The nucleoid-associated protein Gbn binds to GATC sequences and affects sporulation and antibiotic production in *Streptomyces*

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 called nucleoid-associated proteins (NAPs). Many NAPs have been studied in detail in *Streptomyces*, including Lsr2, HupA, HupS, and sIHF. Here, we show that SCO1839 represents a novel family of small NAPs unique to Actinobacteria and recognizes a consensus sequence consisting of GATC followed by (A/T)T. The protein was designated Gbn for GATC-binding NAP. Chromatin immunoprecipitation sequencing (ChIP-Seq) detected more than 2800 binding regions, encompassing some 3600 GATCWT motifs, which comprise 55% of all 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. Despite the huge number of binding sites, the DNA binding profiles were nearly identical during vegetative and aerial growth. The exceptions were SCO1311 and SCOt32, for a tRNA editing enzyme and a tRNA that recognises the rare leucine codon CUA, respectively, which were nearly exclusively bound during vegetative growth. Deletion of *gbn* led to pleiotropic alterations in developmental timing, morphogenesis and antibiotic production. Taken together, our data show that Gbn is a highly pleiotropic NAP that impacts growth and development in 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 *Streptomyces coelicolor* genome encodes some 900 regulatory proteins, of which only a minute fraction has been functionally characterized (9). Many of these affect the control of development and antibiotic production, such as the *bld* and *whi* genes that are responsible for the control of aerial hyphae formation and sporulation, respectively, and global regulatory genes such as *adpA*, *afsR*, *dasR* and *atrA* that pleiotropically control antibiotic production (10).

The control of chromosome structure is an important factor in 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) (11-13). These are generally small DNA binding proteins involved in processes such as controlling gene expression, nucleoid structure or DNA repair. Well-known NAPs in *Streptomyces* include Lsr2, HupA, HupS, and sIHF. Lsr2 binds non-specifically to AT-rich sequences and can globally repress gene expression (14). HupA and HupS are homologs of HU (for histone-like protein from strain U93) proteins, which are differentially regulated depending on the developmental

growth phase (15). sIHF is one of the basic architectural elements conserved in many actinobacteria and is able to influence the regulation of secondary metabolism and cell development (16). IHF binds a well conserved nucleotide sequence, while HU binds to random DNA sequences (17), yet with a preference for bent, distorted or flexible DNA (18). A proteomic survey of *Streptomyces coelicolor* identified 24 proteins with NAP-like properties, namely the known Lsr2, HupA, HupS and sIHF and 20 yet unidentified proteins (19). Although the functions of many NAPs are still not clear, BldC for example has a major impact on the transcriptome (20,21).

We previously showed via pull-down assays that the candidate NAP SCO1839 binds to the promoter region of the cell division regulatory gene *ssgR* (22). SsgR is the transcriptional activator of the cell division activator gene *ssgA* (23). SsgA and its paralogue SsgB are both required for sporulation (24,25), and together coordinate the onset of sporulation-specific cell division in *Streptomyces*, whereby SsgB directly recruits the cell division scaffold protein FtsZ to the future sites of septation (26).

In this study, we show that SCO1839 represents a novel family of small DNA binding proteins, which plays a role in the regulation of *Streptomyces* development and antibiotic production. The protein is specific to the Actinobacteria, with an HTH DNA binding motif containing three helices, and plays a role in the control of morphogenesis. Chromatin immunoprecipitation coupled with massive parallel DNA sequencing (ChIP-Seq) revealed that the protein binds to over 2800 genomic regions with one or more binding sites, recognising a specific DNA binding motif centred around the consensus sequence GATC. Thus, we nominated this protein Gbn for GATC-binding NAP.

MATERIALS AND METHODS

Reagents

All restriction enzymes are ordered from NEB (Massachusetts, U.S.), including *Bam*HI-HF (Cat. R3136), *Xba*I (R0145), *Eco*RI-HF (R3101), *Hin*dIII-HF (R3104), *Bbs*I (R0539), *Nco*I-HF (R3193), *Sna*BI (R0130), *Stu*I (R0187), *Sac*I (R0156), *Nde*I (R0111). Phusion polymerase (M0532) and T4 DNA ligase (M0202) were also obtained from NEB. Plasmid mini-prep kit (Cat. 740727.250) was from BIOKE (Leiden, The Netherlands). DNA purification was achieved using DNA Clean & Concentrator kit (Cat. D4029) from Zymo Research (California, U.S.).

Biological Resources

Strains and growth conditions

All strains used in this study are listed in Table S2. *Escherichia coli* strain JM109 was used for routine cloning, *E. coli* ET12567 (27) for preparing non-methylated DNA, ET12567 containing driver plasmid pUZ8002 (28) was used in conjugation experiments for introducing DNA to *Streptomyces*. *E. coli* strains were grown in Luria broth at 37°C supplemented with the appropriate antibiotics (ampicillin, apramycin, kanamycin and/or chloramphenicol at 100, 50, 25 and 25 $\mu g \cdot mL^{-1}$, respectively) depending on the vector used. *S. coelicolor* A3(2) M145 was the parent strain for all mutants. *Streptomyces* strains were grown on soya flour medium (SFM) for conjugation, SFM or MM agar medium supplemented with 0.5% mannitol for phenotype characterization, and R5 agar plates for protoplast regeneration. Solid cultures were grown in a 30°C incubator unless described specifically. For liquid cultures, approximately 10^6 spores were inoculated in 100 mL Erlenmeyer flask equipped with steel spring containing 15 mL TSBS (tryptone soya broth sucrose) medium (29). The flasks were incubated at 30°C with constant

shaking at 180 rpm. Antibiotics used for screening *Streptomyces* transformants were apramycin and thiostrepton (20 and 10 μg·mL^{-1}, respectively).

Constructs and cloning

Primers used for PCR and short double strand DNA fragment are listed in Table S3. PCR was preformed using Phusion DNA polymerase using standard protocol as described previously (30). All plasmids and constructs described in this study are summarized in Table S4. The constructs generated in this study were verified by sequencing performed in BaseClear (Leiden, The Netherlands).

The *gbn* knock-out strategy was based on the unstable multi-copy vector pWHM3 (31) as described previously (32). Briefly, upstream and downstream region of *gbn* were amplified from genome and cloned into pWHM3. Between these two regions, an apramycin resistance cassette from pGWS728 (32) was inserted as selection marker. The resulting vector pGWS1255 is a knock-out construct that is able to replace nucleotide positions -565 to +228 of *gbn* with apramycin resistance cassette, where +1 refers to the translation start site. The apramycin resistance cassette was subsequently removed using Cre expressing construct pUWL-Cre (33,34) yielding the clean knock-out strain GAD003. For complementation of the *gbn* deletion mutant, the nucleotide positions -565 to +228 relative *gbn* translation start site, containing entire coding region of *gbn* (with stop codon) and its promoting region was cloned into low copy number plasmid pHJL401, yielding complementation construct pGWS1260. This construct was then transformed to GAD003, resulting in strain GAD014.

To perform ChIP-Seq experiment, the 3×FLAG was fused to the end of original copy of *gbn* on genome using codon optimised CRISPR-Cas9 system (35). The spacer sequence located at the end of *gbn* and was inserted into the pCRISPomyces-2 plasmid as described by (35). Template for homology-directed repair (HDR) which ensures the insertion of 3×FLAG

sequence was cloned into pCRISPomyces-2 followed by spacer insertion (Figure S5), yielding *gbn*-3×FLAG knock-in construct pGWS1298. Mutagenesis was done according to (35). A successful 3×FLAG tag knock-in strain was identified by PCR and sequencing, designated GAD043.

For over-expression of *gbn*, the *ermE* promoter was cloned to replace part of the original promoter region of *gbn* using the same CRISPR-Cas9 system. A spacer sequence were designed at the promoter region of *gbn* and this was inserted into pCRISPomyces-2 (35). HDR template was designed to remove -157 to +4 region of *gbn* and replace this region with P*ermE* sequence yielding pGWS1295. P*ermE* was digested from pHM10a (36). Following the same procedure as above, strain GAD039 was obtained, which expresses *gbn* from the *ermE* promoter.

To produce His_6-Gbn for electrophoretic mobility shift assay (EMSA) experiments, *gbn* was cloned into protein expression construct pET28a and transformed into *E. coli* strain BL21 CodonPlus (DE3)-RIPL (Invitrogen, Massachusetts, U.S.). To generate methylated and non-methylated DNA for EMSA, part of the promoter region of *gbn* and a random region of the *gbn* non-binding region were cloned into pUC19 (37). This yielded vector pGWS1300 that contains a Gbn binding domain and pGWS1451 contains a stretch of DNA that is not bound by Gbn.

