Programmable and portable CRISPR-Cas transcriptional activation in bacteria

Hsing-I Ho[1], Jennifer Fang[2], Jacky Cheung[3], Harris H. Wang[1,4*]

Affiliations

[1]Department of Systems Biology, Columbia University, New York, NY, USA.

[2]Department of Biology, Columbia University, New York, NY, USA.

[3]Department of Computer Science and Biology, Columbia University, New York, NY, USA

[4]Department of Pathology and Cell Biology, Columbia University, New York, NY, USA

*Correspondence to: hw2429@columbia.edu

ABSTRACT

Programmable gene activation enables fine-tuned regulation of endogenous and synthetic gene circuits to control cellular behavior. While CRISPR-Cas-mediated gene activation have been extensively developed for eukaryotic systems, similar strategies have been difficult to implement in bacteria. Here, we present a generalizable platform for screening and selection of functional bacterial CRISPR-Cas transcription activators. Using this platform, we identified a novel CRISPR activator, dCas9-AsiA, that could activate gene expression by up to 200-fold across genomic and plasmid targets with diverse promoters after directed evolution. The evolved dCas9-AsiA can simultaneously mediate activation and repression of bacterial regulons in *E. coli*. We further identified hundreds of promoters with varying basal expression that could be induced by dCas9-AsiA, which provides a rich resource of genetic parts for inducible gene activation. Finally, we show that dCas9-AsiA can be ported to other bacteria of clinical and bioindustrial relevance, thus enabling bacterial CRISPRa in more application areas. This work expands the toolbox for programmable gene regulation in bacteria and provides a useful resource for future engineering of other bacterial CRISPR-based gene regulators.

INTRODUCTION

Transcriptional regulation governs almost every cellular process fundamental to life. In response to cellular or external signals, transcription factors (TFs) in the cell interact with specific DNA sequences to mediate gene activation or repression. A potential path for cellular engineering therefore is the rewiring of transcriptional factors to alter gene regulatory networks. Programmable transcriptional activation and repression in principle offers on-demand control of specific biological processes without the need to permanently alter the genome of a cell. As such, significant past efforts have been devoted to developing synthetic transcription activators by fusing DNA-binding proteins with transcription effector domains to recruit the RNA polymerase (RNAP) complex (Dove & Hochschild, 1998; Joung et al, 2000). Unfortunately, these past synthetic TFs generally recognize only predefined DNA sequences and are difficult to reprogram to target other sequences, which greatly limits their utility for transcriptional regulation of diverse endogenous and engineered gene regulatory networks.

With the recent discovery of new DNA-binding proteins such as Zinc-fingers TF (Beerli & Barbas III, 2002), transcription activator-like (TAL) effectors (Joung & Sander, 2013) and CRISPR-Cas systems (Mali et al, 2013), there are now opportunities to develop next-generation synthetic transcription factors with greater activity and programmability. The Cas9 protein, a member of a large class of RNA-guided DNA nucleases, has emerged over the past several years as a promising system for building synthetic TFs (Bikard et al, 2013; Qi et al, 2013). Cas9 utilizes a short guide RNA (gRNA) and a protospacer adjacent motif (PAM) sequence on the target DNA to bind a defined sequence based on RNA-DNA basepairing and for cleavage of the target DNA sequence (Sternberg et al, 2014). Inactivating Cas9 by mutating the catalytic residues in the nuclease domains results in a dead Cas9 (dCas9) that functions solely as a DNA-binding protein. Transcriptional effectors such as activation or repression domains can then be linked to different parts of the dCas9 complex (e.g. dCas9 or gRNA) to enable programmable and targeted transcriptional repression (e.g. CRISPRi (Qi et al, 2013)) or activation (e.g. CRISPRa (Maeder et al, 2013; Perez-Pinera et al, 2013)). While a variety of CRISPRi systems have been successfully demonstrated in bacteria and eukaryotes (Bikard et al, 2013; Qi et al, 2013) and many mammalian CRISPRa approaches exist(Chavez et al, 2016), far fewer successful examples of bacterial CRISPRa have been shown .

In bacteria, sigma factors play a pivotal role in transcriptional initiation machinery (Browning & Busby, 2016). Sigma factors interact with the core RNAP enzyme ($\alpha_2\beta\beta'\omega$) complex and bind to specific promoter sequences. Different types of sigma factors compete for the common pool of core enzymes in bacterial cells and recruit them to corresponding

promoters (Browning & Busby, 2004). Transcription factors further function in trans on the holoenzyme and regulating gene expression. Transcription activators usually bind with specific components of the RNAP complex and direct the complex to the target promoter region (Browning & Busby, 2016). However, most transcriptional activation domains in bacteria are not well-characterized and have not been demonstrated to mediate transcriptional activation when coupled synthetically with DNA binding domains. To our knowledge, only three efforts have been described for engineering bacterial transcriptional activation using CRISPR-Cas. In the first study by Bikard et al (Bikard et al, 2013), dCas9 was fused to the RNAP ω subunit, which interacts with the RNA polymerase to mediate gene activation. However, this CRISPRa system could only function in the ω subunit knockout background (Bikard et al, 2013). Deletion of *rpoZ* that encodes ω subunit is known to lead to altered basal transcription profile and fitness defects (Chatterji et al, 2007; Weiss et al, 2017). Another study used bacterial enhancer binding proteins (bEBPs) as the fused activation domain in a similar approach (Liu et al, 2019), but the bEBPs-mediated CRISPRa is only compatible with σ54 promoters and the deletion of endogenous bEBPs is required. Both systems require modification of the bacterial genome, which limits the portability to genetically tractable microbes. Another study by Dong et al (Dong et al, 2018) used a scaffold RNA (scRNA) containing the gRNA and a MS2 domain, which could bind to a MCP-fused transcription factor SoxS to enable dCas9-mediated transcriptional activation. This system exhibited higher activity after further optimization, but has a narrow targetable region within the promoters. Furthermore, most of these prior studies have only demonstrated CRISPRa in laboratory *E. coli* strains and activity in other bacteria is unknown.

To overcome these challenges, we devised a high-throughput platform to screen and select for bacterial CRISPR-Cas transcriptional activators (CasTAs). We first screened a number of natural bacterial and phage regulatory effectors and identified a phage protein that induced gene activation when fused to dCas9. We characterized the targeting window of this CasTA and further performed rounds of directed evolution using our screening platform to yield higher functioning variants, which can mediate both CRISPRi and CRISPRa of genomic and plasmid targets. We then applied this activator system to a metagenomic promoter library mined from diverse bacteria to build a library of CasTA-inducible promoters of varying basal and induced expression levels that are useful as a resource for the synthetic biology research community. Finally, we describe the successful transfer of our CRISPRa system to other bacterial species of clinical and bioindustrial importance, thus expanding utility to more application areas.

