Title

Global gene expression analysis of the Myxococcus xanthus developmental time course.

Journal

Genomics, 113(1 Pt 1)

ISSN

0888-7543

Authors

Sharma, Gaurav
Yao, Andrew I
Smaldone, Gregory T
et al.

Publication Date

2021

DOI

10.1016/j.ygeno.2020.11.030

Peer reviewed

Genomics 113 (2021) 120–134

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

Global gene expression analysis of the *Myxococcus xanthus* developmental time course

Gaurav Sharma [a,b,1], Andrew I. Yao [c,d,1], Gregory T. Smaldone [a,1], Jennifer Liang [a], Matt Long [a], Marc T. Facciotti [c,d], Mitchell Singer [a,*]

[a] Department of Microbiology and Molecular Genetics, University of California – Davis, One Shields Avenue, Davis, CA 95616, United States of America
[b] Institute of Bioinformatics and Applied Biotechnology, Electronic City, Bengaluru, Karnataka, India
[c] Department of Biomedical Engineering, University of California-Davis, One Shields Avenue, Davis, CA 95616, United States of America
[d] Genome Center, University of California-Davis, One Shields Avenue, Davis CA 95616 Zymergen, Inc., Emeryville, CA, United States of America

ARTICLE INFO

Keywords:
Myxobacteria
Sporulation
Development
Transcriptomics
RNA sequencing

ABSTRACT

To accurately identify the genes and pathways involved in the initiation of the *Myxococcus xanthus* multicellular developmental program, we have previously reported a method of growing vegetative populations as biofilms within a controllable environment. Using a modified approach to remove up to ~90% rRNAs, we report a comprehensive transcriptional analysis of the *M. xanthus* developmental cycle while comparing it with the vegetative biofilms grown in rich and poor nutrients. This study identified 1522 differentially regulated genes distributed within eight clusters during development. It also provided a comprehensive overview of genes expressed during a nutrient-stress response, specific development time points, and during development initiation and regulation. We identified several differentially expressed genes involved in key central metabolic pathways suggesting their role in regulating myxobacterial development. Overall, this study will prove an important resource for myxobacterial researchers to delineate the regulatory and functional pathways responsible for development from those of the general nutrient stress response.

1. Introduction

The Gram-negative bacterium *Myxococcus xanthus* is a ubiquitous soil microorganism, most notably studied for its social and highly organized developmental program [1–3]. *M. xanthus* undergoes a community-wide reprogramming, adapting a vegetative growing biofilm into localized macro-cellular structures called fruiting bodies. During development, cells undergo three potential programmed cell fates: 1) differentiation into stress-resistant metabolically inactive myxospores, 2) environmentally sensing peripheral rod cells that remain outside of the fruiting body, and 3) cell lysis, which is hypothesized to provide energy for the entire morphogenic process [4,5].

Since, before the first Tn*5*transposon screen in *M. xanthus*, researchers have aimed to identify all genes involved in its multicellular developmental program using a variety of genetic screens. [6]. Much of these genetic screens laid the framework for our initial understanding of the major signals [2], transcriptional regulators [7], and temporal progression controlling development [8]. In the years since much has been done to directly identify the genes associated with many of those transposon insertions, and from this, researchers have drawn detailed genetic regulatory maps outlining the transition from nutrient limitation through fruiting body formation and sporulation [9–11]. Yet much remains unknown about this developmental program specifically the understanding of the regulatory networks that control gene expression during growth under high and low nutrient conditions versus those that are necessary for the switch between growth and development.

One approach to deciphering the structures of these regulatory networks is to identify genes and gene clusters whose patterns of transcript abundance suggest they are likely targets of regulation in development and its associated starvation response. Genome-wide transcriptome analysis by RNAseq may help identify such genes and their coordinated behaviors. While the technology has been available for several years, due to various technical and cost considerations, it has been used inadequately in the study of *M. xanthus* development [12,13]. A recent study [13] used RNAseq to study the transcriptome of the 96 h developmental program in *M. xanthus*. They identified 1,415 differentially

* Corresponding author.
E-mail address: mhsinger@ucdavis.edu (M. Singer).
[1] Authors contributed equally to this work.

https://doi.org/10.1016/j.ygeno.2020.11.030
Received 31 August 2020; Received in revised form 6 November 2020; Accepted 29 November 2020
Available online 1 December 2020

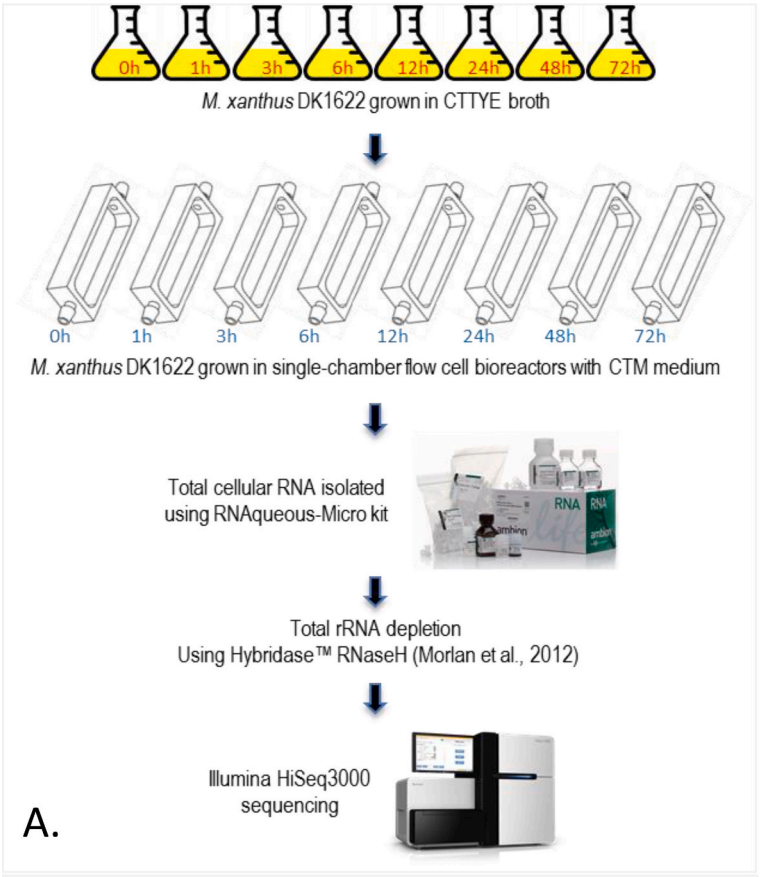
expressed genes whose changing patterns of RNA abundance were well described by 10 discrete clusters of expression profiles.