Database mining and clustering

The protein sequence of Gbn (SCO1839) from *S. coelicolor* was used as query in HMMER web server (38) to obtain all Gbn-like proteins from the database, resulting in 727 hits. Sequences with an E value < 0.01 (684 sequences) were selected to generate a Hidden Markov model (HMM) profile using HMMER suit (v. 3.1b2, 39). This profile was used to search against a

146 custom database containing 146,856 genomes with all available bacteria genomes (access
147 date Feb. 9, 2019). Hits with E-value ≤ 5.5×10^{-9} (2,317 sequences) were aligned to the
148 generated HMM profile using the hmmalign tool from HMMER suit. Using the alignment, a
149 network was built calculating the pairwise distance between all the detected Gbn proteins
150 and the threshold for clustering was settled at 0.8. Network visualizations were constructed
151 using Cytoscape (v. 3.7.1, 40).

152 **Time-lapse confluent plate morphology monitoring**

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

162 **Scanning electron microscopy (SEM)**

163 Mycelia were grown on MM agar supplemented with mannitol and grown for 5 days. Sample
164 preparation and imaging was done as described before (41), using JSM-7600F scanning
165 electron microscope (JEOL, Tokyo, Japan). For each strain, 5 images with 7,500 × magnification
166 were taken in randomly selected spore-rich areas. The length and width of spores in each
167 picture were measured using ImageJ version 1.52p strictly according to a randomized file list,

in order to minimize human error. Only spores which are approximately parallel to the focal plane were measured.

DNA-protein cross-linking and chromatin immunoprecipitation

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

ChIP-Seq data analysis

Clean reads received from the sequencing contractor were aligned to the *S. coelicolor* M145 genome with RefSeq accession number NC_003888.3 using bowtie2 (version 2.3.4, 42). Resulted SAM files are sorted using SAMtools (version 1.9, 43) producing BAM files. MACS2 (version 2.1.2, 44) was then used for binding peak prediction and peak modelling by comparing the chromatin immunoprecipitated DNA sample with the corresponding whole genome sample. The models for both samples are shown in Figure S4. The enrichment data used in Figure 4 was calculated for each nucleotide using MACS2 'bdgcmp' command with '-m FE' switch. The peak summit positions including sub peak positions of each predicted binding region were then extracted, and the ± 150 bp region of each summit was extracted from the genome sequence using a python script dependent on the Biopython module (version 1.70, 45). Extracted sequences were subjected to MEME-ChIP (version 5.0.2, 46), which is suitable for large sequence sets, for binding motif prediction.

The enrichment data of two samples was averaged separately in a moving bin of 20,000 bp and plotted in opposite directions as the middle ring of the circular genome diagram. The G+C content was calculated using the same moving bin and centred at the middle of maximum and minimum value before plotting as the inner ring on the plot (Figure 4). For determining the overlap of low G+C content regions and high enrichment regions, the genome was divided into 1,000 bp long sections, the G+C content and average enrichment levels were calculated. The sections with G+C content below the first quartile was considered low in G+C, and those with average enrichment level above the third quartile were considered high in enrichment. To find genes possibly regulated by Gbn, the locations of promoter regions (-350 to +50) of all genes were extracted from the GenBank file containing annotations and checked for overlap with ± 150 bp location of the summit of Gbn binding peaks. This was done using a

215 python script dependent on the module Biopython and pybedtools (version 0.8, 47) and
216 external BEDTools (version 2.27, 48).

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

218 Gbn-His_6 was expressed and purified as described before (49). Purified protein was dialyzed
219 over-night at 4°C against EMSA buffer (10 mM Tris-HCl, pH 7.9, 0.1 mM EDTA, 50 mM KCl). 50
220 bp double strand DNA was generated by gradual cooling of reverse complemented single
221 strand oligonucleotides in T4 DNA ligase buffer (NEB, Massachusetts, U.S.) from 95°C to 12°C
222 in 45 min. pGWS1300 was extracted from DAM methylation effective *E. coli* strain JM109 and
223 DAM deficient *E. coli* strain ET12567, while pGWS1451 for negative control was extracted from
224 strain ET12567 only. The target fragments were then digested and blunted using DNA
225 polymerase I Klenow fragment (NEB, Massachusetts, U.S.). The *in-vitro* DNA-protein
226 interaction studies were done as described (50) in EMSA buffer in a total reaction volume of
227 10 μL, the reactions were incubated at 30°C for 15 min. The reaction was then loaded on 5%
228 polyacrylamide gels and separated by electrophoresis. The gel was briefly stained with
229 ethidium bromide and imaged in a Gel Doc imaging system (BioRad, California, U.S.).

230 **RNA extraction and quantitative reverse transcription PCR**

231 *S. coelicolor* strains were inoculated on minimal medium agar plates covered with cellophane
232 disks with an inoculum of 10^7 CFU per plate. Biomass from two (for 24 h) or one (48 h) agar
233 plate was collected in 15 mL tube. The RNA was extracted using modified Kirby mix (29). The
234 total nucleic acid was further purified using RNeasy columns (Qiagen, Venlo, The Netherlands)
235 according to vendor's instructions, and the remainder of the DNA was removed by DNase I
236 (NEB, Massachusetts, U.S.) followed by phenol chloroform extraction. The complete removal

of DNA was confirmed by PCR. cDNA was synthesized using iScript cDNA synthesis kit (BioRad, California, U.S.) according to vendor's instructions. Quantitative PCR was done using iTaq Universal SYBR green qPCR kit (BioRad, California, U.S.). House-keeping genes *gyrA* (SCO3873) and *rpmE3* (SCO5359) that are known to show constitutive expression (data not shown) were used as internal controls. The data was analysed using CFX Manager software (version 3.1, BioRad, California, U.S.) using $\Delta\Delta C_q$ method, which is an implementation of the method described by (51).

Tethered particle motion

Tethered particle motion experiments were carried out as described previously (52) 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 for Actinobacteria

Previous work revealed that the small protein SCO1839 is a candidate nucleoid-associated protein (NAP) (19), and it specifically 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 (53) did not yield any significant matches to known protein families, suggesting that SCO1839 is the first member of a novel protein family.

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 contain two paralogues of SCO1839, which divided into two additional clusters. For most genera, more than 90% of the species contain at least one copy of SCO1839-like proteins, and hence proteins are well conserved.

Only *Rhodococcus* stands out, as only 28.7% (92 out of 321) of the sequenced genomes of this genus contain a SCO1839-family protein (Figure 1A), and these proteins divided into three distinct clusters. This could be related to the fact that it is the only non-sporulating genus among all listed genera. The fact that the orthologs from the nonsporulating genus *Rhodococcus* are very distinct from those found in all sporulating Actinobacterial genera, in combination with the low frequency of occurrence, suggests that SCO1839 may have diversified in the non-sporulating *Rhodococcus* and also that in fact SCO1839 may primarily be sporulation-specific.

In silico structural modelling of SCO1839 was performed using the I-TASSER server (54), revealing a putative single DNA binding helix-turn-helix (HTH) motif, in the form of a tri-helical structure (Figure 1C and D). No homology was seen to any other known transcriptional regulator family from bacteria. Very short aa stretches flank the residues belonging to the HTH motif. Nine other protein structures found in PDB (Protein data bank, rcsb.org, 55) share structural analogy to SCO1839, and have similar HTH motifs. These are proteins found in a wide range of organisms and with different functions, including a DNA helicase from the archaeon *Pyrococcus furiosus* (PDB structure ID: 2ZJ8), *E. coli* (2VA8, 2P6R) and from humans (5AGA), a ribosomal protein from yeast (5MRC), a human cell division cycle protein (2DIN), a yeast terminator binding protein (5EYB), a regulator from *Staphylococcus aureus* (2R0Q), and a tRNA synthetase from the archaeon *Archaeoglobus fulgidus* (2ZTG). However, none of these proteins were as small as SCO1839 and neither contained only one DNA binding motif. To the best of our knowledge, no other bacterial protein with similar structure has been reported before. Therefore, we propose that the SCO1839-like proteins form a new family of bacterial DNA binding proteins. SCO1839 is further called Gbn, for GATC-binding protein (see below).

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

To obtain insights into the possible role of Gbn in the life cycle of *S. coelicolor*, a knock-out mutant was generated using a strategy published previously (56). For this, the +1 to +207 region of the gene was replaced by the apramycin resistance cassette *aac(C)IV,* which was flanked by *loxP* sites. The *loxP* sites allowed removal of the cassette using Cre recombinase, important to avoid polar effects. To genetically complement the mutant and see if the wild-type phenotype would be restored, the -565/+228 region of *gbn* was amplified from the *S. coelicolor* chromosome and cloned into pHJL401, a shuttle vector that is useful for genetic complementation due to its low copy number in streptomycetes (57). To also analyse the effect of enhanced expression of *gbn*, a second strain was constructed using CRISPR-Cas9 wherein the native promoter of *gbn* (-157 to +4, start codon modified to ATG) was replaced by the strong constitutive *ermE* promoter (58). For details see the Materials and Methods section.