RESULTS

A screening-selection platform for bacterial CRISPRa development

To expedite the discovery of bacterial CRISPRa components, we developed a screening-selection platform in *Escherichia coli* to identify candidate dCas9-mediated transcription activators. In our CRISPRa design, a *S. pyogenes* dCas9 (Qi et al, 2013) is C-terminally fused with candidate transcription activation domains or proteins via a previously described flexible peptide linker (SAGGGGSGGGGS) (Chen et al, 2013). This CasTA then uses a specific gRNA to target to the regulatory region of a reporter gene for transcriptional activation, gene expression, and production of reporter products (**Fig. 1A**). As such, we separated the three essential components of the platform (i.e. dCas9-activator fusion, the guide RNA, and the reporter gene) into 3 compatible plasmids (**Fig. 1B**). The dCas9-activator was regulated by a P_{tetO} induction system with anhydrotetracycline (aTc) on a p15A medium copy plasmid, while the gRNA was expressed constitutive from a strong promoter (BBa_J23119) on a high copy ColE1 plasmid, and the reporter was placed behind a very weak promoter (BBa_J23117) on a low copy SC101 plasmid (**Fig. S1**). Since different dCas9 activators may have their own respective optimal gRNA binding windows (Bikard et al, 2013; Dong et al, 2018) and possible biases toward targetable promoter sequences (Liu et al, 2019), the screening-selection platform was designed to be highly modular to facilitate combinatorial assessment of system components. As library-scale screening for transcription activators can often be hampered by auto-activators in the population, we further employed a dual screening-selection reporter design by using both fluorescent protein and antibiotic resistance genes to eliminate potential false positive clones. We further engineered the selective reporter to contain multiple separate antibiotic genes with degradation tags (BBa_M0050) to increase the rate of turnover to reach higher stringency and specificity of the selection (see **Methods**, **Fig. S2**).

Using this platform, we first screened a list of transcriptional activator candidates, including phage proteins, transcription factors, and RNAP interacting proteins (**Table S1**), paired with different gRNAs (gRNA-H1, gRNA-H2, gRNA-H3) targeted to different spacing distances to transcriptional start site (TSS) of the reporter gene (60 bp, 85 bp, 120 bp, respectively), for potential dCas9-activators. Among the transcription activation modules screened, we found a phage protein, AsiA, that upregulated the reporter gene expression to a level comparable to the previously identified dCas9-ω activator (Bikard et al, 2013), although at a different optimal spacing distance (**Fig. 1C-D**). AsiA (Audrey Stevens' inhibitor A) is a 90 amino acid anti-σ70 protein from T4 bacteriophage that binds to the host σ70 subunit and suppresses endogenous gene expression (Orsini et al, 1993; Stevens, 1972). In combination with another T4 phage

protein, MotA, the σ70-AsiA-MotA complex specifically binds to T4 phage promoters and activates phage transcription during the T4 viral life cycle (Minakhin & Severinov, 2005).

When directly fused to dCas9 with a peptide linker, AsiA upregulated gene expression of a GFP reporter, with the magnitude of activation tunable via design of the gRNA. Transcriptional activation by dCas9-AsiA (dubbed CasTA1.0) is seen across a wide window along the target regulatory region, reaching up to 12-fold at ~200 base pairs (bp) from the TSS (**Fig. 1E**). In contrast, the optimal gRNA targeting positions for other dCas9 activators (e.g. dCas9-ω and dCas9-MS2/MCP-SoxS) is 100 bp or less from the TSS with a more narrow targetable window (Bikard et al, 2013; Dong et al, 2018), possibly suggesting a distinct mechanism of activation by dCas9-AsiA. Simultaneously targeting with multiple gRNAs can further increase transcriptional activation (**Fig. 1F**), although no synergistic enhancement was observed in contrast to eukaryotic CRISPRa systems (Maeder et al, 2013).

Based on different CRISPRa architectures that have been described in literature (Chavez et al, 2016), we further explored whether AsiA can be tethered to other parts of the dCas9 complex and reach higher activity. The MS2 hairpin RNA has been engineered in the gRNA to enable recruitment of transcription activation domains linked to a MCP domain, such as in the bacterial dCas9-MS2/MCP-SoxS system (Dong et al, 2018) and the eukaryotic Synergistic Activation Mediator (SAM) system (Konermann et al, 2015). We therefore tested CasTA-AsiA where the gRNA contains a MS2 domain in different stem loops and where AsiA is tethered to MCP (i.e. dCas9-MS2/MCP-AsiA). While the MS2 hairpins did not affect the gRNA performance, we did not find that the SAM implementation of AsiA could activate gene activation (**Fig. S3**). These results are in agreement with a previous observation that dCas9-MS2/MCP-AsiA system was not a functional activator (Dong et al, 2018). We also did not find that a G32A mutant (DNA binding disruption) (Griffith & Wolf, 2002) of the previously described SoxS activator in the dCas9-MS2/MCP-SoxS system (Dong et al, 2018) to be functional as a direct dCas9 fusion (i.e. dCas9-SoxS$_{G32A}$) (**Fig. 1C**). These results highlight potential mechanistic and performance differences between CRISPRa systems where the activation domain is directly fused to dCas9 versus tethered via the MS2-MCP system.

Directed evolution and characterization of the dCas9-AsiA transcriptional activator

To increase the dynamic range and performance of dCas9-AsiA-mediated transcriptional activation, we performed a series of directed evolution studies using our screening-selection platform. A dCas9-AsiA variant library was constructed by error-prone PCR of AsiA, with each AsiA variant having on average two randomly distributed residue mutations (**Fig. S4**). We

screened ~5x10^8 AsiA mutant variants for improved transcriptional activation on antibiotic selection plates (**Fig. 2A**, **Fig. S2**). The resulting colonies were individually isolated and plasmids encoding the dCas9-AsiA variants were extracted and transformed into cells expressing a gRNA and GFP reporter for re-validation (**Table S4**). Of 47 colonies isolated and characterized, one variant (m1.1) was found most enriched (>75% of the time), while several other variants (m1.2, m1.3) were also identified at lower frequency (**Fig 2A-B**). The most abundant variant m1.1 after selection also mediated the highest GFP activation (**Fig. 2C**). The m1.1 variant contained two key mutations on AsiA (V58I, E60K). An additional mutation (S1C) on the peptide linker was also found, which likely arose during the cloning steps of the directed evolution protocol. Interestingly, the AsiA mutations occurred within the middle of the anti-σ factor protein and are structurally away from the interface that bind to σ70 (**Fig. 2B**) (Lambert et al, 2004). AsiA binds to sigma factors through the first helix structure (residues 1 to 20) (Minakhin et al, 2001), suggesting that the mutations in m1.1 may not affect direct binding to sigma factors. This m1.1 variant significantly increased the fold in transcriptional activation to ~70 compared to ~10 fold using the wild-type AsiA (**Fig. 2C**). We then performed another round of directed evolution on m1.1 and screened for additional mutants with further improvements (**Fig. 2A**). From 10^7 variants, validation and characterization of the resulting colonies revealed an additional mutant (m2.1) to be significantly enriched in the population with >135-fold activation (**Fig. 2B-C**). The m2.1 variant contained an additional Q51R mutation, which also faced away from σ70 similar to the other m1.1 mutations.

We next explored the activation potential of dCas9-AsiA-m2.1 (CasTA2.1) for targeting promoters with different basal expression levels and at different CasTA2.1 expression levels. We observed that that transcriptional activation across weak to strong promoters reached similar saturating levels and at the same optimal gRNA targeting distance (**Fig. 2D, Fig. S5A-B**). Accordingly, the fold induction inversely correlated with the basal promoter strength. Transcriptionally or translationally enhanced the expression of CasTA1.0 or 2.1 could also increase activation of the target gene (**Fig. S5C**), thus providing different options to tuning the overall system.

