris (pH 8.0), and 2% Triton X-100; and washed twice with TE buffer (10 mM Tris pH 8.0 and 1 mM EDTA). Protein–DNA complexes were eluted from the beads by incubating at 65 °C for 30 min in 50 mM Tris (pH 8.0), 10 mM EDTA, and 1% sodium dodecyl sulfate. The beads were removed by centrifugation, and protein–DNA cross-linking was reversed by incubating the samples for 12 h at 65 °C. DNA was purified using the QIAquick PCR Purification Kit (QIAGEN, Inc., Valencia, CA).

The DNA was amplified via ligation-mediated PCR.⁵⁷ DNA fragments were treated with T4 DNA polymerase, purified by phenol extraction and ethanol precipitation (adding 20 μ g of glycogen), and ligated via T4 DNA ligase to 4 μ M annealed oligonucleotide linkers (oJW102: GCGGTGACCCGGGAGATCTGAATTC; oJW103: GAATTCAGATC) at 16 °C overnight. DNA was ethanol-precipitated before amplification (22 cycles; annealing temperature of 60 °C) with Taq DNA polymerase (New England Biolabs) and 1 μ M oJW102 linker (50- μ l reaction). The amplification reaction was repeated using 15 μ l of the first amplification reaction product as template. After the products had been purified using a QIAquick Kit (QIAGEN), triplicates of control and sample DNA (concentrated to ~250 ng/ μ l) were pooled to obtain ~4- μ g quantities and hybridized to a custom-made NimbleGen microarray (NimbleGen Systems, Madison, WI). For each antibody, triplicate microarrays were analyzed.

The custom-made NimbleGen microarray used for ChIP-chip analysis was designed to tile the *R. sphaeroides* 2.4.1 genome (two chromosomes and five plasmids) with overlapping isothermal probes that ranged from 35 to 65 bases, with an average spacing of 12 bp. The probes on the array alternate between coding and noncoding DNA sequences. The list of probe features contained in this array is available upon request from the authors.

ChIP-chip data analysis

For analysis of each ChIP-chip microarray, quantile normalization was used to obtain the same empirical distribution across the Cy3 and Cy5 channels and across arrays (*limma* package in the R environment) to correct for dye intensity bias and to minimize microarray-to-microarray absolute intensity variations.⁵⁸ The log₂ of the ratio of experimental signals (Cy3) to control signals (Cy5) was calculated. The data from the biological replicates were averaged for visualizations in SignalMap 1.9 software (NimbleGen Systems). Regions of the genome enriched for occupancy by σ^E were identified using TAMALPAIS at $p \leq 0.01$ for a threshold set at the 98th percentile of the log₂ ratio for each chip.¹⁴ Only enriched regions that were

significant in all three replicates were considered. The peaks were ranked by order of intensities calculated from the average of the 10 highest consecutive probe signals for each peak.

Phylogenetic analyses

All genomic DNA sequences were obtained from the Integrated Microbial Genomes system^{||} 59 on January 10, 2007. Groups of orthologous genes were determined using the OrthoMCL software set to the default parameters.⁶⁰ The protein sequences of orthologous gene products were aligned with CLUSTAL W with the default parameters.⁶¹ Phylogenetic trees were constructed from the protein sequence alignments with the MrBayes 3.1.2 software⁶² over 4×10^6 generations (first 25% as burn-in) using the General Time-Reversible model including an estimated proportion of invariable sites and a γ -shaped distribution of rate variation across sites.