In this study, we analyze the transcriptomic data of *M. xanthus* DK1622 cultures grown in steady-state nutrient-rich (1% casitone), nutrient-poor (0.008% casitone) vegetative phases, and a dynamic 72-h developmental cycle. We use a previously described bioreactor growth chamber [14] to carefully control growth and development. We also report an adapted method for the depletion of ribosomal RNA that enables broader coverage and greater read depth of the *M. xanthus* transcriptome per sequencing lane than previously reported. This increased sensitivity has allowed us to identify 1522 genes associated with development out of which only 512 common genes have been reported within the previous analysis [13], therefore, this study uniquely describes the importance of 1010 genes in myxobacterial development. Comparing the RNA abundance profiles of low nutrient vegetative growth and developmental conditions has also allowed us to distinguish between genes primarily involved in development versus those that are responsive to nutrient limitation. Finally, this work has identified several genes whose products may be involved in the regulation of gene expression in development and starvation; these functional hypotheses will be tested in future work.

2. Material and methods

2.1. Bacterial strains, plasmids, growth, and sampling conditions

Yellow isolates of *M. xanthus* strain DK1622 were used as the wild-type strain for all sample collection and experimentation [15]. Wild-



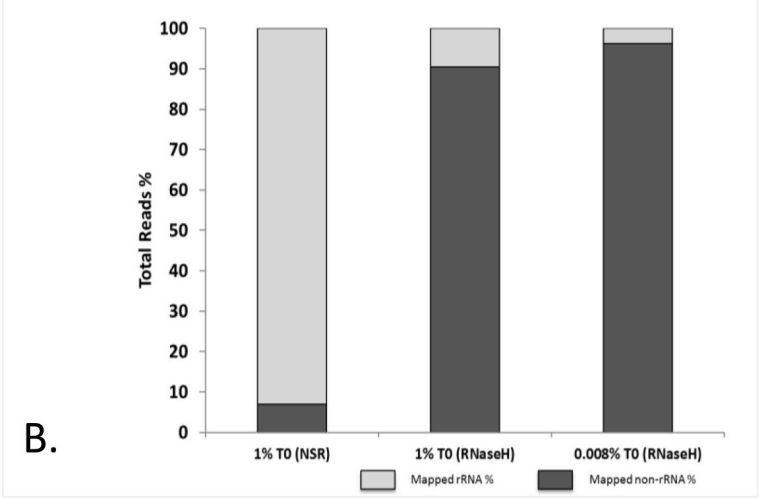


Fig. 1A. Flow chart for the adopted methodology (RNaseH depletion-based Transcriptome analysis). As described in the Methods section, aliquots were removed from a single growth culture and placed into the appropriate growth chambers. Growth chambers were then assayed for bulk RNA at the appropriate time points. Protocol for using the bioreactors can is found in Smaldone et al. [14] Fig. 1B: Comparative read mapping amongst Not-So-Random (NSR) priming and RNaseH depletion method revealing an overwhelming change in functional transcript sequencing

type was grown in CTTYE (1% Casitone [Difco, Franklin Lakes, NJ], 0.2% yeast extract [BD, Sparks, MD], 10 mM Tris-HCl [pH 7.6], 1 mM KH_2PO_4, 8 mM $MgSO_4$) broth or on CTTYE plates containing 1% agar for recover from storage and for general cultivation. Single-chamber flow cell bioreactors supplied with CTM medium (1% or 0.008% casitone (as indicated), 10 mM morpholinepropanesulfonic acid (MOPS) [pH 7.0], 1 mM KH_2PO_4, 8 mM $MgSO_4$, 1 mM $CaCl_2$) were used for the cultivation of all samples in this study, as described earlier [14]. Developmental time courses were initiated by rapid exchange of CTM for MC7 buffer (10 mM MOPS [pH 7.0], 1 mM KH_2PO_4, 8 mM $MgSO_4$, and 1 mM $CaCl_2$) in the bioreactor. Samples were recovered in biological triplicate for all indicated time points

2.2. Growth and development using flow-cell bioreactors

Starter cultures of *M. xanthus* DK1622 wild type were taken from a fresh CTTYE plate and inoculated into 25 mL 1% CTTYE in a 250 mL flask. The culture flask was shaken overnight at 32 °C and grown to a Klett measurement of approximately 80–100, representing mid-exponential growth and approximately 2×10^8 cells/ml.). The culture was pelleted and washed with MC7 buffer (10 mM MOPS (pH 7.0), 1 mM KPO_4, 8 mM $MgSO_4$, and 1 mM $CaCl_2$) and approximately 10^7 cells were inoculated into the flow cell chambers in a volume of 0.25 mL using a 1 mL syringe with a 25 × 5/8″ hypodermic needle. Continuous 1% CTTYE medium flow was turned on at a rate of 2 mL/h after inoculated cells could statically adhere to the coverslips for 2 h. Dense microbial mats were grown for 4 to 5 days as described in Smaldone et al. [14]. The *M. xanthus* developmental program was then initiated by a rapid exchange to the nutrient-deficient buffer MC7 and cells were then allowed to undergo the developmental program as described in the text (Fig. 1A). All RNAseq experiments were done with biological triplicates and in some cases quadruplicate sampling