Since AsiA binds and sequesters the host σ70, overexpression of AsiA may become toxic to the cell (Minakhin & Severinov, 2005). We therefore quantified the toxicity of dCas9-AsiA in our system. Overexpression of CasTA1.0 or 2.1 under aTc induction did not have significant impact on cellular growth rate beyond the basal fitness burden of dCas9 overexpression alone (**Fig. S6**). Doubling times during exponential growth were generally unaffected under CasTA overexpression, while stationary cell density was somewhat impacted.

To gain a higher resolution of the effects of CasTA on the endogenous transcriptome, we performed RNAseq on cells with CasTA1.0 and CasTA2.1, relative to ancestral control cells (**Fig. S7**). We found that CasTA2.1 mediated higher gene activation on the GFP target without loss of specificity genome-wide compared to cells with CasTA1.0 (**Fig. S7A**) or ancestral cells (**Fig. S7B**). Upon overexpression of CasTA2.1, we observed upregulation of some low-expression endogenous genes (**Fig. S7C**). These off-target gene activations may be the result of non-specific dCas9 binding to other genomic loci, which has been reported previously (Cho et al, 2018; Cui et al, 2018). This hypothesis is supported by the fact that strong off-targets (fold change >30) were regulated by not just σ70 but also other σ factors (**Fig. S7C**). Notably, the fold induction of the GFP targets was also higher under significant CasTA2.1 overexpression (**Fig. S5C**), which highlights a trade-off between higher target activation and increased off-targets in this CRISPRa system.

Utility of dCas9-AsiA for multi-gene and library-scale transcriptional regulation

To explore whether CasTA can be used to regulate endogenous genomic targets, we first inserted a GFP reporter into the genome and showed that CasTA2.1 can upregulate the expression of this chromosomal reporter (**Fig. 3A**). We then demonstrated that five genes in the genome could be upregulated (by up to 200-fold) using CasTA2.1 (**Fig. 3B**, **Table S6**). One gRNA was designed for each gene using a search window of 200±20bp from the TSS. We anticipate that optimization of gRNA designs may be necessary for different genomic targets (Bikard et al, 2013; Dong et al, 2018). We further explored whether CasTA2.1 can be used as a transcription repressor and for simultaneous CRISPRa/CRISPRi transcriptional modulation. We found that gRNAs (gRNA-H7 to gRNA-H10) positioned near the TSS or within the gene body of the target GFP reporter could efficiently inhibited gene expression using the CasTA2.1 protein, including both strands of the target DNA (**Fig. 3C**). When two different gRNAs were designed to target two reporter genes for concurrent activation and repression, we observed simultaneous CRISPRa and CRISPRi using CasTA2.1 at efficiencies similar to applying CRISPRa or CRISPRi separately (**Fig. 3D**), which highlights its potential utility for multiplexed gene modulation of regulatory networks in a single cell.

Development of complex synthetic genetic circuits requires diverse regulatory parts with tunable dynamic rage (Slusarczyk et al, 2012). However, the number of inducible promoters with defined expression ranged is currently limited for many applications in synthetic biology. We previously developed a promoter library from metagenomic sequences with varying species-specific constitutive expression levels (Johns et al, 2018). We therefore explored whether such

a constitutive promoter library could be turned into an inducible promoter library using our CRISPRa system (**Fig. 4A**). We designed two gRNAs spaced ~150bp apart targeting the constant regulatory region upstream of the variable regulatory sequences of each promoter and screened for subsets of promoters that could be upregulated by CasTA2.1. The expression level from all promoters in the library with and without CasTA2.1 was quantified by targeted RNAseq (to obtain RNA transcript for each promoter) and DNAseq (to normalize for plasmid copy numbers across the library) as previously described (Yim et al, 2019) (**Fig. S8A**, **Methods**). Of ~8,000 promoters characterized, we identified thousands of promoters that were activated by CasTA2.1 with at least one of the gRNAs (**Fig. 4B**, **S8B**). Among them, several hundred had a high level of induction (>10-fold) across 2-orders of magnitude in basal expression level (**Fig. 4C**). In general, more promoters were activated with the distal gRNA (gRNA-H23), although interestingly the proximal gRNA (gRNA-H22) also resulted in CRISPRi activity in some promoters (**Fig. S8B**). The phylogenetic origin and sequence composition of these inducible promoters were diverse, which will facilitate their use for assembly of large genetic circuits with minimal recurrent sequence motifs (**Fig. S8C**). This library of CasTA-inducible promoters greatly expands the repertoire of regulatory parts that can be activated with one or two gRNAs by CRISPRa for more complex genetic circuits in various synthetic biology applications.

Portability of dCas9-AsiA to other bacteria species

Since homologs of the T4 AsiA protein are widely found in many different phages that infect diverse bacteria (**Fig. 5A**), we hypothesized that our dCas9-AsiA system could be ported to other bacteria with greater possibility of success and minimal re-optimization. We chose two bacterial species *Salmonella enterica and Klebsiella oxytoca* of clinic and bio-industrial significance (Coburn et al, 2007; Kao et al, 2003) to test our CasTA system. Each of the three plasmids (CasTA, gRNA, reporter) was transformed into the two species. We first confirmed that dCas9 was functional in these two species by using a gRNA targeting for repression of a reporter GFP gene (i.e. CRISPRi) activity (**Fig. 5B**). We then tested CRISPRa using the CasTA1.0 and 2.1 systems with the appropriate gRNA and GFP reporter. We were surprised to find that CasTA2.1 showed significant GFP activation in both species, but CasTA1.0 did not. It is interesting to note that AsiA from *Salmonella* phage SG1 shares the same residues at positions 50-61 as the *E. coli* T4 phage, while the *Klebsiella* phage F48 had some differences especially at residues 51-53, 57, and 59, which all face away from the binding surface to σ70. Notably, residues 51-53 and 57-61 of AsiA appear to be more variable across phylogenetically diverse phages (**Fig. 5A**), which are also the key residue regions mutated in m2.1 (Q51R, V58I,

E60K) from our directed evolution experiments. In fact, some of the mutant residues in CasTA2.1 are also found in natural AsiA variants, suggesting that the mutations that we identified might mediate conserved molecular interactions leading to improved gene activation. Together, these results demonstrate that the CasTA system can be ported into other bacteria, although additional mutations may be required through directed evolution or testing of natural homologs.

DISCUSSION

CRISPRa is a powerful approach to elucidate the cellular and genetic basis of various biological phenomena in bacteria and higher organisms (Konermann et al, 2015; Liu et al, 2018). While numerous CRISPRa systems with high efficiency have been established in eukaryotes, reliable bacterial CRISPRa tools have been slow to develop. The screening-selection platform to identify active CRISPR-Cas mediated transcriptional activation established in this work provides a highly modular and portable system to engineering new bacterial CRISPRa systems for enhancing potency of dCas9-activators, modifying gRNA designs, or screening for targetable promoters. We demonstrated here that CasTA2.1 could activate endogenous genes and enable genome-scale gain-of-function studies in bacteria, which has not been shown in prior studies. Given that not all genomic targets are activated with the same potency, we anticipate that some targets may require additional gRNA optimizations to improve activation.

