Cell cycle-associated expression patterns predict gene function in mycobacteria

Running title: Transcriptional compartmentalization of the *Mtb* cell cycle

Aditya C. Bandekar[1], Sishir Subedi[2], Thomas R. Ioerger[2], Christopher M. Sassetti[1]*

[1]Department of Microbiology and Physiological Systems, University of Massachusetts

Medical School, Worcester, MA, 01605, USA

[2]Department of Computer Science and Engineering, Texas A&M University, College Station,

TX, 77843, USA

*Correspondence: christopher.sassetti@umassmed.edu

Summary

While the major events in prokaryotic cell cycle progression are likely to be coordinated

with transcriptional and metabolic changes, these processes remain poorly characterized.

Unlike many rapidly-growing bacteria, DNA replication and cell division are temporally-

resolved in mycobacteria, making these slow-growing organisms a potentially useful

system to investigate the prokaryotic cell cycle. To determine if cell-cycle dependent gene

regulation occurs in mycobacteria, we characterized the temporal changes in the

transcriptome of synchronously replicating populations of *Mycobacterium tuberculosis*

(*Mtb*). By enriching for genes that display a sinusoidal expression pattern, we discover 485

genes that oscillate with a period consistent with the cell cycle. During cytokinesis, the

timing of gene induction could be used to predict the timing of gene function, as mRNA

abundance was found to correlate with the order in which proteins were recruited to the

developing septum. Similarly, the expression pattern of primary metabolic genes could be

used to predict the relative importance of these pathways for different cell cycle processes.

Pyrimidine synthetic genes peaked during DNA replication and their depletion caused a

filamentation phenotype that phenocopied defects in this process. In contrast, the IMP

dehydrogenase *guaB2* dedicated to guanosine synthesis displayed the opposite expression

pattern and its depletion perturbed septation. Together, these data imply obligate

coordination between primary metabolism and cell division, and identify periodically

regulated genes that can be related to specific cell biological functions.

Keywords

Mycobacterium tuberculosis; prokaryotic cell cycle; cell cycle associated transcription; just-

in-time transcription; divisome assembly and cytokinesis; nucleotide metabolism;

cytokinesis

Introduction

Much of prokaryotic cell biology has been elucidated under rapid-growth conditions in

which chromosomal replication takes longer than the doubling time of the cell [1], [2].

Under these conditions, the production of complete chromosomes for daughter cells is

ensured via the simultaneous initiation of multiple rounds of DNA replication, and it is not

possible for cells to temporally segregate DNA replication from cytokinesis as is seen in the

eukaryotic cell cycle.   However, this paradigm may not apply to many, if not most, of the

bacteria in the environment. For example, *Caulobacter crescentus* exploits a specialized

developmental program that produces distinct sessile and motile cells, which is associated

with a strict cell cycle that segregates DNA replication from cytokinesis. More generally,

most of the bacterial biomass in nutrient poor natural environments is likely to persist in

slow-growing states [3]. When these conditions are modeled in nutrient-restricted

Escherichia coli, major cellular events become restricted into distinct cell cycle periods

designated, "B", "C", and "D"; in which DNA replication is restricted to the C period. The B

and D periods occupy the time before or after C, respectively, and are divided by

cytokinesis[4] [5]. Similarly, mycobacteria are slow-growing organisms, with doubling

times that range from 3 hours to several days, which appear to constitutively employ this

type of segregated cell cycle. While an ordered cell cycle may be more common in

prokaryotes than generally appreciated, it has only been investigated in a limited number

of systems.

The prokaryotic cell cycle has been most thoroughly studied in the aquatic bacterium, *C.*

crescentus, largely because it is possible to produce bacterial cultures in which cells are

replicating synchronously with respect to the cell cycle. As a result, the only genome-wide

transcriptional profiles of synchronously-replicating bacteria have been produced in this

species. These studies identified periodic fluctuations in mRNA abundance that correlated

with the cell cycle [6], [7] and led to the elucidation of a regulatory cascade that controls

cell cycle progression and cellular differentiation that appears to be conserved in other

alphaproteobacteria [8]. In addition to these regulators of cell cycle progression, 19% of

the genome was found to be periodically expressed [6]. These genes included those

involved in primary metabolic processes that are not directly associated with cell cycle

progression, suggesting that major cellular events, such as DNA replication and cytokinesis,

may be coordinated with other aspects of cellular physiology. These apparent links

between metabolism and cell cycle are consistent with a number of studies in model

organisms, where metabolites such as NAD(P)H [9] and ATP [10] oscillate according to

the cell cycle in *E. coli*; and UDP-glucose levels coordinate cell division timing with nutrient

availability in both B*. subtilis* [11] and *E. coli* [12]. While these data indicate that cell cycle

progression is likely coupled with some aspects of primary metabolism, it remains unclear

how these processes interact and whether insights from transcriptional profiling in *C.*

crescentus are generalizable to more diverse bacteria.

We sought to extend these paradigms to mycobacteria, a genus that includes relatively

rapid growing environmental organisms, such as *M. smegmatis*, and slower growing

pathogenic species such as *M. tuberculosis* (Mtb) and *M. bovis*. The cell cycle has been

extensively characterized in both fast- and slow-growing mycobacteria using single-cell

analyses of strains engineered to express fluorescent markers of DNA replication and

cytokinesis.   These studies show that DNA replication occurs only once per cycle in the

majority of cells, and re-initiation of replication before division occurs only rarely [13]

[14][15][16][17].   While the relative durations of the cell cycle phases can be influenced by

the environment and stochastic factors such as birth length [16], the average duration of

the B, C, and D periods in Mtb are generally in the range of 6-8 hours, 9-12 hours, and 6-9

hours, respectively [18].

These observations from time-lapse microscopy are consistent with metabolic labeling

studies in synchronously-replicating cultures of *Mtb*, which can be generated using a

mutant strain that harbors a cold-sensitive (cos) allele of the DNA replication initiator

DnaA [19]. This *Mtbcos* strain is unable to initiate a new round of DNA replication at 30ºC.

Upon release into the permissive temperature (37ºC), cultures synchronously incorporate

5,6-^{3}H-uracil into alkali stable DNA for 11 hours, consistent with the C period length

observed in single cells. The ability to produce synchronously replicating cultures that

recapitulate the behavior of single cells makes mycobacteria an attractive system to

investigate cell cycle-associated transcriptional changes.

Using the *Mtbcos* strain, we determined the transcriptional profile of synchronously

replicating cultures across the cell cycle, and report that 485 genes meet strict criteria for

periodic gene expression. Only a small fraction of the cell cycle-regulated gene sets of *Mtb*

and *C. crescentus* overlap, suggesting that the links between cell cycle and metabolism are

species-specific. We demonstrate that mRNA expression patterns in *Mtb* reflect the time at

which the encoded proteins are incorporated into the developing septum, suggesting that

functional information can be inferred from the kinetics of gene expression. Using this

framework, we discover an unanticipated functional specialization of distinct nucleotide

anabolic pathways. These observations show that DNA replication and cytokinesis are

coordinated with different primary metabolic pathways, expanding the processes that are

associated with these essential cellular events.

Results

DNA replication and cytokinesis are segregated in synchronously growing

populations of *Mtb*

We generated synchronously replicating cultures of *Mtb* using the temperature-sensitive

Mtbcos strain [19]. Chromosomal replication was inhibited by incubating this strain at

30ºC for 36 hours. Upon shift to the permissive temperature (37ºC), the optical density

(Absorbance$_{600}$) of both the *Mtbcos* mutant and a wild type control culture increased

throughout a 54 hour time course, demonstrating that nutrients did not become limiting.

While the control culture grew at a constant rate over this period, the *dnaAcos* strain

showed a reproducible multiphasic growth pattern, an initial indication that cellular

metabolism may be linked to cell cycle events (**Fig1A**).

In order to estimate the efficiency of the synchronization and to delineate cell cycle periods,

we collected cells every three hours and monitored chromosomal replication and

cytokinesis. The phosphothreonine-binding protein, FhaA, marks sites of division [20], and

we used a fluorescent allele of this protein to calculate a "septation index" that

corresponded to the fraction of cells with FhaA localization at midcell. While the septation

index of a control culture of asynchronous cells (*MtbRv*) was constant throughout the time

course, this metric varied in a periodic manner in the *Mtbcos* strain. The majority of cells

arrested at the non-permissive temperature had an FhaA focus at midcell, which is likely an

artifact of the DnaA inactivation. The septation index of *Mtbcos* cells quickly decreased

upon shift to the permissive temperature, falling below the index of unsynchronized

cultures by 12 hours. Septation reached a peak in the synchronized cultures between 27

and 33 hours after release, marking the cytokinesis phase of the cell cycle **(Fig1B)**.