2.3. Total RNA isolation

Total cellular RNA was isolated from growth chambers by removing the bottom slide containing the attached biofilm and using the RNAqueous-Micro kit (Thermo Fisher Scientific, Waltham, MA) per the manufacturer's protocol. Total RNA was eluted off the column with 2 × 10 μL RNase-free H_2O at 70 °C. Total RNA samples were treated with RNase-free DNaseI (Thermo Fisher Scientific, Waltham, MA) to eliminate genomic DNA contamination by adding 2 μL 10× DNaseI buffer and 1 μL Invitrogen DNaseI (2 U/μL). The reaction was incubated at 37 °C for 30 min at which time 2 μL Invitrogen DNaseI inactivation reagent was added and incubated at 25 °C for 2 min. The reaction was centrifuged in a microcentrifuge for 1 min and the supernatant was recovered and subjected to ethanol precipitation

2.4. Ribosomal RNA depletion

Ribosomal RNA (rRNA) was digested from total RNA samples via a protocol adapted from a previous report [16]. Briefly, approximately 2 μg total RNA was added to 0.6 μL 10× hybridization buffer (1 M Tris-HCl (pH 7.0), 2 M NaCl, RNase-free H_2O), 2 μL 16S and 23S synthetic DNA oligonucleotide probes (76 total oligos at 0.5 μM/oligo; Table S1), and RNase-free H_2O to a total volume of 6 μL. Hybridization was carried out in a thermocycler (Bio-Rad, Hercules, CA) with the following parameters: 95 °C for 2 min, ramp down to 45 °C at −0.1 °C/s, 45 °C for 5 min, hold at 45 °C. To the hybridization reaction 1 μL pre-heated 10× Hybridase™ buffer (500 mM Tris-HCl (pH 7.4), 1 M NaCl, 200 mM $MgCl_2$), 1 μL pre-heated RNase-free H_2O, and 2 μL pre-heated thermostable Hybridase™ RNaseH (5 U/μL; Epicentre, Madison, WI) was added and incubated at 45 °C for 40 min. The reaction was moved to ice and then treated with RNase-free DNaseI by adding 10 μL 10× DNaseI buffer, 2 μL DNaseI, and bringing the total volume up to 100 μL with RNase-free H_2O. This was incubated at 37 °C for 40 min. The DNaseI treated samples were shifted to ice then followed with an Agencourt RNAClean™ XP bead cleanup (Beckman Coulter, Brea, CA) with a 1.8× volume of beads to sample as per manufacturers protocol. RNA samples depleted of rRNAs were eluted off the beads in 13 μL RNase-free H_2O

2.5. Illumina sequencing library preparation

Library preparation for Illumina HiSeq3000 (Illumina, San Diego, CA) sequencing was performed using the KAPA Stranded RNA-Seq Library Preparation kit (KAPA Biosystems, Wilmington, MA). After rRNA depletion, RNA concentrations were too low to detect by any current methods, so the maximum volume (10 μL) of the depleted RNA sample was used to generate libraries. For multiplex adapter ligation, Illumina TruSeq™ adapters were ordered from IDT (Coralville, IA) and stock concentrations of 350 nM were used for each library prepared. PCR amplification (15 s at 98 °C to denaturation; 30 s at 60 °C for annealing; 30 s at 72 °C for extension (13 rounds), and then final 5 s at 72 °C) was used for library amplification. For the comparison of RNaseH depletion to another commercially available methods of rRNA depletion, libraries were prepared from vegetative zero time point and low nutrition (0.008 CTM) *M. xanthus* total RNA samples using the NuGEN Ovation Complete Prokaryotic RNA-Seq Library System (NuGEN Technologies, Inc., San Carlos, CA)

2.6. Illumina HiSeq3000 sequencing

Libraries were submitted to the UC Davis DNA Technologies Core (Davis, CA) for sequencing. Samples were multiplexed into a single sample in equal molar ratios. Since all libraries had similar fragment distribution at approximately 300 bp (as determined by Bioanalyzer (Agilent Technologies, Santa Clara, CA)), the libraries were quantified with a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) and equal nanograms of each sample were pooled. Quantitative PCR was performed on the pool to calculate the molarity. 216 nM of diluted pooled libraries were loaded onto the cBOT for clustering onto the flow cell as per Illumina's protocol. The pool was sequenced on the HiSeq3000 (HiSeq Control Software v3.3.20) with a 51 bp read/8 bp index run. The raw data was demultiplexed with bcl2fastq (v2.16.0.10) to produce separate 50 bp read files for each library in the pool. Raw reads for all RNAseq experiments can be accessed at the NIH Sequence Read Archive database (http://www.ncbi.nlm.nih.gov/sra) under the number SUB5036818

2.7. RNAseq, transcriptome, differential gene expression (DGE) analysis, and gene filtering

FastQC [17] was used to quality check raw parsed fastq files before further analysis. Reads were cleaned of barcode and adapter sequences and quality trimmed using Trimmomatic [18]. Trimmed reads were then aligned to the *M. xanthus* DK1622 genome using Bowtie [19] and mapped to predicted transcripts using HTSeqcount [20]. All preceding software was run with default settings. DGE analysis was conducted using DESeq2 in R [21]. Cluster analysis was performed on normalized data (using the rlog command built into the DESeq2 package) using the Pearson correlation (using the pheatmap [22] function in RStudio) to calculate clustering distance. Discrete clusters were selected using the "complete" agglomeration method within the 'hclust' package

Given the depth of coverage in the sequencing data (Table S2), more than 95% of all potential structurally annotated genes had at least one read mapped to them. For genes that did not show differential expression, it was necessary to determine a method to evaluate and indicate, for each condition, whether a gene was likely to be functionally constitutively expressed or not. We evaluated several previously published schemes for solving this problem but ultimately determined that the previously described methods were not well suited to the structure of our dataset [23–25]. We, therefore, created a simple heuristic based on

previously reported transcripts known to be present at low abundance when functionally expressed. These genes include many of the known development-related transcription factors and signaling genes including *sdeK, sigC, mrpC,* etc. Examining the numbers of mapped reads for this subset of transcripts suggested that using a threshold of 1000 mapped reads per gene would provide a suitable threshold for filtering and flagging gene lists. We applied this filter post DGE analysis to flag genes that had at least one time point in the developmental time course above the filtering threshold. These data and flags are available in Table S4. We also determined constitutively expressed genes applying the heuristic filter at all developmental and vegetative growth time points (Table S7)

