Systems-wide analysis of the GATC-binding nucleoid-associated protein Gbn and its impact on *Streptomyces* development

Chao Du, Joost Willemse, Amanda M. Erkelens, Victor J. Carrion, Remus T. Dame, and Gilles P. van Wezel*

Centre for Microbial Cell Biology, Leiden University, The Netherlands

* Corresponding author: g.wezel@biology.leidenuniv.nl; Tel +31 71 5274310

Keywords: Actinobacteria; Chromosome structure; nucleoid associated protein; DNA binding studies; development

Running title: A novel nucleoid associated protein in *Streptomyces*

ABSTRACT

Bacterial chromosome structure is organized by a diverse group of proteins collectively referred to as nucleoid-associated proteins (NAPs). Many NAPs have been well studied in *Streptomyces*, including Lsr2, HupA, HupS, and sIHF. Here, we show that SCO1839 represents a novel family of Actinobacteria NAPs and recognizes a consensus sequence consisting of GATC followed by (A/T)T. The protein was designated Gbn for GATC-binding NAP. Deletion of *gbn* led to alterations in development and antibiotic production in *Streptomyces coelicolor*. Chromatin immunoprecipitation sequencing (ChIP-Seq) detected more than 2800 binding regions, encompassing some 3600 GATCWT motifs, which comprise 55% of all such motifs in the *S. coelicolor* genome. DNA binding of Gbn *in vitro* increased DNA stiffness but not compaction, suggesting a role in regulation rather than chromosome organization. Transcriptomics analysis showed that Gbn binding generally leads to reduced gene expression. The DNA binding profiles were nearly identical between vegetative and aerial growth. Exceptions are SCO1311 and SCOt32, for a tRNA editing enzyme and a tRNA that recognises the rare leucine codon CUA, respectively, which nearly exclusively bound during vegetative growth. Taken together, our data show that Gbn is a highly pleiotropic NAP that impacts growth and development in streptomycetes.

IMPORTANCE

A large part of the chemical space of bioactive natural products is derived from Actinobacteria. Many of the biosynthetic gene clusters for these compounds are cryptic, in others words, they are expressed in nature but not in the laboratory. Understanding the global regulatory networks that control gene expression is key to the development of

approaches to activate this biosynthetic potential. Chromosome structure has a major impact on the control of gene expression. In bacteria, the organization of chromosome structure is mediated by a diverse group of proteins referred to collectively as nucleoid-associated proteins (NAPs), which play an important role in the control of gene expression, nucleoid structure and DNA repair. We here present the discovery of a novel and extremely pleiotropic NAP, which we refer to as Gbn. Gbn is a sporulation-specific protein that occurs only in the Actinobacteria and binds to GATC sequences, with a subtle but broad effect on global gene expression. The discovery of Gbn is a new step towards better understanding of how gene expression and chromosome structure is governed in antibiotic-producing streptomycetes.

INTRODUCTION

Streptomycetes are filamentous soil bacteria with a complex life cycle, which are well known for their ability to produce various kinds of antibiotics and other valuable natural products. Thus, they are a major source of clinical drugs (1-3). The life cycle of *Streptomyces* starts with the germination of a spore that grows out to form vegetative hyphae. Exponential growth is achieved via tip extension and branching, eventually resulting in a dense mycelial network (1, 4). When the environmental situation requires sporulation, for example due to nutrient starvation, streptomycetes start their reproductive growth phase by developing aerial hyphae, which eventually differentiate into chains of unigenomic spores (5, 6). The production of antibiotics temporally correlates with the onset of development (7, 8). The complexity of the underlying regulatory networks is underlined by the fact that the

54 *Streptomyces coelicolor* genome encodes some 900 regulatory proteins, of which only a
55 minute fraction has been functionally characterized (9). Many of these affect the control of
56 development and antibiotic production, such as the *bld* and *whi* genes that are responsible
57 for the control of aerial hyphae formation and sporulation, respectively, and global
58 regulatory genes such as *adpA*, *afsR*, *dasR* and *atrA* that pleiotropically control antibiotic
59 production (10).

60 The control of chromosome structure is an important factor in the control of gene
61 expression. In bacteria, the organization of chromosome structure is mediated by a diverse
62 group of proteins referred to collectively as nucleoid-associated proteins (NAPs) (11-13).
63 These are generally small DNA binding proteins involved in processes such as controlling
64 gene expression, nucleoid structure, or DNA repair. Well-known NAPs in *Streptomyces*
65 include Lsr2, HupA, HupS, sIHF, and IHF. Lsr2 binds non-specifically to AT-rich sequences and
66 can globally repress gene expression (14). HupA and HupS are homologs of HU (for histone-
67 like protein from strain U93) proteins, which are differentially regulated depending on the
68 developmental growth phase (15). sIHF is one of the basic architectural elements conserved
69 in many actinobacteria and is able to influence the regulation of secondary metabolism and
70 cell development (16). IHF binds a well conserved nucleotide sequence, while HU binds to
71 random DNA sequences (17), yet with a preference for bent, distorted or flexible DNA (18). A
72 proteomic survey of *Streptomyces coelicolor* identified 24 proteins with NAP-like properties,
73 namely the known Lsr2, HupA, HupS and sIHF and 20 yet unidentified proteins (19).
74 Although the functions of many NAPs are still not clear, BldC for example has a major impact
75 on the transcriptome (20, 21).

76 We previously showed via pull-down assays that the candidate NAP SCO1839 binds
77 to the promoter region of the cell division regulatory gene *ssgR* (22). SsgR is the

78 transcriptional activator of the cell division activator gene *ssgA* (23). SsgA and its paralogue
79 SsgB are both required for sporulation (24-26), and together coordinate the onset of
80 sporulation-specific cell division in *Streptomyces*, whereby SsgB directly recruits the cell
81 division scaffold protein FtsZ to the future sites of septation (27).

82 In this study, we show that SCO1839 represents a novel family of small DNA binding
83 proteins, which plays a role in the regulation of *Streptomyces* development and antibiotic
84 production. The protein is specific to the Actinobacteria, with an HTH DNA binding motif
85 containing three helices, and plays a role in the control of morphogenesis. Chromatin
86 immunoprecipitation coupled with massive parallel DNA sequencing (ChIP-Seq) revealed
87 that the protein binds to over 2800 genomic regions with one or more binding sites,
88 recognising a specific DNA binding motif centred around the consensus sequence GATC.
89 Thus, we designated the protein Gbn for GATC-binding NAP. Transcriptomics data showed
90 that genes bound by Gbn on the promoter regions tends to express more in *gbn* mutant,
91 suggesting a suppressive effect of Gbn.

92 MATERIALS AND METHODS

93 Reagents

94 All restriction enzymes are ordered from NEB (Massachusetts, U.S.), including *Bam*HI-HF (Cat.
95 R3136), *Xba*I (R0145), *Eco*RI-HF (R3101), *Hin*dIII-HF (R3104), *Bbs*I (R0539), *Nco*I-HF (R3193),
96 *Sna*BI (R0130), *Stu*I (R0187), *Sac*I (R0156), *Nde*I (R0111). Phusion polymerase (M0532) and T4
97 DNA ligase (M0202) were also obtained from NEB. Plasmid mini-prep kit (Cat. 740727.250)

98 was from BIOKE (Leiden, The Netherlands). DNA purification was achieved using DNA Clean
99 & Concentrator kit (Cat. D4029) from Zymo Research (California, U.S.).

100 **Biological Resources**

101 ***Strains and growth conditions***

102 All strains used in this study are listed in Table S3. *Escherichia coli* strain JM109 was used for
103 routine cloning, *E. coli* ET12567 (28) for preparing non-methylated DNA, ET12567 containing
104 driver plasmid pUZ8002 (29) was used in conjugation experiments for introducing DNA to
105 *Streptomyces*. *E. coli* strains were grown in Luria broth at 37°C supplemented with the
106 appropriate antibiotics (ampicillin, apramycin, kanamycin and/or chloramphenicol at 100, 50,
107 25 and 25 $\mu g \cdot mL^{-1}$, respectively) depending on the vector used. *S. coelicolor* A3(2) M145 was
108 the parent strain for all mutants. *Streptomyces* strains were grown on soya flour medium
109 (SFM) for conjugation, SFM agar medium or MM agar medium supplemented with 0.5%
110 mannitol for phenotype characterization, R5 agar plates for protoplast regeneration, and
111 MM agar medium supplemented with 0.5% mannitol covered with cellophane for ChIP-Seq
112 and transcriptomics culture growth. Solid cultures were grown in a 30°C incubator unless
113 described specifically. For liquid cultures, approximately 10^6 spores were inoculated in 100
114 mL Erlenmeyer flask equipped with steel spring containing 15 mL TSBS (tryptone soya broth
115 sucrose) medium (30). The flasks were incubated at 30°C with constant shaking at 180 rpm.
116 Antibiotics used for screening *Streptomyces* transformants were apramycin and thiostrepton
117 (20 and 10 $\mu g \cdot mL^{-1}$, respectively).

118

119 ***Constructs and cloning***

120 Primers used for PCR and short double strand DNA fragment are listed in Table S4. PCR was
121 preformed using Phusion DNA polymerase using standard protocol as described previously
122 (31). All plasmids and constructs described in this study are summarized in Table S5. The
123 constructs generated in this study were verified by sanger sequencing performed in
124 BaseClear (Leiden, The Netherlands).

125 The *gbn* knock-out strategy was based on the unstable multi-copy vector pWHM3 as
126 described previously (32). Briefly, up- and down-stream region of *gbn* were amplified from
127 genome and cloned into pWHM3. Between these two regions, an apramycin resistance
128 cassette from pGWS728 (33) was inserted as selection marker. The resulting vector
129 pGWS1255 is a knock-out construct that can replace nucleotide positions the +1 to +207 of
130 *gbn* with apramycin resistance cassette, where +1 refers to the translation start site. The
131 apramycin resistance cassette was subsequently removed using Cre expressing construct
132 pUWL-Cre (34, 35) yielding the clean knock-out strain GAD003. For complementation of the
133 *gbn* deletion mutant, the nucleotide positions -565 to +228 relative *gbn* translation start site,
134 containing entire coding region of *gbn* (with stop codon) and its promoting region was
135 cloned into low copy number plasmid pHJL401, yielding complementation construct
136 pGWS1260. This construct was then transformed to GAD003, resulting in strain GAD014.

137 To perform ChIP-Seq experiment, the 3×FLAG was fused to the end of original copy of
138 *gbn* on genome using codon optimised CRISPR-Cas9 system (36). The spacer sequence
139 located at the end of *gbn* and was inserted into the pCRISPomyces-2 plasmid as described by
140 (36). Template for homology-directed repair (HDR) which ensures the insertion of 3×FLAG
141 sequence was cloned into pCRISPomyces-2 followed by spacer insertion (Figure S3), yielding
142 *gbn*-3×FLAG knock-in construct pGWS1298. Mutagenesis was done according to (36). A

143 successful 3×FLAG tag knock-in strain was identified by PCR and sequencing, designated
144 GAD043.

145 For over-expression of *gbn*, the *ermE* promoter was cloned to replace part of the
146 original promoter region of *gbn* using the same CRISPR-Cas9 system. A spacer sequence
147 were designed at the promoter region of *gbn* and this was inserted into pCRISPomyces-2
148 (36). HDR template was designed to remove -157 to +4 region of *gbn* and replace this region
149 with P*ermE* sequence yielding pGWS1295. P*ermE* was digested from pHM10a (37). Following
150 the same procedure as above, strain GAD039 was obtained which expresses *gbn* from the
151 *ermE* promoter.

152 To produce His_6-Gbn for electrophoretic mobility shift assay (EMSA) experiments,
153 *gbn* was cloned into protein expression construct pET28a and transformed into *E. coli* strain
154 BL21 CodonPlus (DE3)-RIPL (Invitrogen, Massachusetts, U.S.). To generate methylated and
155 non-methylated DNA for EMSA, part of the promoter region of *gbn* and a random region
156 that is not bound by Gbn were cloned into pUC19 (38). This yielded vector pGWS1300 that
157 contains a Gbn binding domain and pGWS1451 contains a stretch of DNA that is not bound
158 by Gbn.

159 **Database mining and clustering**

160 The protein sequence of Gbn (SCO1839) from *S. coelicolor* was used as query in HMMER web
161 server (39) to obtain all Gbn-like proteins from the database, resulting in 727 hits. Sequences
162 with an E value < 0.01 (684 sequences) were selected to generate a Hidden Markov model
163 (HMM) profile using HMMER suit v3.1b2 (40). This profile was used to search against a
164 custom database containing 146,856 genomes with all available bacteria genomes (access
165 date Feb. 9, 2019). Hits with E-value ≤ 5.5×10^{-9} (2,317 sequences) were aligned to the

generated HMM profile using the hmmalign tool from HMMER suit. Using the alignment, a network was built calculating the pairwise distance between all the detected Gbn proteins and the threshold for clustering was settled at 0.8. Network visualizations were constructed using Cytoscape v3.7.1 (41).

Time-lapse confluent plate morphology monitoring

Approximately 10^7 spores were plated on MM agar supplemented with mannitol. The plates were then placed upside down in Perfection V370 scanner (Epson, Nagano, Japan) located inside 30°C incubator. A scanning picture was taken every hour, and images were processed using custom python script to get the brightness value of the plate. Specifically, the pictures were first converted to grey scale. 70% the diameter of the plate from the centre was selected as the region of interest (ROI). The average grey value of all the pixels within ROI was used as the brightness of the mycelium lawn. The measured values from one plate were then normalized to range 0 to 1.

Scanning electron microscopy (SEM)

Mycelia were grown on MM agar supplemented with mannitol and grown for 5 days. Sample preparation and imaging was done as described before (42, 43), using JSM-7600F scanning electron microscope (JEOL, Tokyo, Japan). For each strain, 5 images with 7,500 × magnification were taken in randomly selected spore-rich areas. The length and width of spores in each picture were measured using ImageJ version 1.52p strictly according to a randomized file list, to minimize selection bias. Only spores which are approximately parallel to the focal plane were measured.

187 **DNA-protein cross-linking and chromatin immunoprecipitation**

188 10^8 spores of strain GAD043 were plated on MM agar medium specified above. After 25 h or
189 48 h growth, cellophane disks were soaked up-side-down in PBS solution containing 1%
190 formaldehyde for 20 min allowing DNA-protein crosslink. Ten plates were collected for 25 h
191 samples; four plates were collected for 48 h samples. Then the disks were moved to PBS
192 solution containing 0.5 M glycine for 5 min to stop crosslinking reaction. The mycelium was
193 then collected, washed in PBS and resuspended in 0.5 mL lysis buffer (10 mM Tris-HCl pH 8.0,
194 50 mM NaCl, 15 $mg \cdot mL^{-1}$ lysozyme, 1× protease inhibitor (Roche, Bavaria, Germany) and
195 incubated at 37°C for 20 min. After incubation, 0.5 mL IP buffer (100 mM Tris-HCl pH 8.0,
196 250 mM NaCl, 0.8% v/v Triton-X-100) was added to the sample, and chromosomal DNA
197 sheared to 100 to 500 bp using Bioruptor Plus water bath sonication system (Diagenode,
198 Liège, Belgium). After centrifugation to remove cell debris, lysates were incubated with 40
199 μL Anti-FLAG M2 affinity gel (cat A2220, Sigma-Aldrich, St. Louis, U.S.) suspension according
200 to the manufacturer's instructions and incubated at 4°C overnight. After centrifugation and
201 washing, the pellets and 50 μL of untreated total extracts (controls) were incubated in 100
202 μL IP elution buffer (50 mM Tris-HCl pH7.5, 10 mM EDTA, 1% m/v SDS) at 65°C overnight to
203 reverse cross-links. Beads were then removed by centrifugation before DNA extraction with
204 phenol-chloroform. The DNA sample was then extracted with chloroform and the water
205 layer was further purified using DNA Clean & Concentrator kit (Cat. D4029, Zymo Research,
206 California, U.S.). The samples were then sent for next generation sequencing using the BGI-
207 Seq platform (BGI, Hong Kong, China).