To monitor chromosomal replication, we devised a PCR assay to quantify the relative

abundance of DNA at the origin (ori) and terminus (ter) of replication. Upon initiation of

replication, the cell will have an ori:ter ratio of 2:1, and this ratio should be maintained

until the terminus is duplicated. As we observed for septation index, the ori:ter ratio

remained constant in unsynchronized cultures. In contrast, the ori:ter ratio peaked twice

in the synchronized cultures **(Fig 1C)**. The first peak lasted for a duration of approximately

12 hours (between 15h-27h post release) and second one lasted between 48-hrs post

release and the end of the study. Based on these data, we estimate that our time course

captured approximately 1.5 cell cycles. Both the septation index and ori/ter ratio varied by

approximately 50% of the range expected of fully synchronized cells, indicating that the

synchrony of our cultures was incomplete. Regardless, these observations indicated that

DNA replication is temporally segregated from cytokinesis, and the cultures were

sufficiently synchronized to perform transcriptional profiling.

Periodic gene expression correlates with cell cycle progression

In order to investigate whether gene expression changes are associated with major cellular

events like DNA replication and cytokinesis, we profiled mRNA abundance in synchronized

cultures every 3 hours across a 54 hour time course. After normalization and scaling, we

first assessed correlation patterns in the dataset. The initial time point after temperature

shift to 37ºC (0hr) was uncorrelated with the rest of the datasets, presumably due to an

adjustment to the temperature shift and was omitted from subsequent analyses. For the

remaining data, we found the highest degree of correlation between adjacent time points,

as expected for a time-resolved dataset **(Fig 2A).** While this was generally true for both the

synchronized *Mtbcos* and unsynchronized *MtbRv* strains, the correlation matrix from the

Mtbcos cultures displayed a distinct three-block structure suggesting the presence of

transcriptionally distinct phases. This structure is even more apparent upon hierarchical

clustering, which revealed a pattern of gene expression consistent with an ordered

progression of events throughout the time course (**FigS1A**).

To take advantage of both replicate measurements and the relatedness of adjacent time

points, we used a Gaussian Process (GP) smoothing approach to estimate relative

expression levels of each gene across the time course **(Fig S1B)**. The expression of many

genes with known cell-cycle related functions were found to peak during the appropriate

period. For example, genes that are important for cell division, such as the regulator, *mtrA*

[21] or the septal components, *sepF* [22] and *sepIVA* [23], peak in expression once during

cytokinesis. Similarly, genes important for DNA replication, such as those encoding DNA

primase (*dnaG)* and the replicative polymerase (*polA)* display two expression peaks during

this time course, corresponding to DNA replication **(Fig2B)**. In addition, we found that the

expression pattern of several primary metabolic pathways mirrored these cell-cycle related

genes, and suggested alterations in metabolic flux. For example, genes necessary for

arginine biosynthesis were co-regulated and had opposing expression patterns to genes

involved in arginine catabolism **(Fig2B)**.

We sought to more formally define genes with an expression pattern consistent with cell

cycle progression. First, we removed genes with correlated expression patterns in

synchronized and unsynchronized cultures to minimize the effect of changes in culture

conditions during the time course. Next, we fit the raw data from both replicates for each

gene to a sinusoidal function with the expected period of the *Mtb* cell cycle, optimizing the

parameters for trend, amplitude, period and phase. Genes with a period outside the range

of reasonable expectations based on the *Mtb* cell cycle were filtered out, along with genes

with low overall expression or high variance between replicates. A goodness-of-fit criterion

based on curve fitting residuals was applied, which maximized the difference in genes

discovered in synchronized versus unsynchronized cultures. These criteria were

determined to produce a false discovery rate of 0.35% using a permuted dataset, and a 2.6-

fold enrichment in genes discovered in the synchronized data set, compared to the

unsynchronized set. 485 genes were categorized as periodically expressed (**Fig 2C**;

Supplementary Table 2), and all major functional categories were represented in this

gene set (**Fig2D)**.

Clustering this set of periodically-regulated genes further highlighted the association

between gene function and cell cycle stage. Genes were distributed into 8 clusters using

hierarchical clustering of the *Mtbcos* expression profiles **(FigS2),** producing groups of

coordinately regulated genes with peak expression values ranging across the time course.

Three of these clusters reflected expression profiles that peak during DNA replication

(Fig2E). In these clusters, we find 9 genes with defined roles in DNA replication (*dnaG,*

polA, helY, ercc3) and nucleotide biogenesis (*carA, carB, pyrB, pyrC, purL*), further indicating

that expression patterns were consistent with gene function.

The periodically regulated gene set (485 genes) in *Mtb* represents 12% of the genome,

whereas 19% of the *C crescentus* genome was previously found to be cell cycle associated

[6][7]. In order to investigate whether similar cellular functions are periodically regulated

in these phylogenetically diverse organisms, we compared the expression patterns of

orthologous genes. Out of the 880 mutual orthologs identified as being reciprocal best

BLAST matches, 229 genes were defined as cell cycle regulated in *C. crescentus* by virtue of

being differentially expressed over time in synchronized cultures [7] and we found 142 to

be periodically expressed in *Mtb* with an overlap of 39 genes (**Fig2F, Supplementary**

Table 3). This overlap contains a number of genes with known cell cycle associated

functions, such as the DNA replication initiator, *dnaA*, the DNA primase *dnaG* and

nucleotide biogenesis (*nrdZ, purl, pyrB*). However, the modest degree of overlap also

suggests that independent sets of genes are periodically expressed in these two

phylogenetically distinct organisms.

mRNA abundance predicts the order of assembly of mycobacterial divisome

components and regulators

The order in which large multicomponent structures, such as the flagellum, are assembled

in bacteria can be predicted based on the transcriptional regulation of the corresponding

genes[24]. Based on this "just in time" transcription model [25], we hypothesized that the

assembly of the large complex of proteins necessary for cell division, aka the "divisome",

may follow the same principles and provide a model to test whether mRNA abundance

could be used to predict the timing of gene function in our system. To test this model, we

assessed the temporal coincidence between mRNA abundance and protein localization at

the developing and maturing septum.

To broadly identify genes that are induced during cytokinesis (27-33h), we identified genes

that only peak once in the time course **(Supplementary Table 1**) and clustered genes

based on similar expression patterns **(Fig3A).** Within the clusters that peak in expression

at the appropriate times, we found a number of genes known to be involved in cytokinesis

(**Fig3A**). The first to be induced was *ftsZ*, the tubulin-like nucleator of the septum, and the

septally-localized Ser/Thr kinase, *pknD*. This was followed by the expression of genes

encoding several divisome-associated proteins, FtsW, SepIVA, and LamA. After these, we

found the gene encoding the new pole landmark protein, DivIVA, was induced. To

determine if the timing of expression predicted the order of assembly, we chose three

genes, *pknD*, *ftsW* and *divIVA* which peaked in expression early, in the middle, and late

during the cytokinesis window, respectively. Pairs of these proteins were fused with

fluorescent tags and their cellular location was determined by time lapse microscopy in *M.*

smegmatis, a related mycobacterial species that expresses orthologs of these proteins and

is an experimentally tractable model of mycobacterial division [26]. Dual-color

fluorescence imaging revealed that PknD, FtsW, and DivIVA appear at the developing

septum in the order predicted by mRNA abundance (**Fig 3B-D**). This order was not dictated

by the maturation times of the two different fluorophores. Although not directly addressed

in our experiments, it is likely that order of localization at the septum is independent of

transcriptional regulation because PknD, and DivIVA fusion proteins were expressed from

constitutive promoters. Instead, the transcriptional order correlates with assembly, which

appears to be dictated at the posttranscriptional level.

Guanosine synthesis influences cytokinesis in mycobacteria

Having demonstrated that gene expression can predict the timing of gene function, at least

in the case of the developing septum, we investigated whether obligate coordination exists

between cellular events, such as DNA replication and cytokinesis, and upstream pathways

that produce the precursors for these processes. In particular, we focused on nucleotide

metabolism by analyzing the expression patterns of enzymes catalyzing anabolic reactions

beginning from the tricarboxylic acid cycle precursor, glutamate to the final nucleos(t)ide

products. Pyrimidine biogenesis from the very early stages of the *carAB*-encoded reactions

down to the *pyrBCDEF*-encoded reactions were most highly expressed during the C phase

at ~12 hours and ~48 hours **(Fig4A)**, consistent with previous reports of an increased

production of thymine nucleotides during DNA replication in *E coli* [27]. This also

corroborates reports of nucleotide pool sizes in *E coli* being insufficient for chromosome

replication and requiring *de novo* synthesis before the onset of the DNA replication phase

[28]. Unexpectedly, the converse expression pattern was observed for the first reaction

unique to guanosine synthesis *(guaB2)* . While both adenosine and guanosine purine rings

are synthesized from the common precursor inosine monophosphate (IMP), the IMP

dehydrogenase, GuaB2, catalyses the *first* dedicated reaction unique to GMP synthesis.

guaB2 peaked in expression during cytokinesis at ~21hours (**Fig4A**). Genes dedicated to

synthesizing adenosine from IMP, *purB* and *amk,* did not appear to be cell cycle regulated.