2.8. Homology identification

To identify the closest evolutionary hits per sequence, all *M. xanthus* DK1622 proteins were subjected to BLASTp [26] against the NCBI non-redundant (NR) database with an *E*-value cutoff of $1E^{-5}$. From BLAST output, the taxonomy of each homolog was identified. To know the closest relatives outside of myxobacteria, the homologs belonging to order Myxococcales were removed and the relative distribution of the taxonomy of the top 100 homologs per protein was counted and reported here. We also performed the BLASTp of all *M. xanthus* DK1622 proteins against the COG database to identify their putative functional categories using an *E*-value cutoff of $1E^{-5}$

3. Results and discussion

3.1. An efficient ribosomal RNA depletion method optimized for M. xanthus

One of the key factors to consider when conducting transcriptomics using RNAseq is the depth of sequencing required to obtain reliable and reproducible measurements of low abundance transcripts. This is of particular concern when attempting to monitor changes in the abundance of notoriously low expression genes like those encoding regulatory proteins. While the challenge is particularly acute in large eukaryotic genomes, the issue remains problematic for prokaryotic systems with much smaller genomes. In bacteria, ribosomal RNAs (rRNAs) comprise roughly 95–99% of the total isolated RNA [27], making expression analysis via direct RNA sequencing inefficient and costly, if the rRNA is not efficiently and selectively removed. Failure to remove the rRNA can dramatically lower the relative number of reads that map to mRNA transcripts in each experiment and thus limiting the technology's broad adoption for multi-parameter and time-course experiments

While several commercially available kits promise to deplete the rRNA to undetectable levels [28] in prokaryotic samples, most have been optimized and benchmarked on *E. coli* and in our hands have not been particularly successful at depleting rRNA in *M. xanthus*. To address this problem, we adapted a protocol first described for use in the isolation of rRNA depleted total RNA isolated from Human formalin-fixed, paraffin-embedded tissue [16]. Briefly, DNA oligonucleotides were synthesized to cover the length of the *M. xanthus*16s and 23S rRNA genes (Table S1). These oligonucleotides were then hybridized to the total RNA pool. The pool was treated with thermostable RNaseH to remove the rRNA: DNA hybrids. This protocol (Fig. 1A) leaves a pool of RNA up to >90% free of 23S and 16S rRNA (Fig. 1B and Table S2). We argue that using this protocol will allow more multiplexing for sequence analyses leading to higher sequence coverage for less cost as compared to the other methods

3.2. Validation of differentially expressed genes

When introducing any new technique or approach it is essential to know if these new technologies are as efficient as older technologies in identifying components involved in the processes. Therefore, we first sought to evaluate whether the RNAseq data collected from a time course induced in custom flow cell bioreactors [14] could recapitulate the general behavior of 24 "landmark" developmental genes that have shown consistent behavior in various standard developmental time course assays evaluated using DNA microarrays [29–32], qPCR studies [29,30,33], Tn*5lac* fusions studies [6] and protein levels using immunoprecipitation assays [34–37]. Most of these studies have not provided the exact numeric data for all developmental time points, therefore preventing us to do cross-comparison between our results and previous experimental results. The patterns of transcript abundance using our RNAseq protocols are shown in Fig. 2. Overall, expression values identified in our study are similar according to the timing and expression level in previous studies, therefore, functionally validating the RNAseq data analysis in this study. However, it should be noted that depending on the specific method of induction of development, the specific time at which morphogenic and transcriptional markers appear relative to $t = 0$ can vary between experiments, sometimes by up to 6–10 h. This variation is especially prevalent for the later times in the process of development. Therefore, to validate our RNAseq data against previously published work, we looked for similar expression trends and not exact hour-to-hour matches in the appearance of key markers (Fig. 2)

We were predominantly interested in comparing changes in the RNA abundance of four key developmental genes [FruA (MXAN_3117), MrpC (MXAN_5125), SdeK (MXAN_1014), and MbhA (MXAN_7061)] to their previously reported changes in protein abundance for these gene products as measured using β-galactosidase assays and Western blots [34–37]. It should be noted that transcript abundance reflects the levels of RNA and does not take into account any post-transcriptional regulation that may affect translation and therefore protein expression level. FruA is an early developmental transcriptional regulator induced by A-signal, MrpC, and *E*-signal [2,38]. Previous reports, using CF plates to induce development, have shown that FruA protein reaches its highest level of expression between 12 and 20 h into development [34]. Using our RNAseq protocol, we observe that the peak transcript abundance of *fruA* occurs during the first 24 h of development. MrpC and truncated MrpC2 are transcription factors that promote C-signal production [38], however, the latter has recently been shown to be an artifact of proteolytic degradation during sample preparation [39]. β-galactosidase transcriptional fusion assays demonstrated a steady increase in MrpC and MrpC2, up to 8-fold, expression over the first 24 h of development [40]. We identify the maximum transcript abundance of this gene at 48 h and a 4-fold increase in expression. Similarly, for SdeK, an essential early developmental protein, the expression for both our results and those of previous works suggest that maximum expression occurs ~24 h post-starvation with an increase on the order of 5- to 6-fold [36]. MbhA is a myxobacterial hemagglutinin and is most abundantly found during aggregation. Previous functional studies found its specific activity to be the highest between 12 and 36 h post nutrient depletion (depending on the strain used and method of developmental initiation) and its expression to peak (~150-fold increase) at 48 h after initiation of development [37,41]. Our data shows a maximum induction of transcript abundance during 24–48 h with a 12-fold increase in expression. Twenty more targets were analyzed similarly and revealed no major discrepancies between the patterns of transcript abundance as measured by RNAseq and former gene expression analysis (Fig. 2). Once again it should be noted that we are comparing RNA expression levels to either protein levels or β-galactosidase activity from a Tn5lacZ fusion. The point of these experiments was to demonstrate that using our RNAseq protocol we can identify known genes that were previously identified. We conclude that the RNAseq data accurately captures the canonical elements of *M. xanthus* development and extends the functional relevance of the dataset beyond the validated core genes