When grown on MM agar plates with mannitol as the sole carbon source, colonies of the *gbn* mutant appeared darker in colour and irregularly shaped as compared with the parental strain *S. coelicolor* M145 (Figure 2A, 6). On soya flour mannitol (SFM) agar plates, the mutant colonies showed a smoother surface, devoid of the ridges that are typical of wild-type colonies (Figure 2A, 8). Interestingly, *gbn* mutants failed to produce the blue-pigmented antibiotic actinorhodin on SFM agar when colonies were grown at higher densities (Figure 2B, right). The phenotype of the *gbn* mutant was restored by genetic complementation with a functional copy of the gene, when colonies were grown as single colonies (Figure 2A, 9-12). However, at higher density, actinorhodin production on SFM was not restored (Figure 2B, right). The strain expressing *gbn* from the *ermE* promoter produced large amounts of the red-

pigmented antibiotic undecylprodigiosin on MM agar (Figure 2A, 13, 14) and had a reduced growth rate on SFM agar (Figure 2A, 15, 16).

In order to more precisely monitor the differences in the timing of development between *gbn* mutants and their parent, time-lapse imaging was performed on confluent mycelial lawns. This method allows monitoring multiple morphological characteristics and in particular quantifying differences in timing of the individual developmental stages, based on pigmentation of the mycelia. When aerial hyphae are formed, intensity increases due to the white pigmentation, while brightness decreases again when grey-pigmented spores are produced (see Material and Methods section for details). Deletion of *gbn* resulted in acceleration of development by 2-5 h (Figure 3A). 54 h after inoculation, the light intensity of the *gbn* mutant again increased (Figure 3A). This may be due to premature germination and renewed growth. The complemented mutant had less pronounced acceleration of its life cycle, confirming that this phenomenon was primarily due to the deletion of *gbn*. Conversely, the enhanced and constitutive expression of *gbn* delayed sporulation by approximately 17 h (Figure 3A). Taken together, this strongly suggests that the expression of *gbn* correlates to the timing of sporulation.

Closer examination of the spores by scanning electron microscopy (SEM) revealed that the length to width ratio of the spores increased in both the deletion mutant and the strain with enhanced expression of *gbn* (Mann-Whitney U test, p-value < 0.01, Table 1, Figure 3B). The complemented strain produced spores with a normal ratio. All engineered strains (Δ*gbn*, Δ*gbn*+*gbn*, P*ermE*-*gbn*), including the complementation mutant, had a larger variance in the distribution of spore length to width ratio compared to that of the parent strain (Levene's test, p-value < 0.01), with the *gbn* deletion mutant showing the largest variance. Complementation of the mutant reduced the large variance to a level close to that of the wild-type. Thus, we

348 conclude that deletion of *gbn* alters the timing of sporulation and also leads to alteration of
349 the spore morphology.

350 **_gbn_ expression increases during development**

351 To determine the expression of *gbn* during cell development, analysis of transcription on agar
352 plates using quantitative PCR (qPCR) showed a two-fold increase in *gbn* expression at 48 h as
353 compared to 24 h (Figure 3C). Inspection of time series transcriptomics data available in the
354 NCBI GEO database (59) revealed that our qPCR analysis conforms well to results obtained by
355 others. Specifically, we compared our results with available data that include early
356 developmental stages and that had been performed in different media (14,60,61). The authors
357 of these earlier studies showed that transcription of *gbn* is low at early developmental stages,
358 and increases as development progresses, in a medium-independent manner (14,60,61). This
359 indicates that the timing of SCO1839 depends on the developmental stage.

360 **Gbn recognises thousands of DNA binding sites and specifically binds to a small motif with**
361 **GATC as core sequence**

362 To obtain insights into the genome-wide DNA binding capacity of Gbn, ChIP-Seq analysis was
363 performed on samples harvested after 25 h (vegetative growth) and 48 h (sporulation).
364 Following this approach all binding sites of Gbn on the *S. coelicolor* chromosome can
365 potentially be identified. For this purpose, the original copy of *gbn* on the genome was fused
366 with a sequence encoding a triple FLAG tag at its 5'-terminus using CRISPR-Cas9 (see Materials
367 and Methods section). The strain had a phenotype that was highly similar to that of the parent
368 (Figure S1). The ChIP-Seq data showed a wide distribution of Gbn binding sites (Figure 4). Half
369 of the binding sites with a strong enrichment (50% of the sites at 25 and 38 h) colocalized with

370 low G+C content regions. In total, 2825 and 2919 binding regions were identified using MACS2
371 software (44) for samples obtained from mycelia in the vegetative and sporulation stage,
372 respectively. Interestingly, there was a near complete overlap (> 90%, 2402) between the Gbn
373 binding regions found in the two samples (Pearson correlation coefficient 0.945, Figure 4).
374 This not only shows that the binding specificity of Gbn is largely growth phase-independent,
375 but also that the experiments were highly reproducible. The result also indicates that while
376 the expression of *gbn* is higher during sporulation, the specificity of the protein for its binding
377 sites does not change, as shown by the highly similar binding profiles from the ChIP-seq
378 experiments.

379 To obtain a consensus binding site for Gbn, the binding regions from ChIP-Seq results
380 were extracted and modelled using MACS2 and MEME-ChIP (44,46). In this way, the
381 sequences GATCAT and GATCTT were identified as binding sites, and thus GATCWT represents
382 the Gbn binding site (Figure 5A). The most conserved binding core was GATC, which is a
383 palindrome known as recognition site for DNA methylation (62). The predicted motifs also
384 showed G/C preference on the flanking region separated by gaps of two base pairs (Figure 5A).
385 It is important to note that virtually all binding regions that were identified as significant (>
386 99.8%) contained a GATC motif, and most (88.2% for 25 h, 84.0% for 48 h) contained the
387 consensus sequence GATCWT. The *S. coelicolor* genome contains in total 6,501 GATCWT
388 sequences, 54.5% and 55.5% were present in the predicted binding regions in the 25 h and 48
389 h samples, respectively. Note that many binding sites have more than one copy of the motif.
390 Closer inspection of the raw data of the ChIP-seq data revealed that in fact all GATCWT motifs
391 on the *S. coelicolor* genome are covered by an increased sequencing coverage, but not all of
392 them can be reliably detected by MACS2 without increasing the false positives to an
393 unacceptable level. This was not seen for GATC sequences not followed by AT or TT. We

394 believe that Gbn can bind all GATCWT sequences on the *S. coelicolor* chromosome. However,
395 not all GATCWT sequences may be accessible to Gbn either directly because they are occupied
396 by other proteins or indirectly as a consequence of 3D organization of the region. Further
397 investigation of this phenomenon is needed.

398 The affinity of Gbn for its DNA binding motif was further examined *in vitro* using
399 electrophoretic mobility shift assays (EMSA). Result show that Gbn could bind to the DNA with
400 one GATC motif *in vitro*, but the additional nucleotide (AT)T on one side increased the affinity
401 of Gbn, with the half binding concentration reduced by more than a third (Figure 6B).
402 Furthermore, when more GATC motifs were present in the DNA fragments, simultaneous
403 binding of more Gbn proteins was observed (Figure 6C and D). For the short 50 bp DNA
404 fragment with four target DNA motifs, only three binding events can be observed from EMSA
405 experiment, possibly only one of the two GATC sequences in close proximity could be bound
406 by one Gbn protein as a consequence of steric hindrance. Taken together, Gbn showed good
407 binding to a motif centred around GATC in *in vitro* experiments.

408 The GATC motif is the target sequence for deoxyadenosine (DAM) methylase, which is
409 essential for DNA mismatch repair in *E. coli* (62). Another typical DNA methylation is on the
410 second deoxycytosine of the sequence CCTGG (DCM) (63). *S. coelicolor* lacks both methylation
411 systems and degrades methylated exogenous DNA (64,65). Which proteins are involved in the
412 recognition and restriction system in *Streptomyces* remains unknown (66,67). The fact that
413 Gbn binds to GATC sequences could be correlated with the restriction system in *S. coelicolor*.
414 Therefore, the transformation efficiency of methylated DNA and non-methylated DNA was
415 tested in the Gbn mutant. We did not see a significant difference in the transformation
416 efficiencies for methylated DNA in *gbn* mutants and in the parental strain M145. This indicates
417 that the restriction barrier still exists (data not illustrated). Next, we tested the affinity

difference of Gbn for GATC and GA^{m}TC in an electrophoretic mobility shift assay (EMSA). The results revealed that Gbn had lower affinity for methylated DNA (Figure 6D). This may be caused by steric hindrance by the methylation, which may negatively affect *Gbn* binding affinity.