208 **ChIP-Seq data analysis**

209 Clean reads received from the sequencing contractor were aligned to the *S. coelicolor* M145
210 genome with RefSeq accession number NC_003888.3 using bowtie2 v2.34 (44). Resulted
211 SAM files are sorted using SAMtools v1.9 (45) producing BAM files. MACS2 v2.1.2 (46) was
212 then used for binding peak prediction and peak modelling by comparing the chromatin
213 immunoprecipitated DNA sample with the corresponding whole genome sample. The
214 models for both samples are shown in Figure S2. The enrichment data used in Figure 3 was
215 calculated for each nucleotide using MACS2 'bdgcmp' command with '-m FE' switch. The
216 peak summit positions including sub peak positions of each predicted binding region were
217 then extracted, and the ± 150 bp region of each summit was extracted from the genome
218 sequence using a python script dependent on the Biopython module v1.70 (47). Extracted
219 sequences were subjected to MEME-ChIP v5.02 (48), which is suitable for large sequence
220 sets, for binding motif prediction.

221 The enrichment data of two samples was averaged separately in a moving bin of
222 20,000 bp and plotted in opposite directions as the middle ring of the circular genome
223 diagram. The G+C content was calculated using the same moving bin and centred at the
224 middle of maximum and minimum value before plotting as the inner ring on the plot (Figure
225 3). For determining the overlap of low G+C content regions and high enrichment regions, the
226 genome was divided into 1,000 bp long sections, the G+C content and average enrichment
227 levels were calculated. The sections with G+C content below the first quartile was
228 considered low in G+C, and those with average enrichment level above the third quartile
229 were considered high in enrichment. To find genes possibly regulated by Gbn, the locations
230 of promoter regions (-350 to +50) of all genes were extracted from the genome containing
231 annotations and checked for overlap with ± 150 bp location of the summit of Gbn binding

232 peaks. This was done using a python script dependent on the module Biopython and
233 pybedtools v0.8 (49) and external BEDTools v2.27 (50).

234 **Transcriptomics and data analysis**

235 Spores (10^8 cfu) of *S. coelicolor* M145 and *gbn* knockout strain GAD003 were plated on MM
236 agar plates overlayed with cellophane disks. After 24 h (vegetative growth) or 45 h (aerial
237 growth) of growth, mycelia were scraped from the cellophane disks, snap frozen in liquid N_2
238 and disrupted using TissueLyser (Qiagen, Venlo, The Netherlands) for 3 times 30 s at 30 Hz.
239 Total RNA extracted using the Kirby method described (51). The RNA-seq libraries
240 preparation and sequencing were outsourced to Novogene (Novogene Europe, Cambridge,
241 UK). Ribosomal RNA was removed from the samples using NEBNext® Ultra™ Directional RNA
242 Library Prep Kit (NEB, Massachusetts, U.S.). Sequencing libraries were generated using
243 NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, Massachusetts, U.S.) and
244 sequencing was carried on an Illumina® NovaSeq™ 6000 platform. Raw data was cleaned
245 using fastp v0.12.2 (52), then mapped to the *S. coelicolor* M145 genome (GenBank accession
246 AL645882.2) using bowtie2 v2.4.4 (53). Read counts for each gene were generated using
247 featureCounts v2.0.1 (54). Transcripts per million (TPM) values were generated using custom
248 python script, differently expressed genes and $\log_2$ fold change (LFC) was generated using
249 DESeq2 v1.32.0 (55) with the data shrinkage function "apeglm" (56).

250 **Electrophoretic mobility shift assay (EMSA)**

251 Gbn-His_6 was expressed and purified as described previously (57). Purified protein was
252 dialyzed over-night at 4°C against EMSA buffer (10 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 50
253 mM KCl). 50 bp double strand DNA was generated by gradual cooling of reverse

complemented single strand oligonucleotides in T4 DNA ligase buffer (NEB, Massachusetts, U.S.) from 95°C to 12°C in 45 min. pGWS1300 was extracted from DAM methylation effective *E. coli* strain JM109 and DAM deficient *E. coli* strain ET12567, while pGWS1451 for negative control was extracted from strain ET12567 only. The target fragments were then digested and blunted using DNA polymerase I Klenow fragment (NEB, Massachusetts, U.S.). The *in-vitro* DNA-protein interaction studies were done in EMSA buffer in a total reaction volume of 10 µL, the reactions were incubated at 30°C for 15 min. The whole reaction volume was then loaded on 5% polyacrylamide gels and separated by electrophoresis. The gel was briefly stained with ethidium bromide and imaged in a Gel Doc imaging system (BioRad, California, U.S.).

Tethered particle motion

Tethered particle motion experiments were carried out as described previously (58) with minor modifications. The experimental buffer used was 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 50 mM KCl, 0.5% acetylated BSA. Data was collected for each protein concentration at least in duplicate. An anisotropic ratio cut-off of 1.3 and a standard deviation cut-off of 8% were used to select single-tethered beads. The region of interest (-609 to +33 bp relative to the *gbn* translation start site) was amplified from the genome and then inserted into pUC19 using Gibson assembly yielding plasmid pGWS1462. DNA was then amplified as a 685 bp fragment from this construct using a forward primer labelled with biotin (Biotin-CTGGCTGAAACGGAATAGGT) and a reverse primer labelled with digoxygenin (Digoxygenin-AGCTCAGCGAGAACCGG).

RESULTS AND DISCUSSION

SCO1839 is a small NAP specific to Actinobacteria

We previously identified SCO1839 as a DNA binding protein that binds to the promoter region of the sporulation regulatory gene *ssgR* (22). SsgR activates transcription of *ssgA*, which encodes a pleiotropic developmental regulator and activator of sporulation-specific cell division in streptomycetes. SCO1839 is a small protein of 73 amino acids (7.6 kDa) with a predicted isoelectric point (pI) of 10.53, indicative of an alkalic protein. A Pfam sequence search (59) did not yield any significant matches to known protein families, suggesting that SCO1839 is the first member of a novel protein family. It was suggested that SCO1839 may be a nucleoid-associated protein (NAP) (19).

To obtain more insights into the distribution and phylogeny of SCO1839, a conserved Hidden Markov Model (HMM) domain was constructed using all SCO1839-like proteins from *Streptomyces* species. Consequently, an HMM search against all available bacteria full genomes in the database was performed. No hits were found outside the order Actinomycetales, strongly suggesting that SCO1839 is an Actinobacteria-specific protein (Figure 1A and B). Eight main clusters of similar groups of homologs were found. The largest cluster mainly consists of SCO1839 orthologs from *Streptomyces*, *Amycolatopsis*, *Pseudonocardia*, *Frankia*, and *Actinomadura*. Other major clusters include a cluster of *Nocardia* and *Rhodococcus*, *Micromonospora* and *Salinispora*, *Geodermatophilus* and *Blastococcus*. Orthologs from *Rhodococcus* form two separate clusters. Interestingly, 24 *Actinomadura* species and 38 *Streptomyces* species have two paralogues of SCO1839, which divide into two additional clusters. For most genera, more than 90% of the species encode at least one copy of SCO1839-like proteins. The genus *Rhodococcus* forms an exception, as only

298 28.7% (92 out of 321) of the sequenced genomes of this genus contain a SCO1839-family
299 protein (Figure 1A), and these proteins divided into three distinct clusters. This could be
300 related to the fact that it is the only genus among all listed genera that does not have a true
301 mycelial life style (1). The low conservation in the non-sporulating *Rhodococcus* suggests
302 that SCO1839 may primarily be sporulation specific.

303 *In silico* structural modelling of SCO1839 was performed using the I-TASSER server
304 (60), revealing a putative single DNA binding helix-turn-helix (HTH) motif, in the form of a tri-
305 helical structure (Figure 1C and 1D). No homology was seen to any other known
306 transcriptional regulator family from bacteria. Very short aa stretches flank the residues
307 belonging to the HTH motif. Nine other protein structures found in PDB (Protein data bank,
308 rcsb.org, 61) share structural analogy to SCO1839, and have similar HTH motifs. These are
309 proteins found in a wide range of organisms and with different functions, including a DNA
310 helicase from the archaeon *Pyrococcus furiosus* (PDB structure ID: 2ZJ8), *E. coli* (2VA8, 2P6R)
311 and from humans (5AGA), a ribosomal protein from yeast (5MRC), a human cell division
312 cycle protein (2DIN), a yeast terminator binding protein (5EYB), a regulator from
313 *Staphylococcus aureus* (2R0Q), and a tRNA synthetase from the archaeon *Archaeoglobus*
314 *fulgidus* (2ZTG). However, none of these proteins were as small as SCO1839 and neither
315 contained only one DNA binding motif. To the best of our knowledge, no other bacterial
316 protein with similar structure has been reported before. Therefore, we propose that the
317 SCO1839-like proteins form a new family of bacterial DNA binding proteins. SCO1839 is
318 further nominated Gbn, for GATC-binding NAP (see below).

319 **Gbn binds to thousands of DNA binding sites with GATC as core motif**

320 To obtain insights into the genome-wide DNA binding capacity of Gbn, ChIP-Seq analysis was
321 performed on samples harvested after 25 h (vegetative growth) and 48 h (sporulation).
322 Following this approach all binding sites of Gbn on the *S. coelicolor* chromosome can
323 potentially be identified. For this purpose, the original copy of *gbn* on the genome was fused
324 with a sequence encoding a triple FLAG tag at its 5'-terminus using CRISPR-Cas9 (see
325 Materials and Methods section). The strain had a phenotype that was highly similar to that
326 of the parent (Figure S1). The ChIP-Seq data showed a wide distribution of Gbn binding
327 events (Figure 3). Half of the binding sites with a strong enrichment (50% of the sites at 25
328 and 38 h) colocalized with low G+C content regions. In total, 2825 and 2919 binding regions
329 were identified using MACS2 software (46) for samples obtained from mycelia in the
330 vegetative and sporulation stage, respectively. Interestingly, there was a near complete
331 overlap (> 90%, 2402) between the Gbn binding events found in the two samples (Pearson
332 correlation coefficient 0.945, Figure 3). This not only shows that the binding specificity of
333 Gbn is largely growth phase-independent, but also that the experiments were highly
334 reproducible. The result also indicates that while the expression of *gbn* is higher during
335 sporulation, the specificity of the protein for its binding sites does not change, as shown by
336 the highly similar binding profiles from the ChIP-seq experiments.

337 To obtain a consensus binding site for Gbn, the binding regions from ChIP-Seq results
338 were extracted and modelled using MACS2 and MEME-ChIP (46, 48). In this way, the
339 sequences GATCAT and GATCTT were identified as binding sites, and thus GATCWT
340 represents the Gbn binding site (Figure 4A). The most conserved binding core was GATC,
341 which is a palindrome known as recognition site for DNA methylation (62). The predicted
342 motifs also showed G/C preference on the flanking region separated by gaps of two base

343 pairs (Figure 4A). It is important to note that virtually all binding regions that were identified
344 as significant (> 99.8%) contained a GATC motif, and most (88.2% for 25 h, 84.0% for 48 h)
345 contained the consensus sequence GATCWT. The *S. coelicolor* genome contains in total
346 6,501 GATCWT sequences, 54.5% and 55.5% were present in the predicted binding regions
347 in the 25 h and 48 h samples, respectively. Note that many binding sites have more than one
348 copy of the motif. Closer inspection of the raw data of the ChIP-seq data revealed that in fact
349 all GATCWT motifs on the *S. coelicolor* genome are covered by an increased sequencing
350 coverage, but not all of them can be reliably detected by MACS2 without increasing the false
351 positives to an unacceptable level. This was not seen for GATC sequences not followed by AT
352 or TT. We believe that Gbn can bind most GATCWT sequences on the *S. coelicolor*
353 chromosome. However, not all GATCWT sequences may be accessible to Gbn either directly
354 because they are occupied by other proteins or indirectly because of chromosome
355 organization of the region. Further investigation of this phenomenon is needed.

356 The affinity of Gbn for its DNA binding motif was further examined *in vitro* using
357 electrophoretic mobility shift assays (EMSA). Result show that Gbn could bind to the GATC
358 motif *in vitro*. The additional nucleotide (AT)T on one side increased the affinity of Gbn, with
359 the half binding concentration reduced by more than a third (Figure 5B). Furthermore, when
360 more GATC motifs were present in the DNA fragments, simultaneous binding of more Gbn
361 proteins was observed (Figure 5C and D). For the short 50 bp DNA fragment with four target
362 DNA motifs, only three binding events can be observed from EMSA experiment, possibly
363 only one of the two GATC sequences in close proximity could be bound by one Gbn protein
364 as a consequence of steric hindrance. Taken together, Gbn showed good binding to a motif
365 centred around GATC in experiments *in vitro*.

366 The GATC motif is the target sequence for deoxyadenosine methylase (DAM), which
367 is essential for DNA mismatch repair in *E. coli* (62). Another DNA methylation takes place on
368 the second deoxycytosine of the sequence CCTGG (DCM) (63). *S. coelicolor* lacks both
369 methylation systems and degrades methylated exogenous DNA (51, 64). Which proteins are
370 involved in the recognition and restriction system in *Streptomyces* remains unknown (65, 66).
371 To test whether Gbn play a role in this restriction system, we compared the transformation
372 efficiency of methylated DNA and non-methylated DNA in the *gbn* null and in the parental
373 strain M145. No significant differences were observed in the transformation efficiencies
374 between parent and *gbn* mutant, which suggests that Gbn does not play a role in restriction
375 of methylated DNA (data not illustrated). Next, we tested the affinity difference of Gbn for
376 GATC and GA^{m}TC in an electrophoretic mobility shift assay (EMSA). The results revealed that
377 Gbn had lower affinity for methylated DNA (Figure 5D), which may be caused by a steric
378 effect of the methyl group on Gbn binding.

379 **Gbn does not alter the conformation of the DNA**

380 In order to investigate whether Gbn alters DNA conformation *in vitro*, we performed
381 Tethered Particle Motion (TPM) experiments (67). If the binding of a protein to DNA induces
382 bends or affects DNA stiffness this translates into a reduction, or an increase of the root
383 mean square displacement (RMS) respectively compared to that of bare DNA (68-70). Here
384 we used a 685 bp DNA substrate containing the *gbn* promoter region with 10 GATC(WT)
385 sites. Addition of the Gbn protein resulted in a very small increase in RMS of around 4 nm at
386 ≥ 250 nM (Figure 5E). This could be explained by Gbn occupying the 10 binding sites without
387 causing major changes in the DNA conformation. While Gbn does not deform DNA
388 promoting compaction, the increase in RMS is indicative of the DNA being somewhat

389 stiffened by binding of the protein. Earlier studies indicate that the observed mild increase in
390 RMS corresponds to an increase in persistence length of about 5 nm as compared to bare
391 DNA (70). Note that qualitatively similar effects have been observed for *E. coli* and *P.*
392 *aeruginosa* H-NS like proteins (68, 69). These properties suggest that rather than
393 contributing directly to organization and compaction of the chromosome, Gbn functions in
394 regulation of genome transactions such as transcription.