The reciprocal expression patterns of pyrimidine and guanosine synthetic genes suggested

that the requirement for these metabolites varied across the cell cycle and these

requirements were associated with distinct cellular events. To investigate this hypothesis,

we generated mutant *Mtb* strains in which synthesis of pyrimidines or guanosine was

inhibited via the inducible genetic depletion of the PyrE or GuaB2 proteins, respectively.

Each gene was fused to a C-terminal DAS+4 tag (DAS) that facilitated Clp protease-

mediated degradation upon removal of anhydrotetracycline (aTc). In both cases, protein

depletion inhibited bacterial growth (**Fig4B**), consistent with the essentiality of these

pathways [29]. As *Mtb* expresses three GuaB paralogs, we verified that GuaB2 is essential

for guanosine synthesis by supplementing the *guaB2-DAS* strain with 200uM guanine or

guanosine. Consistent with previous studies [30], guanine partially rescued the growth

defect of the depleted strain, whereas guanosine supplementation led to complete rescue

(**Fig4C**).

The inverse expression patterns of *pyr* genes and *guaB2* implied increased *de novo*

synthesis of pyrimidine nucleotides and guanosine was preferentially required for DNA

synthesis or cytokinesis, respectively. To test this hypothesis, we used morphological

criteria to infer which cellular processes were primarily impacted by the inhibition of these

nucleotide synthetic pathways. PyrE depletion resulted in cell elongation before growth

arrest. The mean cell length increased from 3.3uM to 4.2uM upon PyrE depletion (**Fig4D**).

A similar phenotype was observed upon DNA gyrase GyrB depletion and in cells treated

with the gyrase inhibiting fluoroquinolone, moxifloxacin **(Fig4D)**, which disrupts DNA

replication and causes cell filamentation in *E coli* [33].   Cell elongation upon inhibition of

DNA replication is consistent with a number of previous observations in *B. subtilis* [31] and

M. smegmatis [32].

In contrast to the elongation observed upon PyrE depletion, GuaB2-depleted cells were the

same length as wild type, but many of these growth arrested cells had bulges at midcell or

one pole, suggesting that GuaB2 depletion may influence cell division **(Fig4D)**. To

determine if the polar bulges were derived from misshapen septa, we performed time-

lapse microscopy in *M. smegmatis* cells treated with a specific chemical inhibitor of GuaB2

(VCC234718). Similar to genetic depletion, chemical inhibition of the GuaB2 enzyme also

inhibited growth and produced bulges at midcell or one pole **(Fig4E)**. Time-lapse

microscopy revealed that cells began to bulge at midcell by the completion of 1-2 cell cycles

after the initiation of VCC234718 treatment, and that misshapen poles were derived from

these bulges. Taken together, these observations imply that the cellular requirement for

guanosine increases during cytokinesis, and that this requirement is reflected in the

aberrant septation of guanosine-depleted cells.

FtsZ is among the most abundant proteins in the cell and this tubulin-like protein binds and

hydrolyzes GTP [34] as it undergoes the cycles of polymerization and depolymerization

necessary for septation [35]. This GTP requirement suggested a possible mechanistic

connection between guanosine nucleotide levels and septation. To investigate whether the

effect of guanosine depletion on septation could be attributed to altered FtsZ dynamics, we

determined the effect of inhibiting both processes simultaneously using the GuaB2

inhibitor VCC234718 and C109, an inhibitor of FtsZ GTPase activity and polymerization

[36]. Consistent with the hypothesized mechanistic link, we observed significant

antagonism between these compounds. Even at concentrations of VCC234718 that alone

had no effect on growth (0.5 - 4uM), this compound consistently increased the IC50 of C109

(**Fig4F**). In contrast, we found no interaction between VCC234718 and spectinomycin, an

inhibitor of another major GTP consuming pathway, translation. C109, has previously been

found to act additively with PCI90723, a compound that stabilizes the FtsZ filament. The

opposite antagonistic interaction we observed between C109 and a GuaB2 inhibitor implies

that guanosine inhibition inhibits polymerization, an effect that is consistent with the

known GTP requirement for FtsZ polymerization.[34], [37]. Taken together, these data are

consistent with a model in which transcriptional induction of GuaB2 during cytokinesis

coincides with the increased consumption of GTP by FtsZ, and the septal defects observed

on guanosine depletion are related to defects in FtsZ dynamics.

Discussion

Previously, global transcriptomic studies of the prokaryotic cell cycle have been limited to

the *C. crescentus* model. In this study, we leveraged a genetic strategy to synchronize *Mtb*

cultures, and characterized cell cycle associated transcriptional changes in this organism.

Comparisons between our *Mtb* studies and previously generated *C. cresentus* data are

limited by a number of technical differences, including the method and degree of

synchronization, that likely limited our ability to discern periodic patterns of relatively

small amplitude in the *Mtb* dataset.   Regardless, our findings are broadly comparable to

observations in *C. crescentus*, as we found that a similar fraction of the genome is

differentially expressed across the cell cycle in both systems, and a small fraction of

orthologous genes are periodically-regulated in both. Despite these similarities, the

majority of cell-cycle associated transcriptional changes (82% of *C. crescentus* genes and

72% of *Mtb* genes) were unique to each organism. Thus, despite some similarities, our

analysis indicates that cell cycle progression is associated with distinct transcriptional

networks in these structurally and phylogenetically divergent organisms.

In a number of cases, we found that increases in mRNA abundance could be used to

associate genes with temporally-resolved cell cycle events, such as septation. The

sequential expression of divisome components as cell division progresses has been

observed previously in *C. crescentus* [6]. We provide functional evidence that this

hierarchical expression pattern is associated with the order of divisome assembly in

mycobacteria by demonstrating that the timing of gene induction correlates with the

recruitment of the encoded proteins to the developing septum. Based on transcriptional

data, we inferred the following order of assembly at the mycobacterial septum:

FtsZ>PknD>FtsW>LamA>SepIVA>DivIVA. The recruitment of these proteins spans

sequential processes of divisome assembly and new pole biogenesis. FtsZ initially marks

the future division site [38], facilitating the recruitment of divisome components such as

FtsW [39] and SepIVA [23]. The arrival of LamA at the later stages of assembly is

consistent with its role in delaying septation and thereby promoting asymmetric cell

division[40]. DivIVA is thought to be recruited to the negative curvature of the new pole

after septation[41][42][43] and the segregation of daughter cell cytoplasm [13]. While

these observations highlight that gene expression patterns can be used to predict the order

of complex assembly, the outcome of this coordination remains unclear. As only a subset of

currently known septal components were found to be periodically expressed,

transcriptional regulation is unlikely to be the primary determinant of assembly order.

Instead, this type of hierarchical gene expression has also been proposed as a mechanism

to maximize efficiency by restricting protein expression to the period when it is needed

[24][25]. Regulation of divisome assembly and function likely involves additional

posttranslational mechanisms, as we found that the Ser/Thr kinase, PknD, is recruited

relatively early in septal development and previous work described an important role for

the Ser/Thr phosphatase, PstP, in cell division[44], [45]. While it remains possible that

transcriptional regulation controls some aspects of septation, our data only show that

expression pattern can predict the timing of gene function.

The observed periodic expression of primary metabolic functions suggested the

importance of coordinating cell cycle events with the upstream pathways that provide their

precursors. This model is supported by our finding that pyrimidine synthetic genes peak

during a distinct cell cycle period than the dedicated guanosine synthetic gene, GuaB2; and

that the depletion of these genes primarily disrupts different cellular processes.

It is not surprising that the requirement for pyrimidines would increase during DNA

synthesis, as *de novo* nucleotide synthesis is necessary to replicate the chromosome[28].

However, the distinct cytokinesis defect observed upon GuaB2 depletion was

unanticipated. Guanine nucleotides are important for a myriad of cellular processes,

including DNA replication, transcription, and macromolecular synthesis. We speculate that

the preferential septation defect we observe upon GuaB2 depletion is related to the

relatively low affinity of FtsZ for GTP. An accurate determination of nucleotide pool sizes in

mycobacteria is still an open question [46], but in *E coli,* FtsZ has ∼500-fold lower affinity

for GTP than the DnaE1 replicative DNA polymerase (K_m FtsZ$_{GTP}$ =1mM[47] ;

K_mDnaE1$_{GTP}$=2uM[48]). Thus, the reported intracellular concentration of GTP[49] would

support only one-half of the V_{max} of FtsZ [47]. These data indicate that GTP levels may limit

FtsZ dynamics. It is likely that GuaB2 depletion leads to aberrant FtsZ activity and not a

complete loss of function, since genetic depletion of FtsZ leads to filamentation [50], a

phenotype distinct from the one we observe upon GuaB2 depletion. Instead, alterations in

FtsZ filament length or rate of turnover may underly these defects.

Both DNA replication and cell division are essential processes that have been the focus of

antibacterial drug discovery efforts[46][51]. In most cases, these efforts focus on

inhibiting a limited number of physical components of the bacterial replisome or divisome.

Our transcriptional data identified a wide variety of genes that are coordinately expressed

with the genes encoding these complexes, and therefore may be required for their activity.