The capability to multiplex CRISPRa and CRISPRi on multiple targets can facilitate combinatorial screens, which has been challenging to implement in traditional genetic screens in bacteria with transposon mutagenesis libraries or metagenomic expression libraries (Hu & Coates, 2005; Yaung et al, 2015). The inducible promoter library characterized in this study constitutes a useful catalog of promoters with varying basal expression level and defined fold activation upon CRISPRa induction. This library could serve as a community resource for promoter optimization in metabolic engineering applications or construction of complex genetic circuits. Importantly, multiple promoters could be induced by the same gRNA, which provide opportunities to synchronize modulation of multiple network nodes at the same time.

Sequencing of environmental metagenomic libraries have found many biosynthetic pathways encoding biomolecules with various biotherapeutic potential (Cohen et al, 2017; Donia et al, 2014). However, the majority of these biosynthetic gene clusters remain silent when transformed into common heterologous expression hosts such as *E. coli* (Iqbal et al, 2016). As such, CRISPR activation tools could be useful to reawaken these silent gene clusters. Our results indicate that CasTA2.1 is compatible with promoters from diverse bacterial origins,

suggesting its potential utility for activating cryptic gene clusters. In addition, increasing the portability of bacterial CRISPRa systems can greatly improve our ability to manipulate non-model organisms. In combination with *in situ* microbiome engineering tools (Brophy et al, 2018; Ronda et al, 2019), we envision that CasTA and similar technologies (Peters et al, 2019) could modulate gene expression in a precise and programmable fashion across diverse bacterial communities to elucidate fundamental processes underlying microbial ecology and provide useful applications in the emerging field of microbiome engineering.

MATERIAL AND METHODS

Acknowledgements

We thank members of the Wang lab for comments and discussions on the manuscript and experiments, especially S. S. Yim and N. Johns for advice and discussions on the inducible promoter library experiments. H.H.W. acknowledges specific funding support from NSF (MCB-1453219), NIH (1U01GM110714, 1DP5OD009172), DoD ONR (N00014-15-1-2704), and the Sloan Foundation (FR2015-65795) for this work.

Author contributions

H.H. and H.H.W. developed the initial concept; H.H., J.F., and J.C. performed experiments and analyzed the results; and H.H.W. supervised the work throughout. H.H. and H.H.W. wrote the manuscript with inputs from all authors.

Conflicts of interest

The authors declare no competing financial interests.

Materials and Methods

Strains and culturing conditions

E. coli strains and other bacterial species used in the study are listed in **Table S2** and all *E. coli* strains are derived from the MG1655 parental background. Cells were grown in rich LB medium at 37 °C with agitation unless stated otherwise. For plasmid transformation, general protocols were followed, and plasmids were maintained under antibiotics selection at all times. For constructing genomic insertions, GFP expression cassette amplified from pWJ89 (Bikard et al, 2013) was cloned between multiple cloning sites of pOSIP-Kan and inserted chromosomally following the clonetegration method (St-Pierre et al, 2013). For the antibiotic selection and induction of target genes, the following concentrations were used: Carbenicillin (Carb) 50 µg/ml, Chloramphenicol (Cam) 20 µg/ml, Kanamycin (Kan) 50 µg/ml, Spectinomycin (Spec) 50 µg/ml, Bleocin (Bleo) 5 µg/ml, Anhydrotetracycline (aTc) 100 ng/ml. For induction of target genes, aTc was added to the culture at the exponential growth phase for 4 hours before cells were harvested for characterization.

Construction of plasmids

The dCas9 fusion library was constructed based on the pdCas9-bacteria plasmid (Addgene #44249). Linker sequences (SAGGGGSGGGGS) and fusion candidates were either amplified from DNA synthesized *de novo* (IDT Gblocks®) or *E. coli* genomic DNA and sub-cloned after the dCas9 sequence in the pdCad9-bacteria plasmid (Addgene #44249). All guide RNA plasmids (pgRNA-H1 to pgRNA-H21) were constructed from pgRNA-bacteria plasmid (Addgene #44251), using inverted PCR and blunt-end ligation to modify the N20 seed sequences. For dual gRNA plasmids (pgRNA-H4H5, pgRNA-H4H11), each gRNA was built separately and jointed subsequently. GFP reporter plasmids (pWJ89, pWJ96, pWJ97) were gifts from the Marraffini lab at Rockefeller University (Bikard et al, 2013). The promoter region upstream of the GFP reporter in pWJ89 was amplified for constructing other antibiotic reporter plasmids (pHH34-37). The GFP-mScarlet reporter plasmid (pHH39) was constructed by cloning the mScarlet gene from pEB2-mScarlet-I (Addgene #104007) under WJ89 promoter and joined with the constitutive GFP expression cassette from pWJ97. For screening inducible metagenomic promoter library (RS7003) (Johns et al, 2018), gRNA-H22 and gRNA-H23 expression cassettes were jointed with dCas9-AsiA_m2.1 separately, resulting pHH40, pHH41. Cloning was done by Gibson assembly if not otherwise noted in all cases. Plasmids used and associated details were listed in **Table S3**.

Development of CasTA screening platform

dCas9 fusion library, gRNAs, and reporter genes were built on 3 different compatible plasmids (dCas9: p15A, Cam resistance; gRNA: ColE1, Carb resistance; reporter: SC101, Kan resistance), so they can be transformed and propagated within the same cell (**Fig. S1**). To use a antibiotic resistance gene as a reporter, we tested different antibiotic genes and modulated degradation rate (fusion with ssrA tag: AANDENYALAA) for selective stringency (**Fig. S2A-B**). Dual selective reporters (Kan and Bleo) were constructed, which decrease the escape rate by 10 fold (**Fig. S2C-D**).

Directed evolution of dCas9-AsiA using CasTA screening platform

We mutagenized the wild-type AisA region of dCas9-AsiA using the GeneMorph II EZClone Domain Mutagenesis Kit (Agilent Technologies), following manufacture's protocol. In brief, 50 ng of parental template DNA was used for amplification with error prone DNA polymerase (Mutazyme II). Under this condition, the AsiA region contains on average ~2 nucleotides changes per variant after PCR mutagenesis (**Fig. S4**). In the first round of directed evolution, dCas9-AsiA mutant library was transformed to the cells expressing gRNA-H4 and dual selective reporters (pHH37 and pHH38). ~$5x10^8$ transformants were grown under 0.2x regular Kan concentration and 2x regular Bleo concentration. Grown colonies were harvested and propagated together with Cam selection to maintain solely the dCas9-AsiA variant plasmids. The dCas9-AsiA plasmids were subsequently extracted and transformed to cells containing pgRNA-H4 and pWJ89. Individual colony was Sanger sequenced to identify the mutations in AsiA and characterized based on GFP intensity (**Table S4**). The dCas9-AsiA_m1.1 plasmid from most abundant mutant variant were extracted and transformed to GFP reporter strain (containing pgRNA-H4 and pWJ89) again to verify fluorescent intensity (Fig. 2C). In the second directed evolution round, the dCas9_AsiA_m1.1 variant was used as a template to generate additional variants following the same conditions. The second generation of dCas9-AsiA mutant library was transformed to GFP reporter cells, containing pgRNA-H4 and pWJ89 as described above. We enriched the top 0.1% of highest GFP expression from the population of $1x10^7$ transformants using fluorescence activated cell soring (BD FACS Aria II). Post-sorted cell population was plated on selective LB again to obtain clonal colonies, and individual colony was picked for Sanger sequencing and measurement of GFP intensity.