A test promoter library for each bacterium was constructed by extracting 300 bp upstream of each protein coding sequence. Each test promoter sequence was scanned for its best match to the defined PSWM according to its information content (I_{seq}):

$$I_{\text{seq}} = \sum_i \sum_b f_{b,i} \log_2 \frac{f_{b,i}}{p_b}$$

where i is the position in the matrix, b is one of the four bases, $f_{b,i}$ is the observed frequency of each base at that position, p_b is the background frequency of each base in the target genome ($p_A + p_T = 0.31$ and $p_G + p_C = 0.69$ for *R. sphaeroides*). The motif was defined by two matrices (6 bp for the -35 region and 4 bp for the -10 region). Alternative spacer lengths of 15–17 bp were considered without penalties. Promoter sequences containing a match of at least 0.75 times the maximal information content that could be obtained from the PSWM were considered positive (and assigned a value of 1.0). Genes directly downstream and transcribed in the same direction of genes with a positive promoter were assigned a score of 0.8. The next gene downstream was assigned a score of 0.6 and so on, until either the end of a potential operon was reached (five genes) or another gene transcribed in the opposite was found. The score for the each group of orthologous gene products was calculated by summing individual gene scores.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank R. Alexandridis for help with the gene expression analysis, R. Greenwell for help with *in vitro* transcription assays, J. Rowland for help with the ChIP assay, Y. Wang and E. Ruby for sharing gene expression data for *V. fischeri*, and J. Vierstra and A. Shade for discussion. This work was supported by the National Institutes of General Medical Sciences (GM075273) and the Basic Energy Sciences program of the Department of Energy (DE-FG02-05ER15653). Yann Dufour was supported by BACTER, a computational biology training program funded by the Advanced Scientific Computing Research program of the Department of Energy (ER63232-1018220-0007203).

Abbreviations used

¹ O ₂	reactive oxygen species singlet oxygen
PSWM	

^{||}<http://img.jgi.doe.gov>

position-specific weighted matrix

ChIP-chip

chromatin immunoprecipitation and microarray

Ba

billion years

EDTA

ethylenediaminetetraacetic acid

References

1. Kotelnikova EA, Makeev VJ, Gelfand MS. Evolution of transcription factor DNA binding sites. *Gene* 2005;347:255–263. [PubMed: 15725380]
2. Rhodius VA, Suh WC, Nonaka G, West J, Gross CA. Conserved and variable functions of the sigmaE stress response in related genomes. *PLoS Biol* 2006;4:e2. [PubMed: 16336047]
3. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet* 2005;39:309–338. [PubMed: 16285863]
4. Madan Babu M, Teichmann SA, Aravind L. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* 2006;358:614–633. [PubMed: 16530225]
5. Mackenzie C, Eraso JM, Choudhary M, Roh JH, Zeng X, Bruscella P, et al. Postgenomic adventures with *Rhodobacter sphaeroides*. *Annu Rev Microbiol* 2007;61:283–307. [PubMed: 17506668]
6. Anthony JR, Warczak KL, Donohue TJ. A transcriptional response to singlet oxygen, a toxic byproduct of photosynthesis. *Proc Natl Acad Sci USA* 2005;102:6502–6507. [PubMed: 15855269]
7. Rinalducci S, Pedersen JZ, Zolla L. Formation of radicals from singlet oxygen produced during photoinhibition of isolated light-harvesting proteins of photosystem II. *Biochim Biophys Acta* 2004;1608:63–73. [PubMed: 14741586]
8. Davies MJ. The oxidative environment and protein damage. *Biochim Biophys Acta Proteins Proteomics* 2005;1703:93–109.
9. Newman JD, Falkowski MJ, Schilke BA, Anthony LC, Donohue TJ. The *Rhodobacter sphaeroides* ECF sigma factor, sigma(E), and the target promoters cycA P3 and rpoE P. *J Mol Biol* 1999;294:307–320. [PubMed: 10610760]
10. Bockhorst J, Qiu Y, Glasner J, Liu M, Blattner F, Craven M. Predicting bacterial transcription units using sequence and expression data. *Bioinformatics* 2003;19:i34–i43. [PubMed: 12855435]
11. Price MN, Huang KH, Alm EJ, Arkin AP. A novel method for accurate operon predictions in all sequenced prokaryotes. *Nucleic Acids Res* 2005;33:880–892. [PubMed: 15701760]
12. Liu X, Brutlag DL, Liu JS. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* 2001;6:127–138. [PubMed: 11262934]
13. Anthony JR, Newman JD, Donohue TJ. Interactions between the *Rhodobacter sphaeroides* ECF sigma factor, sigma(E), and its anti-sigma factor, ChrR. *J Mol Biol* 2004;341:345–360. [PubMed: 15276828]
14. Bieda M, Xu X, Singer MA, Green R, Farnham PJ. Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res* 2006;16:595–605. [PubMed: 16606705]
15. Campbell EA, Greenwell R, Anthony JR, Wang S, Lim L, Das K, et al. A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. *Mol Cell* 2007;27:793–805. [PubMed: 17803943]
16. Lane WJ, Darst SA. The structural basis for promoter –35 element recognition by the group IV sigma factors. *PLoS Biol* 2006;4:e269. [PubMed: 16903784]
17. Helmann JD. The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 2002;46:47–110. [PubMed: 12073657]
18. Cronan JE Jr. Phospholipid modifications in bacteria. *Curr Opin Microbiol* 2002;5:202–205. [PubMed: 11934618]