While we have only investigated these functional dependencies in the context of nucleotide

synthesis, our data suggest that many similar dependencies exist and can be predicted from

transcriptional profiles. If so, these data could be used to identify a wealth of new

strategies for inhibiting these specific essential cellular processes.

Acknowledgements

We would like to thank Aashish Srivastava under the guidance of James Sacchettini and the

Texas A&M Genomics and Bioinformatics Service for sequencing the RNASeq libraries.

VCC234718 was a generous gift from Vinayak Singh & Valerie Mizrahi. C109 was a

generous gift from Vadim Makarov. Genetic depletion strains were generated as part of a

larger collaborative effort including the labs of Deborah Hung, Eric Rubin, Sabine Ehrt and

Dirk Schnappinger. Kenan Murphy and Charlotte Reames contributed to strain generation

from the Sassetti lab. We would like to thank Megan Proulx for critical reading of the

manuscript. This work was supported by grant AI095208 to CMS.

Author contributions

Conceptualization- ACB, TRI, CMS

Methodology- ACB, TRI

Investigation- ACB

Validation- ACB

Formal Analysis- ACB, SS, TRI

Writing Original Draft- ACB

Writing Review & Editing- ACB, TRI, CMS

Funding acquisition- CMS

Supervision- CMS

Declaration of interests

The authors declare no competing interests.

Figure Legends

Figure 1: DNA replication and cytokinesis are segregated in synchronously growing

populations of Mtb

(A) Growth of *Mtbcos* (left) and *MtbRv* (right) after release into permissive temperature

37ºC. X axis: hours at 37ºC. Y axis: Absorbance$_{600}$

(B) FhaA septation index assay to determine cytokinesis phase. Percentage of *Mtbcos* (left)

and *MtbRv* (right) populations containing an FhaA-venus focus localized at midcell after

release into permissive temperature. Data points are representative of two biological

replicates. Blue line is smoothed via the Gaussian process. The blue band indicates 95%

confidence interval. Significant difference between *Mtbcos* and *MtbRv* curves was

determined using a likelihood ratio test which determines if the data is fit best by a

combined model (null hypothesis) or separate strain specific models (alternate

hypothesis). Log_likelihood difference between combined and separate models = -38.489,

p-value (chi-squared distribution; df=3) = 1.1e-16.

(C) Origin/terminus assay to determine the DNA replication phase. Relative ori/ter ratio of

Mtbcos (left) and *MtbRv* (right) populations after release in permissive temperature. Data

points are representative of two biological replicates. Blue line is smoothed via the

Gaussian Process. The blue band indicates 95% confidence interval. Significant difference

between *Mtbcos* and *MtbRv* curves was determined using a likelihood ratio test

Log_likelihood difference between combined and separate models = -12.412, p-value (chi-

squared distribution; df=3) = 1.679e-05.

Figure 2: Periodic gene expression correlates with cell cycle progression

(A) Correlation matrix of DESeq normalized counts for single replicates of *Mtbcos* (top) and

MtbRv (bottom) for all 16 timepoints. (Blue: Pearson's correlation coefficient=0; Yellow:

Pearson's correlation coefficient=1).

(B) Relative expression (GP smoothed, DESeq normalized read count for each time point

divided by the mean value for that gene across time) of genes involved in DNA replication &

cell division (top panel); arginine catabolism & anabolism (bottom panel).

(C) Relative expression (standard normalized DESeq counts - each value is subtracted by

the mean for that gene across time and then divided by the standard deviation) of 485

periodically expressed genes in *Mtbcos* (rows) sorted by peak expression time (columns).

(D)Fraction of periodically expressed genes present in different Gene Ontology categories.

(E) Clusters containing periodic genes with expression patterns consistent with a role in

DNA replication. Known DNA replication/nucleotide biogenesis genes found in these

clusters are listed. X axis: hours post release into permissive temperature. Y axis Relative

expression (standard normalized DESeq counts - each value is subtracted by the mean for

that gene across time and then divided by the standard deviation)

(F) Overlap between periodically expressed (*Mtbcos*) and differentially expressed (*C.

crescentus*) mutual orthologs.

Figure 3: mRNA abundance predicts the order of assembly of mycobacterial divisome

components and regulators

(A) (Left) Clusters of *Mtbcos* genes with expression patterns that peak during the

cytokinesis period. (Right) Scaled relative expression (GP smoothed, DESeq normalized

read count for each time point divided by the mean value for that gene across time, then

scaled) of known cytokinesis genes from these clusters.

(B) (Left) Scaled relative expression of PknD and DivIVA. (Right) Time-lapse imaging of *M*

smegmatis expressing PknD-Venus (green) and DivIVA-RFP (red). Time (minutes) before

the arrival of DivIVA at midcell is indicated.

(C) (Left) Scaled relative expression of FtsW and DivIVA. (Right) Time-lapse imaging of *M*

smegmatis expressing FtsW-Venus (green) and DivIVA-RFP (red). Time (minutes) before

the arrival of DivIVA at midcell is indicated.

(D) Time (minutes) between initial arrival of PknD (n=10), FtsW (n=7) and DivIVA at

midcell. Error bars indicate mean ± SD. Statistically significant difference between pknD

and ftsW determined using an unpaired T-test (α =0.05; p=0.0023). Statistically significant

difference between ftsW and divIVA determined using a chi-squared test (α =0.05 ; χ^2=7;

df=1; p<0.01).

Figure 4: Guanosine synthesis influences cytokinesis in mycobacteria

(A) Relative expression (GP smoothed, DESeq normalized read count for each time point

divided by the mean value for that gene across time) of IMP dehydrogenase *guaB2*

compared to pyrimidine biosynthesis genes *carA*, *carB*, *pyrB*, *pyrC, pyrD, pyrF*

(B) Cumulative growth (Absorbance$_{600}$) of *M tuberculosis guaB-2*DAS (top), *gyrB*-DAS

(center), *pyrE*-DAS (bottom) without depletion (solid line) and with depletion (dotted line).

Arrows indicate the time during the pre-depletion period when cultures were back diluted

into fresh growth medium. Data are represented as mean ± SD of two biological replicates

(C) Growth (Absorbance$_{600}$) of *M tuberculosis guaB2*-DAS ±depletion in the presence of

either 200μM guanine or guanosine. Data are represented as mean ± SD of two biological

replicates.

(D) *M tuberculosis* cellular phenotypes upon genetic depletion of GuaB2, PyrE, GyrB.

Images were obtained after the cessation of growth in depleted cells. In the case of GuaB2,

septal bulges (arrowheads) and polar bulges (arrows) are indicated. Histograms indicate

the cell length distribution of cells in which the target was either not depleted (gray) or

depleted (black). *Mtb* was treated with 0.2μM moxifloxacin for 24 hours and imaged.

Histograms indicate the cell length distribution of cells in untreated (gray) or treated cells

(black). MFD = Maximum Feret Diameter (1μM ~ 0.11MFD). Statistically significant/non-

significant difference between the cell length distributions was determined using the

Mann-Whitney test. ($p_{guaB2}=0.214$; $p_{pyrE}<0.001$; $p_{gyrB}<0.001$; $p_{moxifloxacin}<0.001$)

(E) Time-lapse imaging at 20 minute intervals of GFP-expressing *M smegmatis* treated with

2μM VCC234718.

(F) Left: Susceptibility of *M smegmatis* to VCC234718. Data are represented as mean ± SD

Center: Cross titration assay on GFP-expressing *M smegmatis* with the indicated

concentrations of VCC234718 and C109. Statistically significant difference between the

VCC alone curve and other curves was determined using an extra-sum-of-squares F-test

($\alpha=0.05$). $p_{(VCC\ 0.5_\mu m)}= 0.0215$; $p_{(VCC\ 1_\mu M)}= 0.0284$; $p_{(VCC2_\mu M)}= 0.0001$; $p_{(VCC4_\mu M)}= 0.0019$

Right: Cross titration assay on GFP-expressing *M smegmatis* with the indicated

concentrations of VCC234718 and spectinomycin. The differences between the VCC alone

curve and other curves were not significant, as determined using an extra-sum-of-squares

F-test ($\alpha =0.05$). $p_{(VCC0.5_\mu M)}= 0.9989$; $p_{(VCC1_\mu M)}= 0.9999$; $p_{(VCC2_\mu M)}=$ambiguous ; $p_{(VCC4_\mu M)}=$

0.9978

Materials and Methods

Strains

The *Mtbcos* strain was obtained from [19]. *MtbRv* is the H37Rv strain used as an unsynchronized control. Mtbcos and MtbRv expressing FhaA m-venus were transformed with pKP887 (mycobacterial replicating plasmid MEH expressing MSMEG FhaA-Venus expressed from the MSMEG *fhaA* native promoter (from K.P. Sundaram). *M smegmatis* expressing ftsW-mVenus and divIVA-RFP was transformed with ptb21-ftsW-mVenus-MEK and tb21-divIVA-RFP-MCtH. *M smegmatis* expressing pknD-mVenus and DivIVA-RFP was transformed with p16-pknD-mVenus-MEK [52] and tb21-DivIVA-RFP-MCtH. *Mtb* hypomorphs used in this study were generated as part of an earlier study [53] using a controlled protein degradation system described previously[54]. Three strains were used in this study: *Mtb guaB2*-DAS-Hyg^R+Giles-TetON1-sspB-str^R ; *Mtb* gyrB-DAS-Hyg^R+Giles-TetON6-sspB-str^R ; *Mtb* pyrE-DAS-Hyg^R+Giles-TetON1-sspB-str^R. *M smegmatis* expressing green fluorescence contains the plasmid CT161 (m-Venus pMV261 Hyg^R) obtained from the Eric Rubin Lab.