Quantification of gene expression induced by CasTA

To examine CRISPRa on genomic targets, pdCas9-AsiA_m2.1 was transformed along with gRNA constructs (gRNA-H12 to gRNA-H21, **Table S5**) designed for each gene (**Table S6**). Cells expressing dCas9-AsiA_m2.1 and a non-specific gRNA (gRNA-H4) were used as controls. After CRISPRa induction with 100 ng/ml aTc, cells were harvested for RNA extraction following the RNAsnap protocol (Stead et al, 2012). After column purification (RNA Clean & Concentrator Kits, Zymo Research), total RNA was reverse transcribed into cDNA using random hexmers (SuperScript III Reverse Transcriptase, Invitrogen). Quantitative PCR was performed on each sample with gene specific primers (**Table S6**) using the KAPA SYBR FAST qPCR Master Mix (Kapa Biosystems). The *rrsA* gene was selected as the housekeeping gene to normalize expression between samples.

For whole-transcriptome analysis of CRISPRa specificity, we extracted total RNA from the samples as described above and depleted rRNAs using Ribo-Zero rRNA removal-Bacteria kit (Illumina). RNA libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit (New England BioLabs) and sequenced on the Illumina NextSeq platform (Mid-Output Kit, 150 cycles). The raw reads were aligned to the reference genome (BW25113) using Bowtie 2 (Langmead & Salzberg, 2012), and the read counts of each gene were quantified by HTseq (Anders et al, 2015). Expression level of individual gene was normalized by total read counts within each sample.

Screening for CRISPRa mediated inducible promoters

Metagenomic promoter library (RS7003) was derived from Johns et al (Johns et al, 2018). About 8,000 regulatory elements (**Table S7**) was transformed to cells expressing dCas9-AsiA_m2.1 and either gRNA-H22, gRNA-H23 or genomic targeting gRNA-H24 (**Table S5**). After CRISPRa induction, four biological replicates were harvested to measure promoter activity. A constitutive promoter without CRISPRa induction (ID:14076, **Table S7**) was spiked in the cell populations for normalizing expression levels between samples. Total RNA was extracted and purified as previously described. Gene specific primers were used for cDNA generation (Maxima reverse transcriptase, Thermo Scientific), and RNA sequencing library was prepared by ligation with the common adaptor primer for downstream sequencing (Yim et al, 2019) To quantify abundance of each promoter in the library, plasmid DNA from each sample was also extracted using PrepGem bacteria kit (Zygem) and used to generate a DNA amplicon sequencing library. Both RNA and DNA library were sequenced on the Illumina NextSeq platform (Mid-output kit, 300 cycles).

Sequencing reads from DNA and RNA library were merged by BBmerge and filtered out low quality reads (Bushnell et al, 2017). Custom pipeline that previous described (Yim et al, 2019) were adopted to identify sequencing reads corresponding to each promoter. Expression level of each promoter was quantified by determining the ratio of RNA abundance over DNA abundance. To compare across samples, expression levels were normalized to the same spiked-in control promoter in each sample. Fold change in CRISPRa induced gene expression was calculated by dividing by the reporter expression of control cells containing dCas9-AsiA_m2.1 and a genomic targeting gRNA-H24.

REFERENCES

Beerli RR, Barbas III CF (2002) Engineering polydactyl zinc-finger transcription factors. *Nature biotechnology* **20:** 135

Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic acids research* **41:** 7429-7437

Brophy JAN, Triassi AJ, Adams BL, Renberg RL, Stratis-Cullum DN, Grossman AD, Voigt CA (2018) Engineered integrative and conjugative elements for efficient and inducible DNA transfer to undomesticated bacteria. *Nature Microbiology* **3:** 1043-1053

Browning DF, Busby SJW (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology* **2:** 57-65

Browning DF, Busby SJW (2016) Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology* **14:** 638

Chatterji D, Ogawa Y, Shimada T, Ishihama A (2007) The role of the omega subunit of RNA polymerase in expression of the relA gene in Escherichia coli. *FEMS Microbiology Letters* **267:** 51-55

Chavez A, Tuttle M, Pruitt BW, Ewen-Campen B, Chari R, Ter-Ovanesyan D, Haque SJ, Cecchi RJ, Kowal EJK, Buchthal J, Housden BE, Perrimon N, Collins JJ, Church G (2016) Comparison of Cas9 activators in multiple species. *Nature methods* **13:** 563-567

Chen X, Zaro JL, Shen W-C (2013) Fusion protein linkers: property, design and functionality. *Advanced drug delivery reviews* **65:** 1357-1369

Cho S, Choe D, Lee E, Kim SC, Palsson B, Cho B-K (2018) High-Level dCas9 Expression Induces Abnormal Cell Morphology in Escherichia coli. *ACS Synthetic Biology* **7:** 1085-1094

Coburn B, Grassl GA, Finlay BB (2007) Salmonella, the host and disease: a brief review. *Immunology & Cell Biology* **85:** 112-118

Cohen LJ, Esterhazy D, Kim S-H, Lemetre C, Aguilar RR, Gordon EA, Pickard AJ, Cross JR, Emiliano AB, Han SM (2017) Commensal bacteria make GPCR ligands that mimic human signalling molecules. *Nature* **549:** 48

Cui L, Vigouroux A, Rousset F, Varet H, Khanna V, Bikard D (2018) A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. *Nature Communications* **9:** 1912

Dong C, Fontana J, Patel A, Carothers JM, Zalatan JG (2018) Synthetic CRISPR-Cas gene activators for transcriptional reprogramming in bacteria. *Nature Communications* **9:** 2489

Donia MS, Cimermancic P, Schulze CJ, Brown LCW, Martin J, Mitreva M, Clardy J, Linington RG, Fischbach MA (2014) A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158:** 1402-1414

Dove SL, Hochschild A (1998) Conversion of the ω subunit of Escherichia coli RNA polymerase into a transcriptional activator or an activation target. *Genes & development* **12:** 745-754

Griffith KL, Wolf RE (2002) A Comprehensive Alanine Scanning Mutagenesis of the Escherichia coli Transcriptional Activator SoxS: Identifying Amino Acids Important for DNA Binding and Transcription Activation. *Journal of Molecular Biology* **322:** 237-257

Hu Y, Coates ARM (2005) Transposon mutagenesis identifies genes which control antimicrobial drug tolerance in stationary-phase Escherichia coli. *FEMS Microbiology Letters* **243:** 117-124

Iqbal HA, Low-Beinart L, Obiajulu JU, Brady SF (2016) Natural product discovery through improved functional metagenomics in Streptomyces. *Journal of the American Chemical Society* **138:** 9341-9344

Isalan M, Lemerle C Fau - Michalodimitrakis K, Michalodimitrakis K Fau - Horn C, Horn C Fau - Beltrao P, Beltrao P Fau - Raineri E, Raineri E Fau - Garriga-Canut M, Garriga-Canut M Fau - Serrano L, Serrano L Evolvability and hierarchy in rewired bacterial gene networks.