3.3. Identification of nutrient stress transcriptional response

To identify genes involved with the maintenance of growth at low

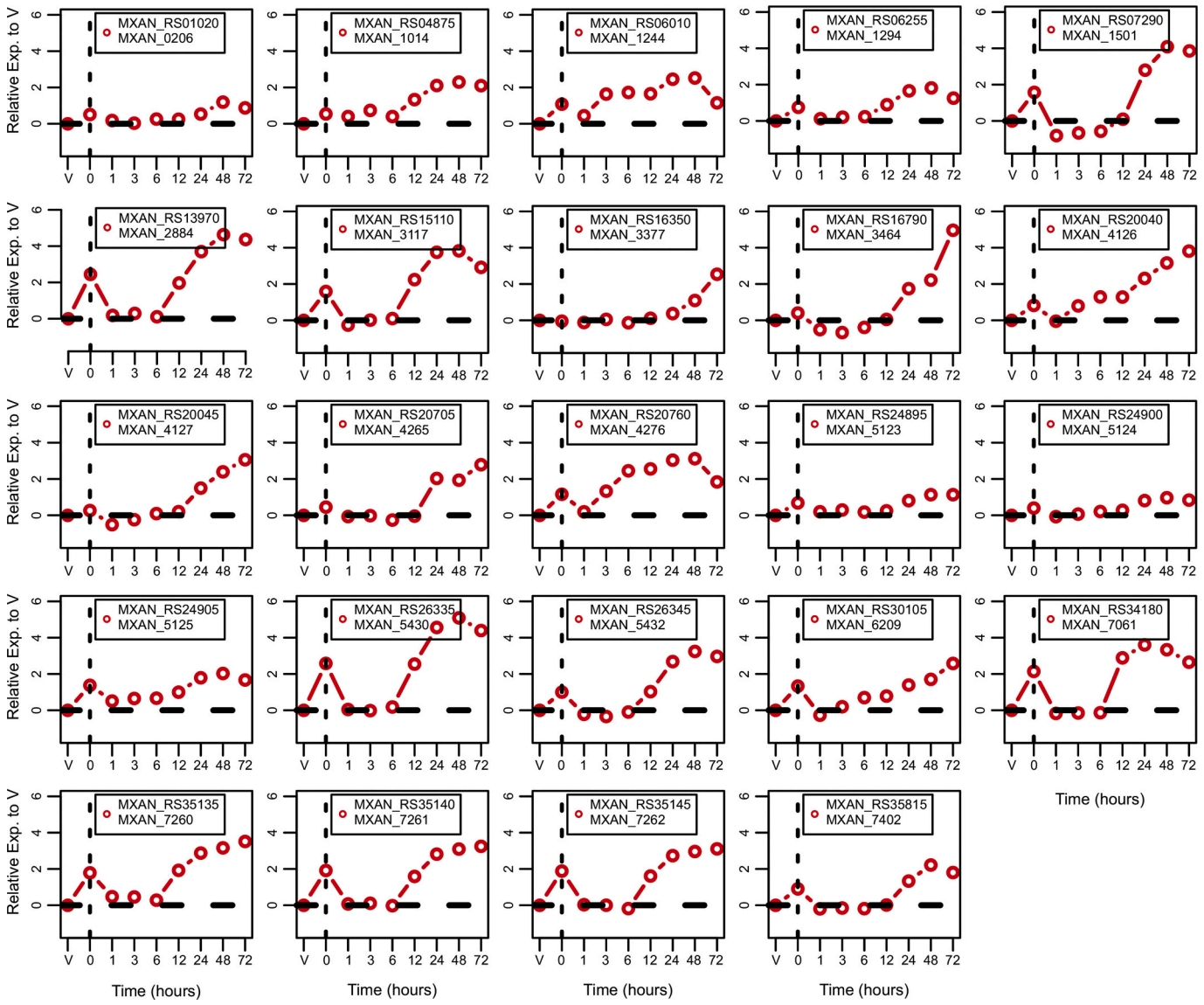


Fig. 2. Time course expression of 24 well-known developmentally induced genes. Rlog normalized (rlog() in DESeq2) expression relative to the vegetative state 'V' is plotted against time in hours. Relative expression is calculated by subtracting the average rlog normalized value of each time point by the average rlog normalized value for the vegetative state

nutrient conditions and growth rates, we performed RNAseq analysis on vegetative biofilms grown and maintained for three days in bioreactors at 1% and 0.008% CTM media as described in materials and methods. Previous work showed that 0.008% of CTM was adequate to grow confluent biofilms without inducing development [14]. The heat map (Fig. 3) depicts the differentially expressed genes (782) between these two vegetative conditions constituting the nutrient stress response of *M. xanthus* DK1622 (Table S3). Considering a *p*-value cutoff of 0.05 with a minimum $\log_2$-fold expression of one, we found 297 genes whose transcript abundance (up to 150-fold change) during low nutrient conditions (category 1) is lower as compared to rich nutrient conditions and 485 genes whose transcript abundance is higher during low nutrient conditions (category 2) as compared to the rich nutrient conditions

As expected, most of the category 1 protein-encoding genes are involved in translation, ribosomal structure, and biogenesis (29 genes), signal transduction mechanisms (23 genes), and transcription (17 genes); all systems highly associated with actively growing cells. Our analysis reported multiple hits with several-fold decreased transcript abundance of which little could be found in previous literature. MXAN_6998 is an XRE family transcriptional regulator that showed a ~ 150-fold decrease in transcript abundance. Similarly, we found a subset of 60 genes encoding proteins with more than a 10-fold decrease in transcript abundance in low nutrient conditions relative to high nutrient conditions. Many of these genes are present together in a cluster in the genome such as MXAN_7370–71 (>70-fold; encoding putative serine/threonine protein kinases), MXAN_5587–88 (>30-fold; genes of unknown function), MXAN_0162–70 (~10-fold for most of the genes; encoding a putative Kdp-ATPase system) which suggest that these genes play a pivotal role in nutritional stress. It has been shown earlier that mutations in MXAN_7370 (*pktF4*) are deficient in fruiting body formation, sporulation, and S-motility [42] although no such data is available for its close neighbor MXAN_7371 (*psd2*). Similarly, the Kdp-ATPase genes have been known to be some of the most highly induced genes due to the presence of prey, such as *E. coli,* suggesting the involvement of these gene products in indirect prey sensing via an osmotic stress response [43]. It has been proposed that this kind of sensing might trigger multicellular fruiting body formation in the myxobacteria [44]