19. Hendrich AK, Braatsch S, Glaeser J, Klug G. The *phrA* gene of *Rhodobacter sphaeroides* encodes a photolyase and is regulated by singlet oxygen and peroxide in a sigma(E)-dependent manner. *Microbiology* 2007;153:1842–1851. [PubMed: 17526841]
20. Sancar A. Photolyase and cryptochrome blue-light photoreceptors. *Adv Protein Chem* 2004;69:73–100. [PubMed: 15588840]
21. Green HA, Donohue TJ. Activity of *Rhodobacter sphaeroides* RpoH_{II}, a second member of the heat shock sigma factor family. *J Bacteriol* 2006;188:5712–5721. [PubMed: 16885439]
22. Sies H. Damage to plasmid DNA by singlet oxygen and its protection. *Mutat Res* 1993;299:183–191. [PubMed: 7683086]
23. Cao M, Kobel PA, Morshedi MM, Wu MF, Paddon C, Helmann JD. Defining the *Bacillus subtilis* sigma(W) regulon: a comparative analysis of promoter consensus search, run-off transcription/microarray analysis (ROMA), and transcriptional profiling approaches. *J Mol Biol* 2002;316:443–457. [PubMed: 11866510]
24. Manganelli R, Voskuil MI, Schoolnik GK, Smith I. The *Mycobacterium tuberculosis* ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol* 2001;41:423–437. [PubMed: 11489128]
25. Paget MS, Molle V, Cohen G, Aharonowitz Y, Buttner MJ. Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the sigmaR regulon. *Mol Microbiol* 2001;42:1007–1020. [PubMed: 11737643]
26. Braatsch S, Moskvina OV, Klug G, Gomelsky M. Responses of the *Rhodobacter sphaeroides* transcriptome to blue light under semiaerobic conditions. *J Bacteriol* 2004;186:7726–7735. [PubMed: 15516587]
27. Alvarez-Martinez CE, Lourenco RF, Baldini RL, Laub MT, Gomes SL. The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*. *Mol Microbiol* 2007;66:1240–1255. [PubMed: 17986185]
28. Shultzaberger RK, Chen Z, Lewis KA, Schneider TD. Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* 2007;35:771–788. [PubMed: 17189297]
29. Weindl J, Hanus P, Dawy Z, Zech J, Hagenauer J, Mueller JC. Modeling DNA-binding of *Escherichia coli* sigma70 exhibits a characteristic energy landscape around strong promoters. *Nucleic Acids Res* 2007;35:7003–7010. [PubMed: 17940097]
30. Wang AY, Grogan DW, Cronan JE Jr. Cyclopropane fatty acid synthase of *Escherichia coli*: deduced amino acid sequence, purification, and studies of the enzyme active site. *Biochemistry* 1992;31:11020–11028. [PubMed: 1445840]
31. Heath RL, Packer L. Photoperoxidation in isolated chloroplasts: I. Kinetics and stoichiometry of fatty acid peroxidation. *Arch Biochem Biophys* 1968;125:189. [PubMed: 5655425]
32. Hayes JD, McLellan LI. Glutathione and glutathione-dependent enzymes represent a co-ordinately regulated defence against oxidative stress. *Free Radical Res* 1999;31:273–300. [PubMed: 10517533]
33. Edge R, McGarvey DJ, Truscott TG. The carotenoids as anti-oxidants—a review. *J Photochem Photobiol B* 1997;41:189–200. [PubMed: 9447718]
34. Ledford HK, Chin BL, Niyogi KK. Acclimation to singlet oxygen stress in *Chlamydomonas reinhardtii*. *Eukaryotic Cell* 2007;6:919–930. [PubMed: 17435007]
35. Krieger-Liszkay A. Singlet oxygen production in photosynthesis. *J Exp Bot* 2005;56:337–346. [PubMed: 15310815]
36. Wagner-Dobler I, Biebl H. Environmental biology of the marine *Roseobacter* lineage. *Annu Rev Microbiol* 2006;60:255–280. [PubMed: 16719716]
37. Miller TR, Belas R. Dimethylsulfoniopropionate metabolism by *Pfiesteria*-associated *Roseobacter* spp. *Appl Environ Microbiol* 2004;70:3383–3391. [PubMed: 15184135]
38. Wojtaszek P. Oxidative burst: an early plant response to pathogen infection. *Biochem J* 1997;322:681–692. [PubMed: 9148737]
39. Bolwell GP. Role of active oxygen species and NO in plant defence responses. *Curr Opin Plant Biol* 1999;2:287–294. [PubMed: 10459001]
40. Thompson FL, Iida T, Swings J. Biodiversity of vibrios. *Microbiol Mol Biol Rev* 2004;68:403. [PubMed: 15353563]

