Transcriptional Regulation in Archaea: From Individual Genes to Global Regulatory Networks

Mar Martinez-Pastor,[1] Peter D. Tonner,[1,2,*] Cynthia L. Darnell,[1,*] and Amy K. Schmid[1,2,3]

[1]Department of Biology, Duke University, Durham, North Carolina 27708, USA

[2]Graduate Program in Computational Biology and Bioinformatics, Duke University, Durham, North Carolina 27708, USA

[3]Center for Genomic and Computational Biology, Duke University, Durham, North Carolina 27708, USA; email: amy.schmid@duke.edu

*These authors contributed equally.

Keywords

Archaea, transcription, gene regulatory networks, systems biology, transcription factor

Abstract

Archaea are major contributors to biogeochemical cycles, possess unique metabolic capabilities, and resist extreme stress. To regulate the expression of genes encoding these unique programs, archaeal cells use gene regulatory networks (GRNs) composed of transcription factor proteins and their target genes. Recent developments in genetics, genomics, and computational methods used with archaeal model organisms have enabled the mapping and prediction of global GRN structures. Experimental tests of these predictions have revealed the dynamical function of GRNs in response to environmental variation. Here, we review recent progress made in this area, from investigating the mechanisms of transcriptional regulation of individual genes to small-scale subnetworks and genome-wide global networks. At each level, archaeal GRNs consist of a hybrid of bacterial, eukaryotic, and uniquely archaeal mechanisms. We discuss this theme from the perspective of the role of individual transcription factors in genome-wide regulation,

how these proteins interact to compile GRN topological structures, and how these topologies lead to emergent, high-level GRN functions. We conclude by discussing how systems biology approaches are a fruitful avenue for addressing remaining challenges, such as discovering gene function and the evolution of GRNs.

1. INTRODUCTION

Gene regulatory networks (GRNs) are defined as groups of interacting regulatory transcription factors (TFs) and their target genes. In response to environmental stimuli, TFs interact within the GRN to promote or inhibit RNA polymerase (RNAP) to differentially regulate the expression of genes encoding proteins that alter physiology (141). Hence, GRNs are central to the process of maintaining dynamic, physiological responses to a variable environment (20). Microorganisms of the domain *Archaea* thrive during environmental variation. They resist stress under conditions at the limits of sustainable life (161), enabled by efficient stress-repair systems (82, 88) and unique metabolic capabilities (35, 122). With this unique metabolic capacity, archaea make major contributions to balance in biogeochemical cycles (120) and produce key chemicals and stress-tolerant enzymes of interest to industry (29, 203). Recent phylogenetic work places archaea as the evolutionary progenitors of eukaryotes on the tree of life (143, 169, 201). GRN studies in archaea, therefore, provide a unique window into understanding how GRN function in environmental response is conserved across the tree of life. In particular, mapping how TFs and their target genes are connected within GRNs (here referred to as GRN topology) across archaeal species enables us to understand (*a*) how GRNs function to adjust physiology during wide fluctuations in environmental conditions, (*b*) how particular GRN structures produce different patterns of dynamic gene expression, and (*c*) how strong environmental forces shape regulatory networks over evolutionary timescales. However, knowledge of archaeal transcriptional regulatory mechanisms has historically lagged behind that of bacterial and eukaryotic model organisms.

Since the first archaeal genome was sequenced more than two decades ago (26), key milestones have accelerated progress toward building and understanding whole-genome GRNs for several archaeal model organisms across phylogenetic lineages (**Figure 1**). Several genetically tractable model archaeal species are available (48, 93); improved genetic tool kits for established genetic model organisms have been developed (2); and established systems have been adapted to enable genetic tractability across closely related species (85, 100). Completion of whole-genome sequences and subsequent comparative genomics studies have revealed a wide diversity of metabolic and stress-response capabilities (54, 90, 94, 107). Systems biology experimental and computational pipelines have been developed for archaea (36), with methods including transcriptomics (47, 145, 146, 163, 184, among others) and analysis of growth phenotypes in TF knockout strains (178), chromatin immunoprecipitation sequencing (ChIP–seq) and ChIP–chip identification of TF binding sites (47, 118, 144, 157, 166, 167, 179, 192), and de novo identification of TF binding *cis* sequences (157, 162) (**Figure 1**). Integrating these methods has enabled the mapping of entire regulons under the direct control of TFs of interest, and GRN computational inference has predicted how multiple TFs interact to function within the global GRN.

Thus, systems biology approaches have revolutionized the investigation of archaeal transcription regulation, rapidly increasing the understanding of how responses to unique environmental challenges are regulated at the genome-wide level in this important yet understudied domain of

Gene regulatory network (GRN): consists of transcription factors and target genes and the interactions between them

Transcription factor (TF): a protein that binds to *cis* sequences of target genes, activating or repressing transcription

Target gene: genes in gene regulatory networks whose transcriptional activity is regulated by the binding of a transcription factor

Topology: the architecture of a gene regulatory network; describes the structure of how edges wire the nodes together

Transcriptomics: an experimental method that measures the expression of all genes in the genome simultaneously using microarray technology or next-generation sequencing (RNA sequencing)
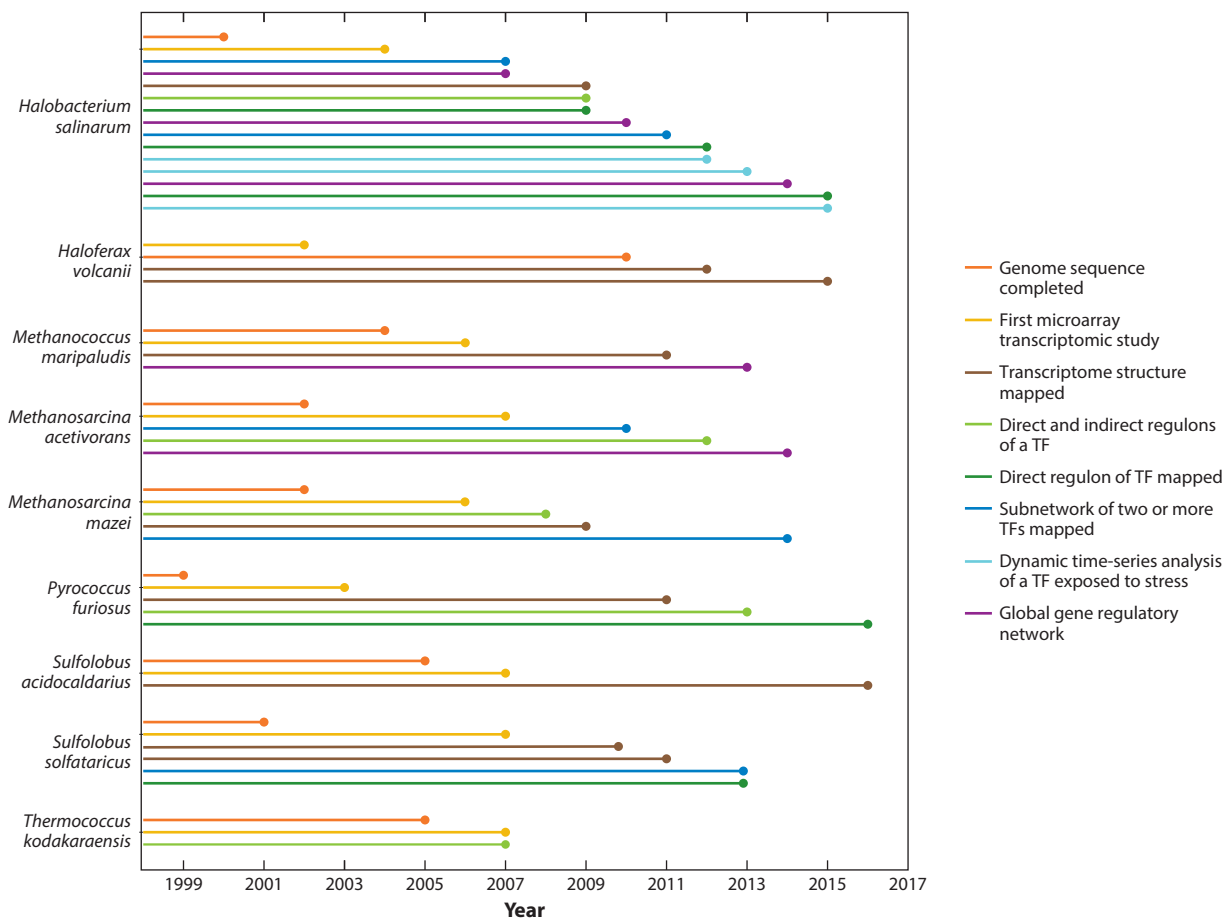
Figure 1

Time line of key milestones in progress investigating genome-wide gene regulatory networks in archaeal model organisms. For brevity, only studies on the most intensively studied, genetically tractable organisms (48, 93) are included. For each species, each colored dot represents the first report of a type of milestone: orange, genome sequence completed (31, 40, 53, 54, 64, 68, 108, 116, 164); yellow, first microarray transcriptomic study (8, 58, 77, 97, 104, 159, 184, 196, 200); brown, transcriptome structure mapped (start sites, 5′ untranslated regions, etc.) using whole-genome microarray or RNA sequencing (4, 6, 33, 76, 86, 195, 197); light green, direct and indirect regulons of a transcription factor (TF) mapped by transcriptomics in TF knockout strains (77, 146, 157, 186, 189, 204); dark green, direct regulon of a TF mapped (117, 144, 157, 167, 179, 192); dark blue, subnetwork of two or more TFs mapped (47, 118, 145, 155, 188); cyan, dynamic time-series analysis of a TF exposed to stress (163, 177, 179); and purple, global gene regulatory network (20, 24, 82, 134, 198).

life. Recent reviews have comprehensively discussed the underlying mechanisms of the entire transcription cycle, from initiation to elongation and termination (56, 60), as well as posttranscriptional mechanisms that affect gene expression (7, 46). We focus here on networks deduced exclusively from measurements of transcription initiation events. In particular, recent progress in understanding how TFs function in the context of archaeal GRNs will be synthesized at three levels: (*a*) how the components of networks (nodes) interact (edges) to initiate transcription at individual genes, (*b*) how subnetworks composed of a few TFs and their target genes function in the dynamic regulation of cellular processes, and (*c*) how global GRNs function in physiological adaptation.

ChIP–chip or ChIP–seq: chromatin immunoprecipitation coupled with microarray or sequencing; locates the binding sites of a transcription factor across the genome

2. TRANSCRIPTIONAL REGULATION OF INDIVIDUAL GENES

Generally speaking, the archaeal proteins required for initiating basal transcription resemble those of eukaryotes at the level of amino acid sequence and function (14). In contrast, the proteins that modulate transcription beyond the basal level (e.g., activator and repressor TFs) typically resemble those of bacteria (128). The combinatorial code of extensive protein–protein interactions between these two classes of proteins binding to DNA at each gene ultimately results in the promotion or blockage of transcription initiation. In this section we describe archaeal initiation mechanisms briefly as a prerequisite for the subsequent discussion of GRNs. Transcription initiation can be divided into basal and regulated mechanisms, which are discussed in turn.

2.1. Basal Transcription

Basal transcription initiation minimally requires transcription factor B (TFB), the homolog of eukaryotic TFIIB, and a TATA-binding protein (TBP) to recruit the eukaryote-like RNAP II-like polymerase (**Figure 2**) (39, 190). These eukaryotic-like TFs are referred to here and elsewhere (47) as general transcription factors (GTFs). Each GTF binds to a specific DNA element in the promoter region (**Figure 2**) (166). TBP first recognizes and binds to the TATA box (14, 62). TFB then binds to the B recognition element (BRE), stabilizing the TBP-TATA box interaction (128, 166). Formation of the TBP–TFB–DNA complex, also called the preinitiation complex (PIC), is required for the recruitment of RNAP (114). The TBP–TFB complex orients the direction of transcription (99). Unlike in eukaryotes, separation of the DNA strands does not require energy in archaea (114). TFE protein assists with open complex formation (18), depending on the promoter and organism (158, 166). The mechanisms that enable the transition to transcriptional elongation remain to be fully characterized (56), although factors important in this process have recently been elucidated (158, 166). Initiation at weak promoters is facilitated by the additional promoter proximal DNA element (PPE), located between the TATA box and the transcription start site, which likely binds TFB (130, 149). The initiator element, located 1 bp downstream of the transcription start site, can also provide additional promoter selectivity (166).
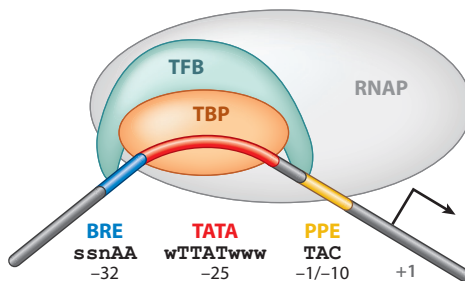
Figure 2

Initiation of basal transcription. DNA is depicted by the gray strand. The transcription start site is represented by the bent arrow. Blue, red, and yellow segments of the strand correspond to several DNA element consensus sequences conserved between *Halobacterium salinarum* and *Pyrococcus furiosus* (149, 162) and their positions relative to the transcription start site (+1). Abbreviations: BRE, B recognition element; n, any nucleotide; PPE, promoter proximal element; RNAP, RNA polymerase; s, G or C; TBP, TATA-binding protein; TFB, transcription factor B; w, A or T.