Mtbcos synchronization

Cultures of Mtb*dnaAcos115* generated in a previous study [19], MtbH37Rv, Mtb*dnaAcos115*-FhaA-Venus and MtbH37Rv-fhaA-Venus were grown in standard culture media at 37ºC under shaking conditions till OD$_{600}$ 0.4. The cells were shifted to 30ºC for 36 hours. The cultures were then shifted to 37ºC and the cultures were processed for either DNA isolation, RNA isolation or fluorescent microscopy at the following times: 0h, 3h, 6.5h, 9h, 12h, 18.5h, 21h, 27h, 31h, 33h, 36h, 39.5h, 42h, 45.5h, 52h and 55h.

Chromosomal DNA isolation

Chromosomal DNA was isolated from the cell pellet of 5ml culture from each timepoint.

Briefly, 0.5 ml of chloroform:methanol (2:1) was added and the mixture was vortexed 5X

1min. 0.5ml of phenol:chloroform was added and the mixture was vortexed for 30 seconds.

Finally, 0.5ml of TE buffer was added. This was centrifuged at 12,000g at4C for 5 minutes.

The upper phase was mixed with 1 volume of chloroform and vortexed. After

centrifugation, the upper phase was added to a new tube and 1/10 volume of 3M sodium

acetate and 1 volume of isopropanol was added. Precipitated DNA was spun out of solution

and resuspended in 20ul of TE buffer.

Origin:terminus assay

Multiple primer sets (designed using the Primer3 design tool) amplifying 150bp at each

location (Origin-0MB region surrounding Rv0001; Terminus –2.2MB region surrounding

Rv1949c) of the MtbH37Rv genome were tested for amplification efficiency. Efficiency was

calculated from the negative slope of the standard curve of C_T v/s template concentration.

The primer sets with the highest and most similar efficiencies for both loci were selected

(95% for the origin and 93% for the terminus). Quantitative PCR was done using SYBR

green (Biorad iQ SYBR Green Supermix) with 2ng of gDNA template per reaction. Delta Ct

values were calculated as $dCt= Ct_{ori}-Ct_{ter}$. 2^-dCt values were then calculated for each

timepoint. These values were then divided by the mean 2^-dCt across all timepoints to

generate a relative ori/ter ratio for each timepoint.

Microscopy

Static imaging

At each time point post release into 37ºC (Fig1B) or timepoint post genetic depletion of

GyrB, GuaB2 and PyrE (Fig4D), 1ml of *Mtb* culture was centrifuged and cells were re-

suspended in a phosphate buffered saline solution containing 0.05% Tween80 and 4%

paraformaldehyde. These fixed cells were then placed onto an agarose pad and DIC (Fig4D)

or wide field fluorescence imaging (Fig1B) was performed with a DeltaVision Personal DV

microscope (GE Healthcare) using a 60X oil immersion objective (AP). Cell lengths in Fig4D

were determined using CellProfiler™ [55] which calculates a MFD (Maximum Feret

Diameter) which is a measurement of the largest number of pixels between the two ends of

the cell obtained while rotating a caliper along all possible angles. The approximate

conversion factor of MFD to microns is 0.11. Calculating an MFD is especially useful for

measuring mycobacteria since all cells are not strict rods (cells undergo V-snapping prior

to resolution of cytokinesis and daughter cell separation). The cell debris observed during

GyrB depletion in Fig4D was excluded from cell length quantification by training

CellProfiler using CP Analyst™.

Live cell imaging

10ul of cells in logarithmic phase (OD$_{600}$ 0.2-0.5) were spotted on a glass bottom 24-well

plate (MatTek Corporation). 500ul of molten Luria Bertani medium (40-50C) was spread

over the cells and allowed to solidify. For experiments with VCC234718, molten LB

containing 2uM final concentration of VCC234718 was prepared before layering over the

cells. Time-resolved imaging was performed with a DeltaVision Personal DV wide field

fluorescence microscope equipped with Ultimate Focus™ capabilities and an

environmental chamber warmed to 37 °C (Applied Precision). Images were taken at 5 or 10

minute intervals.

RNA isolation, library preparation and sequencing

At each timepoint, 45ml culture was pelleted and resuspended in 1ml of TRIzol

(Invitrogen) and transferred to lysing matrix tubes (MP Biomedicals: Lysing Matrix B).

Cells were lysed in a MP Biomedicals Fast Prep-24 homogenizer (maximum power-6.5, 4 X

30s cycles, rest on ice for 5 minutes in between cycles). RNA was purified according to the

manufacturer's directions. RNA cleanup was performed with Qiagen RNeasy Mini kit

(74104) omitting the DNase step. Instead, after elution, in-tube DNase treatment was

performed using Ambion DNase Turbo. RNeasy cleanup was repeated again with double

volumes of RLT and ethanol. RNA was subjected to rRNA removal with Ribozero Bacteria

kit (Illumina-MRZB12424). Deep sequencing library was prepared using KAPA Stranded

RNASeq kit (KK8401). The RNAseq libraries were sequenced on an Illumina 2500

instrument in paired-end mode, using a read-length of 150+150bp. The mean number of

reads per sample was 8.9M (range 4.2-16.5M). The reads were mapped to the H37Rv

genome using Burroughs Wheeler Alignment [56] with default parameter settings. Reads

mapping to each ORF were totaled (sense strand only). Because certain loci were over-

represented (e.g. rrs, rnpB, ssr, Rv3661, which had counts ~0.5-1M), counts were

truncated to a maximum coverage of 10,000 (reads/nt).

Data normalization, filtering and centering

The global expression profiles of *Mtbcos* samples showed a gradual increase in expression

of a few genes that dominate expression at latter time-points. Consequently, a

compensatory decrease was observed in expression of other genes, making normalization

by traditional reads per kilobase per million (RPKM) mis-representative. To correct for the

bias induced by these outliers, the normalization method implemented in DESeq2 [57] was

used, which first normalizes counts by the geometric mean for each gene across samples,

and then scales each dataset to have a common median (which is less sensitive to outliers).

This was applied to all 64 datasets (2 strains X 2 replicates X 16 time-points) in parallel. As

a result, the expression patterns were well-calibrated between time-points, with the

medians matched**.**

To identify a subset of genes with meaningful expression, the average expression over all

time-points was calculated for each gene and divided by gene length (in nucleotides). 1070

genes out of 4018 with coverage<0.25 were dropped because expression patterns for

genes with low expression are inherently noisy, leaving 2948 genes with coverage>0.25.

(Supplementary Table 1). Additionally, we removed 127 genes out of 2948 genes whose

expression was >90% correlated between *Mtbcos* and *MtbRv* from subsequent analysis, as

their expression patterns were assumed to be determined more by time than by difference

in the strains. To center the expression values, the counts were divided by the mean for

that gene across all the time points. This was done independently for *Mtbcos* and *MtbRv.*

Gaussian Process Smoothing

In order to meaningfully integrate the data from the two replicates and to smooth out

profiles over time, we used a Gaussian Process (GP) to fit the raw data (septation index and

ori/ter – Fig1, gene expression- Fig2, Fig3, Fig4).

A GP model is a Bayesian model that estimates the probability distribution over functions

using Gaussian distributions for likelihood functions. The advantage of a GP is that it is

unbiased and therefore does not require assumptions of form of function. Instead, it only

assumes that adjacent time points are better coupled than distant time points and that this

correlation is based on Gaussian distributions.