Johns NI, Gomes AL, Yim SS, Yang A, Blazejewski T, Smillie CS, Smith MB, Alm EJ, Kosuri S, Wang HH (2018) Metagenomic mining of regulatory elements enables programmable species-selective gene expression. *Nature methods* **15:** 323

Joung JK, Ramm EI, Pabo CO (2000) A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proceedings of the National Academy of Sciences* **97:** 7382-7387

Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14:** 49-55

Kao CM, Liu JK, Lou HR, Lin CS, Chen SC (2003) Biotransformation of cyanide to methane and ammonia by Klebsiella oxytoca. *Chemosphere* **50:** 1055-1061

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517:** 583

Lambert LJ, Wei Y, Schirf V, Demeler B, Werner MH (2004) T4 AsiA blocks DNA recognition by remodeling σ70 region 4. *The EMBO Journal* **23:** 2952-2962

Liu Y, Wan X, Wang B (2019) Engineered CRISPRa enables programmable eukaryote-like gene activation in bacteria. *Nature communications* **10:** 1-16

Liu Y, Yu C, Daley TP, Wang F, Cao WS, Bhate S, Lin X, Still C, II, Liu H, Zhao D, Wang H, Xie XS, Ding S, Wong WH, Wernig M, Qi LS (2018) CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23:** 758-771.e758

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK (2013) CRISPR RNA–guided activation of endogenous human genes. *Nature methods* **10:** 977

Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nature Methods* **10:** 957

Minakhin L, Camarero JA, Holford M, Parker C, Muir TW, Severinov K (2001) Mapping the molecular interface between the σ70 subunit of E. coli RNA polymerase and T4 AsiA11Edited by M. Gottesman. *Journal of Molecular Biology* **306:** 631-642

Minakhin L, Severinov K (2005) Transcription regulation by bacteriophage T4 AsiA. *Protein expression and purification* **41:** 1-8

Orsini G, Ouhammouch M, Le Caer J, Brody E (1993) The asiA gene of bacteriophage T4 codes for the anti-sigma 70 protein. *Journal of bacteriology* **175:** 85-93

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW (2013) RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature methods* **10:** 973

Peters JM, Koo B-M, Patino R, Heussler GE, Hearne CC, Qu J, Inclan YF, Hawkins JS, Lu CH, Silvis MR (2019) Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nature microbiology* **4:** 244

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152:** 1173-1183

Ronda C, Chen SP, Cabral V, Yaung SJ, Wang HH (2019) Metagenomic engineering of the mammalian gut microbiome in situ. *Nature methods* **16:** 167

Slusarczyk AL, Lin A, Weiss R (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nature Reviews Genetics* **13:** 406-420

Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507:** 62

Stevens A (1972) New small polypeptides associated with DNA-dependent RNA polymerase of Escherichia coli after infection with bacteriophage T4. *Proc Natl Acad Sci U S A* **69:** 603-607

Weiss A, Moore BD, Tremblay MH, Chaput D, Kremer A, Shaw LN (2017) The ω subunit governs RNA polymerase stability and transcriptional specificity in Staphylococcus aureus. *Journal of bacteriology* **199:** e00459-00416

Yaung SJ, Deng L, Li N, Braff JL, Church GM, Bry L, Wang HH, Gerber GK (2015) Improving microbial fitness in the mammalian gut by in vivo temporal functional metagenomics. *Molecular systems biology* **11:** 788-788

Yim SS, Johns NI, Park J, Gomes ALC, McBee RM, Richardson M, Ronda C, Chen SP, Garenne D, Noireaux V, Wang HH (2019) Multiplex transcriptional characterizations across diverse bacterial species using cell-free systems. *Molecular Systems Biology* **15:** e8875

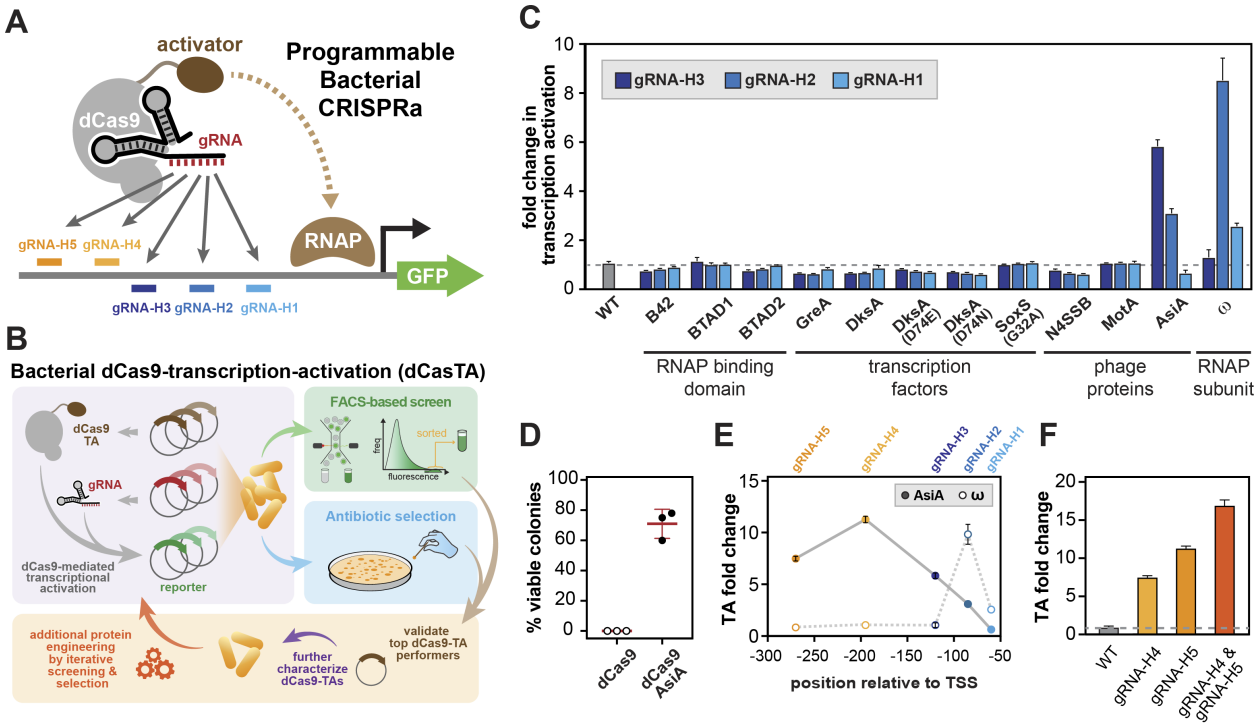

Figure 1. A high-throughput platform to identify and engineer bacterial CRISPR-Cas transcription activators (CasTAs). (A) General strategy for bacterial CRISPRa using a dCas9 fused with a transcriptional activator, a targeting gRNA, and reporter genes. **(B)** System components are constructed in three compatible plasmids. CasTA candidates could be cross validated through GFP and antibiotic resistance gene reporters. **(C)** Fold activation of CasTA candidates using different gRNAs targeting to different locations of the GFP reporter gene compared to a strain without CasTA. **(D)** Survival of cells containing upregulated antibiotic resistance reporter induced by dCas9 or dCas9-AisA with gRNA-H3 under kanamycin selection (2.5 μg/ml). **(E)** Different gRNAs were paired with dCas9-AsiA and dCas9-ω to profile the optimal gRNA binding distance. **(F)** Comparing single and multiple gRNA with dCas9-AsiA. Data in all panels are 3-5 biological replicates with +/- standard error of mean (SEM).