395 **Gbn binding events are found in the regulatory regions of some 10% of all genes**

396 To obtain better insight into which genes are affected by the binding of Gbn, we investigated
397 Gbn binding events in the putative promoter regions (-350 to +50 relative transcription start
398 site of all genes). In total, Gbn bound to the promoter of 769 genes at both 25 h and 48 h
399 (Table S1). Of these genes, 44.5% (342 genes) had at least binding event with more than 10-
400 fold enrichment in the corresponding promoter region. These genes include many genes
401 from cell division and cell wall (*dcw*) cluster (SCO2077-SCO2088), 50S ribosomal protein gene
402 cluster (SCO4702-SCO4727), tRNAs (SCOt02-SCOt50), *atrA*, *redY*, SCO1311 and *gbn* itself
403 (Table 2, Figure 4B). One Gbn binding event was found upstream of *ssgR*, with a peak at
404 around nt position -483 relative to the translation start site (Figure 4C). This is in accordance
405 with the previous observation that Gbn binds to the *ssgR* promoter region (22). Gbn not only
406 bound to the promoter regions of five tRNA genes, but also to the promoter region of
407 SCO1311, which has a tRNA-editing domain responsible for hydrolysing mis-acylated tRNA
408 (coverage 68%, Pfam: PF04073). Interestingly, the promoter regions of SCO1311 and SCOt32
409 showed the largest difference in Gbn binding between 25 h and 48 h samples, with three
410 times higher enrichment at 25 h than at 48 h (Figure 4D, F, Table 2). SCOt23 specifies a
411 leucyl-tRNA with anticodon UAG, which is required for the translation of the rare leucine

412 codon CUA. The rarest codon in *S. coelicolor* is another leucine codon, namely UUA. The
413 tRNA recognizing the UUA codon is specified by *bldA*, and the corresponding TTA codon
414 occurs specifically in many genes involved in development and antibiotic production, making
415 those genes *bldA* dependent (71-74). The CTA leucine codon is also very rare in *Streptomyces*
416 genes, representing only 0.31% of all leucine codons in *S. coelicolor*. In this light, it would be
417 interesting to see if SCOt23 also plays a role in the control of developmental gene expression,
418 and what the role is of Gbn in the control of SCOt23 transcription.

419 Interestingly, many of the strongest enrichments of binding in promoter regions (15-
420 to 20-fold) were found in the *dcw* cluster, which encompasses many key genes for cell
421 division and cell-wall synthesis, including genes for the cell division scaffold FtsZ and DivIVA
422 which is essential for growth (Figure 4G). However, transcriptomics data shows that the
423 expression of *dcw* gene cluster was not significantly altered by deletion of *gbn*. This may be
424 explained by the fact that NAPs typically affect global gene expression through remodelling
425 of DNA structure and not by direct activation or repression of transcription (21). Interestingly,
426 strong binding was also observed in ribosomal protein operons (Figure 4H). In *Streptomyces*,
427 development and the regulation of secondary metabolism are associated with changes in
428 the expression of ribosomal proteins (75, 76). This phenomenon may be related with the
429 resistance mechanism of many antibiotics (77, 78).

430 **Deletion of *gbn* accelerates sporulation of *S. coelicolor***

431 To obtain insights into the possible role of Gbn in the life cycle of *S. coelicolor*, a knock-out
432 mutant was generated using a strategy published previously (79). For this, the +1 to +207
433 region of the gene was replaced by the apramycin resistance cassette *aac(C)IV,* which was
434 flanked by *loxP* sites. The *loxP* sites allowed removal of the cassette using Cre recombinase,

435 important to minimise polar effects. To genetically complement the mutant and see if the
436 wild-type phenotype would be restored, the -565/+228 region of *gbn* was amplified from the
437 *S. coelicolor* chromosome and cloned into pHJL401, a shuttle vector that is useful for genetic
438 complementation due to its low copy number in streptomycetes (80). To also analyse the
439 effect of enhanced expression of *gbn*, a second strain was constructed using CRISPR-Cas9
440 wherein the native promoter of *gbn* (-157 to +4, start codon modified to ATG) was replaced
441 by the strong constitutive *ermE* promoter (81). For details see the Materials and Methods
442 section.

443 The morphology of *gbn* null mutant did not show significant changes comparing to
444 wild type, except for reduced production of the blue-pigmented antibiotic actinorhodin
445 (Figure S1). However, the mutant showed slightly accelerated development in comparison to
446 the parental strain. To investigate this altered timing of development in more detail, time-
447 lapse imaging was performed on confluent mycelial lawns. This method allows monitoring
448 multiple morphological characteristics and in particular quantifying differences in timing of
449 the individual developmental stages, based on pigmentation of the mycelia. When aerial
450 hyphae are formed, the brightness increases due to the increased density of colourless
451 hyphae. The brightness decreases when grey-pigmented spores are produced (see Material
452 and Methods section for details). Deletion of *gbn* led to a 2-5 h acceleration of development
453 as compared to the parental strain (Figure 2A). 54 h after inoculation, the light intensity of
454 the *gbn* mutant again increased (Figure 2A), which may be due to premature germination
455 and renewed growth. The acceleration of development was reversed when a wild-type copy
456 of *gbn* was re-introduced into the null mutant, a genetic complementation experiment that
457 confirmed the accelerated development was indeed primarily due to the deletion of *gbn*.
458 Conversely, the enhanced and constitutive expression of *gbn* delayed sporulation by

459 approximately 17 h (Figure 2A). Taken together, this strongly suggests that the expression of
460 *gbn* correlates to the timing of sporulation.

461 Closer examination of the spores by scanning electron microscopy (SEM) revealed an
462 increased length-to-width ratio of the spores in both the deletion mutant and the strain with
463 enhanced expression of *gbn* (Mann-Whitney U test, p-value < 0.01, Table 1, Figure 2B). The
464 complemented strain produced spores with a normal ratio. The statistical variation in the
465 spore length-to-width ratio showed a larger variance in the engineered strains as compared
466 to that of the parental strain (Levene's test, p-value < 0.01), with the *gbn* deletion mutant
467 showing the largest variance. Complementation of the mutant reduced the variance to a
468 level close to that of the wild type. Thus, deletion of *gbn* altered both the timing of
469 sporulation and the morphology of the spores.

470 **Correlation between Gbn binding and gene expression**

471 To obtain insights into the effect of Gbn on global gene expression, we performed
472 transcriptomics analysis, comparing transcriptional changes between the *gbn* mutant with
473 its parent during vegetative and aerial growth. For this, strains were grown on minimum
474 agar medium covered with cellophane and RNA samples were prepared in triplicate after 24
475 h (vegetative growth) and 45 h (aerial growth). A table with all read counts for each gene can
476 be found at GEO accession GSE186136; differential expression analysis table can be found in
477 Table S2.

478 Table 2 lists the expression of all genes associated with strong Gbn binding in their
479 promoter regions. This surprisingly revealed no significant changes between parent and *gbn*
480 mutant, indicating that Gbn does not directly affect the level of transcription of target genes.
481 As an example, Gbn binds to the promoter region of *atrA* (Figure 4E), a global regulatory

482 gene which among others *trans*-activates the transcription of *actII*-ORF4, the cluster-situated
483 activator of the actinorhodin biosynthetic gene cluster (82). Binding may be reciprocal, as an
484 AtrA binding site was identified in the *gbn* promoter region using the PREDetector algorithm
485 (83), with a high confidence score of 13.3, similar to that for the AtrA binding element
486 upstream of *actII*-ORF4. However, *atrA* transcription did not change in the *gbn* null mutant.
487 In total 13 binding regions were detected in the *red* cluster at both time points, with a very
488 strong binding site upstream of *redY* (Figure 4H). Again, transcription was not altered
489 significantly between wild-type and *gbn* mutant.

490 Interestingly, the average level of gene expression during vegetative growth was
491 significantly higher than during aerial growth for genes with Gbn binding in promoter regions,
492 which was even more pronounced in a *gbn* mutant background (Table 3). Such an effect was
493 not seen for genes where Gbn bound only to the coding regions (+50 to end of gene). Thus,
494 binding of Gbn to the promoter regions of genes has a subtle but significant suppressive
495 effect on overall gene expression, which cannot be explained by direct effect on the
496 expression of specific genes.

497 Transcriptomics data showed a strong increase in the transcription of *gbn* itself at 45
498 h as compared to 24 h (Figure 6A). This increased expression of *gbn* during development was
499 also seen in transcriptome data published by others; all public datasets showed an increase
500 in *gbn* expression over time and in a medium-independent manner (14, 75, 84, 85).
501 Interestingly, repeated ChIP-seq experiments on 24 h old samples or younger failed to pull
502 down any DNA, while after 25 h, some DNA was immunoprecipitated. This correlates to the
503 observation that Gbn expression is initiated after around 24 h after start of growth.

504 **Gbn stimulates secondary metabolic pathways**

505 Principal component analysis (PCA) of transcriptomics data showed an increased distance for
506 aerial growth stage samples (Figure 6B). This suggests that deletion of *gbn* leads to more
507 fundamental transcriptomic changes at a later growth stage. In the *gbn* null mutant, many
508 secondary metabolomic genes were down-regulated, especially during vegetative growth
509 phase (Table 4, Figure 6C), and most of these genes did not show Gbn binding. The down-
510 regulated genes include those of the *act* BGC (SCO5071-SCO5092) and of the SoxR regulon,
511 which can be activated by γ-actinorhodin (86). This is consistent with the observed slower
512 blue pigmentation of the mutant on SFM agar plates (Figure S1). Additionally, genes for a
513 carboxylesterase regulon (SCO0319-SCO0321) and a xylobiose transporter regulon (*bxlEFGA*;
514 SCO7028-SCO7031) (87) were also down-regulated.

515 During aerial growth (45 h), many more genes are affected in the *gbn* null mutant
516 (Table 4, Figure 6D), which correlates well with the increased PCA distance at this time point.
517 An obvious category of down-regulated genes belongs to the GlnR regulon (gene list
518 obtained from (88)), especially for SCO2195, *glnII* (SCO2210), *nirB* (SCO2486, SCO2487),
519 *amtB* (SCO5583), *glnKD* (SCO5584, SCO5585) (Table 4). Among the up-regulated genes
520 during aerial growth in the mutant, SCO4615-SCO4627 were most strongly up regulated.
521 These include phage-related genes, encoding among others an integrase (SCO4615), an
522 excisionase (SCO4616), and the DNA transfer-related TraA1 (SCO4621), TraB1 (SCO4620) and
523 SpdB2 (SCO4625) (89). In the vicinity of this gene cluster lie more genes that are significantly
524 up regulated in the *gbn* mutant during aerial growth. These genes include an atypical two-
525 component system that supresses antibiotic production (SCO4596-SCO4598) (90) and the
526 second NADH dehydrogenase I gene cluster (*nuo2*; SCO4599-SCO4608). The *nuo*2 cluster
527 lacks genes for NuoCDEFG and L subunits and its function is unclear. Conversely, *nuoJKLMN*

528 (SCO4571-SCO4575) from the canonical *nuo* gene cluster (91) were down-regulated in the
529 *gbn* mutant (Table 4, Figure 6D). The gene products of the *nuo2* cluster may cooperates with
530 the proteins derived from the canonical *nuo* operon. The altered balance in the expression
531 of components of the NADH dehydrogenase complex derived from the *nuo* and *nuo2*
532 operons may lead to a hybrid machinery. Whether this is indeed the case, and how this
533 affects energy production during development, remains to be elucidated.

534 In conclusion, transcriptomics data confirmed the early sporulation and suppressed
535 secondary metabolic processes in *gbn* null mutant. It is tempting to propose that these two
536 processes have causal relationship that directly related with Gbn. However, direct links
537 needs to be found to confirm this relationship and understand the regulation pathway that
538 drives these changes related to deletion of *gbn*.

539 **Summary**

540 Gbn (SCO1839) protein is a conserved NAP among Actinobacteria species, it is a highly
541 pleiotropic DNA binding protein that plays a role in the (timing of) development and
542 antibiotic production of *S. coelicolor*. Strains in which *gbn* had been deleted or over-
543 expressed showed accelerated and delayed sporulation, respectively, suggesting that Gbn
544 plays a role in the accurate timing of development. Both *in vivo* and *in vitro* experiments
545 revealed that Gbn binds specifically to a GATC DNA motif and especially those followed by
546 either AT or TT. In addition, we have shown that the methylation of adenine in the GATC
547 sequence reduced affinity of Gbn for its binding sites. Transcriptomics analysis showed Gbn
548 has a suppressive effect on the genes that bound by Gbn on their promoter regions. This
549 suppressive effect, together with thousands genome locations that Gbn binds to, may
550 instead leads to stimulation of secondary metabolic pathways and postponed sporulation.

These results described Gbn as a representative member of a new NAP family that might play important roles in the development and antibiotics production in streptomycetes.

DATA AND CODE AVAILABILITY

Clean ChIP-Seq reads and binding region identification (peak calling) files are available at GEO database (92) with accession number GSE165795. [For reviewers: go to https://www.ncbi.nlm. nih.gov/geo/query/acc.cgi?acc=GSE165795, enter token **ujulcouyzxcfhmn** into the box]. Clean RNA-Seq reads and gene read-counts tables are available at GEO database with (92) with accession number GSE186136. [For reviewers: go to https://www.ncbi.nlm.nih.gov /geo/query/acc.cgi?acc=GSE186136, enter token **qryzegmcrxifrqf** into the box]. Complete ChIP-Seq analysis and transcriptomics analysis code and related data table used in this research can be found at https://github.com/snail123815/Gbn-the-SNP-publication-scripts. Time lapse scanner image analysis code can be found at https://github.com/snail123815/scanLapsePlot.

FUNDING

The work was supported by NWO VICI grant 016.160.613/533 to RTD and by NWO grant 731.014.206 to GPvW.

CONFLICT OF INTEREST

The authors declare no conflict of interests.

ACKNOWLEDGMENTS

We are grateful to Kenny McDowall for stimulating discussions.

572 **REFERENCES**

573 1. Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y, van
574 Wezel GP. 2016. Taxonomy, physiology, and natural products of Actinobacteria. Microbiol Mol Biol Rev
575 80:1-43. doi:10.1128/mmbr.00019-15.

576 2. Hopwood DA. 2007. *Streptomyces* in nature and medicine: the antibiotic makers. Oxford University Press.

577 3. Bérdy J. 2005. Bioactive microbial metabolites. J Antibiot 58:1-26. doi:10.1038/ja.2005.1.

578 4. Chater KF, Losick R. 1997. Mycelial life style of *Streptomyces coelicolor* A3(2) and its relatives., p 149-182.
579 *In* Shapiro JA, Dworkin M (ed), Bacteria as multicellular organisms. Oxford University Press, New York.