A Gaussian Process is specified by a mean function and a covariance function

$$f(x) \sim GP\big(m(x), k(x, x')\big)$$

646   A prior mean m(x)=0 and a covariance function, squared exponential is given as:

$$k(x, x') = \sigma^2 \exp\left(-\frac{1}{2}\sum_{i=1}^{d}\frac{(x_i - x'_i)}{l_i^2}\right)$$

647   where l^2= lengthscale, σ^2 = variance, d = input dimension

648   We normalized the expression value e(g,t) (with addition of pseudocounts of 10) of each

649   gene g at each time point t by dividing the mean across all time points, and then taking log

650   base e transformation so that the normalized value $e'(g,t)$ fluctuates with a mean of 0. The

651   formula is given as:

$$e'(g,t) = \log_e e\,\frac{e(g,t)}{\sum_t^T e(g,t)}$$

652   Gaussian estimation of the expression levels for a gene at different time points, subject to

653   noise is given as:

$$y = f(x) + \varepsilon \quad \text{where: } \varepsilon \sim N(\mu, \sigma_n^2)$$

655   The predictive distribution for 15 test time points ($\sim$ 3 hour intervals, 3-55 hours),

656    $\{x_1, x_2, \dots., x_*\}$ is specified as:

$$p(f_*|x_*, x, y) = N(m(x_*), k(x_*))$$

657   where:

$$m(x_*) = k(x_*, x)^T (k(x, x) + \sigma^2 I)^{-1} y$$

$$k(x_*) = k(x_*, x_*) - k(x_*, x)^T (k(x, x) + \sigma^2 I)^{-1} k(x_*, x) + \sigma^2$$

We utilized the GPy Python package to fit the relative expression data (value for cos1 and

cos2 simultaneously normalized by the mean expression level across all 60 time points for

each gene using the following hyperparameters: variance = 1.0, noise variance = 0.1 and

lengthscale (range 1 ~ 50) optimized to Maximum Likelihood Estimate (MLE) using a grid

search method. After fitting the model, the predicted value (i.e. posterior mean) for each

time point can be extracted. **Fig S1B** shows the GP regression obtained for *polA* (Rv 1629:

DNA polymerase). Not only do the fitted values from the GP model generally interpolate

between the observed data at each time point, they also present a smoother profile by

averaging between adjacent time points to reduce noise. The error bands show the

uncertainty in the model (95% confidence interval which can be denoted as ±1.96*σ, where

σ is the estimated standard deviation at each X-coordinate (time point) from the Gaussian

Process model based on variance of the training data and surrounding points).

Periodicity Analysis

Traditional signal analysis methods like Fourier analysis, Fisher's g-test, etc. as suggested

by Wichert et. al.[58] performed poorly on our dataset because our experiment captured

only about one-and-a-half cell-cycles. Thus, to identify periodic genes, we took an

approach of sinusoidal curve-fitting, reminiscent of the non-linear curve fitting method

described by Straume et.al. (COSIN2NL in COSOPT) [59]. We fit the expression profiles for

each gene to a sin curve with free parameters (including frequency, phase, and trend), and

selected genes with frequencies and amplitudes in a reasonable range. Goodness-of-fit was

measured using residual sum-of-squares (RSS). Importantly, it is difficult to draw an

absolute cutoff for significance based on RSS, since any data can be fit to a sin in some way,

and RSS incorporates intrinsic noise in the data (E.g. between replicate observations).

Hence, we took a comparative approach by also fitting the data to a quadratic curve, which

captures the general trend of the expression profiles. We then compared the RSS of the sin

fit to the RSS of the quadratic fit (which must also pay a similar price for noise in the same

data). Periodic genes are defined as those that exhibit oscillatory behavior above and

beyond the trend that can be represented by a quadratic. The curve fitting for each gene

was applied to the DESeq-normalized read counts (15 time-points, 2 replicates each). The

sinusoidal function implemented is written as:

689                              $y_{sin}(t) = A \sin(\omega t + \Phi) + B + Ct$

where: A = Amplitude; ω= Frequency; B= Mean offset; Φ= phase shift; Ct= a linear term to

capture a net increasing or decreasing trend in the expression. The parameters in this

function were optimized using the *curve_fit*() function in SciPy using non-linear least-

squares. We then selected genes based on period length (27.5 hours< period < 55 hours)

and amplitude ($\geq$ 0.7). We also removed genes with a correlation coefficient of > 0.9

between expression profiles in Cos versus Rv. The residual sum-of-squares (goodness-of-

fit) was calculated as follows:

$$RSS_{sin} = \sum_{t=1..15, \ r=1,2} (y_{t,r} - y_{sin}(t))^2$$

697   where y_{sin} are the sin function estimates for each time point.

698   A similar curve-fitting approach was used to fit the data to a quadratic curve:

699
$$y_{quad}(t) = D\ t^2 + E\ t + F$$

700   using *curve_fit*() to optimize the parameters D, E, and F for each gene, and the residual was

701   calculated as:

$$RSS_{quad} = \sum_{t=1..15,\ r=1,2} (y_{t,r} - y_{quad}(t))^2$$

Finally, a score was calculated for each gene based on the ratio of residuals. To

meaningfully enrich periodic genes in *Mtbcos,* we used a Receiver Operating Characteristic

(ROC) curve to determine the RSS_{sin}/RSS_{quad} range where we optimally enrich for periodic

genes in *Mtbcos.* The RSS_{sin}/RSS_{quad} range was determined to be 0.35-0.45. A threshold of

0.45 was chosen based on examining plots (**Fig S3, S4**) that visually exhibit clear

oscillatory behavior (beyond the general trend). Thus, genes with a ratio of less than 0.45

were identified as periodic:

$$\frac{RSS_{sin}}{RSS_{quad}} < 0.45$$

which means that the sinusoidal fit reduces the residual error by more than two-fold over a

quadratic curve and hence fits the data better.

Using this comparative curve-fitting approach, 485 genes were identified as periodic

in *Mtbcos* (**Supplementary Table 2)**, and only 183 genes in *MtbRv* , a ~2.6 fold

enrichment. To estimate the number of false positives in the set of 485 genes, we

randomized the data (by shuffling the genes and timepoints) and subjected the randomized

dataset to the same analysis as described above. This permutation analysis yielded only 14

periodic genes under the null hypothesis. Thus, we estimate the false discovery rate (FDR)

at approximately 14/4019 = 0.35%.

Clustering

Genes were clustered based on their expression profiles using hierarchical clustering (*hclust()* in R), using the ward.D2 method [60] based on the Euclidean distance between the vectors of expression values averaged between replicates over the 15 time points, which were standard-normalized for each gene (subtract mean and divide by standard deviation) to make the mean expression level equal to zero for each. The optimal number of clusters was determined based on the Bayesian Information Criterion (BIC) using *mclustBIC*() in the *mclust* R package [61], which showed that the optimal number of clusters among the 485 *Mtbcos* periodic genes was 8 (using the 'VEE' model). The dendogram was then divided into 8 disjoint clusters using *cutree()*.

Peak Assignment

Using the GP fit data, we applied the following criteria to assign a peak to a gene's expression profile. The time series T with n observations for each gene with smoothed expression values $\{x_1, x_2, \ldots., x_n\}$ at different time points $\{t_1, t_2, \ldots., t_n\}$ was defined as:

$$T = \{(t_1, x_1), (t_2, x_2), \ldots. (t_n, x_n)\}$$

First, to screen out the increasing or decreasing trend at the beginning and end of the time series, and to focus on the cytokinesis phase in the middle of the time course, we excluded the first and last two time points from the peak assignment. Second, to identify well-spaced major peaks across time points, we defined a point x_i as a peak if it has a greater magnitude than its two nearest neighbors on both sides. This is defined as:

$$x_i > x_{i+1}, x_{i+2}, x_{i-1}, x_{i-2} \quad \forall_i = 3, 4, \ldots., n-2$$

Furthermore, to filter out the genes with lower fluctuations, the difference between the

740   magnitude of the highest peak x_h and the global minimum g_{min} was restricted to be greater

741   than 0.5. Additionally, in the case of more than one peak in the time series, all the peaks

742   were constrained to have at least a half magnitude of the highest peak in the expression

743   profile. Finally, a set of peaks P for a time series was identified as:

$$P = \{(t_i, x_i) | (x_i > x_{i+1}, x_{i+2}, x_{i-1}, x_{i-2}) \wedge (x_i - g_{min} > 0.5) \wedge (x_i \geq 0.5 * x_h)\} \ \forall_i = 3, 4, \dots, n-2$$

Among the significantly expressed genes, the peak assignment identified 1620 genes with a

single peak and 71 genes with two peaks in the *Mtbcos* strain compared to 903 genes with a

single peak and 8 genes with two peaks in *MtbRv*. Similarly, 1222 genes in the cos strain

and 2344 genes in the wild type did not have any major peak. This once again confirmed

that the gene expression levels in the *Mtbcos* strain show significantly higher fluctuations

than *MtbRv*.

Supplemental figure legends

FigS1A: Hierarchical clustering of relative expression of genes (rows) in *Mtbcos* (log

transformed DESeq normalized counts, centered around the mean, similarity metric:

centered correlation, clustering method: centroid linkage). Columns - hours at 37ºC.

FigS1B: Expression profile of the DNA polymerase *polA*. Data from two replicates (yellow

and blue dashed lines) , GP fit (solid blue line) and sinusoidal fit (pink dotted line) are

shown. Y axis: Relative expression (DESeq normalized value for each time point divided by

the mean polA expression value across all time points.

FigS2: Eight clusters containing the 485 periodic genes in *Mtbcos* (Distance matrix:

Euclidean, Clustering method: Hierarchical).