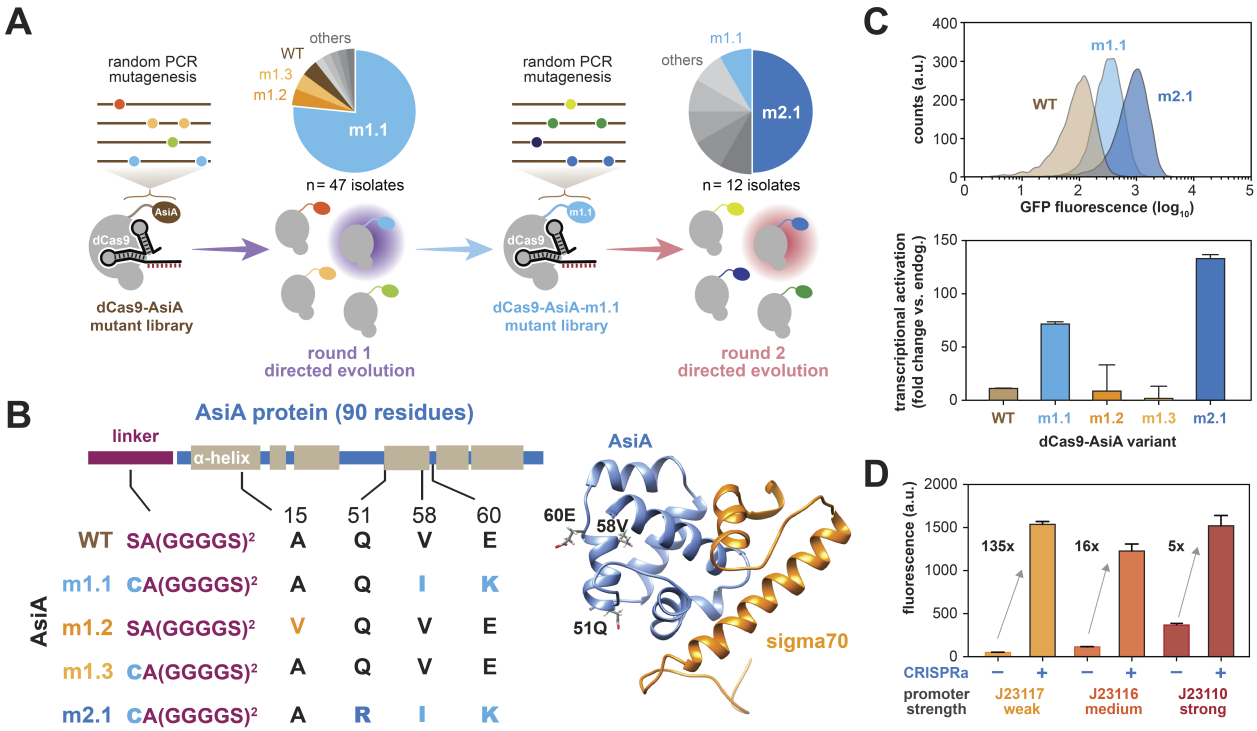

Figure 2. Directed evolution of dCas9-AsiA for higher potency. (A) Schematics of two rounds of directed evolution to improve potency of dCas9-AsiA. Pie charts show frequencies of dCas9-AsiA variants identified from each round. **(B)** Mutations found in enriched AsiA variants and their positions along the Asia secondary structure (left). Crystal structure of wild-type AsiA (blue) interfaced with region 4 of σ70 (orange) are shown (right). **(C)** Distribution of fluorescence signal of the GFP reporter induced by different dCas9-AsiA variants (top). Fold induction by different dCas9-AsiA variants is shown (bottom). **(D)** CRISPRa induction of promoters with varying basal expression levels. CRISPRa-, bassal expression of the promoter; CRISPRa+, expression activated by dCas9-AsiA-m2.1 and associated gRNAs. Data shown are 3 biological replicates with +/- SEM.

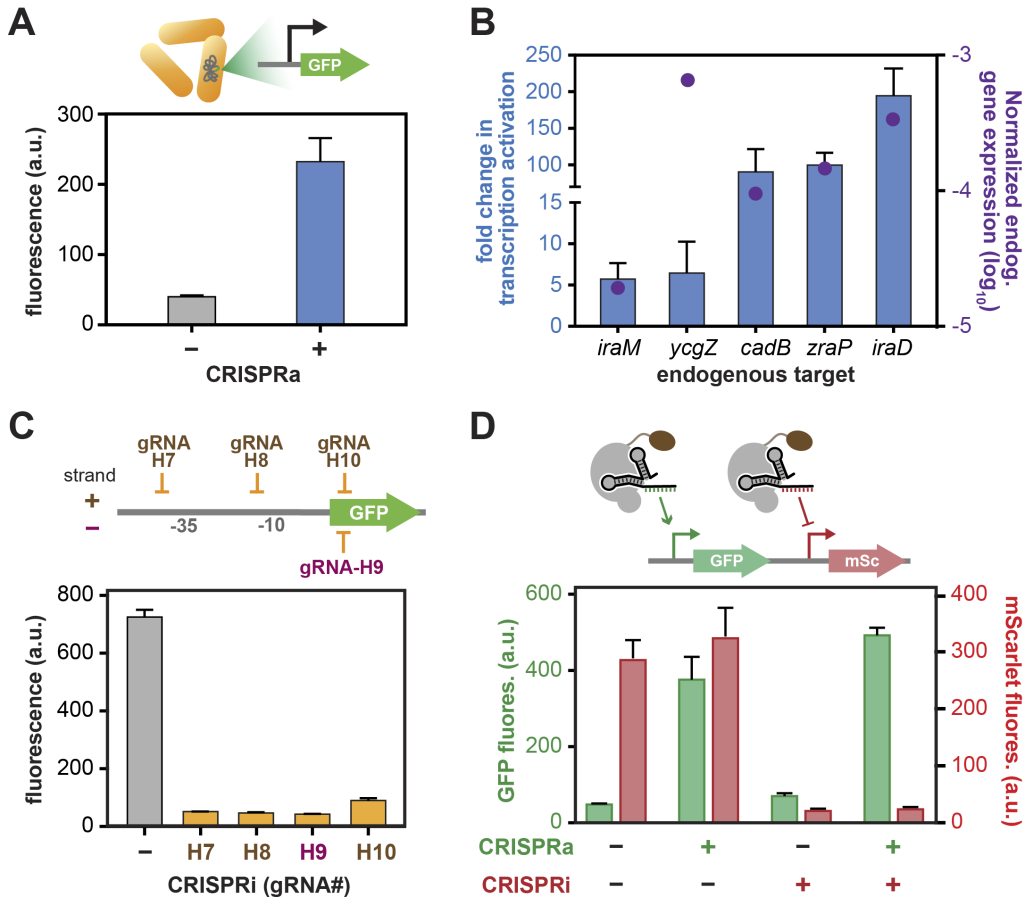

Figure 3. Evolved CasTA2.1 could activate genomic targets and mediate multiplexed gene activation and repression. (A) CasTA2.1 upregulates a genomically inserted GFP reporter. **(B)** Chromosomal gene targets activated by CasTA2.1 with bars showing the activation fold change and dots showing basal expression of each gene. **(C)** CasTA2.1 can mediate CRISPRi with appropriate gRNA designs by positioning different gRNAs relative to the target gene. A non-specific gRNA is used as the negative control. **(D)** Demonstration of multiplexed CRISPRa and CRISPRi using CasTA2.1 on a reporter containing GFP and mScarlet. Parental cells had low basal GFP and high basal mScarlet expression. Data shown are 3-4 biological replicates with +/- SEM.