2.2. Regulated Transcription

The regulatory functions of nearly 60 bacterial-like TFs across a wide variety of species and protein families have been studied experimentally using in vitro biochemistry, genetics, and, more recently, genome-wide approaches (**Figure 3**) (79, 127, 128). According to genome-wide analyses across as many as 52 archaeal genome sequences, the majority of these TF families include helix-turn-helix (HTH), winged HTH, and ribbon-helix-helix domains (30, 132). Eukaryote-like DNA binding domains (e.g., zinc finger) have been detected, although these are present in a smaller proportion (132). Larger genomes tend to encode more TFs, although the overall number of TFs per archaeal genome is lower than that in bacteria (30, 132).

Many archaeal TFs, like those of bacteria, directly sense environmental cues by binding to small-molecule ligands (**Figure 3**) (79, 127). Typically, such chemical signals either release or stimulate TF–DNA binding interactions (101, 157). For example, structural analyses suggest that ligand binding by the leucine-responsive protein (Lrp) family of TFs regulates DNA binding, as well as the oligomeric state of the TF and wrapping of DNA (101, 121, 127). Some TFs can bind multiple effectors with differential effects on DNA binding. For example, the same Lrp protein can bind different sets of amino acids (127). TrmB in *Thermococcus kodakaraensis* binds various substrates (maltose, sucrose, maltotriose, maltodextrin, or trehalose) with different affinities to bind different promoters, to give differential specificity for the same promoter, or both (91). In contrast, the TrmB ligand response in *Halobacterium salinarum* is more binary. TrmB binds to DNA in the absence of glucose, promoting the expression of genes involved in gluconeogenesis and repressing those in glycolysis (157, 177). TrmB dissociates from DNA when glucose is high, which rapidly deactivates gluconeogenic genes and derepresses glycolytic genes (177).

Interestingly, several recent studies have discovered unique classes of archaeal-specific regulators that respond to cellular redox status to regulate diverse pathways, from the oxidative stress response to anaerobic respiration (**Figure 3**) (65, 74, 78, 85). A few of these TFs possess the unique V4R domain that contains multiple redox-active cysteine residues that respond differentially to oxidant. For example, MsvR in *Methanothermobacter thermoautotrophicus* regulates transcription of the oxidative stress response *fpaA-rlp-rub* operon under both oxidizing and reducing conditions (78); however, MsvR in *Methanosarcina acetivorans* undergoes conformational rearrangements that allow DNA binding of targets involved in oxidative stress only under reducing conditions (74).

Extensive in vitro biochemical studies have established the mechanisms by which bacterial-like TFs interact directly with the basal transcription apparatus, the *cis*-regulatory sequence, or both, to activate or repress transcription (128). Transcription is blocked when a TF binds its *cis* sequence overlapping the BRE by inhibiting binding of the PIC to the DNA (16) or RNAP recruitment (15). In contrast, initiation is typically stimulated by facilitating RNAP, TBP, or TFB recruitment (124, 125). These mechanisms of activation and repression of individual genes appear to be general across several species of archaea (14).

More complex mechanisms of regulation have also been discovered. For example, *Sulfolobus solfataricus* Ss-LrpB can either activate or repress its own promoter, depending on which combination of the three binding sites is bound (129). Each of these sites binds LrpB with a different affinity. In *Methanosarcina mazei*, transcription inhibition of the nitrogen response gene *glnK1* requires both NprI and NprII transcriptional corepressors, which form a complex that blocks RNAP recruitment (187). NprI binds DNA downstream of TATA, and NprII serves as a bridge, making protein–protein contacts with the PIC and NprI. 2-Oxoglutarate releases the NprI–NprII complex, allowing rapid RNAP recruitment during nitrogen starvation (187). Across several species, transcriptome mapping and TF–DNA binding studies have detected many

General transcription factor (GTF): includes TATA-binding protein and transcription factor IIB; each GTF binds to a specific DNA site within the promoter region

Promoter: a DNA region upstream of the transcription start site; composed of general transcription factor binding sites

Preinitiation complex (PIC): transcription factor B and TATA-binding protein in the PIC are required for the recruitment of RNA polymerase

Figure 3

An expanding constellation of archaeal transcription factors (TFs) whose DNA binding is regulated by redox activity or ligand binding. Shown are TF families, TFs, the conditions that regulate TF–DNA binding, and the pathways regulated, along with node colors and shapes as defined in the key. The size of the gray TF family nodes corresponds to the number of experimentally characterized TFs belonging to that family. Helix-turn-helix (HTH) and winged HTH (wHTH) families possess a bacterial-type DNA binding domain fused to poorly understood, archaeal-specific catalytic domains. The two (or three) letters following the TF name denote the organism. Evidence for inducing conditions and pathways regulated by each TF are detailed in References 14, 57, 79, 127, and 128, and as follows: AsnC-Hs, CopR-Hs (also termed VNG1179C), Trh2-Hs, Trh3-Hs, Trh6-Hs, and VNG1237C-Hs, Reference 138; BarR-Sa, Reference 101; Bat-Hs, Reference 11; BldR-Ss, Reference 51; CorQR-To, Reference 85; Csa3a-Si, Reference 102; DhcR-My and HhcR-My, Reference 95; DmsR-Hs, Reference 113; DmsR-Hv, Reference 142; EarA-Mma, Reference 41; HrsM-Mv, Reference 172; MalR-Sa, Reference 185; NarO-Hv, Reference 65; RbkR-My, Reference 151; and ThiR-Hv, Reference 72. Abbreviations: Af, *Archaeoglobus fulgidus;* Hs, *Halobacterium salinarum;* Hv, *Haloferax volcanii;* Ma, *Methanosarcina acetivorans;* Mm, *Methanosarcina mazei;* Mma, *Methanococcus maripaludis;* Mt, *Methanothermobacter thermoautotrophicus;* My, *Metallosphaera yellowstonensis;* Pf, *Pyrococcus furiosus;* Sa, *Sulfolobus acidocaldarius;* Si, *Sulfolobus islandicus;* Ss, *Sulfolobus solfataricus;* Tk, *Thermococcus kodakaraensis;* Tl, *Thermococcus litoralis;* To, *Thermococcus onnurineus.*

alternative and condition-dependent transcription start sites within operons and gene coding sequences (86, 179, 197). Together, these complexities in transcriptional regulation could afford regulatory flexibility during environmental variation.

A general strategy for adjusting cellular physiology in response to a given TF is to activate or repress the transcription of a large subset of genes, known as a regulon. Of the TFs whose regulons have been characterized by global TF–DNA binding combined with gene expression in TF knockout strains, nearly all are bifunctional, activating some genes but repressing others (155, 157, 163, 179). Hence, we refer here to transcriptional regulatory proteins collectively as transcription factors (TFs) rather than using the specific terms repressors or activators. *Cis*-regulatory TF binding sequences identified thus far resemble either gapped, inverted repeat palindromes, like those of bacteria (128), or 8-mers, like those of eukaryotes (13). It is unclear which type of *cis* sequence structures are most widely conserved across TFs and species. Nonetheless, a common hypothesis is emerging that the relative positioning of the TF's *cis* sequence and promoter elements (**Figure 2**) predicts the activation or repression of the nearby gene (77, 92, 128, 144, 179).

This hypothesis was tested directly in *H. salinarum* by measuring genome-wide binding locations and the resultant effects of such binding on gene expression, both for GTFs (47, 162) and TFs (157, 163, 179). The integration of such data sets enables a genome-wide view of relative binding positions. For example, for genes regulated by the *H. salinarum* RosR TF in response to oxidative stress, a significant association was detected among the time-dependent binding activity of RosR, resultant gene expression dynamics, and the relative location of RosR binding and PIC binding locations (179). On average, RosR binding occurred within 75 bp of a GTF binding site (upstream in the case of activated genes, overlapping or downstream in the case of repressed genes). A similar significant binding location relationship was observed for TrmB in *H. salinarum* (157). Whether TFs are endowed with both activator and repressor functions through more complex mechanisms [e.g., protein–protein interactions with another TF (187)], or whether the TF alone is sufficient for either function, requires further studies that combine in vitro biochemical approaches, in vivo measurement of genome-wide binding of TFs and GTFs, and transcriptomics in TF knockouts. Taken together, the expanding constellation of archaeal TFs demonstrates a great diversity of mechanisms and gene regulatory functions across species.

3. TRANSCRIPTION SUBNETWORKS AND ENVIRONMENTAL RESPONSE

Free-living microbes, such as the model species of archaea, respond to interrelated environmental regimes that can cause similar types of cellular damage (25, 37). To ensure appropriate adaptation to environmental change, subnetworks composed of multiple TFs and their regulons respond to certain sets of conditions. Studies that map the structure and function of these subnetworks shed light on the regulation of higher order cellular functions. In this section, we illustrate these principles using recent subnetwork case studies.

3.1. Hierarchical Subnetworks Adjust Metabolism During Nutrient Shifts

TrmB family proteins include a diversity of global metabolic regulators. TrmB and TrmBL1 in, respectively, halophiles and thermophiles, directly regulate genes that encode enzymes in central carbon metabolism (e.g., glycolysis and gluconeogenesis) (84, 92, 144, 157, 177). However, outside of central metabolism, targets of TrmB and TrmBL1 differ. Schmid and colleagues (157) showed by ChIP–chip and transcriptomic analysis in knockout strains that TrmB in *H. salinarum* regulates genes involved in peripheral pathways in amino acid, vitamin, and purine biosynthesis, in addition to central carbon metabolism. In contrast, recent ChIP–seq experiments expanded the TrmBL1

regulon in *Pyrococcus furiosus* to include genes that encode functions such as pyruvate-ferredoxin oxidoreductase, proteolysis, and uncharacterized pathways (144). MreA from *Methanosarcina ace-tivorans* is a divergent TrmB family member lacking the C-terminal sugar binding domain. MreA direct and indirect regulons include more than 280 genes involved in acetate metabolism (146).

The global regulatory purview of TrmB in many species includes other TFs. TrmB homologs regulate four other TFs in *H. salinarum* (157), one in *T. kodakaraensis* (77), two in *P. furiosus* (144), and nine, directly and indirectly by MreA, in *Methanosarcina acetivorans* (146). Although the function of these other TFs remains unknown, the TrmB subnetwork structure is consistent with a hierarchical topology observed in other domains of life in which a few master regulators control a large number of genes (199). Such topological simplicity suggests that the nutritional acquisition network could be rapidly rewired during the course of evolution by point mutations in the TrmB ligand binding site (89) or its *cis*-regulatory target sequences (92, 157). The former is a common means of changing regulatory network specificity in bacteria (43, 131), whereas the latter is frequently observed in eukaryotes (96, 168, 194). How archaeal networks are rewired remains to be determined. For example, the sequence motif at the promoter of genes regulated by TrmB in *H. salinarum* TACT-N(7–8)-GAGTA (157) is slightly different than that of the *P. furiosus* TrmBL1 (TATCAC-N5-GTGATA) (144, 182) despite homology between these TF protein homologs. Taken together, the hierarchical TrmB metabolic subnetworks studied so far suggest a flexible strategy by which a single, conserved TF can respond to a diversity of nutrient substrates in various species.

3.2. Combinatorial Control of Gene Expression by Transcription Factor Paralogs

Recent studies of archaeal GRN subnetworks using genome-wide approaches have reported that paralogous TFs within a given species tend to control overlapping regulons. Within such subnetworks, also known as partner networks in the theoretical literature (17), two or more TFs coordinately regulate partially overlapping regulons. Each TF can also regulate its own target genes independently. Frequently, these partner TFs are paralogs. In archaea, partner subnetworks comprising bacterial-like TFs, eukaryote-like TFs, or a combination of both, have been documented. In the ensuing section, we describe examples of each of these types.