580 5. Claessen D, Rozen DE, Kuipers OP, Sogaard-Andersen L, van Wezel GP. 2014. Bacterial solutions to
581 multicellularity: A tale of biofilms, filaments and fruiting bodies. Nat Rev Microbiol 12:115-124.
582 doi:10.1038/nrmicro3178.

583 6. Flärdh K, Buttner MJ. 2009. *Streptomyces* morphogenetics: Dissecting differentiation in a filamentous
584 bacterium. Nat Rev Microbiol 7:36-49. doi:10.1038/nrmicro1968.

585 7. Bibb MJ. 2005. Regulation of secondary metabolism in streptomycetes. Curr Opin Microbiol 8:208-215.
586 doi:10.1016/j.mib.2005.02.016.

587 8. van der Heul HU, Bilyk BL, McDowall KJ, Seipke RF, van Wezel GP. 2018. Regulation of antibiotic
588 production in Actinobacteria: New perspectives from the post-genomic era. Nat Prod Rep 35:575-604.
589 doi:10.1039/C8NP00012C.

590 9. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, James KD, Harris DE, Quail MA,
591 Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A,
592 Hidalgo J, Hornsby T, Howarth S, Huang CH, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch
593 E, Rajandream MA, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K,
594 Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome
595 sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141-7.
596 doi:10.1038/417141a
597 417141a [pii].

598 10. Urem M, Świątek MA, Rigali S, van Wezel GP. 2016. Intertwining nutrient-sensory networks and the
599 control of antibiotic production in *Streptomyces*. Mol Microbiol:183-195. doi:10.1111/mmi.13464.

600 11. Dame RT, Rashid F-ZM, Grainger DC. 2020. Chromosome organization in bacteria: Mechanistic insights
601 into genome structure and function. Nat Rev Genet 21:227-242. doi:10.1038/s41576-019-0185-4.

602 12. Dillon SC, Dorman CJ. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene
603 expression. Nat Rev Microbiol 8:185-195. doi:10.1038/nrmicro2261.

604 13. Dame RT. 2005. The role of nucleoid-associated proteins in the organization and compaction of bacterial
605 chromatin. Mol Microbiol 56:858-870. doi:https://doi.org/10.1111/j.1365-2958.2005.04598.x.

606 14. Gehrke EJ, Zhang X, Pimentel-Elardo SM, Johnson AR, Rees CA, Jones SE, Hindra, Gehrke SS, Turvey S,
607 Boursalie S, Hill JE, Carlson EE, Nodwell JR, Elliot MA. 2019. Silencing cryptic specialized metabolism in
608 *Streptomyces* by the nucleoid-associated protein Lsr2. eLife 8:e47691. doi:10.7554/eLife.47691.

609 15. Salerno P, Larsson J, Bucca G, Laing E, Smith CP, Flärdh K. 2009. One of the two genes encoding nucleoid-
610 associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved
611 in spore maturation. J Bacteriol 191:6489-6500. doi:10.1128/jb.00709-09.

16. Yang Y-H, Song E, Willemse J, Park S-H, Kim W-S, Kim E-j, Lee B-R, Kim J-N, van Wezel GP, Kim B-G. 2012. A novel function of *Streptomyces* integration host factor (sIHF) in the control of antibiotic production and sporulation in *Streptomyces coelicolor*. Antonie van Leeuwenhoek 101:479-492. doi:10.1007/s10482-011-9657-z.

17. Swinger KK, Rice PA. 2004. IHF and HU: Flexible architects of bent DNA. Curr Opin Struct Biol 14:28-35. doi:https://doi.org/10.1016/j.sbi.2003.12.003.

18. Kamashev D, Balandina A, Rouviere-Yaniv J. 1999. The binding motif recognized by HU on both nicked and cruciform DNA. The EMBO Journal 18:5434-5444. doi:https://doi.org/10.1093/emboj/18.19.5434.

19. Bradshaw E, Saalbach G, McArthur M. 2013. Proteomic survey of the *Streptomyces coelicolor* nucleoid. J Proteom 83:37-46. doi:10.1016/j.jprot.2013.02.033.

20. Bush MJ, Chandra G, Al-Bassam MM, Findlay KC, Buttner MJ. 2019. BldC delays entry into development to produce a sustained period of vegetative growth in *Streptomyces venezuelae*. mBio 10:e02812-18. doi:10.1128/mBio.02812-18.

21. Dorman CJ, Schumacher MA, Bush MJ, Brennan RG, Buttner MJ. 2020. When is a transcription factor a NAP? Curr Opin Microbiol 55:26-33. doi:https://doi.org/10.1016/j.mib.2020.01.019.

22. Kim S, Traag B, Hasan A, McDowall K, Kim B-G, van Wezel GP. 2015. Transcriptional analysis of the cell division-related *ssg* genes in *Streptomyces coelicolor* reveals direct control of *ssgR* by AtrA. Antonie van Leeuwenhoek 108:201-213. doi:10.1007/s10482-015-0479-2.

23. Traag BA, Kelemen GH, Van Wezel GP. 2004. Transcription of the sporulation gene *ssgA* is activated by the IclR-type regulator SsgR in a whi-independent manner in *Streptomyces coelicolor* A3(2). Mol Microbiol 53:985-1000.

24. Keijser BJ, Noens EE, Kraal B, Koerten HK, van Wezel GP. 2003. The *Streptomyces coelicolor ssgB* gene is required for early stages of sporulation. FEMS Microbiol Lett 225:59-67.

25. Noens EE. 2007. Control of sporulation-specific cell division in *Streptomyces coelicolor*Leiden University, Leiden.

26. Noens EE, Mersinias V, Traag BA, Smith CP, Koerten HK, van Wezel GP. 2005. SsgA-like proteins determine the fate of peptidoglycan during sporulation of *Streptomyces coelicolor*. Mol Microbiol 58:929-44.

27. Willemse J, Borst JW, de Waal E, Bisseling T, van Wezel GP. 2011. Positive control of cell division: FtsZ is recruited by SsgB during sporulation of *Streptomyces*. Genes Dev 25:89-99.

28. MacNeil DJ, Gewain KM, Ruby CL, Dezeny G, Gibbons PH, MacNeil T. 1992. Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. Gene 111:61-68. doi:http://dx.doi.org/10.1016/0378-1119(92)90603-M.

29. Paget MSB, Chamberlin L, Atrih A, Foster SJ, Buttner MJ. 1999. Evidence that the extracytoplasmic function sigma factor ς^E is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). J Bacteriol 181:204-211.

30. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical *Streptomyces* genetics. John Innes Foundation, Norwich, U.K.

31. Colson S, Stephan J, Hertrich T, Saito A, van Wezel GP, Titgemeyer F, Rigali S. 2007. Conserved *cis*-acting elements upstream of genes composing the chitinolytic system of streptomycetes are DasR-responsive elements. J Mol Microbiol Biotechnol 12:60-66.

32. Vara J, Lewandowska-Skarbek M, Wang YG, Donadio S, Hutchinson CR. 1989. Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). J Bacteriol 171:5872-5881.

33. Zhang L, Willemse J, Hoskisson PA, van Wezel GP. 2018. Sporulation-specific cell division defects in *ylmE* mutants of *Streptomyces coelicolor* are rescued by additional deletion of *ylmD*. Sci Rep 8:7328. doi:10.1038/s41598-018-25782-1.

34. Fedoryshyn M, Welle E, Bechthold A, Luzhetskyy A. 2008. Functional expression of the Cre recombinase in Actinomycetes. Appl Microbiol Biotechnol 78:1065-1070. doi:10.1007/s00253-008-1382-9.

35. Khodakaramian G, Lissenden S, Gust B, Moir L, Hoskisson PA, Chater KF, Smith MCM. 2006. Expression of Cre recombinase during transient phage infection permits efficient marker removal in *Streptomyces*. Nucleic Acids Res 34:e20-e20. doi:10.1093/nar/gnj019.

36. Cobb RE, Wang Y, Zhao H. 2015. High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. ACS Synth Biol 4:723-728. doi:10.1021/sb500351f.

37. Motamedi H, Shafiee A, Cai S-J. 1995. Integrative vectors for heterologous gene expression in *Streptomyces* spp. Gene 160:25-31. doi:http://dx.doi.org/10.1016/0378-1119(95)00191-8.

38. Sambrook J, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold spring harbor laboratory press, New York.

39. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res 46:W200-W204. doi:10.1093/nar/gky448.

40. Eddy SR. 2011. Accelerated profile HMM searches. PLoS Comput Biol 7:e1002195. doi:10.1371/journal.pcbi.1002195.

41. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498-2504. doi:10.1101/gr.1239303.

42. Piette A, Derouaux A, Gerkens P, Noens EEE, Mazzucchelli G, Vion S, Koerten HK, Titgemeyer F, De Pauw E, Leprince P, van Wezel GP, Galleni M, Rigali S. 2005. From dormant to germinating spores of *Streptomyces coelicolor* A3(2): New perspectives from the *crp* null mutant. J Proteome Res 4:1699-1708. doi:10.1021/pr050155b.

43. Keijser BJ, Noens EE, Kraal B, Koerten HK, van Wezel GP. 2003. The *Streptomyces coelicolor ssgB* gene is required for early stages of sporulation. FEMS Microbiol Lett 225:59-67.

44. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357. doi:10.1038/nmeth.1923
https://www.nature.com/articles/nmeth.1923#supplementary-information.

45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079. doi:10.1093/bioinformatics/btp352.

46. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137. doi:10.1186/gb-2008-9-9-r137.

47. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B, de Hoon MJL. 2009. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25:1422-1423. doi:10.1093/bioinformatics/btp163.

694 48. Machanick P, Bailey TL. 2011. MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics 27:1696-
695 1697. doi:10.1093/bioinformatics/btr189.

696 49. Dale RK, Pedersen BS, Quinlan AR. 2011. Pybedtools: A flexible Python library for manipulating genomic
697 datasets and annotations. Bioinformatics 27:3423-3424. doi:10.1093/bioinformatics/btr539.

698 50. Quinlan AR. 2014. BEDTools: The swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics
699 47:11.12.01-11.12.34. doi:10.1002/0471250953.bi1112s47.

700 51. Kieser T, Hopwood DA. 1991. Genetic manipulation of *Streptomyces*: integrating vectors and gene
701 replacement, p 430-458, Methods Enzymol, vol 204. Academic Press.

702 52. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics
703 34:i884-i890. doi:10.1093/bioinformatics/bty560.

704 53. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357-359.
705 doi:10.1038/nmeth.1923.

706 54. Liao Y, Smyth GK, Shi W. 2013. featureCounts: An efficient general purpose program for assigning
707 sequence reads to genomic features. Bioinformatics 30:923-930. doi:10.1093/bioinformatics/btt656.

708 55. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data
709 with DESeq2. Genome Biol 15. doi:10.1186/s13059-014-0550-8.

710 56. Zhu A, Ibrahim JG, Love MI. 2018. Heavy-tailed prior distributions for sequence count data: removing the
711 noise and preserving large differences. Bioinformatics 35:2084-2092. doi:10.1093/bioinformatics/bty895.

712 57. Mahr K, van Wezel GP, Svensson C, Krengel U, Bibb MJ, Titgemeyer F. 2000. Glucose kinase of
713 Streptomyces coelicolor A3(2): large-scale purification and biochemical analysis. Antonie Van
714 Leeuwenhoek 78:253-61.

715 58. Henneman B, Heinsman J, Battjes J, Dame RT. 2018. Quantitation of DNA-binding affinity using tethered
716 particle motion, p 257-275, Bacterial Chromatin. Springer.

717 59. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA,
718 Smart A, Sonnhammer EL L, Hirsh L, Paladin L, Piovesan D, Tosatto SC E, Finn RD. 2018. The Pfam protein
719 families database in 2019. Nucleic Acids Res 47:D427-D432. doi:10.1093/nar/gky995.

720 60. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: Protein structure and function
721 prediction. Nat Methods 12:7-8. doi:10.1038/nmeth.3213.

722 61. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The
723 protein data bank. Nucleic Acids Res 28:235-242. doi:10.1093/nar/28.1.235.

724 62. Barras F, Marinus MG. 1989. The great GATC: DNA methylation in *E. coli*. Trends Genet 5:139-143.
725 doi:https://doi.org/10.1016/0168-9525(89)90054-1.

726 63. May MS, Hattman S. 1975. Analysis of bacteriophage deoxyribonucleic acid sequences methylated by
727 host- and R-factor-controlled enzymes. J Bacteriol 123:768-770.

728 64. Flett F, Mersinias V, Smith CP. 1997. High efficiency intergeneric conjugal transfer of plasmid DNA from
729 *Escherichia coli* to methyl DNA-restricting Streptomycetes. FEMS Microbiol Lett 155:223-229.
730 doi:10.1111/j.1574-6968.1997.tb13882.x.

731 65. Liu G, Ou H-Y, Wang T, Li L, Tan H, Zhou X, Rajakumar K, Deng Z, He X. 2010. Cleavage of
732 phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. PLoS
733 Genet 6:e1001253. doi:10.1371/journal.pgen.1001253.

734 66. González-Cerón G, Miranda-Olivares OJ, Servín-González L. 2009. Characterization of the methyl-specific
735 restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases.
736 FEMS Microbiol Lett 301:35-43. doi:10.1111/j.1574-6968.2009.01790.x.

737 67. van der Valk RA, Laurens N, Dame RT. 2017. Tethered particle motion analysis of the DNA binding
738 properties of architectural proteins, p 127-143. *In* Espéli O (ed), The bacterial nucleoid: methods and
739 protocols. Springer New York, New York, NY.

740 68. Qin L, Bdira FB, Sterckx YGJ, Volkov AN, Vreede J, Giachin G, van Schaik P, Ubbink M, Dame Remus T. 2020.
741 Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. Nucleic Acids Res
742 48:2156-2172. doi:10.1093/nar/gkz1226.

743 69. van der Valk RA, Vreede J, Qin L, Moolenaar GF, Hofmann A, Goosen N, Dame RT. 2017. Mechanism of
744 environmentally driven conformational changes that modulate H-NS DNA-bridging activity. eLife 6:e27369.
745 doi:10.7554/eLife.27369.

746 70. Driessen RPC, Sitters G, Laurens N, Moolenaar GF, Wuite GJL, Goosen N, Dame RT. 2014. Effect of
747 temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. Biochemistry
748 53:6430-6438. doi:10.1021/bi500344j.

749 71. Leskiw BK, Lawlor EJ, Fernandez-Abalos JM, Chater KF. 1991. TTA codons in some genes prevent their
750 expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. Proc Natl Acad Sci USA
751 88:2461-2465. doi:10.1073/pnas.88.6.2461.

752 72. Li W, Wu J, Tao W, Zhao C, Wang Y, He X, Chandra G, Zhou X, Deng Z, Chater KF, Tao M. 2007. A genetic
753 and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for
754 regulation by a developmentally significant tRNA. FEMS Microbiol Lett 266:20-28. doi:10.1111/j.1574-
755 6968.2006.00494.x.

756 73. Lawlor EJ, Baylis HA, Chater KF. 1987. Pleiotropic morphological and antibiotic deficiencies result from
757 mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). Genes Dev 1:1305-
758 1310. doi:10.1101/gad.1.10.1305.