Genes belonging to Category 2 correspond to stress response genes whose products might help maintain vegetative biofilms during low nutrient environments (Fig. 3). Interestingly, 27% of genes in this

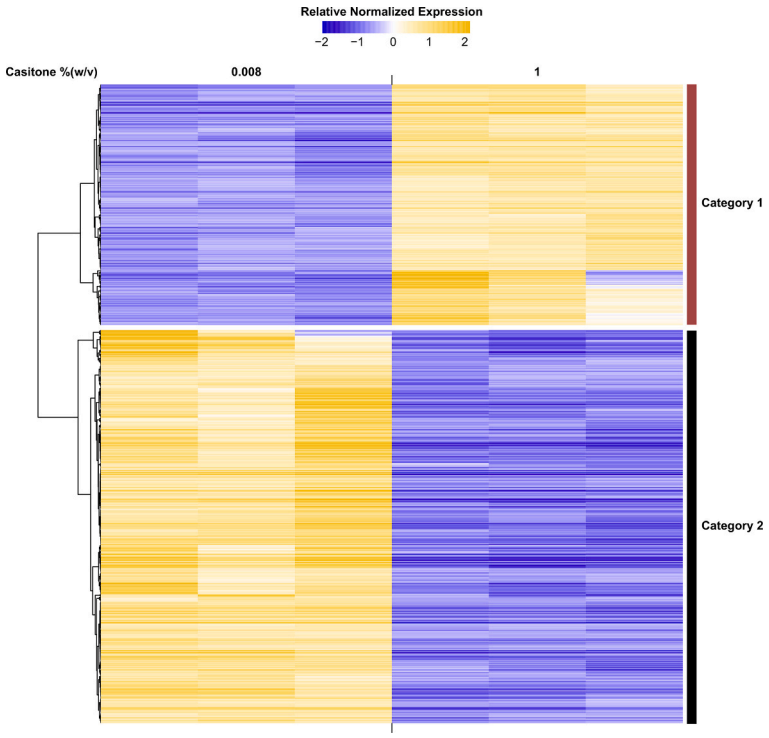


Fig. 3. Heat map of genes constituting nutrient stress response (overexpressed and underexpressed) in *M. xanthus* DK1622: Here the topmost differentially expressed genes ($log_2FC > 1$) between low and normal nutrient conditions with adjusted *p*-values are sorted according to hierarchical clustering based on rlog-transformed reads

category are involved in cellular processes and signaling, 24% in metabolism, and 9% of genes encode proteins responsible for information storage and processing. We found that most of the upregulated genes were involved in signal transduction (33 genes), secondary metabolites biosynthesis (29 genes), lipid transport-metabolism (25 genes), posttranslational modification (24 genes), transcription (24 genes), and cell wall biogenesis (22 genes). MXAN_3941–42 (encoding a putative polyketide synthase and an acyl-carrier-protein) and MXAN_3943 (encoding cytochrome P450) are involved in secondary metabolites synthesis. MXAN_4937 (encoding an EGF domain-containing protein) is most likely secreted and acting as secondary metabolites in some survival-promoting capacity

Forty percent of the identified up/down-regulated genes are annotated as hypothetical proteins, however, only in a few cases, we were able to trace their likely function. A group of proteins ranging from MXAN_4488 to MXAN_4506 has putative phage-associated domains, suggesting their role in the activation of a lysogenic virus-associated gene cluster. Besides that, many hypothetical genes are present in many of the operons identified, such as MXAN_1017–1021 (COG analysis suggests their role in cell motility). Our study revealed the maximum expression change for MXAN_0504 (~490-fold), which encodes a hypothetical protein with no COG or PFAM domain mapping

3.4. Transcriptomic analysis of the myxobacterial developmental time course

Next, we defined the "development regulon" as the collection of all genes whose expression is altered during development. We calculated the union of all genes found to be differentially regulated (at both low and high nutrient conditions) that pass the filter of *p*-value <0.05 and minimum log_2 fold change = 1.0 at any time point t_x (where x = 1, 3, 6, 12, 24, 48, or 72 h). Based on this methodology, we identified that the expression patterns of 1522 genes showed significant changes at some point during the time course

3.5. Analysis of the development regulon

Genes in the development regulon were hierarchically clustered based on the similarity of transcript abundance profiles (Fig. 4A). Broadly, expression patterns (Table S4A) could be grouped into eight major clusters (Fig. 4B). Functional assignment by COG categories [45] suggests that these DGE proteins per cluster account for 10–36% of the total MxDK1622 genes in the respective COG category (Table S4B). COG categories i.e. signal transduction mechanisms (8.48%), cell wall/membrane/envelope biogenesis (5.72%), and energy production and conversion (5.52%) represent the maximum number of DGE proteins

Genes belonging to cluster-1 are highly expressed during vegetative growth in the rich medium as compared to vegetative growth in the nutrient-poor medium. During development, these 134 genes are highly expressed during the early stages of development, 1, 3, 6, and 12-h post-induction. After 12 h, their expression decreases and remains low for the rest of the development. This cluster includes 134 proteins amongst which 15 are involved in signal transduction function, 17 in transcription, and 11 in inorganic ion transport and metabolism

The genes representing cluster-2 are highly expressed during vegetative growth in rich medium and then down-regulated once development is initiated. This cluster includes 15 genes, all of which are downregulated during the entire developmental cycle