3.2.1. Eukaryote-like general transcription factor partner networks.

Recent work is consistent with the hypothesis that within a given species, GTF paralogs function together in partner networks. Approximately 70% of archaeal genomes encode two or more copies of TFB and/or TBP family proteins (**Figure 4**) (132, 180). TFB family expansion events likely occur through duplication and divergence (10). The number of GTFs encoded in archaeal genomes appears to differ by lineage. Of the genomes sequenced so far, expansion is detectable specifically in the *Halobacteriaceae* family of the *Euryarchaeota* phylum, as well as across the *Thaumarchaeota* phylum (**Figure 4**). Genes encoding the TBP (PF00352) (50) and TFE (PF02002) families have undergone modest expansion within both eukaryotes and archaea (**Figure 4a,b**). However, a larger proportion of sequenced archaeal genomes than eukaryotic genomes harbor more than six genes encoding TFBs (PF00382) (**Figure 4c,d**).

For example, in *H. salinarum*, seven TFBs and six TBPs are encoded, enabling 42 possible combinations for PIC assembly and RNAP recruitment at promoters (10). At least seven of these combinations appear to be active under standard growth conditions (47). Additional combinations may be functional under stress conditions; for example, in vitro studies have implicated TFBb in gene regulation under heat shock (103). Some promoters rely on one TFB–TBP pair for transcriptional activation. For instance, TFBf is the exclusive TFB regulator of genes encoding ribosome biogenesis functions (47). In contrast, multiple TFB–TBP pairs bind overlapping sets
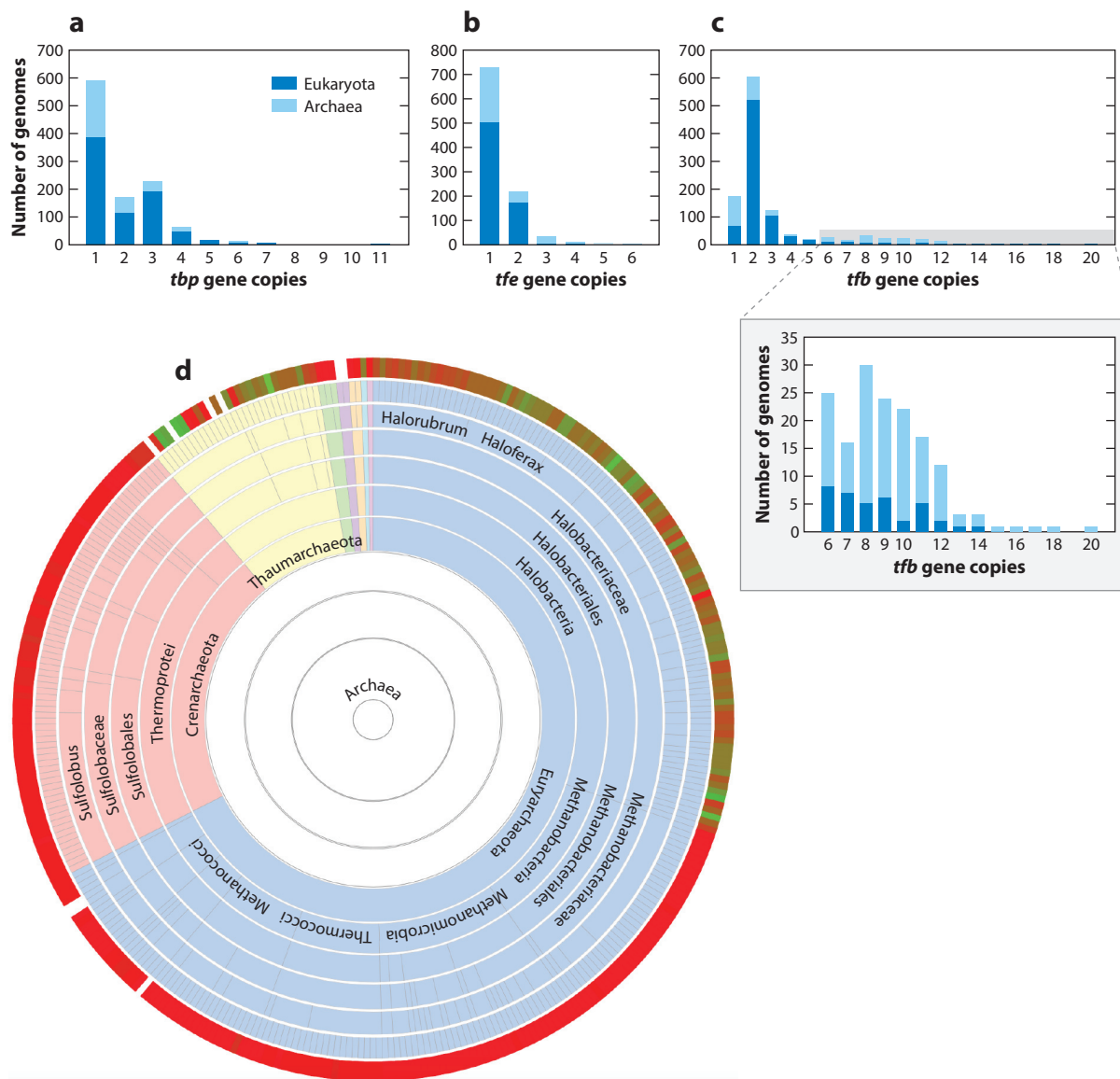
Figure 4

Quantification of the expansion of general transcription factor family proteins in eukaryotic and archaeal genomes. In panels *a–c*, Eukaryota are shown in dark blue and Archaea are shown in light blue. (*a*) The number of genes encoding TATA-binding protein (TBP) is shown on the *x* axis and the number of genomes on the *y* axis. (*b*) Expansion of Transcription factor E (TFE)-encoding genes. (*c*) Transcription factor B (TFB) expansion. The inset illustrates that a larger fraction of archaeal than eukaryotic genomes encodes more than six *tfb* genes. (*d*) Detailed depiction of how TFB family expansion is distributed among archaeal lineages. The innermost rings show the phylum level, with each concentric ring moving outward depicting successively more specific levels. The second ring from the outside is the species level. Colors represent phylum-level designations: blue, Euryarchaeota; light red, Crenarchaeota; yellow, Thaumarchaeota; green, Nanoarchaeota; purple, Parvarchaeota; orange, Nanohaloarchaeota; cyan, uncategorized archaea; magenta, Korarchaeota. In the outermost ring, the TFB copy number is represented by colors: bright red, 1 copy; bright green, 20 copies. Gradations of red and green represent intermediate numbers. Data were accessed from the Pfam database (European Molecular Biology Laboratory, **http://pfam.xfam.org**) (50), and the diagram was drawn using the Aquerium web tool (**http://aquerium.utk.edu**) (1).

of target genes for approximately one-third of genes throughout the genome (47). Computational analysis of GTF-binding sites revealed that slight differences in base composition at individual positions of the BRE and PPE consensus sequences specify which of the seven TFB proteins will bind to which promoters (162).

Genetic knockout studies have demonstrated that each TFB–TBP pair is required for growth under a specific subset of environmental conditions (180); however, some GTF-coding genes are not amenable to knockout, given their essential functions in growth (e.g., TFB-f and -g, and TBP-a, -c, -d) (47). Although some discrepancies exist in the literature regarding which GTFs of *H. salinarum* are essential (34), this perhaps points to conditional essentiality and the sensitivity of such genetic tests to differences in culturing conditions.

In *Methanosarcina acetivorans*, TBP1 appears to be essential, whereas TBPs 2 and 3 are required for growth under low-acetate conditions (145). Transcriptomic analysis revealed that TBP2 and TBP3 regulate an overlapping set of 28 genes, but each also independently regulates approximately 50 genes involved in a range of cellular processes (145).

Together, these studies are consistent with previous arguments that posited archaeal GTFs as the functional analog of bacterial σ factors (10, 47), regulating partially overlapping regulons in response to different conditions (61, 67). However, differentially associating pairs of TFB–TBP proteins, as opposed to a single σ factor, may afford extended combinatorial possibilities for gene regulation during transcription initiation.

3.2.2. Bacterial-like transcription factors control overlapping regulons.
Paralogous bacterial-like TFs also coordinately regulate overlapping sets of genes to integrate multiple environmental signals. For example, in *S. solfataricus*, the metabolic Lrp family regulators LrpB and LysM bind together to 29 promoters throughout the genome (118). LysM alone regulates approximately 40 genes involved in the synthesis and transport of several amino acids (23, 167). Interestingly, the shared LysM–LrpB targets appear to be regulated through protein–protein interactions, with only LysM binding DNA (118), an observation concordant with the observation that LrpB alone regulates fewer than 10 genes (117, 118, 126). In *H. salinarum*, genome-wide binding-site location analyses for the eight Lrp paralogs encoded in the genome and transcriptomics analyses for subsets of Lrps demonstrated that overlapping regulons encode essential cellular functions, whereas independent regulons are responsive to specific stressors (138, 160). Computational analysis suggested that different effector molecule specificities, divergence in DNA binding domains, and differences in the expression levels of *lrp* genes may have contributed to divergence in regulon functions among these paralogous TFs (138).

In *H. salinarum*, two paralogous DtxR TFs, Idr1 and Idr2, directly regulate approximately 20 common targets encoding putative metal binding proteins in response to iron availability (155). These shared target genes also include a third DtxR TF, SirR, which Idr1 activates during iron sufficiency, but Idr2 represses during iron starvation (155). SirR was previously described as a repressor of manganese transport genes (81), but regulation of *sirR* expression during iron imbalance suggests additional roles for SirR in iron homeostasis (81, 111). For example, SirR represses the expression of a gene encoding a fourth iron-dependent TF, TroR, under iron replete conditions (111). In addition to these shared targets, Idr1 and Idr2 each independently regulate different iron uptake systems, such that the overall iron-responsive regulon contains nearly 200 genes (155). Such overlap in regulons is reminiscent of the paralogous iron TFs in yeast Aft1p and Aft2p, which regulate partially overlapping regulons in response to iron starvation (152).

3.2.3. Hybrid partner networks confer stress resistance.
Many archaeal species are extremophiles that thrive at the extremes of temperature (28, 80), pH (22), salinity (123), or radiation