41. Heidelberg JF, Paulsen IT, Nelson KE, Gaidos EJ, Nelson WC, Read TD, et al. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. *Nat Biotechnol* 2002;20:1118–1123. [PubMed: 12368813]
42. Battistuzzi FU, Feijao A, Hedges SB. A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evol Biol* 2004;4. [PubMed: 15018651]
43. Sheridan PP, Freeman KH, Brenchley JE. Estimated minimal divergence times of the major bacterial and archaeal phyla. *Geomicrobiol J* 2003;20:1–14.
44. Bekker A, Holland HD, Wang PL, Rumble D III, Stein HJ, Hannah JL, et al. Dating the rise of atmospheric oxygen. *Nature* 2004;427:117–120. [PubMed: 14712267]
45. Wille M, Kramers JD, Nagler TF, Beukes NJ, Schroder S, Meisel T, et al. Evidence for a gradual rise of oxygen between 2.6 and 2.5 Ga from Mo isotopes and Re-PGE signatures in shales. *Geochim Cosmochim Acta* 2007;71:2417–2435.
46. Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol* 2003;57:441–466. [PubMed: 14527287]
47. Sambrook, J.; Fritsch, EF.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2. Cold Spring Harbor Laboratory; Cold Spring Harbor, NY: 1989. three volumes
48. Siström WR. A requirement for sodium in the growth of *Rhodospseudomonas sphaeroides*. *J Gen Microbiol* 1960;22:778–785. [PubMed: 14447230]
49. Karls RK, Wolf JR, Donohue TJ. Activation of the *cycA* P2 promoter for the *Rhodobacter sphaeroides* cytochrome *c₂* gene by the photosynthesis response regulator. *Mol Microbiol* 1999;34:822–835. [PubMed: 10564521]
50. Flory JE, Donohue TJ. Transcriptional control of several aerobically induced cytochrome structural genes in *Rhodobacter sphaeroides*. *Microbiology* 1997;143:3101–3110. [PubMed: 9353915]
51. Pappas CT, Sram J, Moskvina OV, Ivanov PS, Mackenzie RC, Choudhary M, et al. Construction and validation of the *Rhodobacter sphaeroides* 2.4.1 DNA microarray: transcriptome flexibility at diverse growth modes. *J Bacteriol* 2004;186:4748–4758. [PubMed: 15231807]
52. Moskvina OV, Gomelsky L, Gomelsky M. Transcriptome analysis of the *Rhodobacter sphaeroides* PpsR regulon: PpsR as a master regulator of photosystem development. *J Bacteriol* 2005;187:2148–2156. [PubMed: 15743963]
53. Tavano CL, Podevels AM, Donohue TJ. Identification of genes required for recycling reducing power during photosynthetic growth. *J Bacteriol* 2005;187:5249–5258. [PubMed: 16030219]
54. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003;4:249–264. [PubMed: 12925520]
55. Kendziorski CM, Newton MA, Lan H, Gould MN. On parametric empirical Bayes methods for comparing multiple groups using replicated gene expression profiles. *Stat Med* 2003;22:3899–3914. [PubMed: 14673946]
56. Zhang X, Bremer H. Control of the *Escherichia coli* *rrnB* P1 promoter strength by ppGpp. *J Biol Chem* 1995;270:11181–11189. [PubMed: 7538113]
57. Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* 2006;1:729–748. [PubMed: 17406303]
58. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19:185–193. [PubMed: 12538238]
59. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen IM, et al. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res* 2008;36:D528–D533. [PubMed: 17933782]
60. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 2003;13:2178–2189. [PubMed: 12952885]
61. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]

62. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003;19:1572–1574. [PubMed: 12912839]

Fig. 1.

Expression profiles of σ^E -ChrR regulon members. Coregulated genes identified from the hierarchical clustering of genes differentially expressed in response to increased σ^E activity ($\Delta chrR$ mutant) compared to wild-type cells. Colors represent the relative abundance of transcripts in different conditions. Red indicates high expression level and green indicates low expression level, compared to the mean expression level. Genes are identified by their locus ID number. Brackets denote known or predicted transcriptional units.

Fig. 2.

Sequence logos of the predicted σ^E promoter motif in *R. sphaeroides* (a) and other bacteria (b). (a) The -35 and -10 motif logos were derived by sequence alignment of six σ^E -dependent promoters from *R. sphaeroides*. The two conserved regions are separated by a spacer of 13–14 bp. The information content (I_{seq}) of each motif is indicated. (b) The indicated -35 and -10 motifs were found upstream of the *rpoE* gene in 57 of the 73 selected microbial genomes. In both panels, the logos were produced with WebLogo [<http://weblogo.berkeley.edu/>].

Fig. 3.

In vitro transcription of target promoters using reconstituted *R. sphaeroides* σ^E . (a) Transcripts derived from the known σ^E -dependent promoter RSP1092 (*rpoE*) and a promoter identified in this study (RSP1852) in the presence of σ^E and ChrR. (b) Transcripts derived from RSP6222 and RSP3336 as a function of increasing σ^E levels. The amount of σ^E used in (b) is indicated relative to the amount used in the reaction shown in (a).

Fig. 4.

Activity of σ^E -dependent promoters in *R. sphaeroides*. Promoter activity is reported as the level of β -galactosidase activity (Miller units) in wild-type (gray), $\Delta chrR$ (black), and $\Delta rpoEchrR$ (white) cells. All assays were performed in triplicate, with bars representing standard deviations.

Fig. 5.

Localization of σ^E -dependent promoters by ChIP-chip. Shown are three representative genomic regions enriched by the immunoprecipitation of DNA fragments using anti- σ^E antibodies (red) or anti- β' antibodies to assess RNA polymerase occupancy (blue). The data plot the $\log_2$ of the ratio of the immunoprecipitated sample to the control sample as a function of probe location along the *R. sphaeroides* genome (coordinates are indicated in base pairs). Regions significantly enriched by anti- σ^E immunoprecipitation ($p \leq 0.01$) are indicated by red blocks. The locations of the annotated open reading frames are indicated by black blocks (read forward or reverse if above or below the baseline, respectively). The positions and orientations of the known or predicted σ^E promoter are indicated by a green arrow. The data were plotted using SignalMapTM v1.9 (NimbleGen Systems).