Genes belonging to cluster-3 show high expression during vegetative growth regardless of nutrient level (with a few exceptions) and the first 6 h of development. At approximately 12 h, expression decreases and remains low for the remainder of development. This group includes 323 genes encoded proteins, involved in COG functions related to mostly metabolism (50%), 27 proteins in signal transduction, and 17 in post-translational modification, protein turnover, and chaperones

Cluster-4 genes show a maximum expression during growth in the nutrient-poor medium as compared to the rich medium. In several cases, their expression is low during early development, however, it increases around 6–12 h. Usually their expression decreases during development, however, in a few cases, it increases towards the end of development

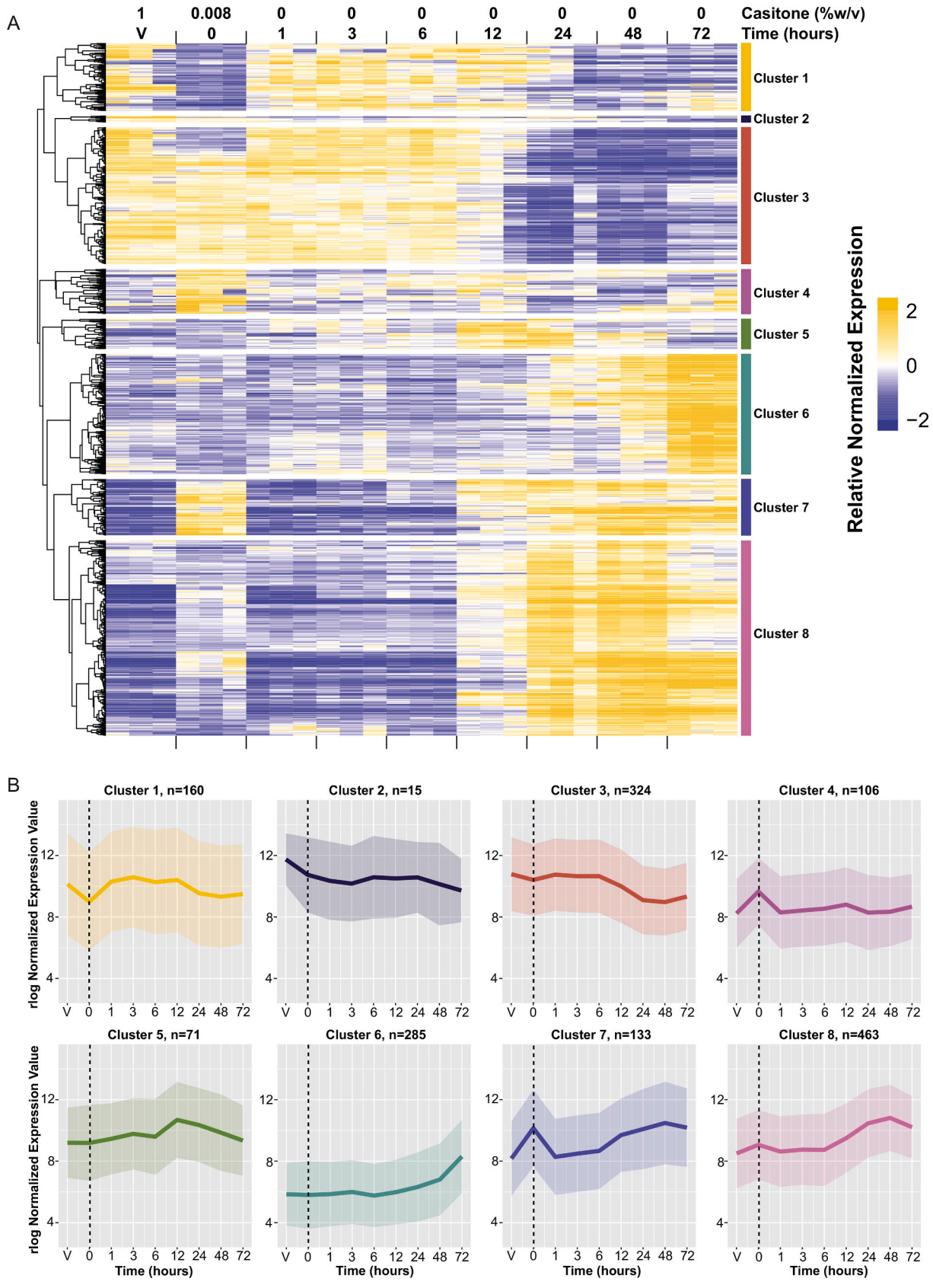


Fig. 4A. Developmental time course heat map of differentially expressed genes ($log_2FC > 1$, $p_{adj} < 0.05$) genes in *M. xanthus* DK1622. Casitone concentration (% *w/v*) and time (hours) are reported on the horizontal axis. 'V' in the time course indicates a vegetative steady state. Fig. 4B: Trend charts of the eight identified clusters representing the differentially expressed genes ($log_2FC > 1$, $p_{adj} < 0.05$) over the *M. xanthus* DK1622 developmental time course reported as normalized expression vs. time (hours). 'V' in the time course indicates a vegetative steady state. The bolded line represents the mean rlog expression of genes in each cluster. Shaded areas represent plus or minus one standard deviation from the mean

(48–72 h). Out of 106 genes identified in this cluster, ~50% of the genes encode hypothetical proteins of unknown function and 28 are involved in the cellular processes and signaling category. Only a small fraction of genes in the cluster encode known metabolic proteins. Within the cellular processes and signaling category, we identified six genes encoding proteins putatively involved in defense mechanisms and nine involved with cell membrane biogenesis. Within the "defense mechanism" associated protein set, we highlight the presence of CRISPR associated proteins (MXAN_7014-MXAN_7020), whose expression is also high during nutrient limitation and in the late sporulation cycle. This could suggest that cells are more sensitive and vulnerable to phage-attack during the developmental process and nutrient stress. Previous reports have demonstrated that glycerol induced myxospores are resistant to the MX-1 phage [46], however, it remains a possibility that cells undergoing the development program (as opposed to the chemical-induced sporulation) may represent a window of vulnerability for a range of potential phages, before spore formation. This vulnerable window may, therefore, require the upregulation of such defense proteins