Fig S3: Sinusoidal fits (red) and quadratic fits (gray) to standard normalized DESeq read

counts (count-mean/SD) for all genes in *Mtbcos*

Fig S4: Sinusoidal fits (red) and quadratic fits (gray) to standard normalized DESeq read

counts (count-mean/SD) for all genes in *MtbRv*

Supplemental tables

Table 1: Normalized DESeq2 values for all detected 2948 genes which were deemed to be

significantly expressed. Each value represents relative expression, the DESeq2 value

normalized to the mean value of that gene across time. Genes determined to have one or

two peaks in expression are indicated along with their peak cluster assignment.

Table 2: Standard normalized DESeq2 values for the 485 periodic genes (Count -mean/SD).

Periodic gene cluster ID is indicated.

Table3: List of overlapping periodically expressed *Mtb* (this study) and differentially

expressed *C crescentus* [7] orthologs.

References

[1] C. E. Helmstetter and S. Cooper, "DNA synthesis during the division cycle of rapidly growing
Escherichia coli Br," *Journal of Molecular Biology*, vol. 31, no. 3, pp. 507–518, Feb. 1968,
doi: 10.1016/0022-2836(68)90424-5.
[2] S. Cooper and C. E. Helmstetter, "Chromosome replication and the division cycle of
Escherichia coli Br," *Journal of Molecular Biology*, vol. 31, no. 3, pp. 519–540, Feb. 1968,
doi: 10.1016/0022-2836(68)90425-7.
[3] B. Gibson, D. J. Wilson, E. Feil, and A. Eyre-Walker, "The distribution of bacterial doubling
times in the wild," *Proc Biol Sci*, vol. 285, no. 1880, Jun. 2018, doi:
10.1098/rspb.2018.0789.
[4] H. E. Kubitschek' and C. N. NEWMANt, "Chromosome Replication During the Division Cycle
in Slowly Growing, Steady-State Cultures of Three Escherichia coli B/r Strains," vol. 136, p.
12, 1978.

[5]   K. Skarstad, H. B. Steen, and E. Boye, "Cell cycle parameters of slowly growing Escherichia
coli B/r studied by flow cytometry.," *J Bacteriol*, vol. 154, no. 2, pp. 656–662, May 1983.
[6]   M. T. Laub, H. H. McAdams, T. Feldblyum, C. M. Fraser, and L. Shapiro, "Global Analysis of
the Genetic Network Controlling a Bacterial Cell Cycle," *Science*, vol. 290, no. 5499, pp.
2144–2148, Dec. 2000, doi: 10.1126/science.290.5499.2144.
[7]   G. Fang *et al.*, "Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals
strong associations between gene co-expression and evolution," *BMC Genomics*, vol. 14,
no. 1, p. 450, 2013, doi: 10.1186/1471-2164-14-450.
[8]   M. Brilli *et al.*, "The diversity and evolution of cell cycle regulation in alpha-proteobacteria:
a comparative genomic analysis," *BMC Systems Biology*, vol. 4, no. 1, p. 52, Apr. 2010, doi:
10.1186/1752-0509-4-52.
[9]   Z. Zhang, A. Milias-Argeitis, and M. Heinemann, "Dynamic single-cell NAD(P)H
measurement reveals oscillatory metabolism throughout the E. coli cell division cycle," *Sci*
Rep, vol. 8, no. 1, p. 2162, Dec. 2018, doi: 10.1038/s41598-018-20550-7.
[10]   H. Yaginuma *et al.*, "Diversity in ATP concentrations in a single bacterial cell population
revealed by quantitative single-cell imaging," *Sci Rep*, vol. 4, no. 1, p. 6522, May 2015, doi:
10.1038/srep06522.
[11]   R. B. Weart, A. H. Lee, A.-C. Chien, D. P. Haeusser, N. S. Hill, and P. A. Levin, "A Metabolic
Sensor Governing Cell Size in Bacteria," *Cell*, vol. 130, no. 2, pp. 335–347, Jul. 2007, doi:
10.1016/j.cell.2007.05.043.
[12]   N. S. Hill, P. J. Buske, Y. Shi, and P. A. Levin, "A Moonlighting Enzyme Links Escherichia coli
Cell Size with Central Metabolism," *PLoS Genet*, vol. 9, no. 7, p. e1003663, Jul. 2013, doi:
10.1371/journal.pgen.1003663.
[13]   I. Santi, N. Dhar, D. Bousbaine, Y. Wakamoto, and J. D. McKinney, "Single-cell dynamics of
the chromosome replication and cell division cycles in mycobacteria," *Nat Commun*, vol. 4,
no. 1, p. 2470, Dec. 2013, doi: 10.1038/ncomms3470.
[14]   I. Santi and J. D. McKinney, "Chromosome Organization and Replisome Dynamics in
Mycobacterium smegmatis," *mBio*, vol. 6, no. 1, pp. e01999-14, Feb. 2015, doi:
10.1128/mBio.01999-14.
[15]   D. Trojanowski *et al.*, "Choreography of the Mycobacterium Replication Machinery during
the Cell Cycle," *mBio*, vol. 6, no. 1, pp. e02125-14, Feb. 2015, doi: 10.1128/mBio.02125-14.
[16]   M. M. Logsdon *et al.*, "A Parallel Adder Coordinates Mycobacterial Cell-Cycle Progression
and Cell-Size Homeostasis in the Context of Asymmetric Growth and Organization,"
Current Biology, vol. 27, no. 21, pp. 3367-3374.e7, Nov. 2017, doi:
10.1016/j.cub.2017.09.046.
[17]   D. Trojanowski, J. Hołówka, K. Ginda, D. Jakimowicz, and J. Zakrzewska-Czerwińska,
"Multifork chromosome replication in slow-growing bacteria," *Sci Rep*, vol. 7, no. 1, p.
43836, Apr. 2017, doi: 10.1038/srep43836.
[18]   M. M. Logsdon and B. B. Aldridge, "Stable Regulation of Cell Cycle Events in Mycobacteria:
Insights From Inherently Heterogeneous Bacterial Populations," *Front. Microbiol.*, p. 514,
2018, doi: 10.3389/fmicb.2018.00514.
[19]   N. Nair, R. Dziedzic, R. Greendyke, S. Muniruzzaman, M. Rajagopalan, and M. V. Madiraju,
"Synchronous replication initiation in novel Mycobacterium tuberculosis dnaA cold-

sensitive mutants," *Molecular Microbiology*, vol. 71, no. 2, pp. 291–304, 2009, doi:
10.1111/j.1365-2958.2008.06523.x.

[20] C. L. Gee *et al.*, "A Phosphorylated Pseudokinase Complex Controls Cell Wall Synthesis in
Mycobacteria," *Sci. Signal.*, vol. 5, no. 208, pp. ra7–ra7, Jan. 2012, doi:
10.1126/scisignal.2002525.

[21] R. Plocinska *et al.*, "Septal localization of the Mycobacterium tuberculosis MtrB sensor
kinase promotes MtrA regulon expression," *J. Biol. Chem.*, vol. 287, no. 28, pp. 23887–
23899, Jul. 2012, doi: 10.1074/jbc.M112.346544.

[22] S. Gola, T. Munder, S. Casonato, R. Manganelli, and M. Vicente, "The essential role of SepF
in mycobacterial division," *Molecular Microbiology*, vol. 97, no. 3, pp. 560–576, 2015, doi:
10.1111/mmi.13050.

[23] K. J. Wu *et al.*, "Characterization of Conserved and Novel Septal Factors in Mycobacterium
smegmatis," *Journal of Bacteriology*, vol. 200, no. 6, Mar. 2018, doi: 10.1128/JB.00649-17.

[24] S. Kalir *et al.*, "Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics
from Living Bacteria," *Science*, vol. 292, no. 5524, pp. 2080–2083, Jun. 2001, doi:
10.1126/science.1058758.

[25] A. Zaslaver *et al.*, "Just-in-time transcription program in metabolic pathways," *Nat Genet*,
vol. 36, no. 5, pp. 486–491, May 2004, doi: 10.1038/ng1348.

[26] K. J. Kieser and E. J. Rubin, "How sisters grow apart: mycobacterial growth and division,"
Nat Rev Microbiol, vol. 12, no. 8, pp. 550–562, Aug. 2014, doi: 10.1038/nrmicro3299.

[27] K. G. Lark, "Variation in bacterial acid-soluble deoxyribotides during discontinuous
deoxyribonucleic acid synthesis," *Biochimica et Biophysica Acta*, vol. 51, no. 1, pp. 107–
116, Jul. 1961, doi: 10.1016/0006-3002(61)91021-6.

[28] L. Huzyk and D. J. Clark, "Nucleoside Triphosphate Pools in Synchronous Cultures of
Escherichia coli," *Journal of Bacteriology*, vol. 108, no. 1, pp. 74–81, Oct. 1971.

[29] M. A. DeJesus *et al.*, "Comprehensive Essentiality Analysis of the Mycobacterium
tuberculosis Genome via Saturating Transposon Mutagenesis," *mBio*, vol. 8, no. 1, Mar.
2017, doi: 10.1128/mBio.02133-16.