Fig. 6. Phylogeny of the σ^E -ChrR pair. Phylogenetic tree constructed with the concatenated amino acid sequence of σ^E and ChrR (a) or RuvB, RpoD, and GyrB (b) using Bayesian inference. Proteobacteria classes are indicated by color (pale blue for γ -proteobacteria and salmon for α -proteobacteria). Numbers on branches indicate Bayesian posterior probabilities. Branches of the tree for closely related species were collapsed for clarity.

Fig. 7. Alignments of the amino acid sequences of σ^E homologs. Alignment of the regions 2.3–2.4 and 4.2 of σ^E homologs from a set of representative species containing σ^E -ChrR proteins. The histogram below the alignment represents the relative conservation score for each position based on alignment of all 73 σ^E homologs considered in this study. Arrows indicate residues predicted to contact DNA in the $\alpha 35$ region of the promoter based on alignment with *E. coli* σ^E .¹⁶ Sequences are identified by their locus ID.

Fig. 8. Potential σ^E target genes across selected bacteria. (a) Groups of orthologous genes (columns) that contain the putative σ^E binding motif in their promoter regions across the indicated α -proteobacteria. (b) Same as in (a), but for the indicated γ -proteobacteria. A gray box means that the organism does not possess a homolog for the corresponding group; black means that it possesses a homolog for the group but no conserved σ^E binding motif was found in its promoter region; yellow means that a σ^E binding motif was found in the promoter region for the homologous gene. Shades of yellow represent the degree to which the gene is connected to the σ^E binding motif (see the text). A description and an annotation of these orthologs are presented in Table 3 and Supplemental Table 2.

Table 1
Candidate and confirmed σ^E target operons and their putative promoters

Gene ID	Name	Score	Normalized score	-35	Spacer	-10	Distance to start codon	Correlation coefficient
<i>A. Putative promoters identified from the sequence analysis of a library of promoter regions</i>								
RSP1087-1091		20.14	1.00	TGATCCG	ccttgaggagacag	TCCGTAT	-54	0.91
RSP0601	<i>rpoH_H</i>	19.50	0.97	TGATCCG	gacatgtgtttt	TCCGTAG	-66	0.88
RSP2143-2144	<i>phrBcdS</i>	18.95	0.91	TGATCCG	ggaagaggagccg	CGCGTAA	-49	0.85
RSP1092-1093	<i>rpoEchrR</i>	18.32	0.91	TGATCCA	gactgcccgcgc	CGCGTAA	-130	1
RSP1409		18.14	0.90	TCATCCG	caggagccgcttc	TGCGTAG	-71	0.84
RSP1852	★	18.06	0.90	TGATCTG	aaccgtcgcttaa	CCCGTAT	-103	0.93
RSP6076		17.89	0.89	TGATCTT	caagtgagaccgga	TCCGTAA	-43	NA
RSP3284		17.51	0.87	TGATCCG	gaggtcggggctc	TCCGAAG	-103	-0.30
RSP0296	<i>cycA</i>	16.83	0.84	TGATCCG	gaacgagcggccc	GCAGTAG	-88	-0.12
RSP4129		16.37	0.82	TGATCTT	aaactaaagctt	GCCATAT	-180	-0.22
RSP6222		16.26	0.80	TGATCTT	catggggatctct	CCCGTAG	-72	NA
RSP1207	<i>hslO</i>	16.15	0.80	CGATCCG	cccgcacgggtc	CCCGTAT	-110	0.58
RSP3336	★	15.66	0.78	AGATCTG	acgtgaacagat	ACCGTAA	-172	-0.25
RSP1521		15.72	0.78	TGATCCA	gaccgtaccggc	GCCGGAT	-140	-0.20
RSP0357		15.72	0.78	TGATCCA	gtcgcgcgcac	GCCGTGA	-183	0.63
<i>B. Putative additional promoters identified within the regions significantly enriched in the ChIP-chip experiment using anti-σ^E antibodies</i>								
RSP3832		19.49	0.97	TGATCAG	gggtgcggagacc	TGCGTAA	NA	NA
RSP2324		16.80	0.85	TGATCCG	ggcgcgatgca	ATCGTAG	-434	0.03
RSP2978	<i>mrcA</i>	16.54	0.84	AGATCCG	gcgtgctgcgc	GGCGTAT	NA	NA
RSP1955		16.44	0.83	TGGTCCG	gagcggctcgcg	TGCGTAG	NA	NA
RSP1222	<i>hamI</i>	15.99	0.81	AGATCCA	gcacggctggcc	CCCGTAG	NA	NA
RSP2047		15.87	0.80	GCATCCG	gtacctcttgc	TGCGTAT	+39	-0.05
RSP3101		15.47	0.78	CGATCCA	ccttcacatct	TTCGTAT	-68	-0.24
RSP1612		14.79	0.74	CGATCAG	ctggccgcgag	CCTGTAG	NA	NA
RSP4003	<i>dhaL</i>	14.58	0.74	CGATCCA	gatggtcttcagc	TGGGTAT	NA	NA
RSP2401		13.67	0.69	TGGACCG	gatcgactctcc	ACCGTAG	-56	0.27
RSP2793		13.66	0.69	GGATTTC	ccaaggaacagag	CCCGTAA	NA	NA
RSP2940		12.38	0.63	CCATCAG	ccggcgccggcat	CCCGCAT	NA	NA
RSP3007		11.05	0.56	CAATCTC	gaaggaaagtcca	GGCGTAT	NA	NA