759 74. Chater KF, Chandra G. 2008. The use of the rare UUA codon to define "Expression Space" for genes
760 involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. J
761 Microbiol 46:1-11. doi:10.1007/s12275-007-0233-1.

762 75. Huang J, Lih C-J, Pan K-H, Cohen SN. 2001. Global analysis of growth phase responsive gene expression
763 and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays.
764 Genes Dev 15:3183-3192. doi:10.1101/gad.943401.

765 76. Blanco G, Rodicio MR, Puglia AM, Méndez C, Thompson CJ, Salas JA. 1994. Synthesis of ribosomal proteins
766 during growth of *Streptomyces coelicolor*. Mol Microbiol 12:375-385. doi:https://doi.org/10.1111/j.1365-
767 2958.1994.tb01027.x.

768 77. Xu J, Tozawa Y, Lai C, Hayashi H, Ochi K. 2002. A rifampicin resistance mutation in the *rpoB* gene confers
769 ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). Mol Genet Genomics
770 268:179-189. doi:10.1007/s00438-002-0730-1.

771 78. Shima J, Hesketh A, Okamoto S, Kawamoto S, Ochi K. 1996. Induction of actinorhodin production by *rpsL*
772 (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans*
773 and *Streptomyces coelicolor* A3(2). J Bacteriol 178:7276-7284. doi:10.1128/jb.178.24.7276-7284.1996.

774 79. Świątek MA, Tenconi E, Rigali S, van Wezel GP. 2012. Functional analysis of the N-acetylglucosamine
775 metabolic genes of *Streptomyces coelicolor* and role in the control of development and antibiotic
776 production. J Bacteriol 194:1136-1144. doi:10.1128/JB.06370-11.

80. van Wezel GP, White J, Hoogvliet G, Bibb MJ. 2000. Application of *redD*, the transcriptional activator gene of the undecylprodigiosin biosynthetic pathway, as a reporter for transcriptional activity in *Streptomyces coelicolor* A3 (2) and *Streptomyces lividans*. J Mol Microbiol Biotechnol 2:551-556.

81. Bibb MJ, Janssen GR, Ward JM. 1985. Cloning and analysis of the promoter region of the erythromycin resistance gene *(ermE*) of *Streptomyces erythraeus*. Gene 38:215-26.

82. Uguru GC, Stephens KE, Stead JA, Towle JE, Baumberg S, McDowall KJ. 2005. Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. Mol Microbiol 58:131-150. doi:10.1111/j.1365-2958.2005.04817.x.

83. Tocquin P, Naome A, Jourdan S, Anderssen S, Hiard S, van Wezel GP, Hanikenne M, Baurain D, Rigali S. 2016. PREDetector 2.0: Online and enhanced version of the prokaryotic regulatory elements detector tool. bioRxiv. doi:10.1101/084780.

84. Castro-Melchor M, Charaniya S, Karypis G, Takano E, Hu W-S. 2010. Genome-wide inference of regulatory networks in *Streptomyces coelicolor*. BMC Genom 11:578. doi:10.1186/1471-2164-11-578.

85. Nieselt K, Battke F, Herbig A, Bruheim P, Wentzel A, Jakobsen OM, Sletta H, Alam MT, Merlo ME, Moore J, Omara WA, Morrissey ER, Juarez-Hermosillo MA, Rodriguez-Garcia A, Nentwich M, Thomas L, Iqbal M, Legaie R, Gaze WH, Challis GL, Jansen RC, Dijkhuizen L, Rand DA, Wild DL, Bonin M, Reuther J, Wohlleben W, Smith MC, Burroughs NJ, Martin JF, Hodgson DA, Takano E, Breitling R, Ellingsen TE, Wellington EM. 2010. The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. BMC Genomics 11:10. doi:10.1186/1471-2164-11-10.

86. Shin J-H, Singh AK, Cheon D-J, Roe J-H. 2011. Activation of the SoxR regulon in *Streptomyces coelicolor* by the extracellular form of the pigmented antibiotic actinorhodin. J Bacteriol 193:75-81. doi:doi:10.1128/JB.00965-10.

87. Tsujibo H, Kosaka M, Ikenishi S, Sato T, Miyamoto K, Inamori Y. 2004. Molecular characterization of a high-affinity xylobiose transporter of *Streptomyces thermoviolaceus* OPC-520 and its transcriptional regulation. J Bacteriol 186:1029-1037. doi:doi:10.1128/JB.186.4.1029-1037.2004.

88. Sola-Landa A, Rodríguez-García A, Amin R, Wohlleben W, Martín JF. 2012. Competition between the GlnR and PhoP regulators for the *glnA* and *amtB* promoters in *Streptomyces coelicolor*. Nucleic Acids Res 41:1767-1782. doi:10.1093/nar/gks1203.

89. Servn-Gonzlez L, Sampieri A, Cabello J, Galvn L, Jurez V, Castro C. 1995. Sequence and functional analysis of the *Streptomyces phaeochromogenes* plasmid pJV1 reveals a modular organization of *Streptomyces* plasmids that replicate by rolling circle. Microbiology 141:2499-2510. doi:https://doi.org/10.1099/13500872-141-10-2499.

90. Rodríguez H, Rico S, Yepes A, Franco-Echevarría E, Antoraz S, Santamaría RI, Díaz M. 2015. The two kinases, AbrC1 and AbrC2, of the atypical two-component system AbrC are needed to regulate antibiotic production and differentiation in *Streptomyces coelicolor*. Front Microbiol 6. doi:10.3389/fmicb.2015.00450.

91. Friedrich T, Dekovic DK, Burschel S. 2016. Assembly of the *Escherichia coli* NADH:ubiquinone oxidoreductase (respiratory complex I). Biochim Biophys Acta - Bioenerg 1857:214-223. doi:https://doi.org/10.1016/j.bbabio.2015.12.004.

92. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: Archive for functional genomics data sets - update. Nucleic Acids Res 41:D991-5. doi:10.1093/nar/gks1193.

Table 1. Spore length to width ratio of all strains comparing with parent strain M145

Strain	n	Median	Mann–Whitney *U* test of difference		Variance	Levene's test of equal variance	
			U	*p*-value		*W*	*p*-value
M145	560	1.49	-	-	0.069	-	-
Δ*gbn*	332	1.69	1.2×10^{5}	9.3×10^{-16}	0.145	49.51	3.93×10^{-12}
Δ*gbn*+*gbn*	556	1.50	1.6×10^{5}	0.40	0.099	13.49	2.52×10^{-4}
PermE-*gbn*	441	1.60	1.5×10^{5}	1.4×10^{-7}	0.125	34.89	4.78×10^{-9}

Table 2. Genes with strong Gbn binding event at the promoter region.

Gene ID	Gene product	Binding width		Binding summit position**		Fold enrichment	
		25h	48h	25h	48h	25h	48h
SCO1311*	hypothetical protein with tRNA edit domain	271	ND†	-76	-72†	13.41	2.42†
SCO1839*	transcriptional regulator	547	530	-326	-324	67.75	58.18
SCO2077*	DivIVA	251	300	-90	-91	17.19	13.48
SCO2081*	YlmD	271	378	-348	-348	10.44	17.18
SCO2084*	MurG (UDPdiphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase)	280	313	-280	-273	13.22	13.41
SCO2086*	MurD (UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase)	482	471	-90	-100	9.63	10.62
SCO2088*	MurF (UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate- D-alanyl-alanyl ligase)	227	285	-5	-10	7.37	14.21
SCO3213	TrpG (anthranilate synthase component II)	247	299	-240	-236	7.85	10.69
SCO3224	ABC transporter ATP-binding protein	349	367	+33	+48	14.23	15.90
SCO3225	AbsA1 (two component sensor kinase)	Δ	Δ	-168	-183	Δ	Δ
SCO4702	RplC (50S ribosomal protein L3)	459	511	-34	-37	12.19	22.16
SCO4707	RplV (50S ribosomal protein L22)	290	388	-198	-196	15.80	25.17
SCO4713	RplX (50S ribosomal protein L24)	372	406	-7	-5	58.66	44.69
SCO4714	RplE (50S ribosomal protein L5)	Δ	Δ	-330	-328	Δ	Δ
SCO4717	RplF (50S ribosomal protein L6)	372	418	-324	-326	25.78	23.66
SCO4718	RplR (50S ribosomal protein L18)	252	641	+46	+45	10.27	19.35
SCO4721	RplO (50S ribosomal protein L15)	380	432	+22	+21	34.82	29.13
SCO4726	RpmJ (50S ribosomal protein L36)	310	761	-3	+21	6.69	15.32
SCO4727	RpsM (30S ribosomal protein S13)	Δ	Δ	-305	-282	Δ	Δ
SCO5880*	RedY (RedY protein)	286	347	-168	-161	14.59	18.85
SCOt02	tRNA Val (anticodon CAC)	509	559	-136	-138	56.47	55.65
SCOt17	tRNA Gly (anticodon UCC)	295	348	-325	-320	20.40	28.92
SCOt23*	tRNA Leu (anticodon UAG)	338	280	-78	-90	25.89	8.10
SCOt49	tRNA Thr (anticodon GGU)	261	294	-170	-168	15.21	13.95
SCOt50	tRNA Met (anticodon CAU)	Δ	Δ	-289	-287	Δ	Δ

* Shown in Figure 4

** Relative to the start of the gene (+1)

† Binding region not detected by the software, local summit position and corresponding fold enrichment are shown instead

Δ same as above, shared binding region as the gene above

Table 3. Expression differences between genes where Gbn bound in the promoter region.

Conditions		bias due to Gbn promoter binding #			Mann–Whitney *U* test	
		N	Mean	Median	*U*	*p*-value
Gene expression (TPM)	WT 24 h	983 / 6852	195.2 / 110.1	30.0 / 22.9	$3.6 / 3.1 \times 10^6$	2.8×10^{-5}
	WT 45 h	954 / 6881	202.5 / 107.9	33.6 / 29.3	$3.5 / 3.1 \times 10^6$	2.3×10^{-3}
	Δ*gbn* 24 h	983 / 6852	200.1 / 109.1	32.3 / 22.8	$3.6 / 3.1 \times 10^6$	1.5×10^{-6}
	Δ*gbn* 45 h	954 / 6881	189.4 / 109.9	37.9 / 30.6	$3.5 / 3.0 \times 10^6$	1.6×10^{-4}
Log2 fold change (Δ*gbn* / WT)	24 h	1185 / 6665	0.019 / -0.012	0.006 / 0.004	$4.3 / 3.6 \times 10^6$	3.3×10^{-6}
	45 h	1185 / 6665	0.047 / 0.011	-0.014 / -0.040	$4.1 / 3.7 \times 10^6$	9.7×10^{-4}

expression ratio between genes where Gbn binds in the promoter region and those where Gbn did either not bind at all, or not in the promoter region.

Table 4. Genes with high expression that changes the most by deletion of *gbn*

SCO ID*	Gene name / product	Group	Log_2 fold change Δ*gbn* vs. WT+		Posterior SD		Mean**
			24h	45h	24h	45h	
SCO0901	hypothetical protein		0.89	*2.58*	0.12	0.11	849.76
SCO0902	hypothetical protein		1.71	*3.24*	0.13	0.12	1290.94
SCO1293	hypothetical protein		-0.03	*-3.90*	0.16	0.32	643.65
SCO1838	enoyl-CoA hydratase/isomerase		*-3.00*	-1.66	0.09	0.08	2505.38
SCO1840	ABC transporter ATP-binding protein		0.20	*2.36*	0.08	0.08	2144.41
SCO1904	transcriptional regulator		0.36	*2.07*	0.48	0.36	1575.40
SCO1905	hypothetical protein		0.15	*2.63*	0.26	0.52	7180.96
SCO2063	small hydrophilic protein		-0.05	*-2.07*	0.17	0.12	2198.63
SCO2100	transcriptional regulator		-0.12	*2.93*	0.13	0.13	621.86
SCO2195	hypothetical protein	GlnR regulon	-0.11	*-2.59*	0.10	0.11	1012.27
SCO2205	hypothetical protein		0.01	*2.21*	0.13	0.15	1347.51
SCO2210	*glnII*	GlnR regulon	-0.08	*-5.89*	0.20	0.58	7109.18
SCO2211	hypothetical protein		-0.10	*-4.25*	0.21	0.60	3739.16
SCO2473	nitrate reductase		-0.22	*-6.37*	0.41	0.44	594.99
SCO2486	NAD(P)H-nitrite reductase, large subunit	GlnR regulon	-0.14	*-7.71*	0.26	0.56	1160.61
SCO2487	NAD(P)H-nitrite reductase, large subunit	GlnR regulon	-0.07	*-7.54*	0.20	0.84	1965.02
SCO2958	bifunctional uroporphyrinogen-III synthetase/response regulator domain		-0.11	*-3.55*	0.22	0.53	506.48
SCO3022	hypothetical protein		0.11	*2.63*	0.13	0.15	610.77
SCO3169	hypothetical protein		-0.79	*-4.37*	0.18	0.12	6467.30
SCO4265	transport integral membrane protein	SoxR regulon	*-3.20*	0.13	0.37	0.28	2148.27
SCO4266	oxidoreductase	SoxR regulon	*-4.18*	-0.59	0.53	0.56	873.71
SCO4597	*abrC2*		-0.04	*2.31*	0.11	0.12	648.14
SCO4598	*abrC1*		0.10	*2.04*	0.11	0.12	721.45
SCO4601	dehydrogenase		0.08	*2.13*	0.08	0.09	1556.77
SCO4602	*nuoH2*	*nuo2* cluster	0.17	*2.66*	0.12	0.11	1328.48
SCO4603	*nuoI2*	*nuo2* cluster	-0.05	*2.14*	0.11	0.12	538.95
SCO4604	*nuoJ2*	*nuo2* cluster	0.11	*2.61*	0.12	0.13	1008.31
SCO4605	*nuoK2*	*nuo2* cluster	0.14	*2.69*	0.14	0.15	583.61
SCO4606	*nuoL2*	*nuo2* cluster	0.38	*2.86*	0.12	0.11	4379.23
SCO4607	*nuoM2*	*nuo2* cluster	0.23	*2.57*	0.13	0.12	2117.78
SCO4615	integrase	Possible HGT cluster	0.33	*4.18*	0.30	0.16	513.54
SCO4620	*traB1*	Possible HGT cluster	0.17	*4.10*	0.18	0.13	1594.84
SCO4625	hypothetical protein	Possible HGT cluster	0.09	*4.48*	0.18	0.19	568.36
SCO4949	*narJ3*		-0.22	*-2.12*	0.12	0.13	592.98
SCO5082	*actII-1*	*act* cluster	*-2.14*	0.15	0.21	0.19	1069.56
SCO5083	*actII-2*	*act* cluster	*-4.72*	0.11	0.65	0.35	4584.97
SCO5084	*actII-3*	*act* cluster	*-4.55*	0.07	0.65	0.34	4538.40
SCO5085	*actII-4*	*act* cluster	*-2.78*	0.12	0.38	0.28	838.97
SCO5583	*amtB*	GlnR regulon	-0.11	*-5.45*	0.23	0.73	7383.07
SCO5584	*glnK*	GlnR regulon	-0.12	*-5.34*	0.24	0.68	3010.06
SCO5585	*glnD*	GlnR regulon	-0.13	*-4.66*	0.25	0.61	2335.98
SCO5840	transcriptional regulator		0.02	*-2.84*	0.13	0.20	696.17
SCO6149	ribosome-associated GTPase		0.16	*2.32*	0.27	0.50	5127.49
SCO6150	ADA-like regulatory protein		0.58	*2.09*	0.47	0.36	840.27
SCO6197	hypothetical protein		0.83	*2.08*	0.12	0.13	5621.53
SCO6593	hypothetical protein		1.19	*2.19*	0.13	0.14	2282.06
SCO6682	*ramS*		-0.28	*-2.22*	0.21	0.22	823.66
SCO6729	hypothetical protein		0.21	*2.06*	0.13	0.13	664.67
SCO6811	hypothetical protein		0.18	*2.19*	0.16	0.14	871.26

* Genes with log2 fold change ≥ 2 or ≤-2, mean of expression ≥ 500.