[30] V. Singh *et al.*, "The Inosine Monophosphate Dehydrogenase, GuaB2, Is a Vulnerable New
Bactericidal Drug Target for Tuberculosis," *ACS Infect. Dis.*, vol. 3, no. 1, pp. 5–17, Jan.
2017, doi: 10.1021/acsinfecdis.6b00102.

[31] H. A. Arjes, A. Kriel, N. A. Sorto, J. T. Shaw, J. D. Wang, and P. A. Levin, "Failsafe
mechanisms couple division and DNA replication in bacteria," *Curr. Biol.*, vol. 24, no. 18,
pp. 2149–2155, Sep. 2014, doi: 10.1016/j.cub.2014.07.055.

[32] K. M. Mann *et al.*, "Rv0004 is a new essential member of the mycobacterial DNA
replication machinery," *PLOS Genetics*, vol. 13, no. 11, p. e1007115, Nov. 2017, doi:
10.1371/journal.pgen.1007115.

[33] J. M. Diver and R. Wise, "Morphological and biochemical changes in Escherichia coli after
exposure to ciprofloxacin," *J Antimicrob Chemother*, vol. 18, no. Supplement_D, pp. 31–41,
Nov. 1986, doi: 10.1093/jac/18.Supplement_D.31.

[34] P. de Boer, R. Crossley, and L. Rothfield, "The essential bacterial cell-division protein FtsZ is
a GTPase," *Nature*, vol. 359, no. 6392, pp. 254–256, Sep. 1992, doi: 10.1038/359254a0.

[35] A. W. Bisson-Filho *et al.*, "Treadmilling by FtsZ filaments drives peptidoglycan synthesis
and bacterial cell division," *Science*, vol. 355, no. 6326, pp. 739–743, Feb. 2017, doi:
10.1126/science.aak9973.
[36] A. M. Hogan *et al.*, "Competitive Fitness of Essential Gene Knockdowns Reveals a Broad-
Spectrum Antibacterial Inhibitor of the Cell Division Protein FtsZ," *Antimicrob Agents*
Chemother, vol. 62, no. 12, pp. e01231-18, /aac/62/12/e01231-18.atom, Oct. 2018, doi:
10.1128/AAC.01231-18.
[37] A. Mukherjee and J. Lutkenhaus, "Dynamic assembly of FtsZ regulated by GTP hydrolysis,"
The EMBO Journal, vol. 17, no. 2, pp. 462–469, Jan. 1998, doi: 10.1093/emboj/17.2.462.
[38] E. Bi and J. Lutkenhaus, "FtsZ ring structure associated with division in Escherichia coli,"
Nature, vol. 354, no. 6349, pp. 161–164, Nov. 1991, doi: 10.1038/354161a0.
[39] L. Wang, M. K. Khattar, W. D. Donachie, and J. Lutkenhaus, "FtsI and FtsW Are Localized to
the Septum inEscherichia coli," *Journal of Bacteriology*, vol. 180, no. 11, pp. 2810–2816,
Jun. 1998.
[40] E. H. Rego, R. E. Audette, and E. J. Rubin, "Deletion of a mycobacterial divisome factor
collapses single-cell phenotypic heterogeneity," *Nature*, vol. 546, no. 7656, pp. 153–157,
Jun. 2017, doi: 10.1038/nature22361.
[41] R. Lenarcic *et al.*, "Localisation of DivIVA by targeting to negatively curved membranes,"
EMBO J, vol. 28, no. 15, pp. 2272–2282, Aug. 2009, doi: 10.1038/emboj.2009.129.
[42] K. S. Ramamurthi and R. Losick, "Negative membrane curvature as a cue for subcellular
localization of a bacterial protein," *PNAS*, vol. 106, no. 32, pp. 13541–13545, Aug. 2009,
doi: 10.1073/pnas.0906851106.
[43] X. Meniche *et al.*, "Subpolar addition of new cell wall is directed by DivIVA in
mycobacteria," *PNAS*, vol. 111, no. 31, pp. E3243–E3251, Aug. 2014, doi:
10.1073/pnas.1402158111.
[44] A. K. Sharma *et al.*, "Serine/Threonine Protein Phosphatase PstP of *Mycobacterium*
tuberculosis Is Necessary for Accurate Cell Division and Survival of Pathogen," *J. Biol.*
Chem., vol. 291, no. 46, pp. 24215–24230, Nov. 2016, doi: 10.1074/jbc.M116.754531.
[45] Iswahyudi *et al.*, "Mycobacterial phosphatase PstP regulates global serine threonine
phosphorylation and cell division," *Sci Rep*, vol. 9, no. 1, p. 8337, Dec. 2019, doi:
10.1038/s41598-019-44841-9.
[46] D. F. Warner, T. Tønjum, and V. Mizrahi, "DNA Metabolism in Mycobacterial
Pathogenesis," *Curr. Top. Microbiol. Immunol.*, May 2013, doi: 10.1007/82_2013_328.
[47] H. A. Arjes, B. Lai, E. Emelue, A. Steinbach, and P. A. Levin, "Mutations in the bacterial cell
division protein FtsZ highlight the role of GTP binding and longitudinal subunit interactions
in assembly and function," *BMC Microbiol*, vol. 15, no. 1, p. 209, Dec. 2015, doi:
10.1186/s12866-015-0544-z.
[48] J. M. Rock *et al.*, "DNA replication fidelity in Mycobacterium tuberculosis is mediated by an
ancestral prokaryotic proofreader," *Nat Genet*, vol. 47, no. 6, pp. 677–681, Jun. 2015, doi:
10.1038/ng.3269.
[49] M. H. Buckstein, J. He, and H. Rubin, "Characterization of Nucleotide Pools as a Function of
Physiological State in Escherichia coli," *Journal of Bacteriology*, vol. 190, no. 2, pp. 718–
726, Jan. 2008, doi: 10.1128/JB.01020-07.

[50] S. Ehrt *et al.*, "Controlling gene expression in mycobacteria with anhydrotetracycline and
Tet repressor," *Nucleic Acids Res*, vol. 33, no. 2, p. e21, 2005, doi: 10.1093/nar/gni013.
[51] P. Sass and H. Brötz-Oesterhelt, "Bacterial cell division as a target for new antibiotics,"
Current Opinion in Microbiology, vol. 16, no. 5, pp. 522–530, Oct. 2013, doi:
10.1016/j.mib.2013.07.006.
[52] C. E. Baer, A. T. Iavarone, T. Alber, and C. M. Sassetti, "Biochemical and Spatial Coincidence
in the Provisional Ser/Thr Protein Kinase Interaction Network of Mycobacterium
tuberculosis," *J. Biol. Chem.*, p. jbc.M114.559054, Jun. 2014, doi:
10.1074/jbc.M114.559054.
[53] E. O. Johnson *et al.*, "Large-scale chemical–genetics yields new M. tuberculosis inhibitor
classes," *Nature*, vol. 571, no. 7763, pp. 72–78, Jul. 2019, doi: 10.1038/s41586-019-1315-z.
[54] J.-H. Kim, J.-R. Wei, J. B. Wallach, R. S. Robbins, E. J. Rubin, and D. Schnappinger, "Protein
inactivation in mycobacteria by controlled proteolysis and its application to deplete the
beta subunit of RNA polymerase," *Nucleic Acids Research*, vol. 39, no. 6, pp. 2210–2220,
Mar. 2011, doi: 10.1093/nar/gkq1149.
[55] A. E. Carpenter *et al.*, "CellProfiler: image analysis software for identifying and quantifying
cell phenotypes," *Genome Biology*, vol. 7, no. 10, p. R100, Oct. 2006, doi: 10.1186/gb-
2006-7-10-r100.
[56] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows–Wheeler
transform," *Bioinformatics*, vol. 25, no. 14, pp. 1754–1760, Jul. 2009, doi:
10.1093/bioinformatics/btp324.
[57] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion
for RNA-seq data with DESeq2," *Genome Biol*, vol. 15, no. 12, 2014, doi: 10.1186/s13059-
014-0550-8.
[58] S. Wichert, K. Fokianos, and K. Strimmer, "Identifying periodically expressed transcripts in
microarray time series data," *Bioinformatics*, vol. 20, no. 1, pp. 5–20, Jan. 2004, doi:
10.1093/bioinformatics/btg364.
[59] M. Straume, "DNA microarray time series analysis: automated statistical assessment of
circadian rhythms in gene expression patterning," *Meth. Enzymol.*, vol. 383, pp. 149–166,
2004, doi: 10.1016/S0076-6879(04)83007-6.
[60] J. H. W. Jr, "Hierarchical Grouping to Optimize an Objective Function," *Journal of the
American Statistical Association*, vol. 58, no. 301, pp. 236–244, Mar. 1963, doi:
10.1080/01621459.1963.10500845.
[61] L. Scrucca, M. Fop, T. B. Murphy, and A. E. Raftery, "mclust 5: Clustering, Classification and
Density Estimation Using Gaussian Finite Mixture Models," *R J*, vol. 8, no. 1, pp. 289–317,
Aug. 2016.
954

S1

A

Genes >>

Time >>

B