Previously identified σ^E target operons/genes^{6,9} are shown in bold; starred genes were shown to be members of the σ^E -ChrR regulon in this study. Promoters upstream of the remaining candidate genes/operons promoters failed to produce detectable levels of transcripts with reconstituted $E\sigma^E$ *in vitro*.

The normalized score is relative to the maximal score that can be obtained from the constructed PSWM for the σ^E promoter motif. The correlation coefficient represents the Pearson's correlation coefficient of each transcript level pattern relative to the *rpoE* transcript level pattern from the expression microarray dataset used in this study.

Table 2

Regions occupied by Eo^E *in vivo* as assayed by ChIP

Scaffold	Genomic coordinates		σ^E enrichment	β' enrichment	Associated locus
	Start	End			
Chromosome 1	744077	745110	35.09	19.02	RSP2143, upstream
Chromosome 1	2342719	2344030	34.25	14.78	RSP0601, upstream
Chromosome 1	210584	211536	33.71	7.80	RSP1612–1613, intergenic antisense
Chromosome 2	28320	29366	31.08	8.00	RSP3832–3833, intergenic antisense
Chromosome 1	2848771	2853927	30.86	13.37	RSP1091–1092, upstream
Chromosome 1	445724	446747	26.36	10.93	RSP1852, upstream
Plasmid P002	44685	45230	25.65	2.05	RSP4003, within coding sequence antisense
Chromosome 2	145759	146248	20.57	6.21	RSP3101, upstream
Chromosome 1	642229	642880	19.11	5.34	RSP2047, upstream
Chromosome 1	3186065	3186861	18.73	1.05	RSP1409, upstream
Chromosome 1	1030027	1030565	17.55	17.09	RSP2401, upstream
Chromosome 1	1668034	1668547	16.92	4.58	RSP2978, within coding sequence antisense
Chromosome 1	2994575	2994971	14.90	0.66	RSP1221, within coding sequence antisense
Chromosome 1	946459	946982	14.55	1.07	RSP2324, upstream
Chromosome 1	1528441	1528994	14.36	0.44	RSP2793, within coding sequence antisense
Chromosome 1	1619624	1620253	13.97	-1.69	RSP2940, within coding sequence antisense
Chromosome 1	551623	552034	11.94	3.52	RSP1955–1956, intergenic antisense
Chromosome 2	42996	43226	11.53	0.47	RSP3007–3008, intergenic antisense

σ^E enrichment represents the average of the log₂ ratio of the 10 best consecutive probes in each selected region. The β' signal represents the average of the log₂ ratio of the same probes from the β' immunoprecipitation experiment (see Materials and Methods). Previously validated σ^E -dependent promoters are indicated in bold.