** Mean of expression (normalized counts) for all samples

+ Colour scale of log_2 fold change: *-5* *-3* -1 1 *3* *5*

FIGURE LEGENDS

Figure 1. Distribution, phylogenetic network and structural analysis of Gbn-family proteins. **(A)** Bar plot representing the distribution of Gbn-like proteins in all genomes used in this study. Light colours indicate the number of sequenced genomes containing Gbn-like proteins in specific genera, dark colours represent the total number of sequenced genomes in each genus. **(B)** Sequence similarity network built with a threshold of 0.8 using all Gbn-like proteins detected in this study. Nodes represent Gbn-like proteins. Edges connecting the nodes represent phylogenetic distances. Colours indicate the taxonomic affiliation of each orthologue. Shapes denote whether Gbn is the primary copy in one genome. **(C)** Predicted secondary structure of Gbn, obtained from the I-TASSER output. **(D)** *In silico* structural model of the Gbn protein generated with I-TASSER.

Figure 2. Growth and spore shape analysis of *gbn* mutants. (A) scanner measurements of confluent plate brightness of *S. coelicolor* M145 and mutants (Δ*gbn*, Δ*gbn*+*gbn*, P*ermE-gbn*) grown as confluent lawns. The X-axis represents the time after inoculation and the Y-axis the normalized brightness measured from time lapse scan pictures. Mutants are presented in different colours; the timing of curve peaks are marked on top. Black vertical line illustrates 24 h time point. **(B)** Violin plot showing the distributions of the spore length to width ratio as derived from SEM images. Each strain was grown on MM agar for 5 days before imaging. ▯ with significantly ($p < 0.01$) different length to width ratio compared to *S. coelicolor* M145

(Mann–Whitney U test), **#** with significantly ($p < 0.01$) different variance compared to *S. coelicolor* M145 (Levene's test).

Figure 3. Genome-wide distribution of Gbn protein binding sites along the *S. coelicolor* genome. The outer ring shows the genome location; the middle ring shows the local average fold enrichment from ChIP-Seq analysis, 48 h sample oriented outwards, 25 h sample oriented inwards; the inner ring shows the local average G+C content. The G+C percentage above median is plotted outwards, below median inwards. Bin size for local averaging was 20,000 bp. *gbn* location have been indicated by black arrow.

Figure 4. Gbn protein binding sites analysis. (A) Gbn DNA binding motifs predicted by combining the programs MACS2 and MEME-ChIP. **(B-I)** Enrichment level (Y axis) of Gbn in target region as measured by ChIP-Seq experiment, corresponding samples of blue and orange lines are indicated by the legend in B. B-F, targeted genes with flanking regions (1000 bp) are included. Red arrow indicating target gene, other genes are coloured blue. G-I, *dcw* gene cluster, ribosomal protein gene cluster, and *red* gene cluster, respectively. Note that the y-axes of the plots are in automatic scale.

Figure 5. Binding specificity and affinity of Gbn protein. (A) Sequence of short DNA fragments used in the Electrophoretic mobility shift assays (EMSA). The motifs included in the EMSA sequences were designed with different degrees of binding strength, GATC motif is indicated by "+", GATCWT by "++", and the DNA sequence containing two GATCWTs and two GATCs is indicated by "Quadruple". GATC motifs are highlighted with the trailing (A/T)T sequence coloured red. **(B, C)** Electrophoretic mobility shift assays (EMSA) using 6×His-

tagged Gbn on synthetic 50 bp dsDNAs containing the different binding strengths GATC motifs described in (A). **(D)** EMSA using 6×His-tagged Gbn protein performed on DNA extracted from DNA methylation-deficient *E. coli* ET12567 and methylation-positive *E. coli* JM109. The *gbn* promoter region (-561 to -239) contains eight GATC motifs and was used as test sequence. Negative Control (NC) sequence derived from a random location with no detectable Gbn protein binding as shown by ChIP-Seq analysis, specifically the position 5,052,200 to 5,052,548 (349 bp). In panels B), C), and D), the protein-to-DNA molar ratios are shown on top of each gel picture, 10 µL of mixed samples was loaded per lane with DNA concentrations: (B) 5 nM; (C) 7.5 nM; and (D) 2 nM. **(E)** Root mean square displacement (RMS) of the bead for Gbn on a 685 bp DNA substrate containing the *gbn* promoter region as shown as a function of protein concentration. Error bars indicate the standard error of the data points (N ≥ 100). On the right, a close-up of the dashed region is shown.

Figure 6. Transcriptomics data of vegetative (24 h) and aerial (45 h) growth stage cells of Δgbn and wild type strain. (A) Expression of *gbn* gene in wild type strain. Error bars indicate the standard error of triplicates. **(B)** Principal component analysis of transcriptomics samples with *gbn* excluded from analysis. **(C, D)** Volcano plot showing changes in gene expression due to deletion of *gbn* (Δ*gbn* vs. wild type) at vegetative (C) and aerial growth stage (D). Interested genes are highlighted in colours other than blue. Both (C) and (D) share the same colour scheme which is shown in the legend to the right. Lighter colours indicating the change is not significant (fold change ≤ 2, p-value ≥ 0.05); size of each dot represents the normalized average expression of related conditions. Except (A), all plot data was calculated using DESeq2 with shrinkage function (55, 56).

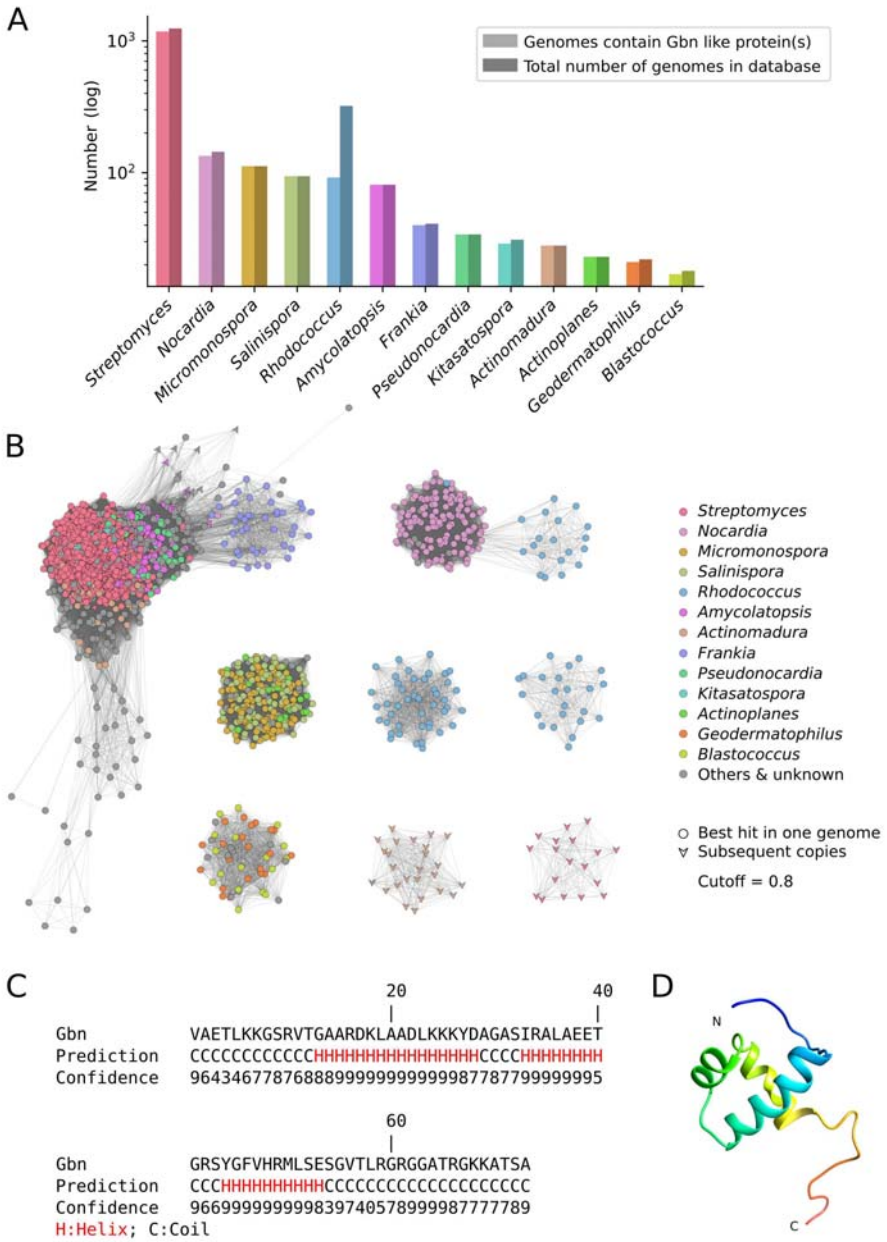
A
Genomes contain Gbn like protein(s)
Total number of genomes in database
Number (log)
10^3
10^2
Streptomyces
Nocardia
Micromonospora
Salinispora
Rhodococcus
Amycolatopsis
Frankia
Pseudonocardia
Kitasatospora
Actinomadura
Actinoplanes
Geodermatophilus
Blastococcus
B
Streptomyces
Nocardia
Micromonospora
Salinispora
Rhodococcus
Amycolatopsis
Actinomadura
Frankia
Pseudonocardia
Kitasatospora
Actinoplanes
Geodermatophilus
Blastococcus
Others & unknown
Best hit in one genome
Subsequent copies
Cutoff = 0.8
C
20 40
Gbn VAETLKKGSRVTGAARDKLAADLKKKYDAGASIRALAEET
Prediction CCCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCHHHHHHHH
Confidence 964346778768889999999999999987787799999995
60
Gbn GRSYGFVHRMLSESGVTLRGRGGATRGKKATSA
Prediction CCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC
Confidence 966999999999839740578999987777789
H:Helix; C:Coil
D
N
C

Figure 1.

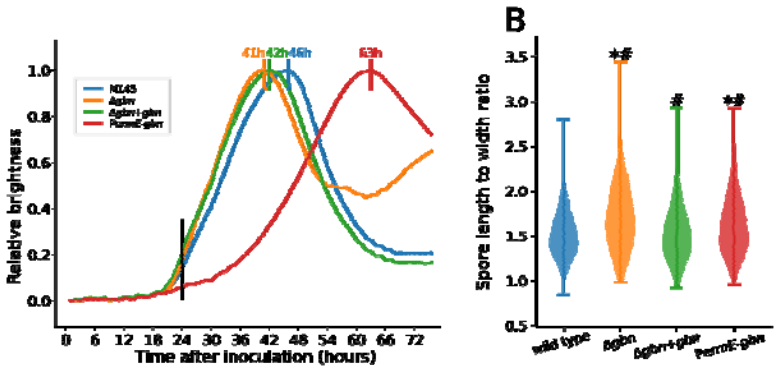
B
Relative brightness
Time after inoculation (hours)
Spore length to width ratio
wild type
Δgbn
Δgbn+gbn
PermE-gbn

Figure 2

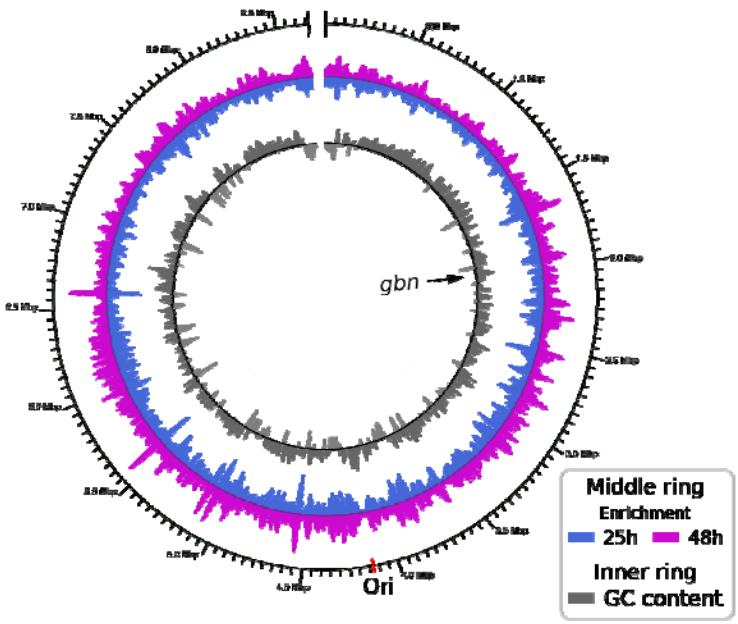
gbn
Ori
Middle ring
Enrichment
25h
48h
Inner ring
GC content