(88, 173), among other factors endemic to extreme environments. Therefore, intensive efforts to understand transcriptional regulation in archaea have been targeted toward understanding the TFs, GRNs, and inducible damage-repair pathways that extremophiles use to regulate stress responses (8, 81, 103, 191, 193). Many of these conditions, especially radiation, produce oxidative radicals through radiolysis of water (27, 73), and so archaea have evolved unique GRNs to deal with reactive oxygen species (ROS). For example, *H. salinarum* is nearly an order of magnitude more resistant to oxidative stress than mesophilic bacteria (163). A central regulator in a GRN subnetwork specific to hypersaline-adapted archaea, RosR, binds and directly regulates the expression of genes encoding 19 other bacterial-type TFs, TFBb, and oxidative stress repair (163, 179). In turn, under optimum growth conditions, TFBb regulates the gene encoding RosR combinatorially with TFBf, TFBg, and TFBd (47, 162, 179). The regulons for these GTFs also overlap with that of RosR. Resistance to oxidative stress is impaired in deletion strains of *H. salinarum* lacking bacterial-like TFs that are combinatorially regulated by RosR and TFBs (37, 179). Together, these results suggest that extensive inter-TF feedback regulation and a high degree of connectivity in the network is important for surviving extremely high levels of oxidative stress. Intriguingly, in this oxidative stress subnetwork, a unique hybrid of eukaryote-like GTF and bacterial-type TFs interact to control each other and the oxidative stress response.
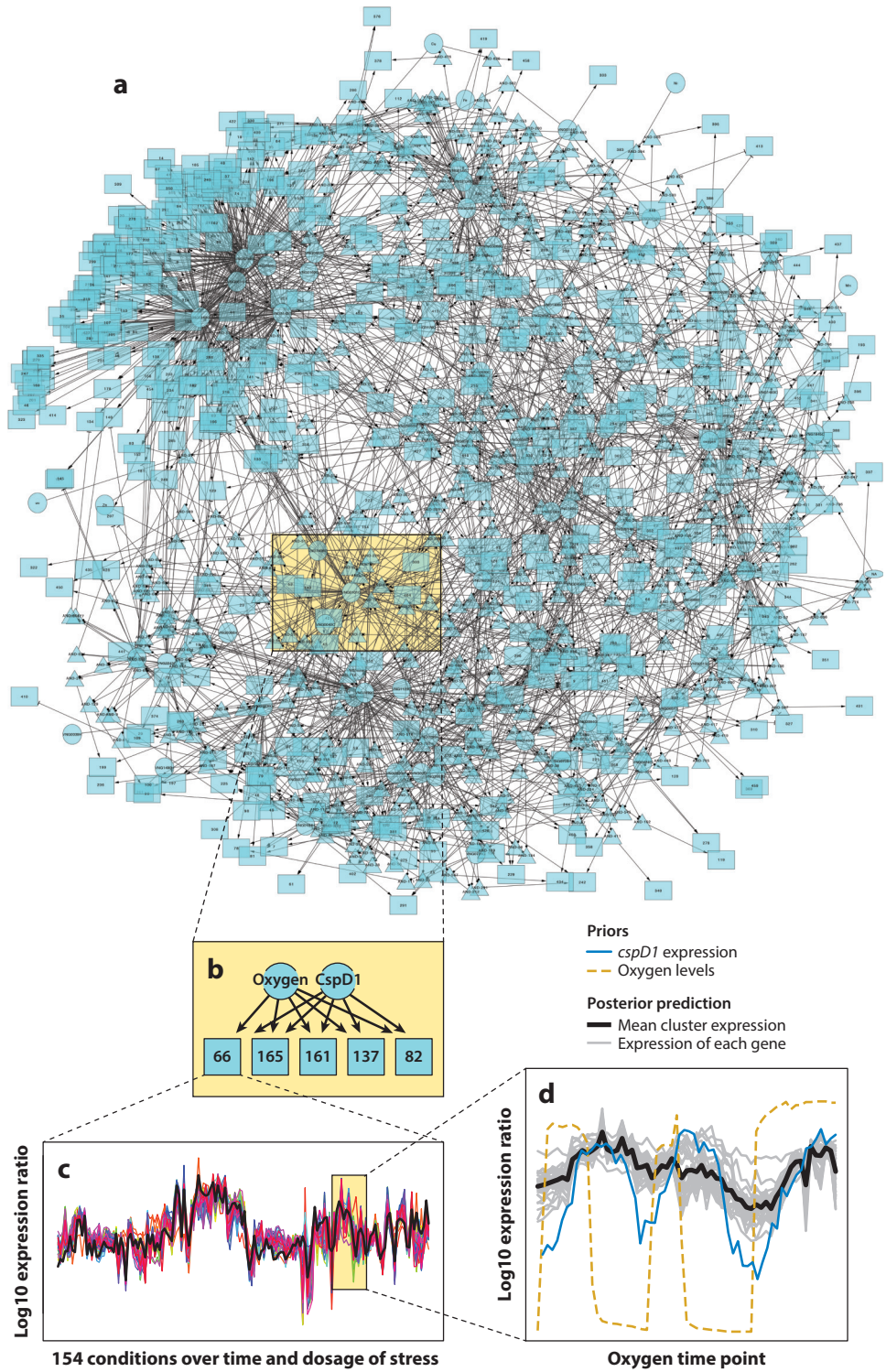
3.2.4. Partner gene regulatory network structures may be universally conserved. In bacteria and eukaryotes, overlapping regulons have been reported to endow GRNs with the ability to regulate gene expression in response to multiple environmental signals (43, 183). This appears to hold true for archaeal subnetworks as well. For example, because hypersaline environments are quite limited in iron (42), the expanded Idr1–Idr2 regulon controlled by the more complex partner network in *H. salinarum*, in addition to transcriptional feedback between TFs themselves, may be an important mechanism enabling the observed resistance to iron starvation (70, 111, 155). Partner networks can expand the regulatory repertoire of TFs (17) and innovate new TF–target gene connections over evolutionary time (96, 168, 194). For example, in *H. salinarum*, slight functional differences in paralogous Lrp proteins with extensively overlapping regulons may provide fine-tuning of gene regulation in a variable environment (138). Taken together, the studies described in this section suggest that partial redundancy of gene regulation appears to be a general theme across the tree of life (17, 43, 96, 152).

4. GLOBAL NETWORKS AND EMERGENT PROPERTIES

4.1. Computational Inference of Global Gene Regulatory Networks in Archaea

Global GRNs include all TFs encoded in the genome, and their *cis* sequences, target genes, and the interactions among them (**Figure 5**). Bayesian statistical methods have been used to generate predictive, global GRN models from gene expression and genome sequencing data for *H. salinarum* (20, 21, 24, 82) and *Methanococcus maripaludis* (198). Although both archaeal global GRN models were constructed using the same computational protocol, many other methods to address the GRN inference challenge have been developed for eukaryotes and bacteria. The literature on the statistical inference of global GRNs is large and is not the focus of this review (38, 110). Rather, in this section we describe the predictive computational GRN models that exist for archaea and the biological insights gained from them.

The global GRN model for *H. salinarum* has been continually refined through iterative rounds of experimentation and computation (20, 24, 82). The original model used more than 300 transcriptome data sets to predict how 72 of the 130 putative TFs influence the expression of

a

b
Oxygen   CspD1

66 165 161 137 82

Priors
— *cspD1* expression
— — Oxygen levels

Posterior prediction
— Mean cluster expression
— Expression of each gene

c
Log10 expression ratio

154 conditions over time and dosage of stress

d
Log10 expression ratio

Oxygen time point

approximately 80% of the 2,400 genes in the genome in response to a wide variety of environmental conditions (20). This network, called the environmental gene regulatory influence network (EGRIN) (20), was inferred using a two-step statistical inference procedure (**Figure 5**). First, genes were grouped into coregulated sets called biclusters using the cMonkey algorithm (147, 148). These groups were optimized according to coexpression under subsets of conditions, prediction of shared TF binding sites, and gene functional annotations (**Figure 5b,c**) (19, 116). In the second step, TFs and environmental factors that control these biclusters were inferred using a statistical Bayesian regression-based algorithm called the Inferelator (21, 59). The Inferelator uses TFs and environmental factors as priors to predict the posterior distribution of the mean expression profile for each bicluster. The Inferelator also incorporates time-series and stress-dose information to infer how the regulation of genes by certain TFs (i.e., activation or repression edges) changes in response to stress over time (**Figure 5b–d**). When gene expression profiles predicted from the GRN model are compared with data collected under new environmental conditions, the correlation is 0.8, meaning that the model can predict transcriptome behavior with reasonable accuracy (20). These predictions are consistent with known TF functions. For instance, the EGRIN model of *H. salinarum* accurately predicted the coregulation of genes involved in phototrophy, the TF controlling them (Bat), and the known Bat-binding site (9, 20). Follow-up modeling studies have refined these original predictions by integrating new experimental data that query the transcriptional response to individual stressors, such as oxidative damage (82), ionizing radiation (191), fluctuations in oxygen (156), nutrient deprivation (157), and other conditions relevant to the salt flat environment (155). These studies have increased the number of experiments in the *H. salinarum* transcriptome database to nearly 1,500. Modeling these new data sets refined the predictions of the original EGRIN (24, 82, 148).

The *Methanococcus maripaludis* EGRIN network was constructed by using a similar modeling framework, and it predicts how 46 of the 57 TFs encoded in the genome influence the regulation of 1,661 genes in response to changes in macronutrients (198). This GRN was more accurate than that of *H. salinarum*, most likely due to the inclusion of chemostats rather than standard culture flasks to more closely control experimental conditions. In both EGRIN models, important factors in maintaining high predictive accuracy included using (*a*) transcriptome measurements from a common reference control condition for normalization across genetic and environmental perturbations, (*b*) common procedures across experiments within each species, and (*c*) expression data from TF knockouts (20, 198).

Figure 5

Computational inference of the global gene regulatory network (GRN): the environmental gene regulatory influence network (EGRIN) of *Halobacterium salinarum* and an example transcription factor (TF) functional prediction. (*a*) The network graph for the global GRN is depicted. Node circles represent TFs; triangles, combinatorial logic gates; and rectangles, biclusters of coregulated genes. Edge arrows represent activation; bars represent repression. The yellow box indicates the inset in panel *b*. The network is shown in a spring-embedded layout in which dense regions represent highly interconnected nodes. The network diagram was generated using the Cytoscape program (32). (*b*) Example function for the CspD1 TF. Each bicluster is numbered, corresponding to details given at **http://db.systemsbiology.net:8080/biclusterviewer**. (*c*) Example data that fueled the prediction for bicluster 66. The graph depicts the expression of all 34 genes (*colored lines*) and the mean expression of those genes (*black line*) over 154 conditions in bicluster 66. This bicluster is significantly enriched for functions in ribosome biogenesis [hypergeometric test on archaeal clusters of orthologous genes categories (109), $p \leq 3.63 \times 10^{-25}$]. Panel *c* adapted from Reference 20. (*d*) Expression of the gene encoding CspD1 (Bayesian prior; *blue*) compared with genes in bicluster 66 (posterior prediction; *gray*) during oxygen fluctuation over time (*gold dashed line*).

Progress toward a global GRN model has been made by collating published experimental data for *Methanosarcina acetivorans* (134). Specifically, the direct TF–DNA interaction network was compiled from known binding interactions (e.g., in vitro mobility shift and footprinting assays), as well as from inferred direct interactions, by searching the genome for known TF *cis* sequences. A total of 10 TFs and 248 edges were included (134). An expanded indirect network also counted TF–target gene interactions based on correlated expression profiles from genome-wide expression data sets (e.g., 134, 145, 146, among others). These GRNs were constructed for the purpose of building an ordinary differential equation–based kinetic model that incorporated several other data types (protein–protein interactions, enzyme kinetics, and others). This model accurately predicted the rate of consumption of methanogenic substrates (e.g., acetate); however, the growth rate predictions on these substrates require further refinements (134). In the context of this model, the TF MreA was confirmed as a key regulator of the switch between different methanogenic growth substrates (134, 146).

4.2. Discovering the Regulatory Hubs of Gene Regulatory Networks

By studying all components of global GRNs and their interactions as a whole, dynamical and/or phenomenological emergent properties—that is, higher-order functions—of the GRN may be deduced. For instance, research in bacterial and eukaryotic model organisms has established that a surprisingly small set of TFs function as hubs in the global GRN (5, 55, 199), eliciting global, coordinated gene expression response to a wide variety of stimuli [e.g., σ^B in *Bacillus subtilis* (66), σ^S in *Escherichia coli* (12)]. These TF hubs are themselves extensively regulated at the levels of transcription, protein stability, and DNA binding activity (12, 49, 66). Despite the successes of the first few global GRN modeling efforts in archaea, experimental tests of model predictions are still required to verify these TF hubs. Such experiments differentiate between the direct and indirect gene targets of each TF, identify inter-TF regulatory loops, and determine how TF hubs are regulated. The use of experimental and computational systems biology pipelines specially adapted for archaea (**Figure 1**) (36) has led to the mapping of the regulons of many TFs (**Figure 3**). For example, in *H. salinarum* such pipelines have enabled the discovery of TF hubs such as TrmB, which coordinates central metabolism with cofactor biosynthesis (157, 175–177; see also Section 3.1), and RosR, which regulates many other TFs to resist oxidative stress (163, 179; see also Section 3.2.3). TFs that control essential cellular processes, such as ribosome biogenesis, in response to a wide range of environmental cues have also been discovered (**Figure 5b–d**; 37). In *Methanococcus maripaludis*, transcriptomics experiments in knockout strains were used to test EGRIN predictions. These tests revealed the surprising coordinate regulation of genes encoding essential methanogenesis enzymes by the novel TFs MMP0719 and MMP1100 (TrmB family homologs) in response to a phosphate signal (198). In summary, using genome-wide methods to test global GRN predictions can reveal key TFs that function as network hubs, rapidly elucidate complete sets of target genes, and delineate the relationships among TF regulons.

4.3. Novel Transcription Factor Discovery by Phylogenetic Profiling

An orthogonal computational method known as phylogenetic profiling harnesses comparative genomics to predict TF functions, their target regulons, and *cis*-regulatory binding sequences (94, 95, 151). For example, a recent study used phylogenetic profiling to reconstruct regulons for the DtxR family TFs across archaea (94). By comparing more than 100 DtxR family proteins from available genomes across four phyla, nearly 600 candidate binding sites for DtxR TFs were inferred de novo by searching for conserved palindromic sequences in upstream regions of genes (94). As expected, these motifs were typically located upstream of genes with putative functions

in metal homeostasis. A similar approach has been used to computationally predict the regulons of novel TFs involved in autotrophic metabolism (95) and vitamin biosynthesis (150, 151).