Gbn does not alter the conformation of the DNA

In order to investigate whether Gbn alters DNA conformation *in vitro*, we performed Tethered Particle Motion (TPM) experiments (68). If the binding of a protein to DNA induces bends or affects DNA stiffness this translates into a reduction or an increase of the root mean square displacement (RMS) respectively compared to that of bare DNA (69-71). Here we used a 685 bp DNA substrate containing the *gbn* promoter region with 10 GATC(WT) sites. Addition of the Gbn protein resulted in a very small increase in RMS of around 4 nm at ≥ 250 nM (Figure 6E). This could be explained by Gbn occupying the 10 binding sites without causing major changes in the DNA conformation. While Gbn does not deform DNA promoting compaction, the increase in RMS is indicative of the DNA being somewhat stiffened by binding of the protein. Earlier studies indicate that the observed mild increase in RMS corresponds to an increase in persistence length of about 5 nm as compared to bare DNA (71). Note that qualitatively similar effects have been observed for *E. coli* and *P. aeruginosa* H-NS like proteins (69,70). These properties suggest that rather than contributing directly to organization and compaction of the chromosome, Gbn functions in regulation of genome transactions such as transcription.

Gbn binding sites are found in the regulatory regions of some 10% of all genes

To find the genes affected by the binding of Gbn, the common binding regions were mapped to the translational start site of all genes (-350 to +50 of transcription start site). After removing ambiguous low binding regions, 736 of the Gbn binding regions were found in promoter regions, corresponding to 862 genes (Table S1). Of these genes, 47% (406 genes) had at least one binding region with more than 10-fold enrichment in the corresponding promoter region. These genes include the cell division and cell wall (*dcw*) cluster (SCO2077-SCO2088), a 50S ribosomal protein gene cluster (SCO4702-SCO4727), tRNAs (SCOt02-SCOt50), *atrA*, *redY*, SCO1311 and *gbn* itself (Table 2, Figure 5B). One binding region was found upstream of *ssgR*, with a peak at around nt position -483 relative to the translation start site (Figure 5C). This is in accordance with the previous observation that Gbn binds to the *ssgR* promoter region (22). Interestingly, Gbn not only bound to the promoter regions of five tRNA genes, but also to the promoter region of SCO1311, which has a tRNA-editing domain responsible for hydrolysing mis-acylated tRNA (coverage 68%, Pfam: PF04073). The promoter region of SCO1311 and SCOt32 showed the largest difference in Gbn binding between 25 h and 48 h samples, with three times higher enrichment at 25 h than at 48 h (Figure 5D, F, Table 2). SCOt23 specifies the leucyl-tRNA with anticodon UAG, which is required for the translation of the very rare leucine codon CUA. The rarest codon in *S. coelicolor* is another leucine codon, namely UUA. The tRNA recognizing the UUA codon is specified by *bldA*, and the corresponding TTA codon occurs specifically in many genes involved in development and antibiotic production, making those genes *bldA* dependent (72-75). The CTA leucine codon is also very rare in *Streptomyces* genes, representing only 0.31% of all leucine codons in *S. coelicolor*. In this light, it would be interesting to see if SCOt23 also plays a role in the control of

developmental gene expression, and what the role is of Gbn in the control of SCOt23 transcription.

Interestingly, a large number of the top hits in terms of Gbn binding (enrichment level 15-20 times) are located in the *dcw* cluster, which encompasses many key genes for cell division and cell-wall synthesis, including genes for the cell division scaffold FtsZ and DivIVA which is essential for growth (Figure 5G). It is unclear whether this has implications for the expression of these genes, as NAPs typically affect global gene expression through remodelling of DNA structure and not by direct activation or repression of transcription (21). Still, the large number of Gbn binding sites and the very high scores within the *dcw* cluster may explain at least in part the effects of *gbn* deletion on *Streptomyces* development. Interestingly, strong binding was also observed in ribosomal protein gene clusters (Figure 5H). In *Streptomyces*, development and the regulation of secondary metabolism are associated with changes in the expression of ribosomal proteins (76,77). This phenomenon may be related with the resistance mechanism of many antibiotics (78,79).

The phenotype of *gbn* null mutants suggests a direct correlation between Gbn and the level of antibiotic production in particular for actinorhodin (Act) and undecylprodigiosin (Red). Mutants produced much less Act, while a strain overexpressing Gbn produces significantly more Act and Red than the parent (Figure 2B). This may correlate to binding of Gbn to the promoter region of *atrA* (Figure 5E), a global regulatory gene which is essential for the activation of the *act* gene cluster (80). Interestingly, we identified an AtrA binding site in the *gbn* promoter region using the PREDetector algorithm (81), with a high confidence score of 13.3. This score is the same as for the confirmed AtrA binding site upstream of *actII*-ORF4. Additionally, over-expression of *gbn* also enhanced Red production on MM agar (Figure 2A). In total 13 binding regions were detected in the *red* cluster at both time points, with a very

strong binding site upstream of *redY* (Figure 5H). The precise role of Gbn in the control for antibiotic production needs to be investigated further.

In conclusion, our work highlights Gbn as a highly pleiotropic DNA binding protein that plays a role in the (timing of) development and antibiotic production of *S. coelicolor*. Strains in which *gbn* had been deleted or over-expressed showed accelerated and delayed sporulation, respectively, suggesting that Gbn plays a role in the timing (and in particular delay) of development. Both *in vivo* and *in vitro* experiments revealed that Gbn binds specifically to a GATC DNA motif followed by either AT or TT. In addition, we have shown that the methylation of adenine in the GATC sequence reduced affinity of Gbn for its binding sites. These results described Gbn as a representative member of a new NAP family that might play important roles in the growth and development in streptomycetes.

DATA AVAILABILITY

Clean reads and binding region identification (peak calling) files are available at GEO database (82) with the accession number GSE165795. [For reviewers: Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165795 and enter token **ujulcouyzxcfhmn** into the box].

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.

REFERENCES

1. Barka, E.A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y. and van Wezel, G.P. (2016) Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol. Mol. Biol. Rev.*, **80**, 1-43.
2. Hopwood, D.A. (2007) *Streptomyces in nature and medicine: the antibiotic makers*. Oxford University Press.
3. Bérdy, J. (2005) Bioactive microbial metabolites. *J. Antibiot.*, **58**, 1-26.
4. Chater, K.F. and Losick, R. (1997) In Shapiro, J. A. and Dworkin, M. (eds.), *Bacteria as multicellular organisms*. Oxford University Press, New York, pp. 149-182.
5. Claessen, D., Rozen, D.E., Kuipers, O.P., Sogaard-Andersen, L. and van Wezel, G.P. (2014) Bacterial solutions to multicellularity: A tale of biofilms, filaments and fruiting bodies. *Nat. Rev. Microbiol.*, **12**, 115-124.
6. Flärdh, K. and Buttner, M.J. (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. *Nat. Rev. Microbiol.*, **7**, 36-49.
7. Bibb, M.J. (2005) Regulation of secondary metabolism in streptomycetes. *Curr. Opin. Microbiol.*, **8**, 208-215.
8. van der Heul, H.U., Bilyk, B.L., McDowall, K.J., Seipke, R.F. and van Wezel, G.P. (2018) Regulation of antibiotic production in Actinobacteria: New perspectives from the post-genomic era. *Nat. Prod. Rep.*, **35**, 575-604.
9. Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L., Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H., Harper, D. *et al.* (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature*, **417**, 141-147.
10. Urem, M., Świątek, M.A., Rigali, S. and van Wezel, G.P. (2016) Intertwining nutrient-sensory networks and the control of antibiotic production in *Streptomyces*. *Mol. Microbiol.*, 183-195.
11. Dame, R.T., Rashid, F.-Z.M. and Grainger, D.C. (2020) Chromosome organization in bacteria: mechanistic insights into genome structure and function. *Nat. Rev. Genet.*, **21**, 227-242.
12. Dillon, S.C. and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat. Rev. Microbiol.*, **8**, 185-195.
13. Dame, R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Mol. Microbiol.*, **56**, 858-870.
14. Gehrke, E.J., Zhang, X., Pimentel-Elardo, S.M., Johnson, A.R., Rees, C.A., Jones, S.E., Hindra, Gehrke, S.S., Turvey, S., Boursalie, S. *et al.* (2019) Silencing cryptic specialized metabolism in *Streptomyces* by the nucleoid-associated protein Lsr2. *eLife*, **8**, e47691.
15. Salerno, P., Larsson, J., Bucca, G., Laing, E., Smith, C.P. and Flärdh, K. (2009) One of the two genes encoding nucleoid-associated HU proteins in *Streptomyces coelicolor* is developmentally regulated and specifically involved in spore maturation. *J. Bacteriol.*, **191**, 6489-6500.
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, G.P. and 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.
17. Swinger, K.K. and Rice, P.A. (2004) IHF and HU: flexible architects of bent DNA. *Curr. Opin. Struct. Biol.*, **14**, 28-35.
18. Kamashev, D., Balandina, A. and Rouviere-Yaniv, J. (1999) The binding motif recognized by HU on both nicked and cruciform DNA. *The EMBO Journal*, **18**, 5434-5444.