Figure 3

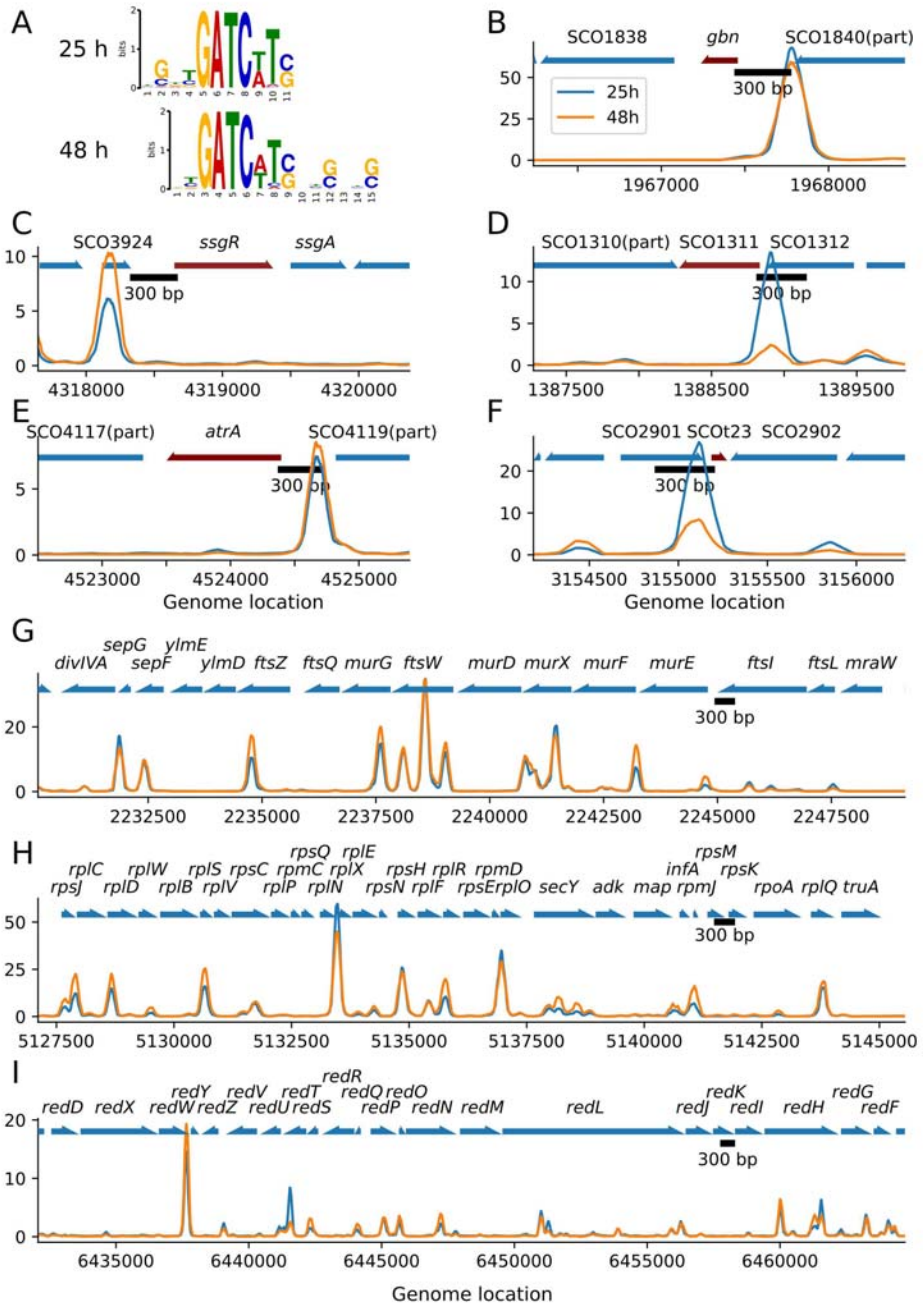
A
25 h
48 h
B
SCO1838
gbn
SCO1840(part)
300 bp
25h
48h
1967000
1968000
C
SCO3924
ssgR
ssgA
300 bp
4318000
4319000
4320000
D
SCO1310(part)
SCO1311
SCO1312
300 bp
1387500
1388500
1389500
E
SCO4117(part)
atrA
SCO4119(part)
300 bp
4523000
4524000
4525000
Genome location
F
SCO2901
SCOt23
SCO2902
300 bp
3154500
3155000
3155500
3156000
Genome location
G
sepG
ylmE
divIVA
sepF
ylmD
ftsZ
ftsQ
murG
ftsW
murD
murX
murF
murE
ftsI
ftsL
mraW
300 bp
2232500
2235000
2237500
2240000
2242500
2245000
2247500
H
rplC
rplW
rplS
rpsC
rpsQ
rplE
rpmC
rplX
rpsH
rplR
rpmD
rpsM
infA
rpsK
rpsJ
rplD
rplB
rplV
rplP
rplN
rpsN
rplF
rpsE
rplO
secY
adk
map
rpmJ
rpoA
rplQ
truA
300 bp
5127500
5130000
5132500
5135000
5137500
5140000
5142500
5145000
I
redR
redY
redV
redT
redQ
redO
redK
redG
redD
redX
redW
redZ
redU
redS
redP
redN
redM
redL
redJ
redI
redH
redF
300 bp
6435000
6440000
6445000
6450000
6455000
6460000
Genome location

Figure 4

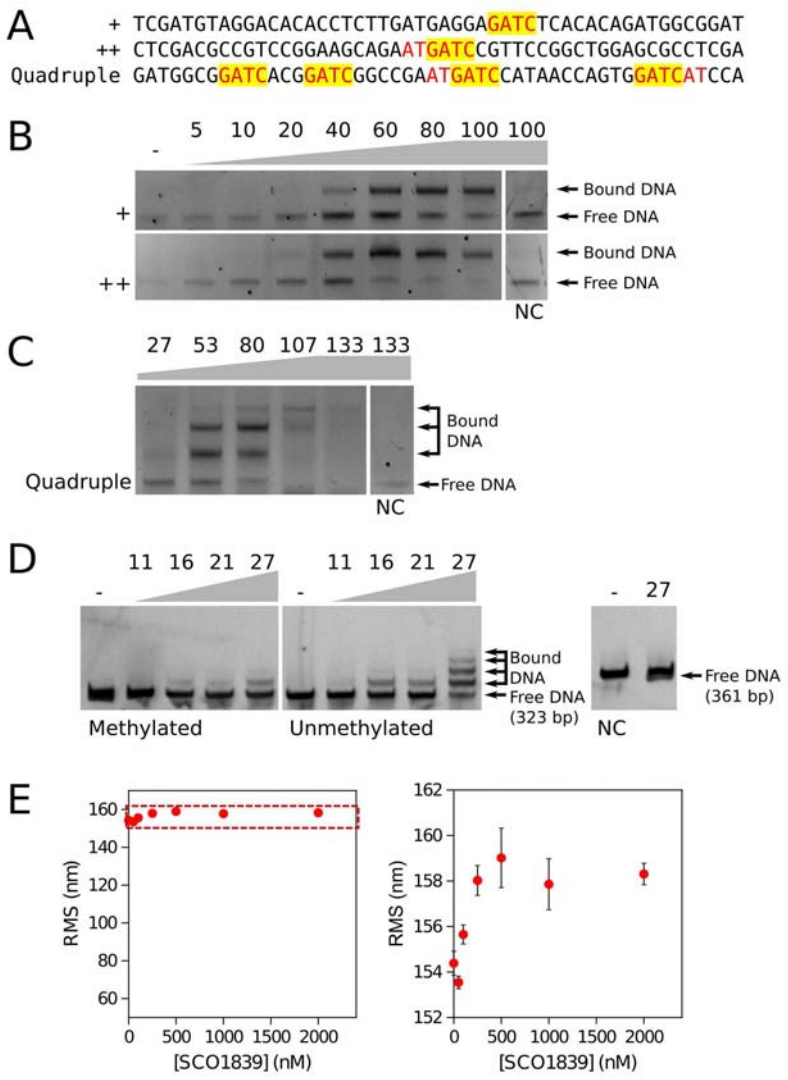
A
+ TCGATGTAGGACACACCTCTTGATGAGGAGATCTCACACAGATGGCGGAT
++ CTCGACGCCGTCCGGAAGCAGAATGATCCGTTCCGGCTGGAGCGCCTCGA
Quadruple GATGGCGGATCACGGATCGGCCGAATGATCCATAACCAGTGGATCATCCA
B
5 10 20 40 60 80 100 100
-
+
++
Bound DNA
Free DNA
Bound DNA
Free DNA
NC
C
27 53 80 107 133 133
Bound DNA
Quadruple
Free DNA
NC
D
11 16 21 27
11 16 21 27
- 27
Bound DNA
Free DNA (323 bp)
Free DNA (361 bp)
Methylated
Unmethylated
NC
E
RMS (nm)
[SCO1839] (nM)
RMS (nm)
[SCO1839] (nM)

Figure 5

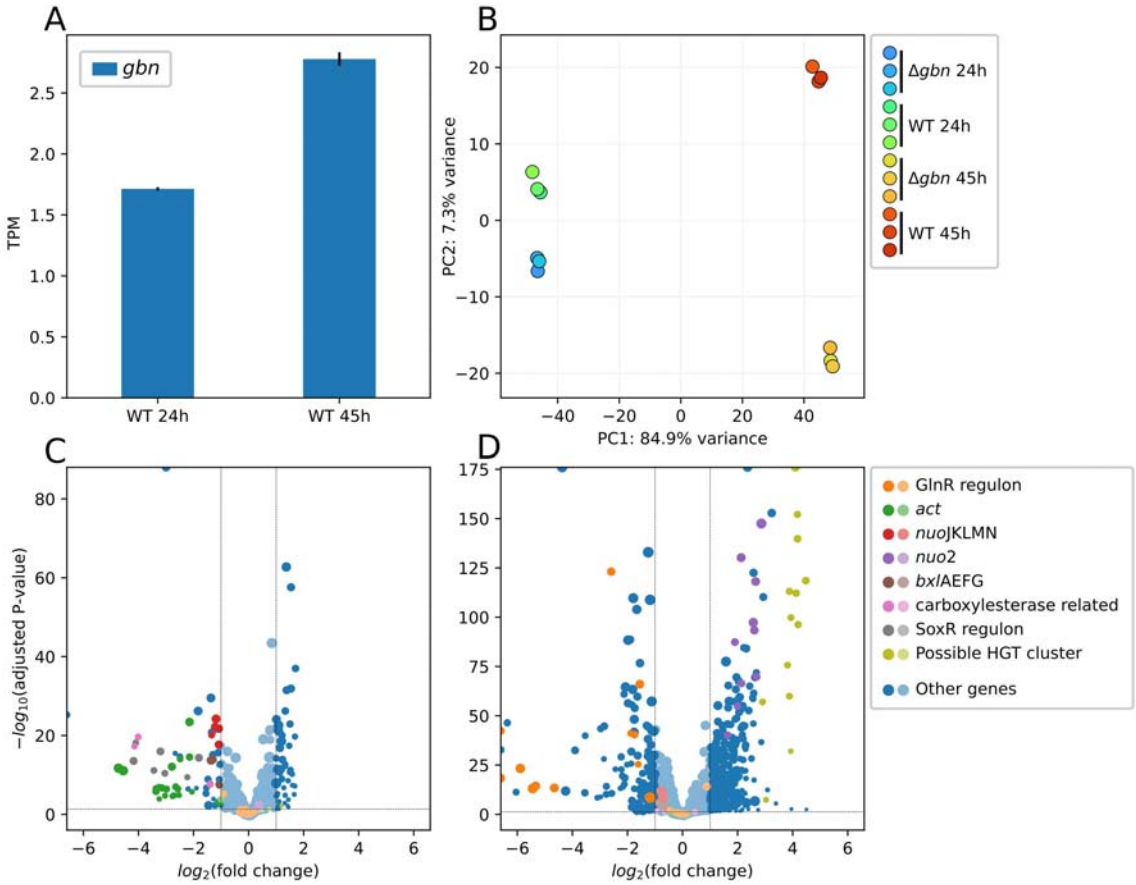
A
gbn
TPM
WT 24h
WT 45h
B
PC2: 7.3% variance
PC1: 84.9% variance
Δgbn 24h
WT 24h
Δgbn 45h
WT 45h
C
−log10(adjusted P-value)
log2(fold change)
D
log2(fold change)
GlnR regulon
act
nuoJKLMN
nuo2
bxlAEFG
carboxylesterase related
SoxR regulon
Possible HGT cluster
Other genes

Figure 6

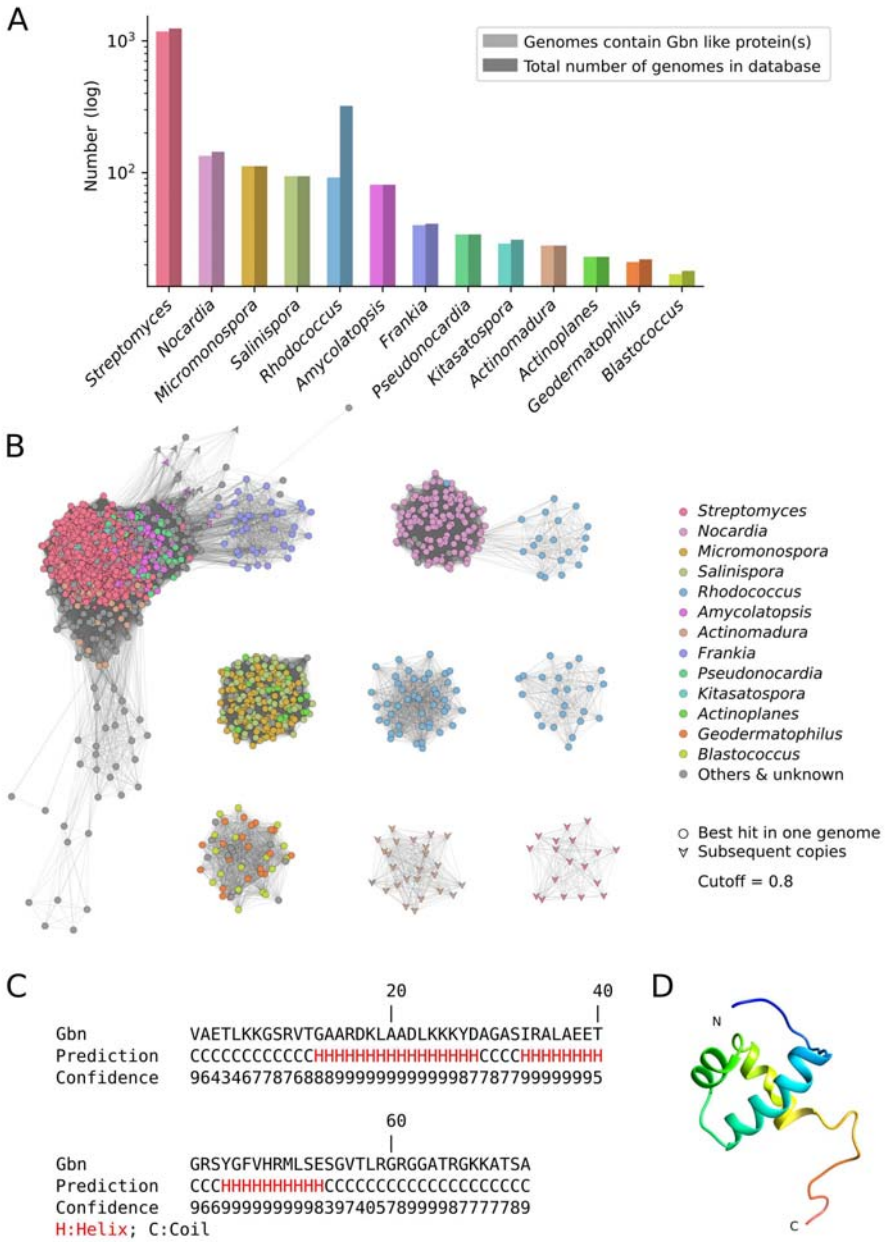
A
Genomes contain Gbn like protein(s)
Total number of genomes in database
Number (log)
10^3
10^2
Streptomyces
Nocardia
Micromonospora
Salinispora
Rhodococcus
Amycolatopsis
Frankia
Pseudonocardia
Kitasatospora
Actinomadura
Actinoplanes
Geodermatophilus
Blastococcus
B
Streptomyces
Nocardia
Micromonospora
Salinispora
Rhodococcus
Amycolatopsis
Actinomadura
Frankia
Pseudonocardia
Kitasatospora
Actinoplanes
Geodermatophilus
Blastococcus
Others & unknown
Best hit in one genome
Subsequent copies
Cutoff = 0.8
C
20 40
Gbn VAETLKKGSRVTGAARDKLAADLKKKYDAGASIRALAEET
Prediction CCCCCCCCCCCCHHHHHHHHHHHHHHHHCCCCHHHHHHHH
Confidence 964346778768889999999999999987787799999995
60
Gbn GRSYGFVHRMLSESGVTLRGRGGATRGKKATSA
Prediction CCCHHHHHHHHHHCCCCCCCCCCCCCCCCCCCC
Confidence 966999999999839740578999987777789
H:Helix; C:Coil
D
N
C

Figure 1.