Experimental tests of these predictions demonstrate that such profiling approaches are successful in discovering novel TFs and their regulons. For example, the DtxR family TF SirR was predicted to perform dual roles in manganese homeostasis and iron regulation (94). Transcriptomics data from varying metal conditions and the *sirR* knockout strain of *H. salinarum* validated this prediction (81, 94, 111, 155). In addition, the TrmB family homolog HhcR bound in vitro to target genes encoding enzymes in unusual carbon dioxide fixation pathways (95). In another striking example, phylogenetic profiling predicted that, unlike posttranscriptional control of riboflavin biosynthesis genes by riboswitches such as occurs in bacteria, the archaeal pathway is under the transcriptional control of the novel TF RbkR (151). This prediction was confirmed by in vitro binding experiments in which RbkR bound CTP to specify DNA binding to promoters of genes encoding riboflavin biosynthesis and uptake functions (151). These studies demonstrate the power of comparative genome-wide computational approaches to accurately predict novel mechanisms for the regulation and function of unknown genes.

4.4. Recurrent Subnetwork Topologies Within Global Gene Regulatory Networks Predict Gene Expression Dynamics

In bacteria and eukaryotes, the computational deconstruction of global GRNs has been a fruitful approach for discovering how TFs function together in subnetworks. For instance, in the model organisms *E. coli* and *Saccharomyces cerevisiae*, statistical comparisons of the experimentally characterized global GRN with randomized GRNs revealed that certain subnetwork structures occur significantly more often than others (105, 112, 165). The topology and dynamical properties of such overrepresented subnetworks, termed network motifs (112, 165), have been reviewed elsewhere (3, 106); however, we describe these briefly for the sake of comparison with archaeal network motifs.

The same subnetwork types recur in the networks of both bacteria and yeast (112): feedforward loops (FFLs), bifans, and single-input modules (SIMs) (**Figure 6**) (165). Given a certain structure, or topology, of a subnetwork, mathematical theory from the fields of physics and engineering predicts the dynamic effect of the subnetwork on downstream gene expression (3). FFLs allow for noise filtering such that the regulated gene is induced or repressed only in the presence of a persistent stimulus (3). SIMs allow for precise temporal ordering of gene induction by the same TF. Such ordering depends upon the threshold concentration of the TF (i.e., the binding affinity of the TF for the *cis* sequence) required to activate or repress each promoter in the SIM (3). Bifans serve as signal integrators from dual stimuli and ensure appropriate gene expression output. The prediction of dynamics from topology is powerful in focusing experiments to enable understanding of the transcriptional regulation of complex cellular behaviors.

Although a comprehensive analysis of statistically overrepresented motifs has not yet been conducted for the known global GRNs in archaea, the exemplar subnetworks discussed in Section 3 reveal striking similarities to the recurrent GRN motifs from other domains of life. For instance, TrmB is a master regulator of an extensive SIM in *H. salinarum* (**Figure 6**) (177). This topology was deduced from microarray transcriptomics and genome-wide TrmB–DNA binding in the presence and absence of glucose (157). The dynamical properties of the SIM were further tested by measuring TrmB–DNA binding and RNA abundance over time in response to a bolus of glucose delivered to cultures during gluconeogenic growth on amino acids (177). Switch-like dynamics in expression were observed within 2 minutes of the glucose addition (**Figure 6**). The phosphoenolpyruvate synthase–encoding gene is shown in **Figure 6** as an example of the expression
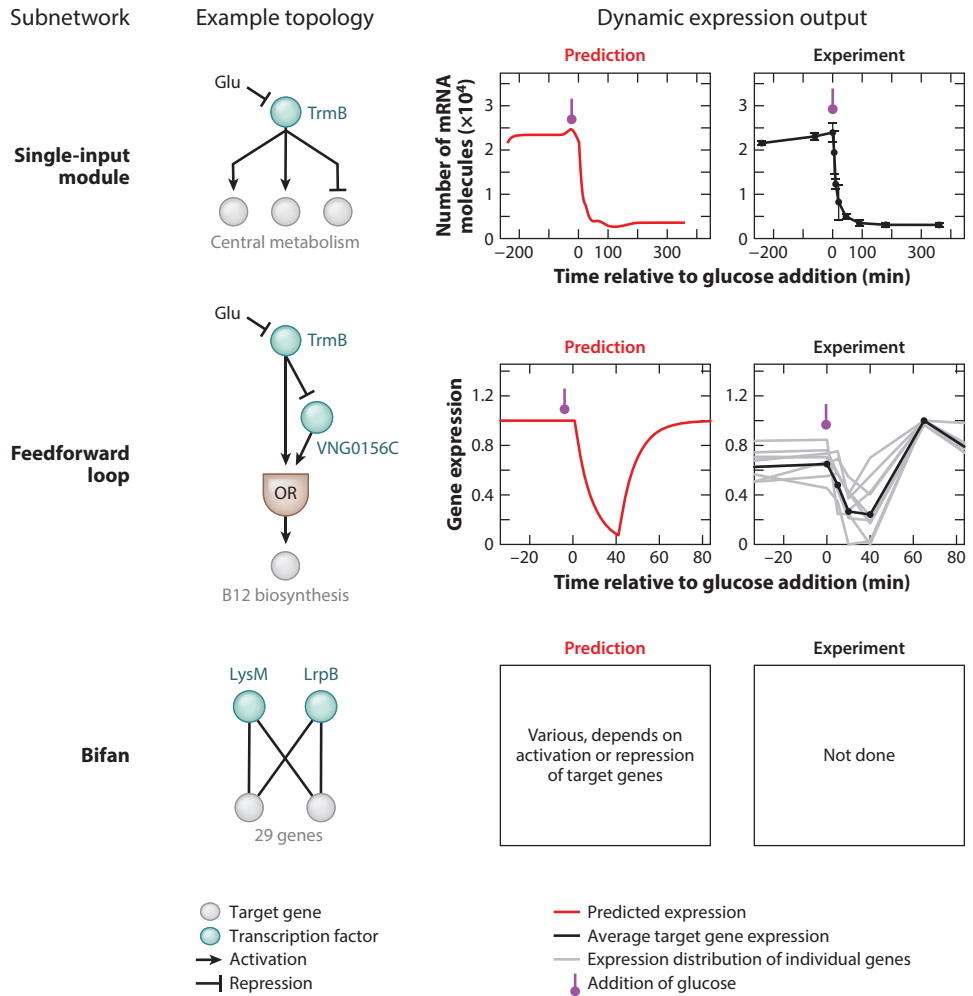
Figure 6

Subnetwork motifs that are overrepresented in bacterial and eukaryotic global gene regulatory networks have also been identified in archaea. (*Left*) Subnetwork types with example topologies. Node VNG0156C was computationally predicted to participate in the feedforward loop with TrmB (177). (*Right*) Gene expression predictions (*red lines*) for target genes and experimental data (*gray* and *black lines*). Abbreviation: Glu, glucose.

dynamics of all gluconeogenic enzyme–coding genes (177). Ordinary differential equation models accurately predicted these gene expression dynamics directly from TrmB–DNA binding time-course data, suggesting that TrmB alone is sufficient to regulate genes in the SIM (177). Surprisingly, a transient wave of expression following glucose addition was observed in genes involved in peripheral metabolic pathways (**Figure 6**) (177). Logic models (3) predicted that such pulses of gene expression may be regulated by an FFL subnetwork in which TrmB represses another TF and, together, these TFs generate pulses of expression (**Figure 6**) (177). Clues to the identity of the other TFs were provided by the EGRIN model (20) and global TrmB–DNA binding data (157).

Multiple bifan motifs are possible within each of the subnetworks coordinately regulated by LrpB and LysM in *S. solfataricus* (118) and by Idr1 and Idr2 in *H. salinarum* (**Figure 6**;

Section 3.2.2) (155). The eight Lrp family TFs in *H. salinarum* also regulate extensively overlapping regulons, providing significant combinatorial potential for different dynamical responses to environmental cues (138, 160). How these bifan motifs affect the dynamic expression of the target genes requires further experimentation. Nevertheless, taking together the studies discussed in this section, subnetwork motifs appear to be shared across global GRNs of bacteria, eukaryotes, and archaea (55, 63, 112, 165).

5. FUTURE CHALLENGES

5.1. Gene Functions Remain to Be Discovered in Archaea

GRN studies enable TF discovery based on the predicted function of genes directly regulated by the TF of interest (36). However, in some archaeal genomes, the functions encoded by up to 60% of genes remain unknown (19, 26, 116). Traditional methods such as the phenotypic analysis of gene knockout strains and biochemical characterization of purified enzymes are powerful for discovery and in-depth study of the function of gene products. However, such methods investigate only one gene at a time and require significant time and financial investment. Therefore, the need for improved methods to conduct simultaneous functional discovery for thousands of unknown genes remains a major bottleneck in archaeal genomics. Computational methods—such as de novo protein structure prediction and the subsequent construction of a gene functional association network—enabled a substantial improvement in gene functional knowledge in *H. salinarum* (19). The archaeal clusters of orthologous genes (arCOG) functional classification for archaea has also vastly improved gene function annotations by including uniquely archaeal enzymes, processes, and pathways (109). The arCOG classification has advanced understanding of the function of TF regulons by enabling statistical analyses of gene functional enrichment in lists of differentially expressed genes from genomics studies (36).

New experimental methods also show promise for rapid discovery of gene functions. For instance, transposon mutagenesis libraries have been constructed for genetically amenable model archaeal organisms (83, 154). Coupling such mutagenesis to sequencing (known as Tn-seq) in *Methanococcus maripaludis* enabled the genome-wide discovery of essential genes, more than 100 of which are unique to methanogens (154). Tools using targeted CRISPR (clustered regularly interspaced short palindromic repeat)-mediated gene silencing have been employed to characterize the function of an essential gene in halophiles (170). CRISPR technology was also used to knock down β-galactosidase activity in *Sulfolobus* (202). These CRISPR methods are extensible to genome-wide studies, which have been conducted recently in bacteria (133). Genome-wide transcriptome start site mapping (**Figure 1**) (6, 76, 86, 195, 197) and quantitative proteomics (156, 174, 181, 196) have improved genome annotations, enabling the discovery of new open reading frames, the correction of start codons, and improvements in databases (86, 136, 181). Therefore, genome sequence annotations can be viewed as incomplete and constantly improving data sets, even for the most well-understood model organisms.

5.2. How Do Posttranscriptional Mechanisms Contribute to the Function of Gene Regulatory Networks?

This review has focused exclusively on the transcriptional regulation of gene expression. However, additional layers of regulation undoubtedly play important parts in GRN function. For instance, a surprisingly high number of small, untranslated antisense RNA transcripts, on the order of 50–500 bp, have been detected in RNA-seq studies across numerous species (7). Many small RNAs

(sRNAs) in bacteria regulate the expression of global TFs at the posttranscriptional level (171) and, therefore, have central roles in the dynamical properties of GRNs (119). Genetic analyses have revealed that archaeal sRNAs are required for survival across a variety of environmental conditions (7, 45, 69). In *Methanosarcina mazei*, $sRNA_{154}$ is regulated in response to nitrogen starvation by the TF NprR (45). $sRNA_{154}$ can posttranscriptionally either activate or repress expression of nitrogen-responsive pathways (139). Two other sRNAs, including $sRNA_{162}$, target multiple mRNAs, repressing translation during nutrient shifts (75). However, the regulatory function is unknown for the majority of recently identified sRNAs.

Recent evidence has also revealed that the stability of many archaeal proteins, including TFB homologs (52), is regulated by a ubiquitin-like pathway in which small archaeal-specific modifier proteins (SAMPs) are covalently conjugated to protein substrates (71). SAMPylated proteins are then often targeted for degradation by the eukaryote-like proteasome. In the future, systems biology approaches could reveal the topology and function of composite transcriptional/posttranscriptional GRNs.

5.3. How Do Archaeal Gene Regulatory Networks Evolve?

Archaeal genomes are profoundly dynamic. Insertion sequence (IS) and repeat elements are enriched in archaeal genomes [e.g., 91 in *H. salinarum* (116)]. Differential GC-content bias between megaplasmids and the main chromosome has also been observed (116). These features provide evidence of past viral infections and horizontal gene transfer (87). Through recombination, IS and repeats lead to high rates of amplification, deletion, and rearrangement events. For example, these rates are as high as 1×10^{-3} for spontaneous gas vesicle mutants in *H. salinarum* (137, 153), and different strains of *P. furiosus* spontaneously lost *trmB* during in-lab cultivation (144). When exposed to stress, IS element mobility is enhanced (135). Genomic dynamism is exacerbated in haloarchaea during fusion between cells of different species, which exchange hundreds of kilobase pairs of DNA to generate interspecific hybrid strains (115).