19. Bradshaw, E., Saalbach, G. and McArthur, M. (2013) Proteomic survey of the *Streptomyces coelicolor* nucleoid. *J. Proteom.*, **83**, 37-46.
20. Bush, M.J., Chandra, G., Al-Bassam, M.M., Findlay, K.C. and Buttner, M.J. (2019) BldC delays entry into development to produce a sustained period of vegetative growth in *Streptomyces venezuelae*. *mBio*, **10**, e02812-02818.
21. Dorman, C.J., Schumacher, M.A., Bush, M.J., Brennan, R.G. and Buttner, M.J. (2020) When is a transcription factor a NAP? *Curr. Opin. Microbiol.*, **55**, 26-33.
22. Kim, S., Traag, B., Hasan, A., McDowall, K., Kim, B.-G. and van Wezel, G.P. (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.
23. Traag, B.A., Kelemen, G.H. and Van Wezel, G.P. (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, B.J., Noens, E.E., Kraal, B., Koerten, H.K. and van Wezel, G.P. (2003) The *Streptomyces coelicolor ssgB* gene is required for early stages of sporulation. *FEMS Microbiol. Lett.*, **225**, 59-67.
25. van Wezel, G.P., van der Meulen, J., Kawamoto, S., Luiten, R.G., Koerten, H.K. and Kraal, B. (2000) *ssgA* is essential for sporulation of *Streptomyces coelicolor* A3(2) and affects hyphal development by stimulating septum formation. *J. Bacteriol.*, **182**, 5653-5662.
26. Willemse, J., Borst, J.W., de Waal, E., Bisseling, T. and van Wezel, G.P. (2011) Positive control of cell division: FtsZ is recruited by SsgB during sporulation of *Streptomyces*. *Genes Dev.*, **25**, 89-99.
27. MacNeil, D.J., Gewain, K.M., Ruby, C.L., Dezeny, G., Gibbons, P.H. and MacNeil, T. (1992) Analysis of *Streptomyces avermitilis* genes required for avermectin biosynthesis utilizing a novel integration vector. *Gene*, **111**, 61-68.
28. Paget, M.S.B., Chamberlin, L., Atrih, A., Foster, S.J. and Buttner, M.J. (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.
29. Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F. and Hopwood, D.A. (2000) *Practical Streptomyces genetics*. John Innes Foundation, Norwich, U.K.
30. Colson, S., Stephan, J., Hertrich, T., Saito, A., van Wezel, G.P., Titgemeyer, F. and 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.
31. Vara, J., Lewandowska-Skarbek, M., Wang, Y.G., Donadio, S. and Hutchinson, C.R. (1989) Cloning of genes governing the deoxysugar portion of the erythromycin biosynthesis pathway in *Saccharopolyspora erythraea* (*Streptomyces erythreus*). *J. Bacteriol.*, **171**, 5872-5881.
32. Zhang, L., Willemse, J., Hoskisson, P.A. and van Wezel, G.P. (2018) Sporulation-specific cell division defects in *ylmE* mutants of *Streptomyces coelicolor* are rescued by additional deletion of *ylmD*. *Sci. Rep.*, **8**, 7328.
33. Fedoryshyn, M., Welle, E., Bechthold, A. and Luzhetskyy, A. (2008) Functional expression of the Cre recombinase in Actinomycetes. *Appl. Microbiol. Biotechnol.*, **78**, 1065-1070.
34. Khodakaramian, G., Lissenden, S., Gust, B., Moir, L., Hoskisson, P.A., Chater, K.F. and Smith, M.C.M. (2006) Expression of Cre recombinase during transient phage infection permits efficient marker removal in *Streptomyces*. *Nucleic Acids Res.*, **34**, e20-e20.
35. Cobb, R.E., Wang, Y. and Zhao, H. (2015) High-efficiency multiplex genome editing of *Streptomyces* species using an engineered CRISPR/Cas system. *ACS Synth. Biol.*, **4**, 723-728.
36. Motamedi, H., Shafiee, A. and Cai, S.-J. (1995) Integrative vectors for heterologous gene expression in *Streptomyces* spp. *Gene*, **160**, 25-31.
37. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold spring harbor laboratory press, New York.
38. Potter, S.C., Luciani, A., Eddy, S.R., Park, Y., Lopez, R. and Finn, R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, **46**, W200-W204.
39. Eddy, S.R. (2011) Accelerated profile HMM searches. *PLoS Comput. Biol.*, **7**, e1002195.

40. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498-2504.
41. Piette, A., Derouaux, A., Gerkens, P., Noens, E.E.E., Mazzucchelli, G., Vion, S., Koerten, H.K., Titgemeyer, F., De Pauw, E., Leprince, P. *et al.* (2005) From dormant to germinating spores of *Streptomyces coelicolor* A3(2): New perspectives from the *crp* null mutant. *J. Proteome Res.*, **4**, 1699-1708.
42. Langmead, B. and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods*, **9**, 357.
43. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing, S. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
44. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
45. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. *et al.* (2009) Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, **25**, 1422-1423.
46. Machanick, P. and Bailey, T.L. (2011) MEME-ChIP: Motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696-1697.
47. Dale, R.K., Pedersen, B.S. and Quinlan, A.R. (2011) Pybedtools: A flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*, **27**, 3423-3424.
48. Quinlan, A.R. (2014) BEDTools: The swiss-army tool for genome feature analysis. *Curr. Protoc. Bioinformatics*, **47**, 11.12.01-11.12.34.
49. Mahr, K., van Wezel, G.P., Svensson, C., Krengel, U., Bibb, M.J. and Titgemeyer, F. (2000) Glucose kinase of Streptomyces coelicolor A3(2): large-scale purification and biochemical analysis. *Antonie Van Leeuwenhoek*, **78**, 253-261.
50. Urem, M., van Rossum, T., Bucca, G., Moolenaar, G.F., Laing, E., Świątek-Połatyńska, M.A., Willemse, J., Tenconi, E., Rigali, S., Goosen, N. *et al.* (2016) OsdR of Streptomyces coelicolor and the dormancy regulator DevR of Mycobacterium tuberculosis control overlapping regulons. *mSystems*, **1**, e00014-00016.
51. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, research0034.0031.
52. Henneman, B., Heinsman, J., Battjes, J. and Dame, R.T. (2018), *Bacterial Chromatin*. Springer, pp. 257-275.
53. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A. *et al.* (2018) The Pfam protein families database in 2019. *Nucleic Acids Res.*, **47**, D427-D432.
54. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J. and Zhang, Y. (2015) The I-TASSER Suite: Protein structure and function prediction. *Nat. Methods*, **12**, 7-8.
55. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235-242.
56. Świątek, M.A., Tenconi, E., Rigali, S. and van Wezel, G.P. (2012) Functional analysis of the N-acetylglucosamine metabolic genes of *Streptomyces coelicolor* and role in the control of development and antibiotic production. *J. Bacteriol.*, **194**, 1136-1144.
57. van Wezel, G.P., White, J., Hoogvliet, G. and Bibb, M.J. (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.
58. Bibb, M.J., Janssen, G.R. and Ward, J.M. (1985) Cloning and analysis of the promoter region of the erythromycin resistance gene *(ermE*) of *Streptomyces erythraeus*. *Gene*, **38**, 215-226.

59. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2012) NCBI GEO: Archive for functional genomics data sets – update. *Nucleic Acids Res.*, **41**, D991-D995.
60. Castro-Melchor, M., Charaniya, S., Karypis, G., Takano, E. and Hu, W.-S. (2010) Genome-wide inference of regulatory networks in *Streptomyces coelicolor*. *BMC Genom.*, **11**, 578.
61. Nieselt, K., Battke, F., Herbig, A., Bruheim, P., Wentzel, A., Jakobsen, O.M., Sletta, H., Alam, M.T., Merlo, M.E., Moore, J. *et al.* (2010) The dynamic architecture of the metabolic switch in *Streptomyces coelicolor*. *BMC Genomics*, **11**, 10.
62. Barras, F. and Marinus, M.G. (1989) The great GATC: DNA methylation in *E. coli*. *Trends Genet.*, **5**, 139-143.
63. May, M.S. and Hattman, S. (1975) Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes. *J. Bacteriol.*, **123**, 768-770.
64. Flett, F., Mersinias, V. and Smith, C.P. (1997) High efficiency intergeneric conjugal transfer of plasmid DNA from *Escherichia coli* to methyl DNA-restricting Streptomycetes. *FEMS Microbiol. Lett.*, **155**, 223-229.
65. Kieser, T. and Hopwood, D.A. (1991), *Methods Enzymol.* Academic Press, Vol. 204, pp. 430-458.
66. Liu, G., Ou, H.-Y., Wang, T., Li, L., Tan, H., Zhou, X., Rajakumar, K., Deng, Z. and He, X. (2010) Cleavage of phosphorothioated DNA and methylated DNA by the type IV restriction endonuclease ScoMcrA. *PLoS Genet.*, **6**, e1001253.
67. González-Cerón, G., Miranda-Olivares, O.J. and Servín-González, L. (2009) Characterization of the methyl-specific restriction system of *Streptomyces coelicolor* A3(2) and of the role played by laterally acquired nucleases. *FEMS Microbiol. Lett.*, **301**, 35-43.
68. van der Valk, R.A., Laurens, N. and Dame, R.T. (2017) In Espéli, O. (ed.), *The Bacterial Nucleoid: Methods and Protocols*. Springer New York, New York, NY, pp. 127-143.
69. Qin, L., Bdira, F.B., Sterckx, Y.G.J., Volkov, A.N., Vreede, J., Giachin, G., van Schaik, P., Ubbink, M. and Dame, Remus T. (2020) Structural basis for osmotic regulation of the DNA binding properties of H-NS proteins. *Nucleic Acids Res.*, **48**, 2156-2172.
70. van der Valk, R.A., Vreede, J., Qin, L., Moolenaar, G.F., Hofmann, A., Goosen, N. and Dame, R.T. (2017) Mechanism of environmentally driven conformational changes that modulate H-NS DNA-bridging activity. *eLife*, **6**, e27369.
71. Driessen, R.P.C., Sitters, G., Laurens, N., Moolenaar, G.F., Wuite, G.J.L., Goosen, N. and Dame, R.T. (2014) Effect of temperature on the intrinsic flexibility of DNA and its interaction with architectural proteins. *Biochemistry*, **53**, 6430-6438.
72. Leskiw, B.K., Lawlor, E.J., Fernandez-Abalos, J.M. and Chater, K.F. (1991) TTA codons in some genes prevent their expression in a class of developmental, antibiotic-negative, *Streptomyces* mutants. *Proc. Natl. Acad. Sci. USA*, **88**, 2461-2465.
73. Li, W., Wu, J., Tao, W., Zhao, C., Wang, Y., He, X., Chandra, G., Zhou, X., Deng, Z., Chater, K.F. *et al.* (2007) A genetic and bioinformatic analysis of *Streptomyces coelicolor* genes containing TTA codons, possible targets for regulation by a developmentally significant tRNA. *FEMS Microbiol. Lett.*, **266**, 20-28.
74. Lawlor, E.J., Baylis, H.A. and Chater, K.F. (1987) Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor* A3(2). *Genes Dev.*, **1**, 1305-1310.
75. Chater, K.F. and Chandra, G. (2008) The use of the rare UUA codon to define “Expression Space” for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *J. Microbiol.*, **46**, 1-11.
76. Huang, J., Lih, C.-J., Pan, K.-H. and Cohen, S.N. (2001) Global analysis of growth phase responsive gene expression and regulation of antibiotic biosynthetic pathways in *Streptomyces coelicolor* using DNA microarrays. *Genes Dev.*, **15**, 3183-3192.
77. Blanco, G., Rodicio, M.R., Puglia, A.M., Méndez, C., Thompson, C.J. and Salas, J.A. (1994) Synthesis of ribosomal proteins during growth of *Streptomyces coelicolor*. *Mol. Microbiol.*, **12**, 375-385.

78. Xu, J., Tozawa, Y., Lai, C., Hayashi, H. and Ochi, K. (2002) A rifampicin resistance mutation in the *rpoB* gene confers ppGpp-independent antibiotic production in *Streptomyces coelicolor* A3(2). *Mol. Genet. Genomics*, **268**, 179-189.
79. Shima, J., Hesketh, A., Okamoto, S., Kawamoto, S. and Ochi, K. (1996) Induction of actinorhodin production by *rpsL* (encoding ribosomal protein S12) mutations that confer streptomycin resistance in *Streptomyces lividans* and *Streptomyces coelicolor* A3(2). *J. Bacteriol.*, **178**, 7276-7284.
80. Uguru, G.C., Stephens, K.E., Stead, J.A., Towle, J.E., Baumberg, S. and McDowall, K.J. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. *Mol. Microbiol.*, **58**, 131-150.
81. Tocquin, P., Naome, A., Jourdan, S., Anderssen, S., Hiard, S., van Wezel, G.P., Hanikenne, M., Baurain, D. and Rigali, S. (2016) PREDetector 2.0: Online and enhanced version of the prokaryotic regulatory elements detector tool. *bioRxiv*.
82. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.*, **41**, D991-995.

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 strongly bound by Gbn at their promoter regions.

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 5

** 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

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. Colony morphology and antibiotic production of *S. coelicolor* M145 and derivatives. **(A)** *S. coelicolor* M145 and its mutants Δ*gbn*, P*ermE-gbn* and Δ*gbn*+*gbn*. Strains were grown as single colonies on MM agar supplemented with mannitol or on SFM agar at 30°C for 3 and 4 days. All pictures were taken at the same scale. **(B)** *S. coelicolor* M145 and *gbn* mutants compared on SFM medium.

Figure 3. 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 calculated 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). **(C)** qPCR experiment on RNA isolated after 24h and 48 h from cultures of *S. coelicolor* M145 grown on MM agar. Error bars show the standard error of two independent biological replicates.

Figure 4. 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 5. Gbn protein binding sites analysis. (A) Gbn DNA binding motifs predicted by combining MEME-ChIP and MACS2 program. **(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 red. G-I, *dcw* gene cluster, ribosomal protein gene cluster, and *red* gene cluster, respectively. Note that the y-axes of the plots are not in the same scale.

Figure 6. 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 AT 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 lo(76)cation 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. (**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.

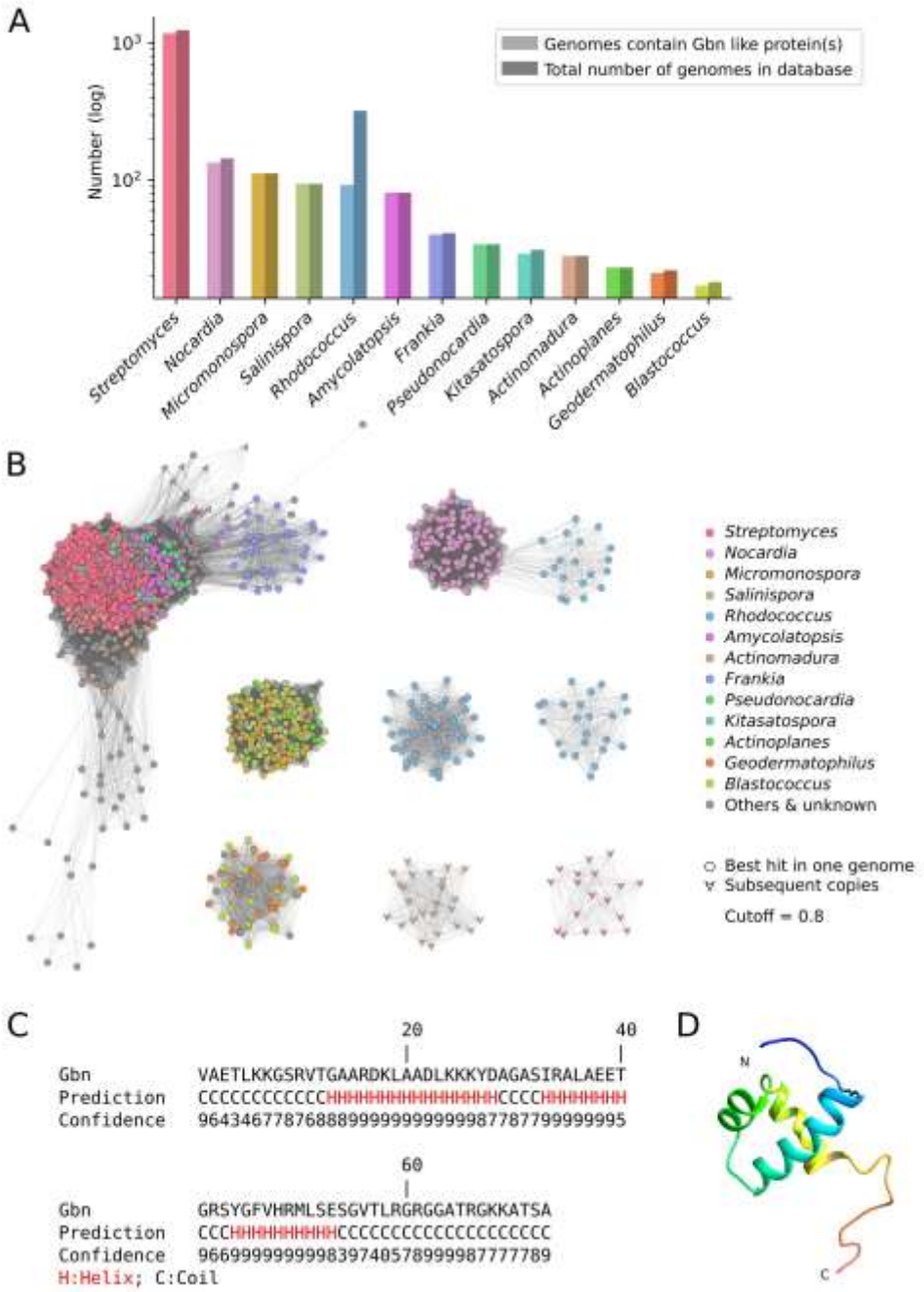
A
Genomes contain Gbn like protein(s)
Total number of genomes in database
Number (log)
10^3
10^2
Streptomyces
Nocardia
Micromonospora
Salinispora
Rhodococcus
Amycolatopsis
Frankia
Pseudonocardia
Kitasatospora
Actinomadura
Actinoplanes
Geodermatophilus
Blastococcus
B
Streptomyces
Nocardia
Micromonospora
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.