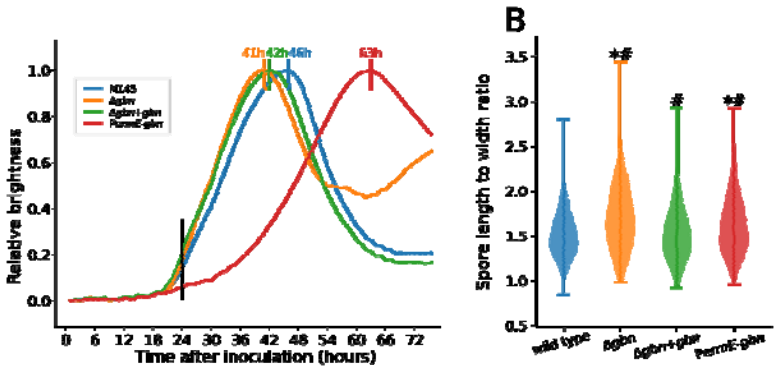
B
Relative brightness
Time after inoculation (hours)
Spore length to width ratio
wild type
Δgbn
Δgbn+gbn
PermE-gbn

Figure 2

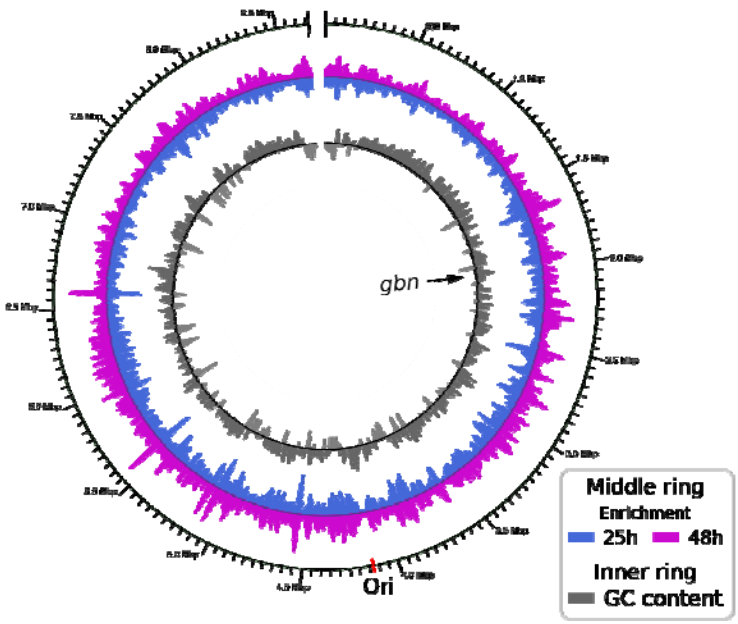
gbn
Ori
Middle ring
Enrichment
25h
48h
Inner ring
GC content

Figure 3

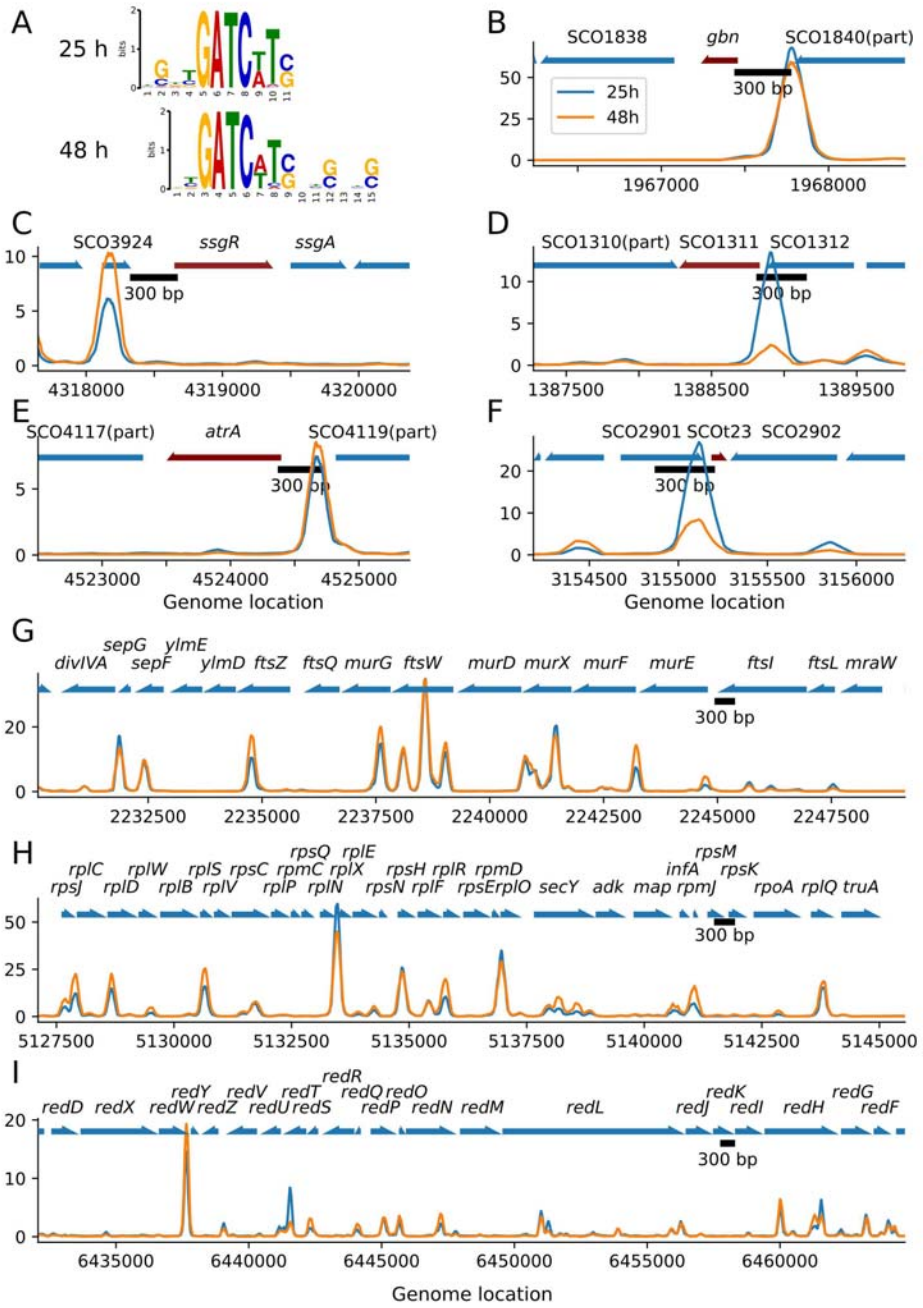
A
25 h
48 h
B
SCO1838
gbn
SCO1840(part)
300 bp
25h
48h
1967000
1968000
C
SCO3924
ssgR
ssgA
300 bp
4318000
4319000
4320000
D
SCO1310(part)
SCO1311
SCO1312
300 bp
1387500
1388500
1389500
E
SCO4117(part)
atrA
SCO4119(part)
300 bp
4523000
4524000
4525000
Genome location
F
SCO2901
SCOt23
SCO2902
300 bp
3154500
3155000
3155500
3156000
Genome location
G
sepG
ylmE
divIVA
sepF
ylmD
ftsZ
ftsQ
murG
ftsW
murD
murX
murF
murE
ftsI
ftsL
mraW
300 bp
2232500
2235000
2237500
2240000
2242500
2245000
2247500
H
rplC
rplW
rplS
rpsC
rpsQ
rplE
rpmC
rplX
rpsH
rplR
rpmD
rpsM
infA
rpsK
rpsJ
rplD
rplB
rplV
rplP
rplN
rpsN
rplF
rpsE
rplO
secY
adk
map
rpmJ
rpoA
rplQ
truA
300 bp
5127500
5130000
5132500
5135000
5137500
5140000
5142500
5145000
I
redR
redY
redV
redT
redQ
redO
redK
redG
redD
redX
redW
redZ
redU
redS
redP
redN
redM
redL
redJ
redI
redH
redF
300 bp
6435000
6440000
6445000
6450000
6455000
6460000
Genome location

Figure 4

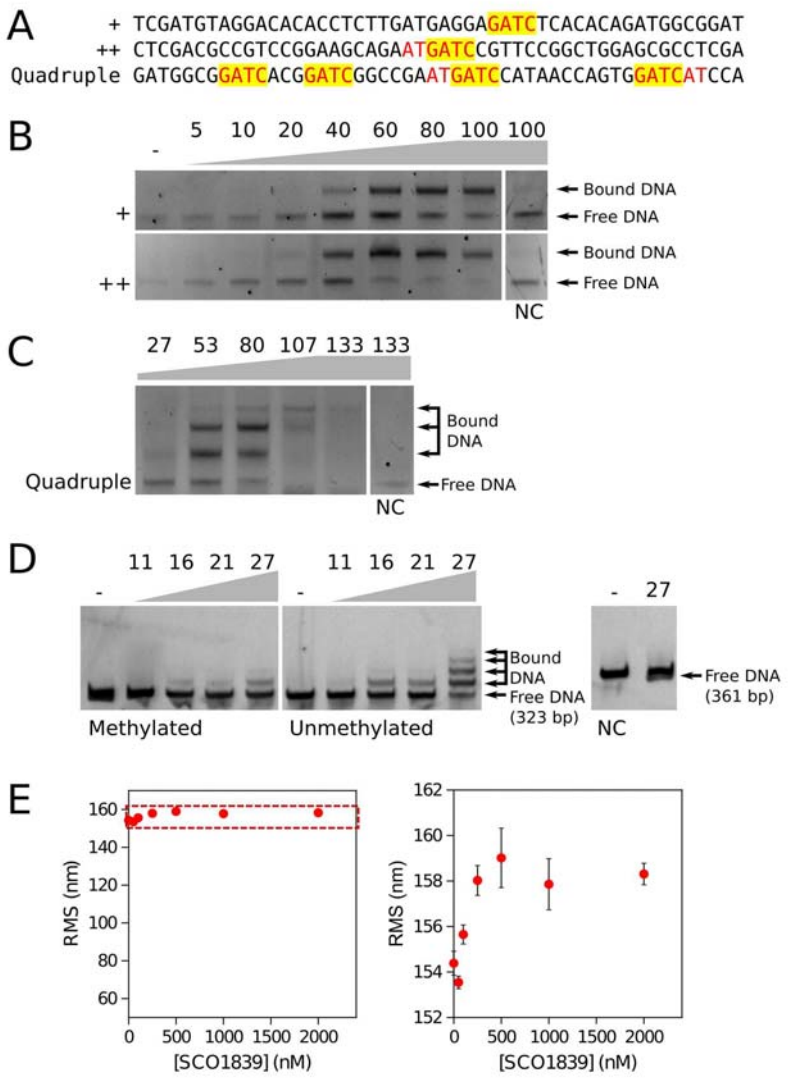
A
+ TCGATGTAGGACACACCTCTTGATGAGGAGATCTCACACAGATGGCGGAT
++ CTCGACGCCGTCCGGAAGCAGAATGATCCGTTCCGGCTGGAGCGCCTCGA
Quadruple GATGGCGGATCACGGATCGGCCGAATGATCCATAACCAGTGGATCATCCA
B
5 10 20 40 60 80 100 100
-
+
++
Bound DNA
Free DNA
Bound DNA
Free DNA
NC
C
27 53 80 107 133 133
Bound DNA
Quadruple
Free DNA
NC
D
11 16 21 27
11 16 21 27
- 27
Bound DNA
Free DNA (323 bp)
Free DNA (361 bp)
Methylated
Unmethylated
NC
E
RMS (nm)
[SCO1839] (nM)
RMS (nm)
[SCO1839] (nM)

Figure 5

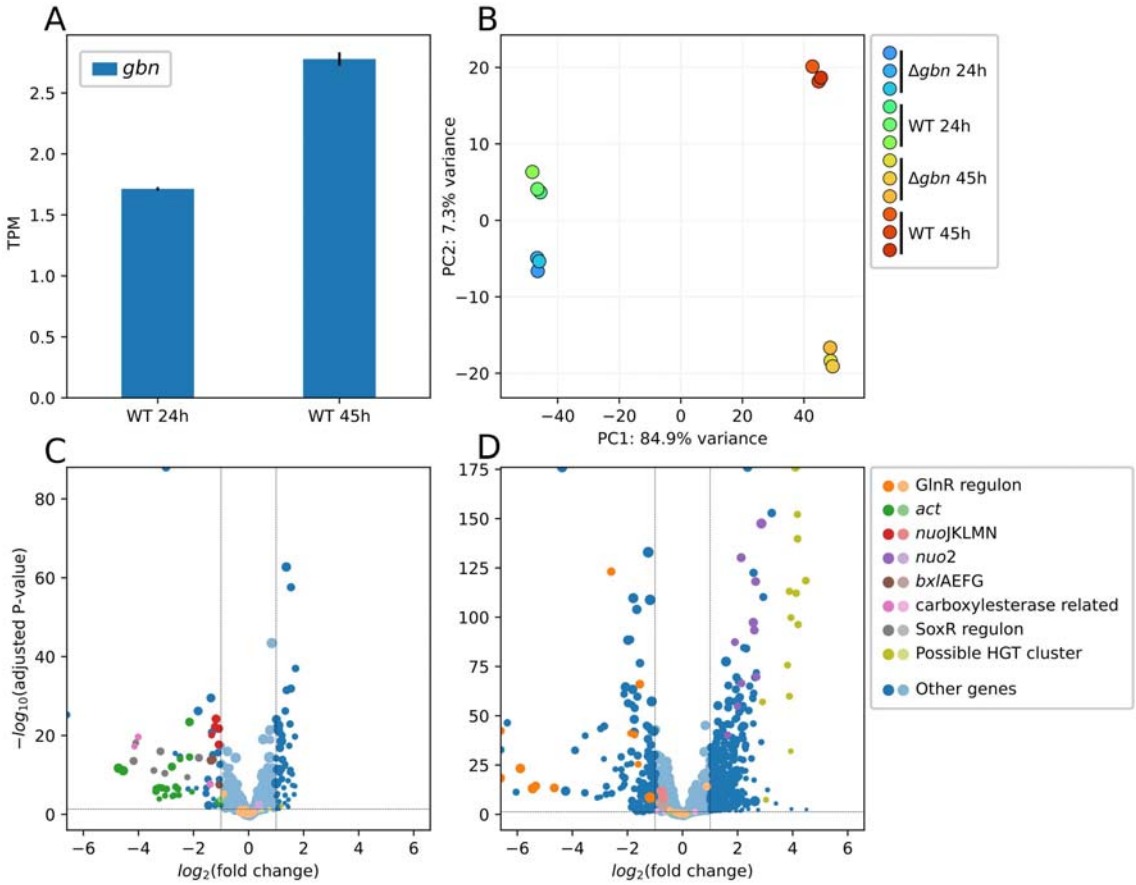
A
gbn
TPM
WT 24h
WT 45h
B
PC2: 7.3% variance
PC1: 84.9% variance
Δgbn 24h
WT 24h
Δgbn 45h
WT 45h
C
−log10(adjusted P-value)
log2(fold change)
D
log2(fold change)
GlnR regulon
act
nuoJKLMN
nuo2
bxlAEFG
carboxylesterase related
SoxR regulon
Possible HGT cluster
Other genes

Figure 6