What is the effect of such extensive genomic plasticity on the topology, function, and evolution of GRNs in archaea? Through rapid gene gain and loss, horizontal gene transfer provides a constant source of potentially novel regulatory features (87, 140), such as co-option of a newly acquired TF for regulating the response to a new environment (43). Extensive functional redundancy in the GRN can increase the regulatory repertoire and enable colonization of new niches (180). Because many model archaeal species thrive in extreme environments, genomic plasticity may be a mechanism for innovating new stress-response strategies, forcing rapid network rewiring and regulation during environmental fluctuation. In these environments, organisms are faced with simultaneous, chemically interrelated stressors. One study suggested that the relative evolutionary rate for microbial communities residing in extreme environments is higher than that for mesophilic environments (98). Recent studies in bacteria have posited that correlated signals may shape the GRN such that different TFs sense different stressors but regulate overlapping regulons (44), perhaps enabling anticipation of future stress (43). Future work to test whether these phenomena hold for archaeal GRNs is of great interest.

SUMMARY POINTS

1. In Archaea, proteins of eukaryotic and bacterial ancestry interact to regulate the initiation of transcription at the promoters of individual genes.

2. Subnetworks composed of three to four transcription factors (TFs) work together to regulate gene expression in response to fluctuating environments. The different topologies of these subnetworks produce characteristic dynamical patterns of gene expression.

3. In archaea, subnetworks have been observed that are composed of bacterial-like TFs, eukaryote-like TFs, or a hybrid of the two.

4. Experimental validation of predicted global gene regulatory networks (GRNs) has enabled the discovery of emergent network properties and novel TF functions.

FUTURE ISSUES

1. Systems biology approaches hold promise for understanding the structure and function of gene regulatory networks (GRNs) across archaea.

2. Which subnetwork motifs and resultant dynamics of gene expression are overrepresented in global archaeal GRNs?

3. The discovery of the function of unknown genes through genome-wide analyses will enable future understanding of GRN function.

4. How do additional layers of gene regulation at the posttranscriptional level contribute to GRN function?

5. How do network structure and function change over evolutionary timescales in response to extreme stress?

6. What are the mechanisms by which archaeal genome dynamics affect GRN function and evolution?

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

1. Adebali O, Zhulin IB. 2017. Aquerium: a web application for comparative exploration of domain-based protein occurrences on the taxonomically clustered genome tree. *Proteins* 85:72–77

2. Allers T, Barak S, Liddell S, Wardell K, Mevarech M. 2010. Improved strains and plasmid vectors for conditional overexpression of His-tagged proteins in *Haloferax volcanii*. *Appl. Environ. Microbiol.* 76:1759–69

3. Alon U. 2006. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, FL: CRC Press, Taylor & Francis Group

4. Ammar R, Torti D, Tsui K, Gebbia M, Durbic T, et al. 2012. Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife* 1:e00078

5. Arrieta-Ortiz ML, Hafemeister C, Bate AR, Chu T, Greenfield A, et al. 2015. An experimentally supported model of the *Bacillus subtilis* global transcriptional regulatory network. *Mol. Syst. Biol.* 11:839

6. Babski J, Haas KA, Näther-Schindler D, Pfeiffer F, Förstner KU, et al. 2016. Genome-wide identification of transcriptional start sites in the haloarchaeon *Haloferax volcanii* based on differential RNA-Seq (dRNA-Seq). *BMC Genom.* 17:629

7. Babski J, Maier LK, Heyer R, Jaschinski K, Prasse D, et al. 2014. Small regulatory RNAs in Archaea. *RNA Biol.* 11:484–93

8. Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, et al. 2004. Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium NRC-1*. *Genome Res.* 14:1025–35

9. Baliga NS, DasSarma S. 1999. Saturation mutagenesis of the TATA box and upstream activator sequence in the haloarchaeal *bop* gene promoter. *J. Bacteriol.* 181:2513–18

10. Baliga NS, Goo YA, Ng WV, Hood L, Daniels CJ, DasSarma S. 2000. Is gene expression in *Halobacterium* NRC-1 regulated by multiple TBP and TFB transcription factors? *Mol. Microbiol.* 36:1184–85

11. Baliga NS, Kennedy SP, Ng WV, Hood L, DasSarma S. 2001. Genomic and genetic dissection of an archaeal regulon. *PNAS* 98:2521–25

12. Battesti A, Majdalani N, Gottesman S. 2011. The RpoS-mediated general stress response in *Escherichia coli*. *Annu. Rev. Microbiol.* 65:189–213

13. Bauer M, Marschaus L, Reuff M, Besche V, Sartorius-Neef S, Pfeifer F. 2008. Overlapping activator sequences determined for two oppositely oriented promoters in halophilic Archaea. *Nucleic Acids Res.* 36:598–606

14. Bell SD. 2005. Archaeal transcriptional regulation—variation on a bacterial theme? *Trends Microbiol.* 13:262–65

15. Bell SD, Cairns SS, Robson RL, Jackson SP. 1999. Transcriptional regulation of an archaeal operon in vivo and in vitro. *Mol. Cell* 4:971–82

16. Bell SD, Jackson SP. 2000. Mechanism of autoregulation by an archaeal transcriptional repressor. *J. Biol. Chem.* 275:31624–29

17. Bhardwaj N, Carson MB, Abyzov A, Yan KK, Lu H, Gerstein MB. 2010. Analysis of combinatorial regulation: scaling of partnerships between regulators with the number of governed targets. *PLOS Comput. Biol.* 6:e1000755

18. Blombach F, Smollett KL, Grohmann D, Werner F. 2016. Molecular mechanisms of transcription initiation—structure, function, and evolution of TFE/TFIIE-like factors and open complex formation. *J. Mol. Biol.* 428:2592–606

19. 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

20. Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, et al. 2007. A predictive model for transcriptional control of physiology in a free living cell. *Cell* 131:1354–65

21. Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, et al. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol.* 7:R36

22. Bowers KJ, Wiegel J. 2011. Temperature and pH optima of extremely halophilic archaea: a mini-review. *Extremophiles* 15:119–28

23. Brinkman AB, Bell SD, Lebbink RJ, de Vos WM, van der Oost J. 2002. The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J. Biol. Chem.* 277:29537–49

24. Brooks AN, Reiss DJ, Allard A, Wu WJ, Salvanha DM, et al. 2014. A system-level model for the microbial regulatory genome. *Mol. Syst. Biol.* 10:740

25. Brooks AN, Turkarslan S, Beer KD, Lo FY, Baliga NS. 2011. Adaptation of cells to new environments. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 3:544–61

26. Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1058–73

27. Buxton GV. 1987. Radiation chemistry of the liquid state: (1) water and homogeneous aqueous solutions. In *Radiation Chemistry: Principles and Applications*, ed. Farhataziz, MAJ Rodgers, pp. 321–50. New York: VCH Publ.

28. Canganella F, Wiegel J. 2014. Anaerobic thermophiles. *Life* 4:77–104

29. Charlesworth JC, Burns BP. 2015. Untapped resources: biotechnological potential of peptides and secondary metabolites in archaea. *Archaea* 2015:282035

30. Charoensawan V, Wilson D, Teichmann SA. 2010. Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res*. 38:7364–77

31. Chen L, Brugger K, Skovgaard M, Redder P, She Q, et al. 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. *J. Bacteriol*. 187:4992–99

32. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, et al. 2007. Integration of biological networks and gene expression data using Cytoscape. *Nature Protoc*. 2:2366–82

33. Cohen O, Doron S, Wurtzel O, Dar D, Edelheit S, et al. 2016. Comparative transcriptomics across the prokaryotic tree of life. *Nucleic Acids Res*. 44:46–53

34. Coker JA, DasSarma S. 2007. Genetic and transcriptomic analysis of transcription factor genes in the model halophilic Archaeon: coordinate action of TbpD and TfbA. *BMC Genet*. 8:61

35. Costa KC, Leigh JA. 2014. Metabolic versatility in methanogens. *Curr. Opin. Biotechnol*. 29:70–75

36. Darnell CL, Schmid AK. 2015. Systems biology approaches to defining transcription regulatory networks in halophilic archaea. *Methods* 86:102–14

37. Darnell CL, Tonner PD, Gulli JG, Schmidler S, Schmid AK. 2017. Systematic discovery of archaeal transcription factor functions in regulatory networks through quantitative phenotyping analysis. *mSystems* 2:e00032-17

38. De Smet R, Marchal K. 2010. Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol*. 8:717–29

39. Decker KB, Hinton DM. 2013. Transcription regulation at the core: similarities among bacterial, archaeal, and eukaryotic RNA polymerases. *Annu. Rev. Microbiol*. 67:113–39

40. Deppenmeier U, Johann A, Hartsch T, Merkl R, Schmitz RA, et al. 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol*. 4:453–61

41. Ding Y, Nash J, Berezuk A, Khursigara CM, Langelaan DN, et al. 2016. Identification of the first transcriptional activator of an archaellum operon in a euryarchaeon. *Mol. Microbiol*. 102:54–70

42. Domagalski JL Eugster HP, Jones BF. 1990. Trace metal geochemistry of Walker, Mono, and Great Salt Lakes. In *Fluid–Mineral Interactions: A Tribute to H. P. Eugster*, ed. RJ Spencer, I-M Chou, pp. 315–53. Washington, DC: Geochem. Soc.

43. 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

44. 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

45. Ehlers C, Jäger D, Schmitz RA. 2011. Establishing a markerless genetic exchange system for *Methanosarcina mazei* strain Gö1 for constructing chromosomal mutants of small RNA genes. *Archaea* 2011:439608

46. Esser D, Hoffmann L, Pham TK, Brasen C, Qiu W, et al. 2016. Protein phosphorylation and its role in archaeal signal transduction. *FEMS Microbiol. Rev*. 40:625–47

47. Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, et al. 2007. General transcription factor specified global gene regulation in archaea. *PNAS* 104:4630–35

48. Farkas JA, Picking JW, Santangelo TJ. 2013. Genetic techniques for the Archaea. *Annu. Rev. Genet*. 47:539–61

49. Fiebig A, Herrou J, Willett J, Crosson S. 2015. General stress signaling in the Alphaproteobacteria. *Annu. Rev. Genet*. 49:603–25

50. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res*. 34:D247–51

51. Fiorentino G, Ronca R, Cannio R, Rossi M, Bartolucci S. 2007. MarR-like transcriptional regulator involved in detoxification of aromatic compounds in *Sulfolobus solfataricus*. *J. Bacteriol*. 189:7351–60

52. Fu X, Liu R, Sanchez I, Silva-Sanchez C, Hepowit NL, et al. 2016. Ubiquitin-like proteasome system represents a eukaryotic-like pathway for targeted proteolysis in archaea. *mBio* 7:e00379-16

53. Fukui T, Atomi H, Kanai T, Matsumi R, Fujiwara S, Imanaka T. 2005. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.* 15:352–63

54. Galagan JE, Nusbaum C, Roy A, Endrizzi MG, Macdonald P, et al. 2002. The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* 12:532–42

55. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado L, et al. 2016. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* 44:D133–43

56. Gehring AM, Walker JE, Santangelo TJ. 2016. Transcription regulation in Archaea. *J. Bacteriol.* 198:1906–17

57. Gindner A, Hausner W, Thomm M. 2014. The TrmB family: a versatile group of transcriptional regulators in Archaea. *Extremophiles* 18:925–36

58. Götz D, Paytubi S, Munro S, Lundgren M, Bernander R, White MF. 2007. Responses of hyperthermophilic crenarchaea to UV irradiation. *Genome Biol.* 8:R220

59. Greenfield A, Hafemeister C, Bonneau R. 2013. Robust data-driven incorporation of prior knowledge into the inference of dynamic regulatory networks. *Bioinformatics* 29:1060–67

60. Grohmann D, Werner F. 2011. Recent advances in the understanding of archaeal transcription. *Curr. Opin. Microbiol.* 14:328–34

61. Gruber TM, Gross CA. 2003. Multiple σ subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.* 57:441–66

62. Hain J, Reiter WD, Hudepohl U, Zillig W. 1992. Elements of an archaeal promoter defined by mutational analysis. *Nucleic Acids Res.* 20:5423–28

63. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* 431:99–104

64. Hartman AL, Norais C, Badger JH, Delmas S, Haldenby S, et al. 2010. The complete genome sequence of *Haloferax volcanii* DS2, a model archaeon. *PLOS ONE* 5:e9605

65. Hattori T, Shiba H, Ashiki K, Araki T, Nagashima YK, et al. 2016. Anaerobic growth of haloarchaeon *Haloferax volcanii* by denitrification is controlled by the transcription regulator NarO. *J. Bacteriol.* 198:1077–86