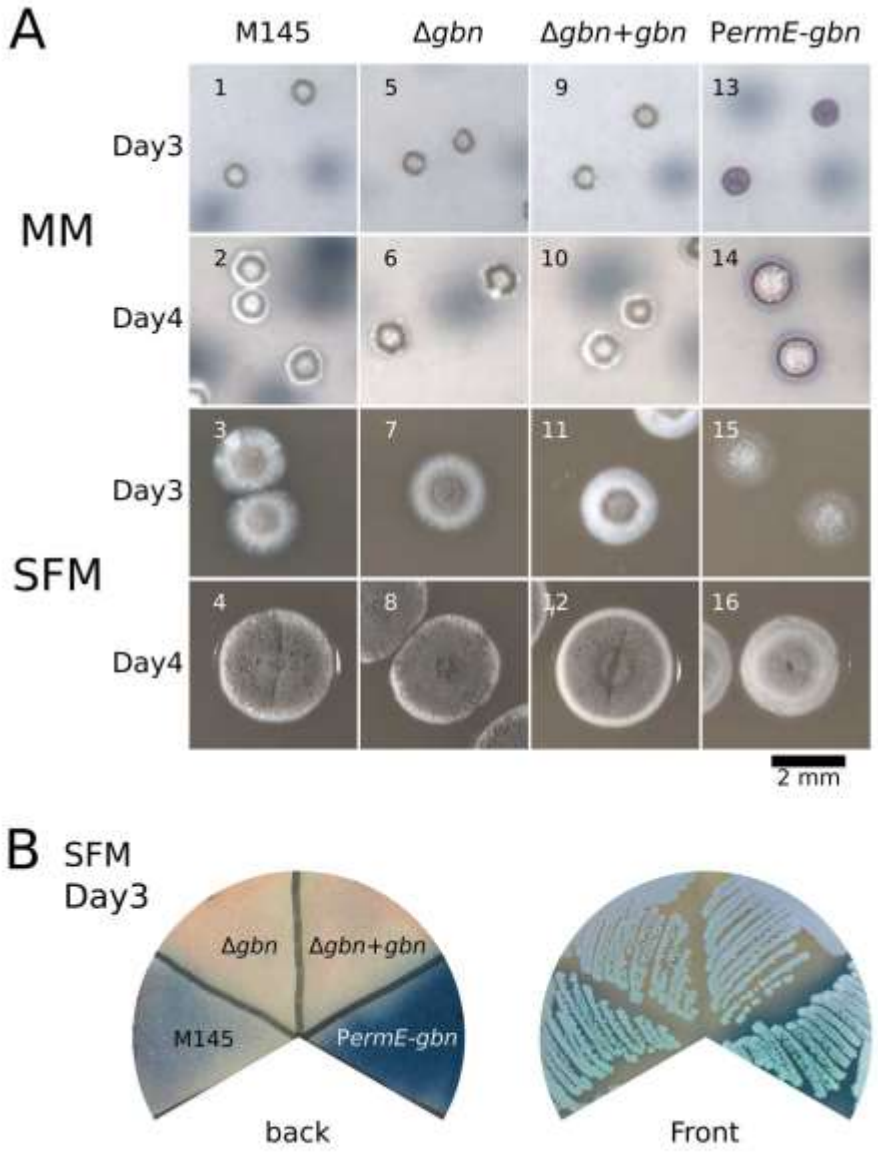
A
M145
Δgbn
Δgbn+gbn
PermE-gbn
MM
Day3
Day4
SFM
Day3
Day4
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
2 mm
B
SFM
Day3
Δgbn
Δgbn+gbn
M145
PermE-gbn
back
Front

Figure 2

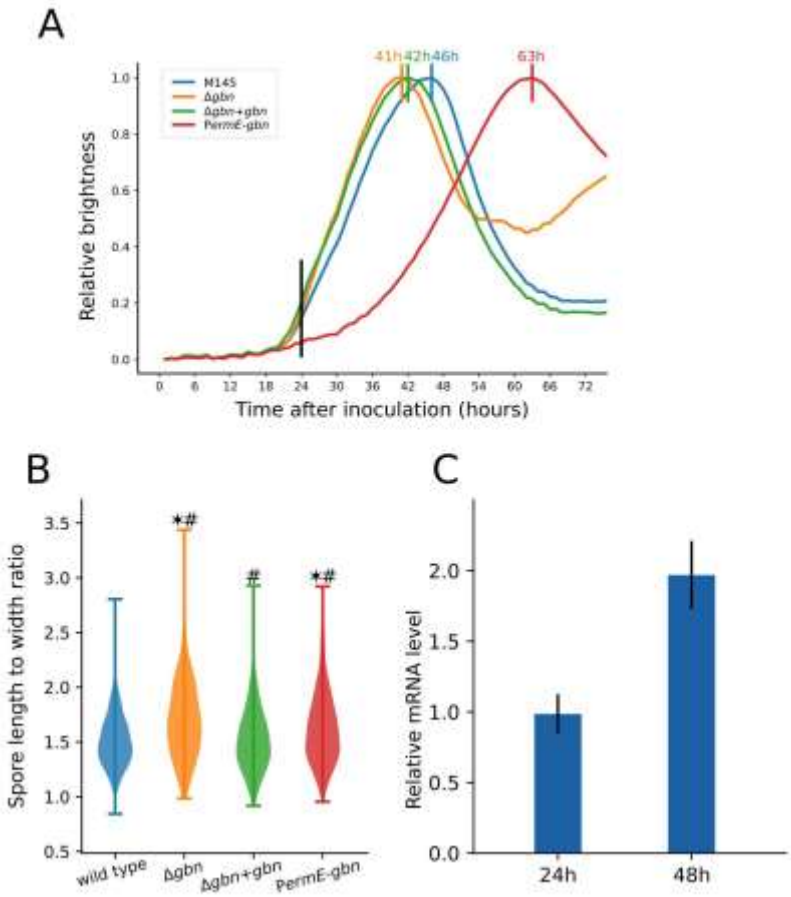
A
41h
42h
46h
63h
M145
Δgbn
Δgbn+gbn
PermE-gbn
Relative brightness
Time after inoculation (hours)
B
Spore length to width ratio
wild type
Δgbn
Δgbn+gbn
PermE-gbn
C
Relative mRNA level
24h
48h

Figure 3

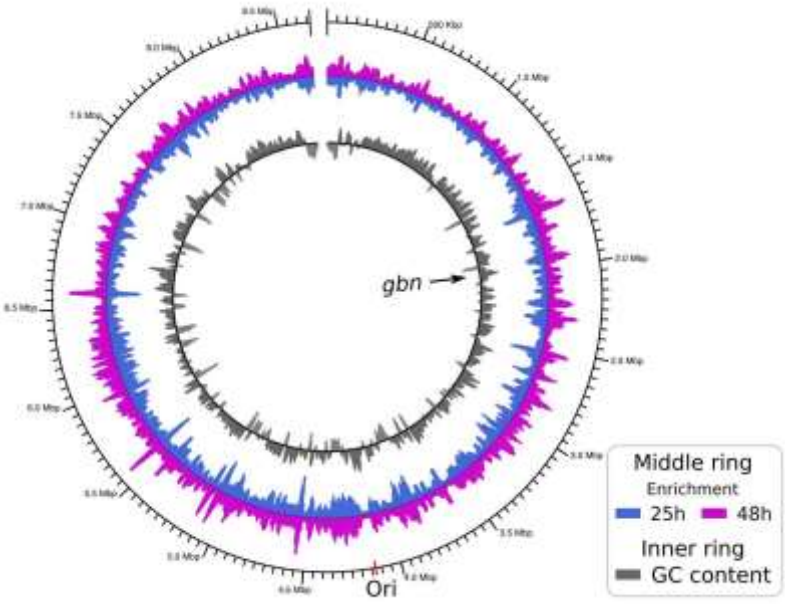
gbn
Ori
Middle ring
Enrichment
25h
48h
Inner ring
GC content