Table 3
Functional annotation of selected groups of homologous genes across the species considered in this study

Ortho ID	<i>Rsph</i> locus	<i>Rsph</i> gene name	Annotation
The conserved core σ^E -ChrR regulon			
254	RSP2144	<i>cfaS</i>	Cyclopropane-fatty acyl-phospholipid synthase
20013	RSP1090		Uncharacterized conserved protein
20348	RSP1087		Putative short-chain dehydrogenase/reductase
20362	RSP1093	<i>chrR</i>	Anti- σ factor, anti-RpoE
21307	RSP1092	<i>rpoE</i>	Alternative σ factor, group IV
21508	RSP1088		Uncharacterized protein
21520	RSP1091		Putative protein binding nicotinamide adenine dinucleotide/flavin adenine dinucleotide
21876	RSP2143	<i>phrB</i>	Deoxyribodipyrimidine photolyase
The extended σ^E -ChrR regulon			
In <i>R. sphaeroides</i>			
2073	RSP1089		Putative Na ⁺ /melibiose symporter
19916	RSP1409		Protein containing fasciclin-like repeats
20947	RSP0601	<i>rpoH_{II}</i>	Alternative σ factor from the σ^{32} family
21322	RSP1852		Uncharacterized conserved protein
21508	RSP1088		Uncharacterized protein
21842	RSP0296	<i>cycA</i>	Cytochrome <i>c</i> ₂
22229	RSP3336		ABC transporter, inner membrane subunit
22678	RSP6222		Uncharacterized protein
In <i>Shewanella</i> or <i>Vibrio</i> species			
71	RSP2389		Putative glutathione peroxidase
127	RSP2633	<i>ccmF</i>	Cytochrome <i>c</i> maturation protein
157	RSP1805	<i>ccmG</i>	Thiol-disulfide isomerase and thioredoxins
791			Uncharacterized protein
884			Putative sodium/glutamate symporter
2176			Uncharacterized conserved protein
19769	RSP3077		Putative deoxyribodipyrimidine photolyase
20054	RSP1803	<i>ccmC</i>	ABC heme exporter, inner membrane subunit
20172			Putative acyl-CoA dehydrogenases
21430			Putative nucleoside diphosphate sugar epimerase
21552			Uncharacterized conserved protein
22038	RSP2945	<i>ccmE</i>	Cytochrome- <i>c</i> -type biogenesis protein
22231			Putative small polyketide cyclase or steroid isomerase
23701			Adenosylmethionine-8-amino-7-oxononanoate aminotransferase
23867	RSP3424		Putative dehydrogenase

The corresponding *R. sphaeroides* locus ID and gene names are indicated when existing in the respective groups.

Table 4
Evidence used to define the members of the *R. sphaeroides* σ^E -ChrR regulon

Gene locus	Name	Annotation	Sequence Motif	Expression correlation	ChIP-chip	In vitro activity	In vivo reporter activity
RSP1092	<i>rpoE</i>	Group IV σ factor	+	+	+	++	++
RSP1093	<i>chrR</i>	Anti- σ factor		+			
RSP1091		Putative dehydrogenase	+	+	+	++	++
RSP1090				+			
RSP1089				+			
RSP1088				+			
RSP1087		Putative short-chain alcohol dehydrogenase		+			
RSP2143	<i>phrB</i>	Deoxyribodipyrimidine photolyase	+	+	+	++	++
RSP2144	<i>cfaS</i>	Cyclopropane fatty acyl phospholipid synthase		+			
RSP1852			+	+	+	++	++
RSP0601	<i>rpoH_H</i>	σ^{32} family σ factor	+	+	+	++	++
RSP1409			+	+	+	+	++
RSP0296	<i>cycA</i>	Cytochrome <i>c</i> ₂	+	-	-	+	+
RSP3336			+	-	-	+	+
RSP6222			+	-	-	+	+

Genes are grouped by predicted transcriptional units.