66. Hecker M, Pane-Farre J, Volker U. 2007. SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria. *Annu. Rev. Microbiol.* 61:215–36

67. Helmann JD. 2016. *Bacillus subtilis* extracytoplasmic function (ECF) σ factors and defense of the cell envelope. *Curr. Opin. Microbiol.* 30:122–32

68. Hendrickson EL, Kaul R, Zhou Y, Bovee D, Chapman P, et al. 2004. Complete genome sequence of the genetically tractable hydrogenotrophic methanogen *Methanococcus maripaludis. J. Bacteriol.* 186:6956–69

69. Heyer R, Dorr M, Jellen-Ritter A, Spath B, Babski J, et al. 2012. High throughput sequencing reveals a plethora of small RNAs including tRNA derived fragments in *Haloferax volcanii. RNA Biol.* 9:1011–18

70. Hubmacher D, Matzanke BF, Anemuller S. 2003. Effects of iron limitation on the respiratory chain and the membrane cytochrome pattern of the Euryarchaeon *Halobacterium salinarum. Biol. Chem.* 384:1565–73

71. Humbard MA, Miranda HV, Lim JM, Krause DJ, Pritz JR, et al. 2010. Ubiquitin-like small archaeal modifier proteins (SAMPs) in *Haloferax volcanii. Nature* 463:54–60

72. Hwang S, Cordova B, Abdo M, Pfeiffer F, Maupin-Fantiurlow JA. 2017. ThiN as a versatile domain of transcriptional repressors and catalytic enzymes of thiamine biosynthesis. *J. Bacteriol.* 199:e00810-16

73. Imlay JA. 2003. Pathways of oxidative damage. *Annu. Rev. Microbiol.* 57:395–418

74. Isom CE, Turner JL, Lessner DJ, Karr EA. 2013. Redox-sensitive DNA binding by homodimeric *Methanosarcina acetivorans* MsvR is modulated by cysteine residues. *BMC Microbiol.* 13:163

75. Jäger D, Pernitzsch SR, Richter AS, Backofen R, Sharma CM, Schmitz RA. 2012. An archaeal sRNA targeting *cis-* and *trans-*encoded mRNAs via two distinct domains. *Nucleic Acids Res.* 40:10964–79

76. Jäger D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA. 2009. Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *PNAS* 106:21878–82

77. Kanai T, Akerboom J, Takedomi S, van de Werken HJ, Blombach F, et al. 2007. A global transcriptional regulator in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme–encoding genes. *J. Biol. Chem.* 282:33659–70

78. Karr EA. 2010. The methanogen-specific transcription factor MsvR regulates the *fpaA-rlp-rub* oxidative stress operon adjacent to *msvR* in *Methanothermobacter thermautotrophicus*. *J. Bacteriol.* 192:5914–22

79. Karr EA. 2014. Transcription regulation in the third domain. *Adv. Appl. Microbiol.* 89:101–33

80. Kashefi K, Lovley DR. 2003. Extending the upper temperature limit for life. *Science* 301:934

81. 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.* 16:841–54

82. Kaur A, Van PT, Busch CR, Robinson CK, Pan M, et al. 2010. Coordination of frontline defense mechanisms under severe oxidative stress. *Mol. Syst. Biol.* 6:393

83. Kiljunen S, Pajunen MI, Dilks K, Storf S, Pohlschroder M, Savilahti H. 2014. Generation of comprehensive transposon insertion mutant library for the model archaeon, *Haloferax volcanii*, and its use for gene discovery. *BMC Biol.* 12:103

84. Kim M, Park S, Lee SJ. 2016. Global transcriptional regulator TrmB family members in prokaryotes. *J. Microbiol.* 54:639–45

85. Kim MS, Choi AR, Lee SH, Jung HC, Bae SS, et al. 2015. A novel CO-responsive transcriptional regulator and enhanced H_2 production by an engineered *Thermococcus onnurineus* NA1 strain. *Appl. Environ. Microbiol.* 81:1708–14

86. Koide T, Reiss DJ, Bare JC, Pang WL, Facciotti MT, et al. 2009. Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol. Syst. Biol.* 5:285

87. Koonin EV. 2016. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research* 5(F1000 Faculty Rev.):1805

88. Kottemann M, Kish A, Iloanusi C, Bjork S, DiRuggiero J. 2005. Physiological responses of the halophilic archaeon *Halobacterium* sp. strain NRC1 to desiccation and γ irradiation. *Extremophiles* 9:219–27

89. Krug M, Lee SJ, Boos W, Diederichs K, Welte W. 2013. The three-dimensional structure of TrmB, a transcriptional regulator of dual function in the hyperthermophilic archaeon *Pyrococcus furiosus* in complex with sucrose. *Protein Sci.* 22:800–8

90. Lecompte O, Ripp R, Puzos-Barbe V, Duprat S, Heilig R, et al. 2001. Genome evolution at the genus level: comparison of three complete genomes of hyperthermophilic archaea. *Genome Res.* 11:981–93

91. Lee SJ, Moulakakis C, Koning SM, Hausner W, Thomm M, Boos W. 2005. TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers. *Mol. Microbiol.* 57:1797–807

92. Lee SJ, Surma M, Hausner W, Thomm M, Boos W. 2008. The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Arch. Microbiol.* 190:247–56

93. Leigh JA, Albers SV, Atomi H, Allers T. 2011. Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales*. *FEMS Microbiol. Rev.* 35:577–608

94. Leyn SA, Rodionov DA. 2015. Comparative genomics of DtxR family regulons for metal homeostasis in Archaea. *J. Bacteriol.* 197:451–58

95. Leyn SA, Rodionova IA, Li X, Rodionov DA. 2015. Novel transcriptional regulons for autotrophic cycle genes in *Crenarchaeota*. *J. Bacteriol.* 197:2383–91

96. Li H, Johnson AD. 2010. Evolution of transcription networks—lessons from yeasts. *Curr. Biol.* 20:R746–53

97. Li L, Li Q, Rohlin L, Kim U, Salmon K, et al. 2007. Quantitative proteomic and microarray analysis of the archaeon *Methanosarcina acetivorans* grown with acetate versus methanol. *J. Proteome Res.* 6:759–71

98. Li SJ, Hua ZS, Huang LN, Li J, Shi SH, et al. 2014. Microbial communities evolve faster in extreme environments. *Sci. Rep.* 4:6205

99. Littlefield O, Korkhin Y, Sigler PB. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *PNAS* 96:13668–73

100. Liu H, Han J, Liu X, Zhou J, Xiang H. 2011. Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *J. Genet. Genom.* 38:261–69

101. Liu H, Orell A, Maes D, van Wolferen M, Lindås AC, et al. 2014. BarR, an Lrp-type transcription factor in *Sulfolobus acidocaldarius*, regulates an aminotransferase gene in a β-alanine responsive manner. *Mol. Microbiol.* 92:625–39

102. Liu T, Li Y, Wang X, Ye Q, Li H, et al. 2015. Transcriptional regulator–mediated activation of adaptation genes triggers CRISPR *de novo* spacer acquisition. *Nucleic Acids Res.* 43:1044–55

103. Lu Q, Han J, Zhou L, Coker JA, DasSarma P, et al. 2008. Dissection of the regulatory mechanism of a heat-shock responsive promoter in Haloarchaea: a new paradigm for general transcription factor directed archaeal gene regulation. *Nucleic Acids Res.* 36:3031–42

104. Lundgren M, Bernander R. 2007. Genome-wide transcription map of an archaeal cell cycle. *PNAS* 104:2939–44

105. Ma HW, Kumar B, Ditges U, Gunzer F, Buer J, Zeng AP. 2004. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res.* 32:6643–49

106. Macneil LT, Walhout AJ. 2011. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* 21:645–57

107. Maeder DL, Anderson I, Brettin TS, Bruce DC, Gilna P, et al. 2006. The *Methanosarcina barkeri* genome: comparative analysis with *Methanosarcina acetivorans* and *Methanosarcina mazei* reveals extensive rearrangement within methanosarcinal genomes. *J. Bacteriol.* 188:7922–31

108. Maeder DL, Weiss RB, Dunn DM, Cherry JL, Gonzalez JM, et al. 1999. Divergence of the hyperthermophilic archaea *Pyrococcus furiosus* and *P. horikoshii* inferred from complete genomic sequences. *Genetics* 152:1299–305

109. Makarova KS, Wolf YI, Koonin EV. 2015. Archaeal clusters of orthologous genes (arCOGs): an update and application for analysis of shared features between *Thermococcales*, *Methanococcales*, and *Methanobacteriales*. *Life* 5:818–40

110. Marbach D, Costello JC, Kuffner R, Vega NM, Prill RJ, et al. 2012. Wisdom of crowds for robust gene network inference. *Nat. Methods* 9:796–804

111. Martinez-Pastor M, Lancaster WA, Tonner PD, Adams MWW, Schmid AK. 2017. A transcription network of interlocking positive feedback loops maintains intracellular iron balance in archaea. *Nucleic Acids Res.* 45:9990–10001

112. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. 2002. Network motifs: simple building blocks of complex networks. *Science* 298:824–27

113. 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.* 187:1659–67

114. Nagy J, Grohmann D, Cheung ACM, Schulz S, Smollett K, et al. 2015. Complete architecture of the archaeal RNA polymerase open complex from single-molecule FRET and NPS. *Nature Commun.* 6:6161

115. Naor A, Lapierre P, Mevarech M, Papke RT, Gophna U. 2012. Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* 22:1444–48

116. Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, et al. 2000. Genome sequence of *Halobacterium* species NRC-1. *PNAS* 97:12176–81

117. Nguyen-Duc T, Peeters E, Muyldermans S, Charlier D, Hassanzadeh-Ghassabeh G. 2013a. Nanobody®-based chromatin immunoprecipitation/micro-array analysis for genome-wide identification of transcription factor DNA binding sites. *Nucleic Acids Res.* 41:e59

118. Nguyen-Duc T, van Oeffelen L, Song N, Hassanzadeh-Ghassabeh G, Muyldermans S, et al. 2013b. The genome-wide binding profile of the *Sulfolobus solfataricus* transcription factor Ss-LrpB shows binding events beyond direct transcription regulation. *BMC Genom.* 14:828

119. Nitzan M, Fechter P, Peer A, Altuvia Y, Bronesky D, et al. 2015. A defense–offense multi-layered regulatory switch in a pathogenic bacterium. *Nucleic Acids Res.* 43:1357–69

120. Offre P, Spang A, Schleper C. 2013. Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67:437–57

121. Okamura H, Yokoyama K, Koike H, Yamada M, Shimowasa A, et al. 2007. A structural code for discriminating between transcription signals revealed by the feast/famine regulatory protein DM1 in complex with ligands. *Structure* 15:1325–38

122. Oren A. 2008. Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2

123. Oren A. 2014. Halophilic archaea on Earth and in space: growth and survival under extreme conditions. *Philos. Trans. R. Soc. A* 372:20140194

124. Ouhammouch M, Dewhurst RE, Hausner W, Thomm M, Geiduschek EP. 2003. Activation of archaeal transcription by recruitment of the TATA-binding protein. *PNAS* 100:5097–102

125. Ouhammouch M, Geiduschek EP. 2005. An expanding family of archaeal transcriptional activators. *PNAS* 102:15423–28

126. Peeters E, Albers SV, Vassart A, Driessen AJ, Charlier D. 2009. Ss-LrpB, a transcriptional regulator from *Sulfolobus solfataricus*, regulates a gene cluster with a pyruvate ferredoxin oxidoreductase–encoding operon and permease genes. *Mol. Microbiol.* 71:972–88

127. Peeters E, Charlier D. 2010. The Lrp family of transcription regulators in archaea. *Archaea* 2010:750457

128. Peeters E, Peixeiro N, Sezonov G. 2013. *Cis*-regulatory logic in archaeal transcription. *Biochem. Soc. Trans.* 41:326–31

129. Peeters E, Thia-Toong TL, Gigot D, Maes D, Charlier D. 2004. Ss-LrpB, a novel Lrp-like regulator of *Sulfolobus solfataricus* P2, binds cooperatively to three conserved targets in its own control region. *Mol. Microbiol.* 54:321–36

130. Peng N, Xia Q, Chen Z, Liang YX, She Q. 2009. An upstream activation element exerting differential transcriptional activation on an archaeal promoter. *Mol. Microbiol.* 74:928–39