Figure 4

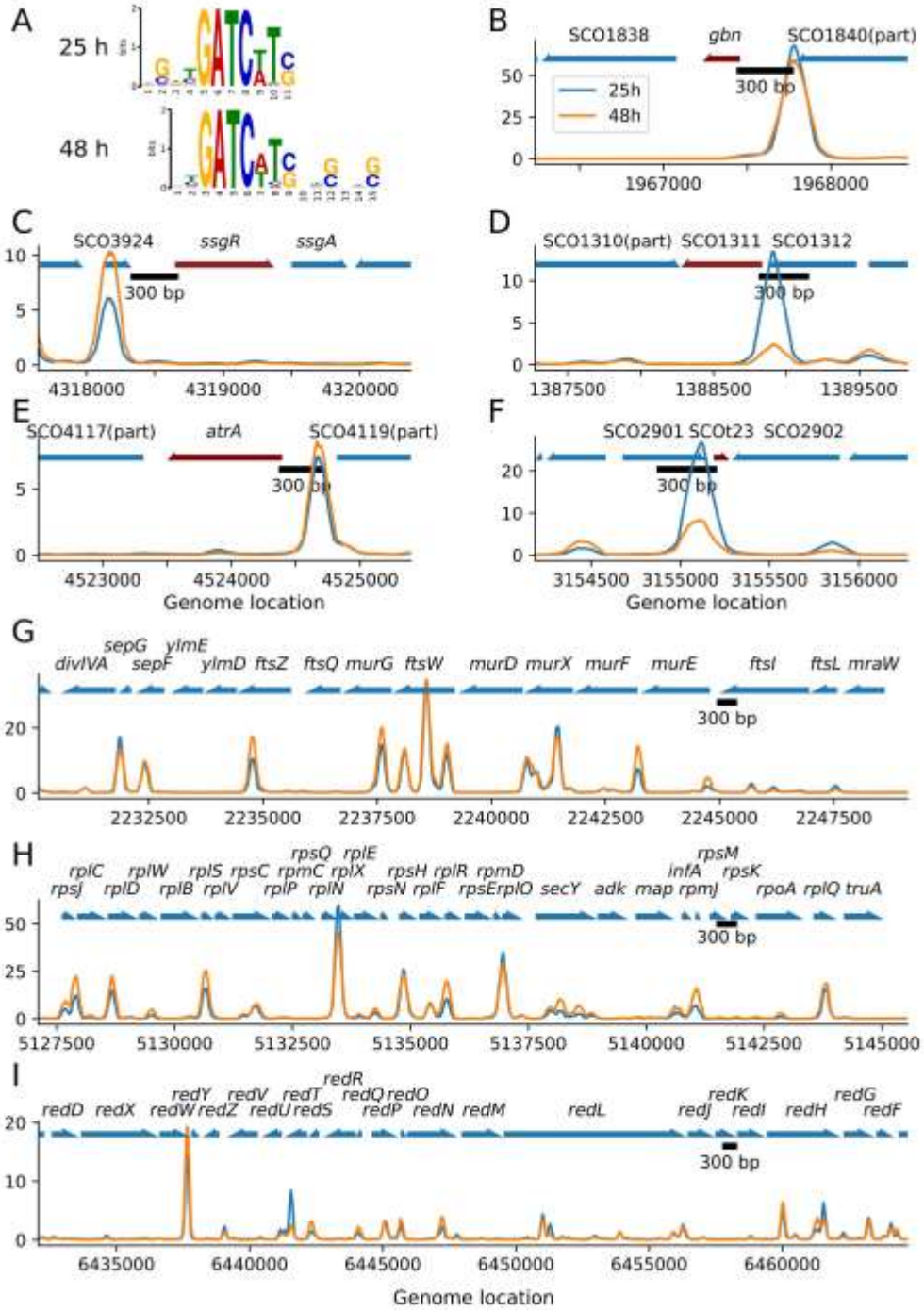
A
25 h
48 h
B
SCO1838
gbn
SCO1840(part)
300 bp
25h
48h
1967000
1968000
C
SCO3924
ssgR
ssgA
300 bp
4318000
4319000
4320000
D
SCO1310(part)
SCO1311
SCO1312
300 bp
1387500
1388500
1389500
E
SCO4117(part)
atrA
SCO4119(part)
300 bp
4523000
4524000
4525000
Genome location
F
SCO2901
SCOt23
SCO2902
300 bp
3154500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

Figure 5.

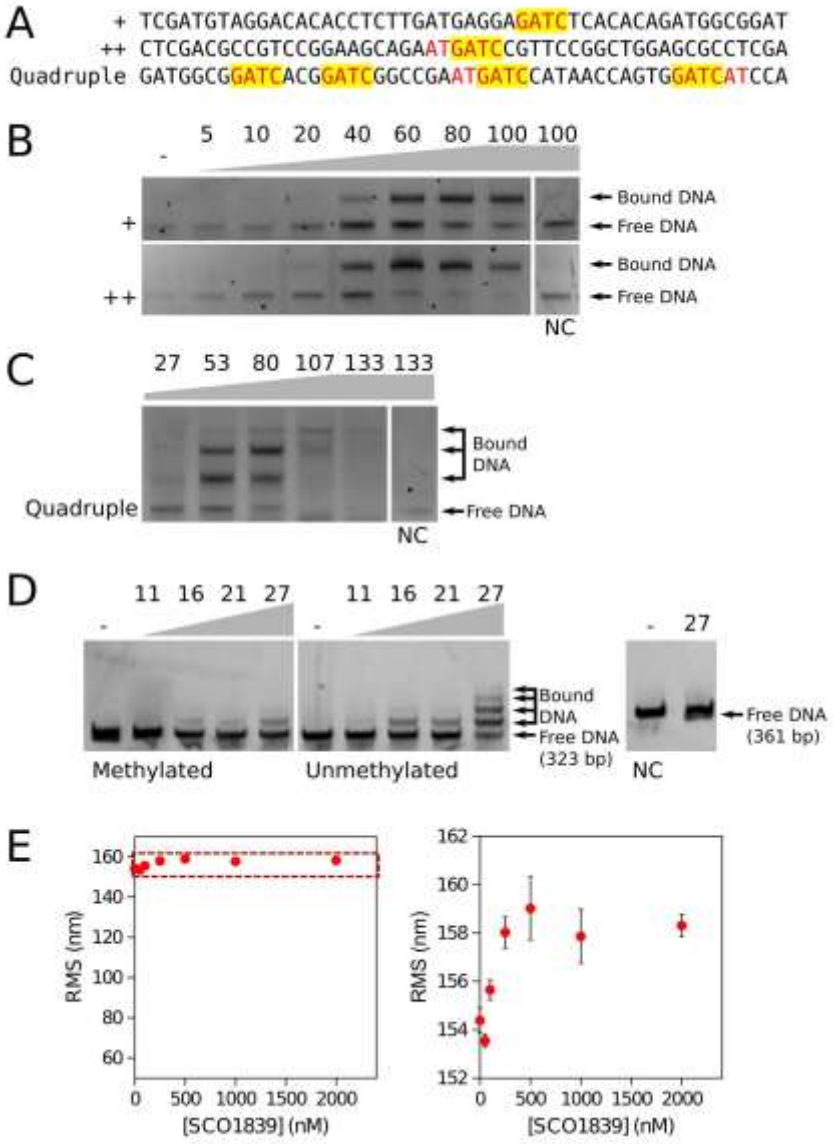
A
+ TCGATGTAGGACACACCTCTTGATGAGGAGATCTCACACAGATGGCGGAT
++ CTCGACGCCGTCCGGAAGCAGAATGATCCGTTCCGGCTGGAGCGCCTCGA
Quadruple GATGGCGGATCACGGATCGGCCGAATGATCCATAACCAGTGGATCATCCA
B
5 10 20 40 60 80 100 100
Bound DNA
Free DNA
Bound DNA
Free DNA
NC
C
27 53 80 107 133 133
Bound DNA
Quadruple
Free DNA
NC
D
11 16 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 6