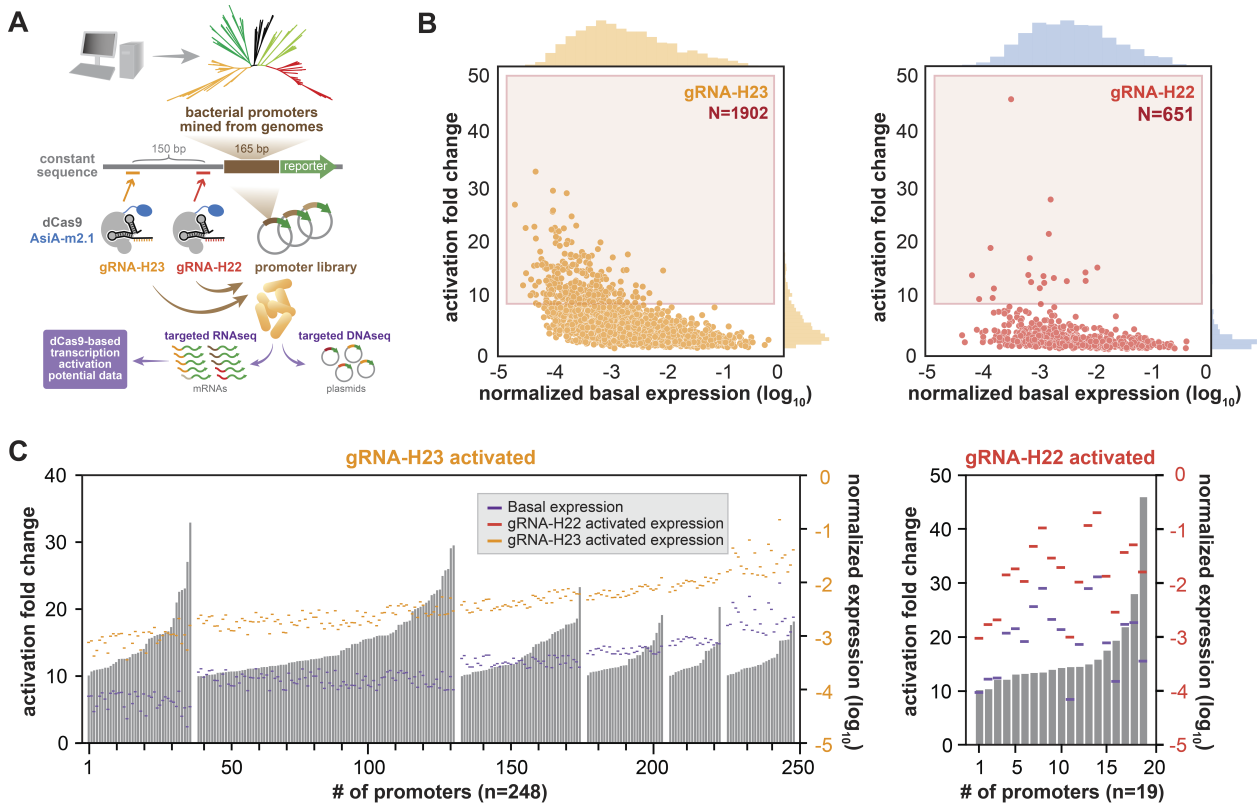

Figure 4. Multiplex reporter assay to identify inducible promoters using CasTA2.1. **(A)** Schematics of construction and screening platform to characterize a library of CRISPRa-mediated inducible promoters by targeted RNAseq and DNAseq. **(B)** Promoters significantly activated by CasTA2.1 using gRNA-H23 (left) or gRNA-H22 (right) are plotted with basal expression level on x-axis against fold activation by CRISPRa on y-axis. N is the total number of promoters shown. Red box correspond to highly activated promoters (fold change >10). **(C)** Highly activated promoters (>10 fold) using gRNA-H23 (left) or gRNA-H22 (right) are plotted. Basal expression levels (purple lines), induced expression level (orange lines on left or red lines on right, activated with gRNA-H23 or gRNA-H22, respectively) and induced fold changes (gray bars) are shown.

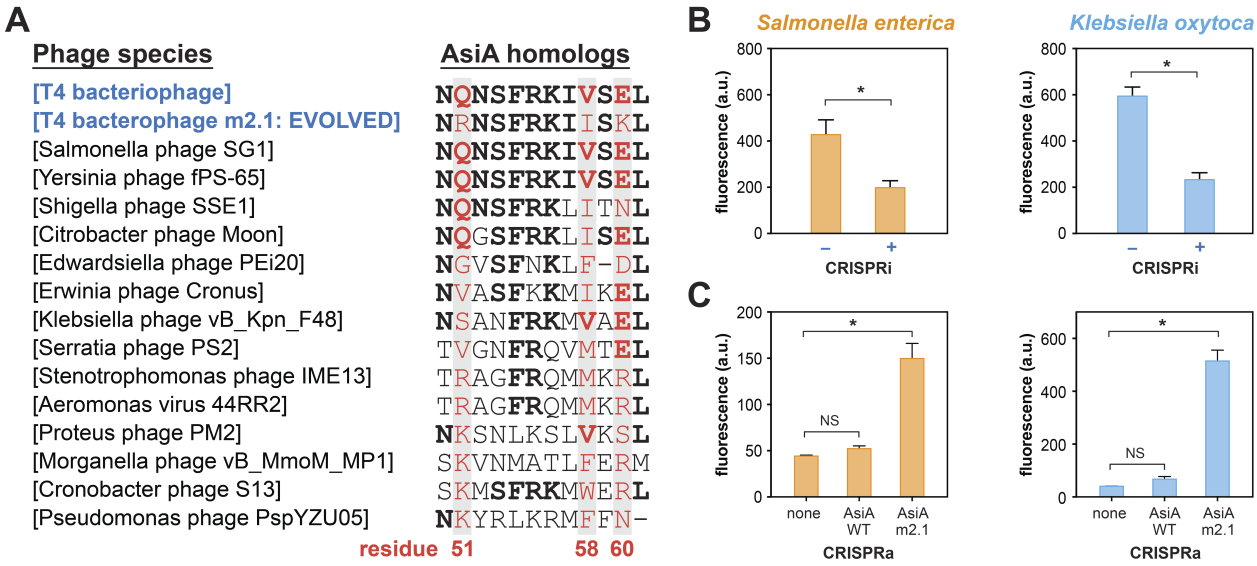

Figure 5. Evolved CasTA functions in multiple bacterial species. (A) Multiple sequence alignment of AsiA homologs from different phage genomes at residue positions 50-61. Highlighted red residues indicate positions that are mutated in AsiA-m2.1. **(B)** CRISPRi in *S. enterica* and *K. oxytoca* using CasTA2.1. **(C)** CRISPRa in *S. enterica* and *K. oxytoca* using CasTA1.0 (dCas9-AsiA_wt), CasTA2.1 (dCas9-AsiA_m2.1) or ancestral strain with basal promoter expression (none). All data are 3-4 biological replicates with +/- SEM. * Student's t-test, p<0.0001 NS non-significant.