131. Perez JC, Groisman EA. 2009. Evolution of transcriptional regulatory circuits in bacteria. *Cell* 138:233–44

132. Perez-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.* 27:1449–59

133. Peters JM, Colavin A, Shi H, Czarny TL, Larson MH, et al. 2016. A comprehensive, CRISPR-based functional analysis of essential genes in bacteria. *Cell* 165:1493–506

134. Peterson JR, Labhsetwar P, Ellermeier JR, Kohler PR, Jain A, et al. 2014. Towards a computational model of a methane producing archaeum. *Archaea* 2014:898453

135. Pfeifer F, Blaseio U. 1990. Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. *Nucleic Acids Res.* 18:6921–25

136. Pfeiffer F, Broicher A, Gillich T, Klee K, Mejia J, et al. 2008a. Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* 190:281–99

137. Pfeiffer F, Schuster SC, Broicher A, Falb M, Palm P, et al. 2008b. Evolution in the laboratory: the genome of *Halobacterium salinarum* strain R1 compared to that of strain NRC-1. *Genomics* 91:335–46

138. Plaisier CL, Lo FY, Ashworth J, Brooks AN, Beer KD, et al. 2014. Evolution of context dependent regulation by expansion of feast/famine regulatory proteins. *BMC Syst. Biol.* 8:122

139. Prasse D, Förstner KU, Jäger D, Backofen R, Schmitz RA. 2017. sRNA$_{154}$ a newly identified regulator of nitrogen fixation in *Methanosarcina mazei* strain Gö1. *RNA Biol.* 15:1–5

140. Price MN, Dehal PS, Arkin AP. 2008. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* 9:R4

141. Ptashne M, Gann A. 2002. *Genes and Signals*. Cold Spring Harbor, NY: Cold Spring Harb. Lab.

142. Qi Q, Ito Y, Yoshimatsu K, Fujiwara T. 2016. Transcriptional regulation of dimethyl sulfoxide respiration in a haloarchaeon, *Haloferax volcanii*. *Extremophiles* 20:27–36

143. Raymann K, Brochier-Armanet C, Gribaldo S. 2015. The two-domain tree of life is linked to a new root for the Archaea. *PNAS* 112:6670–75

144. Reichelt R, Gindner A, Thomm M, Hausner W. 2016. Genome-wide binding analysis of the transcriptional regulator TrmBL1 in *Pyrococcus furiosus*. *BMC Genom.* 17:40

145. Reichlen MJ, Murakami KS, Ferry JG. 2010. Functional analysis of the three TATA binding protein homologs in *Methanosarcina acetivorans*. *J. Bacteriol.* 192:1511–17

146. Reichlen MJ, Vepachedu VR, Murakami KS, Ferry JG. 2012. MreA functions in the global regulation of methanogenic pathways in *Methanosarcina acetivorans*. *mBio* 3:e00189-12

147. Reiss DJ, Baliga NS, Bonneau R. 2006. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinform.* 7:280

148. 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

149. Renfrow MB, Naryshkin N, Lewis LM, Chen HT, Ebright RH, Scott RA. 2004. Transcription factor B contacts promoter DNA near the transcription start site of the archaeal transcription initiation complex. *J. Biol. Chem.* 279:2825–31

150. Rodionov DA, Leyn SA, Li X, Rodionova IA. 2017. A novel transcriptional regulator related to thiamine phosphate synthase controls thiamine metabolism genes in Archaea. *J. Bacteriol.* 199:e00743-16

151. Rodionova IA, Vetting MW, Li X, Almo SC, Osterman AL, Rodionov DA. 2017. A novel bifunctional transcriptional regulator of riboflavin metabolism in Archaea. *Nucleic Acids Res.* 45:3785–99

152. Rutherford JC, Jaron S, Winge DR. 2003. Aft1p and Aft2p mediate iron-responsive gene expression in yeast through related promoter elements. *J. Biol. Chem.* 278:27636–43

153. Sapienza C, Rose MR, Doolittle WF. 1982. High-frequency genomic rearrangements involving archae-bacterial repeat sequence elements. *Nature* 299:182–85

154. Sarmiento F, Mrazek J, Whitman WB. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *PNAS* 110:4726–31

155. 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.* 39:2519–33

156. Schmid AK, Reiss DJ, Kaur A, Pan M, King N, et al. 2007. The anatomy of microbial cell state transitions in response to oxygen. *Genome Res.* 17:1399–413

157. 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

158. Schulz S, Gietl A, Smollett K, Tinnefeld P, Werner F, Grohmann D. 2016. TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *PNAS.* 113:E1816–25

159. Schut GJ, Brehm SD, Datta S, Adams MW. 2003. Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J. Bacteriol.* 185:3935–47

160. Schwaiger R, Schwarz C, Furtwangler K, Tarasov V, Wende A, Oesterhelt D. 2010. Transcriptional control by two leucine-responsive regulatory proteins in *Halobacterium salinarum* R1. *BMC Mol. Biol.* 11:40

161. Seckbach J. 2013. *Polyextremophiles: Life Under Multiple Forms of Stress*. Dordrecht, Neth./New York: Springer

162. Seitzer P, Wilbanks EG, Larsen DJ, Facciotti MT. 2012. A Monte Carlo–based framework enhances the discovery and interpretation of regulatory sequence motifs. *BMC Bioinform.* 13:317

163. 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 Genom.* 13:351

164. She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, et al. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *PNAS* 98:7835–40

165. Shen-Orr SS, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31:64–68

166. Smollett K, Blombach F, Reichelt R, Thomm M, Werner F. 2017. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nature Microbiol.* 2:17021

167. Song N, Nguyen Duc T, van Oeffelen L, Muyldermans S, Peeters E, Charlier D. 2013. Expanded target and cofactor repertoire for the transcriptional activator LysM from *Sulfolobus*. *Nucleic Acids Res.* 41:2932–49

168. Sorrells TR, Johnson AD. 2015. Making sense of transcription networks. *Cell* 161:714–23

169. Spang A, Saw JH, Jørgensen SL, Zaremba-Niedzwiedzka K, Martijn J, et al. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521:173–79

170. Stachler AE, Marchfelder A. 2016. Gene repression in haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats)-Cas I-B system. *J. Biol. Chem.* 291:15226–42

171. Storz G, Altuvia S, Wassarman KM. 2005. An abundance of RNA regulators. *Annu. Rev. Biochem.* 74:199–217

172. Sun J, Klein A. 2004. A lysR-type regulator is involved in the negative regulation of genes encoding selenium-free hydrogenases in the archaeon *Methanococcus voltae*. *Mol. Microbiol.* 52:563–71

173. Tapias A, Leplat C, Confalonieri F. 2009. Recovery of ionizing-radiation damage after high doses of γ ray in the hyperthermophilic archaeon *Thermococcus gammatolerans*. *Extremophiles* 13:333–43

174. Tebbe A, Schmidt A, Konstantinidis K, Falb M, Bisle B, et al. 2009. Life-style changes of a halophilic archaeon analyzed by quantitative proteomics. *Proteomics* 9:3843–55

175. 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.* 93:1172–82

176. 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

177. Todor H, Sharma K, Pittman AM, Schmid AK. 2013. Protein–DNA binding dynamics predict transcriptional response to nutrients in archaea. *Nucleic Acids Res.* 41:8546–58

178. Tonner PD, Darnell CL, Engelhardt BE, Schmid AK. 2017. Detecting differential growth of microbial populations with Gaussian process regression. *Genome Res.* 27:320–33

179. 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

180. Turkarslan S, Reiss DJ, Gibbins G, Su WL, Pan M, et al. 2011. Niche adaptation by expansion and reprogramming of general transcription factors. *Mol. Syst. Biol.* 7:554

181. Van PT, Schmid AK, King NL, Kaur A, Pan M, et al. 2008. *Halobacterium salinarum* NRC-1 PeptideAtlas: toward strategies for targeted proteomics and improved proteome coverage. *J. Proteome Res.* 7:3755–64

182. van de Werken HJ, Verhees CH, Akerboom J, de Vos WM, van der Oost J. 2006. Identification of a glycolytic regulon in the archaea *Pyrococcus* and *Thermococcus*. *FEMS Microbiol. Lett.* 260:69–76

183. van Wageningen S, Kemmeren P, Lijnzaad P, Margaritis T, Benschop JJ, et al. 2010. Functional overlap and regulatory links shape genetic interactions between signaling pathways. *Cell* 143:991–1004

184. Veit K, Ehlers C, Ehrenreich A, Salmon K, Hovey R, et al. 2006. Global transcriptional analysis of *Methanosarcina mazei* strain Gö1 under different nitrogen availabilities. *Mol. Genet. Genom.* 276:41–55

185. Wagner M, Wagner A, Ma X, Kort JC, Ghosh A, et al. 2014. Investigation of the *malE* promoter and MalR, a positive regulator of the maltose regulon, for an improved expression system in *Sulfolobus acidocaldarius*. *Appl. Environ. Microbiol.* 80:1072–81

186. Weidenbach K, Ehlers C, Kock J, Ehrenreich A, Schmitz RA. 2008a. Insights into the NrpR regulon in *Methanosarcina mazei* Gö1. *Arch. Microbiol.* 190:319–32

187. Weidenbach K, Ehlers C, Kock J, Schmitz RA. 2010. NrpRII mediates contacts between NrpRI and general transcription factors in the archaeon *Methanosarcina mazei* Gö1. *FEBS J.* 277:4398–411

188. Weidenbach K, Ehlers C, Schmitz RA. 2014. The transcriptional activator NrpA is crucial for inducing nitrogen fixation in *Methanosarcina mazei* Gö1 under nitrogen-limited conditions. *FEBS J.* 281:3507–22

189. Weidenbach K, Gloer J, Ehlers C, Sandman K, Reeve JN, Schmitz RA. 2008b. Deletion of the archaeal histone in *Methanosarcina mazei* Gö1 results in reduced growth and genomic transcription. *Mol. Microbiol.* 67:662–71

190. Werner F, Grohmann D. 2011. Evolution of multisubunit RNA polymerases in the three domains of life. *Nat. Rev. Microbiol.* 9:85–98

191. Whitehead K, Kish A, Pan M, Kaur A, Reiss DJ, et al. 2006. An integrated systems approach for understanding cellular responses to γ radiation. *Mol. Syst. Biol.* 2:47

192. Wilbanks EG, Larsen DJ, Neches RY, Yao AI, Wu CY, et al. 2012. A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq. *Nucleic Acids Res.* 40:e74

193. Williams E, Lowe TM, Savas J, DiRuggiero J. 2007. Microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus* exposed to γ irradiation. *Extremophiles* 11:19–29

194. Wohlbach DJ, Thompson DA, Gasch AP, Regev A. 2009. From elements to modules: regulatory evolution in *Ascomycota* fungi. *Curr. Opin. Genet. Dev.* 19:571–78

195. Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R. 2010. A single-base resolution map of an archaeal transcriptome. *Genome Res.* 20:133–41

196. Xia Q, Hendrickson EL, Zhang Y, Wang T, Taub F, et al. 2006. Quantitative proteomics of the archaeon *Methanococcus maripaludis* validated by microarray analysis and real time PCR. *Mol. Cell Proteom.* 5:868–81

197. Yoon SH, Reiss DJ, Bare JC, Tenenbaum D, Pan M, et al. 2011. Parallel evolution of transcriptome architecture during genome reorganization. *Genome Res.* 21:1892–904

198. Yoon SH, Turkarslan S, Reiss DJ, Pan M, Burn JA, et al. 2013. A systems level predictive model for global gene regulation of methanogenesis in a hydrogenotrophic methanogen. *Genome Res.* 23:1839–51

199. Yu H, Gerstein M. 2006. Genomic analysis of the hierarchical structure of regulatory networks. *PNAS* 103:14724–31

200. Zaigler A, Schuster SC, Soppa J. 2003. Construction and usage of a onefold-coverage shotgun DNA microarray to characterize the metabolism of the archaeon *Haloferax volcanii. Mol. Microbiol.* 48:1089–105

201. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Backstrom D, Juzokaite L, et al. 2017. *Asgard* archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541:353–58

202. Zebec Z, Manica A, Zhang J, White MF, Schleper C. 2014. CRISPR-mediated targeted mRNA degradation in the archaeon *Sulfolobus solfataricus. Nucleic Acids Res.* 42:5280–88

203. Zhao B, Yan Y, Chen S. 2014. How could haloalkaliphilic microorganisms contribute to biotechnology? *Can. J. Microbiol.* 60:717–27

204. Zhu Y, Kumar S, Menon AL, Scott RA, Adams MW. 2013. Regulation of iron metabolism by *Pyrococcus furiosus. J. Bacteriol.* 195:2400–7