Genes belonging to cluster-5 shows a relatively higher expression during 12 and 24 h on average. Out of 70 genes encoded proteins identified in this cluster, 47% are hypothetical, 31% are involved in cellular processes and signaling, 17% in metabolism, and a mere 4% in information processing

Cluster-6 consists of 285 genes highly expressed late in development, though their expression begins to increase at approximately 24 h. Besides the 42% of genes encoding hypothetical proteins, 26% and 25% of genes in this cluster encode proteins that are involved in cellular signaling and metabolism, respectively. Also, 29 genes encode proteins involved in cell wall/membrane/envelope biogenesis, which is expected because of major morphogenesis during spore formation. Need for sporulation (NFS) genes [47] i.e. *nfsA*, *nfsB*, *nfsC*, *nfsF*, *nfsG*, and *nfsH* are a part of this cluster. We also found *pkn1*, which encodes a putative eukaryotic-like protein kinase, within this cluster. Deletions of *pkn1* result in premature differentiation and poor spore production [48]

In cluster-7, we identified 133 genes that are comparatively upregulated during growth in the nutrient-poor medium as compared to the rich medium. These genes are also highly expressed during late development starting from 12 to 72 h. Of the genes in this cluster, 28% and 21% encode proteins involved in cellular signaling and metabolism respectively. We found several known developmental proteins in this cluster such as the *devTRS* operon and *mbhA*. It has been suggested that proteins encoded by *devTRS* work as a switch and activate the completion of aggregation [49], however, *devTRS* mutants fail to sporulate, therefore suggesting that dev operon regulate the timing of sporulation [50]. Our results support previous studies where the peak expression of *mbhA* has been shown to occurs at 24 h [51,52]

Finally, the 456 genes in cluster-8, show higher expression during late development starting from 12 to 72 h. The primary difference between cluster-7 and 8 is the relatively higher expression observed during growth in the nutrient-poor medium in cluster-7. Cluster-8 genes putatively encode 34% cellular processes and signaling proteins, 18% metabolic proteins, 8% information processing proteins, and 39% hypothetical proteins. Several genes, known to be involved in the development, such as *relA* [53], *fruA* [34], *espABC*, *asgB*, *nsd*, *pkn13*, *sigC*, *tps*, *agmU*, *aglT*, *nfsDE*, etc., map to this cluster

We also performed a comparative analysis between our transcriptomics analysis and the recently published report [13] as provided in Table S4C. We have used 3–4 replicates per time point whereas the other study has used only 2 replicates per time point. We have only ~10% average rRNA contamination in our data whereas the latter has ~98% rRNA contamination. We have identified 1522 whereas the previous study identified 1415 significantly expressed genes and amongst these two, only 512 genes are shared accounting for a mere ~30% of total genes identified in each study. In Table S7, we have shown all MxDK1622 proteins and their expression values at different time points as identified in our study along with the categorization (including fail correlation and fail data points) suggested by Muñoz-Dorado et al. 2019

3.6. Identification of maximum and minimum expressed genes during developmental time course

The analysis described above identified those genes whose expression patterns were most significantly altered during development. Many of these genes may play significant roles in both development and vegetative growth. However, we are also interested to know which genes are exclusively up/downregulated during the developmental time. Therefore, this study also identified 686 genes (maximum relative expression as compared to the vegetative phase) and 429 genes (minimum relative expression as compared to the vegetative phase) during various stages of development (Tables S5A and S5B). Experimentally, development can be induced by rapid nutrient depletion [14], activating both nutrient stress-induced genes as well as genes required for the development. Separating genes involved in the nutrient stress response from those involved in development can, therefore, be difficult. By using our bioreactors, we can identify genes that are expressed under low nutrient conditions (0.008% casitone) as compared to vegetative grown under standard lab conditions as well as development. Table S5A and S5B identify 95 genes expressed higher under low nutrient conditions as compared to other development timepoints, representing a nutrient stress regulon(s). Of these 95 genes, 50% are represented by genes of unknown function/poorly characterized, 25% are involved in cellular processes and signaling, 12.5% are involved in metabolism, and 10.5% are involved in information storage and processing (transcriptional/translational processing). Several protein kinases, proteases, transcription factors, and 9 out of 21 CRISPR-associated proteins are expressed amongst these regulons suggesting a vulnerable state of the cell under low nutrient biofilm conditions

3.7. Pathways involved in the control of the initiation of development

Previously, we have been interested in the transition from vegetative growth to the initiation of the developmental program, identifying cues, conditions, and genetic pathways that control the initiation process [53,54]. We have shown that the primary mechanism *M. xanthus* uses to examine nutrient availability is by monitoring protein synthesis via the stringent response. The primary regulator of the stringent response is RelA, the ppGpp(p) synthase. Several labs have identified genes that either control or regulate (p)ppGpp levels, the secondary molecule that controls the stringent response, under different developmental and growth conditions. A previous review [44] of these genetic pathways organized these genes into the following distinct categories with genes involved in 1) the stringent response 2) controlling the levels of (p)ppGpp, 3) nutrient sensing, 4) developmental timing, 5) the cellular response to starvation (Global regulators) and 6) intercellular signaling. Fig. 5 depicts the expression profiles of genes belonging to these categories studied during the whole development time course

Most of the genes belonging to these six categories have a higher expression in lower nutrient conditions and during the first six hours of development, suggesting that with the decrease in nutrient concentration, these genes become activated to control the developmental program. However, there are several notable exceptions. First, *relA* expression is higher under rich nutrient conditions compared to low nutrient and early development. One explanation is that *relA* is under the control of growth rate, which would keep the levels of transcription-coupled to that of the growth rate (Singer and Kaiser, 1995); since cells growing under poorer nutrient conditions grow at a slower rate, *relA* transcript levels would be lower. The *M. xanthus* genome encodes four *dksA* homologs, (MXAN_3200, MXAN_3006, MXAN_5718, MXAN_7086) and their expression patterns vary. Two of them, MXAN_3200 and MXAN_5718, have the maximum expression during vegetative growth, however, MXAN_3006 has approximately similar expression during the

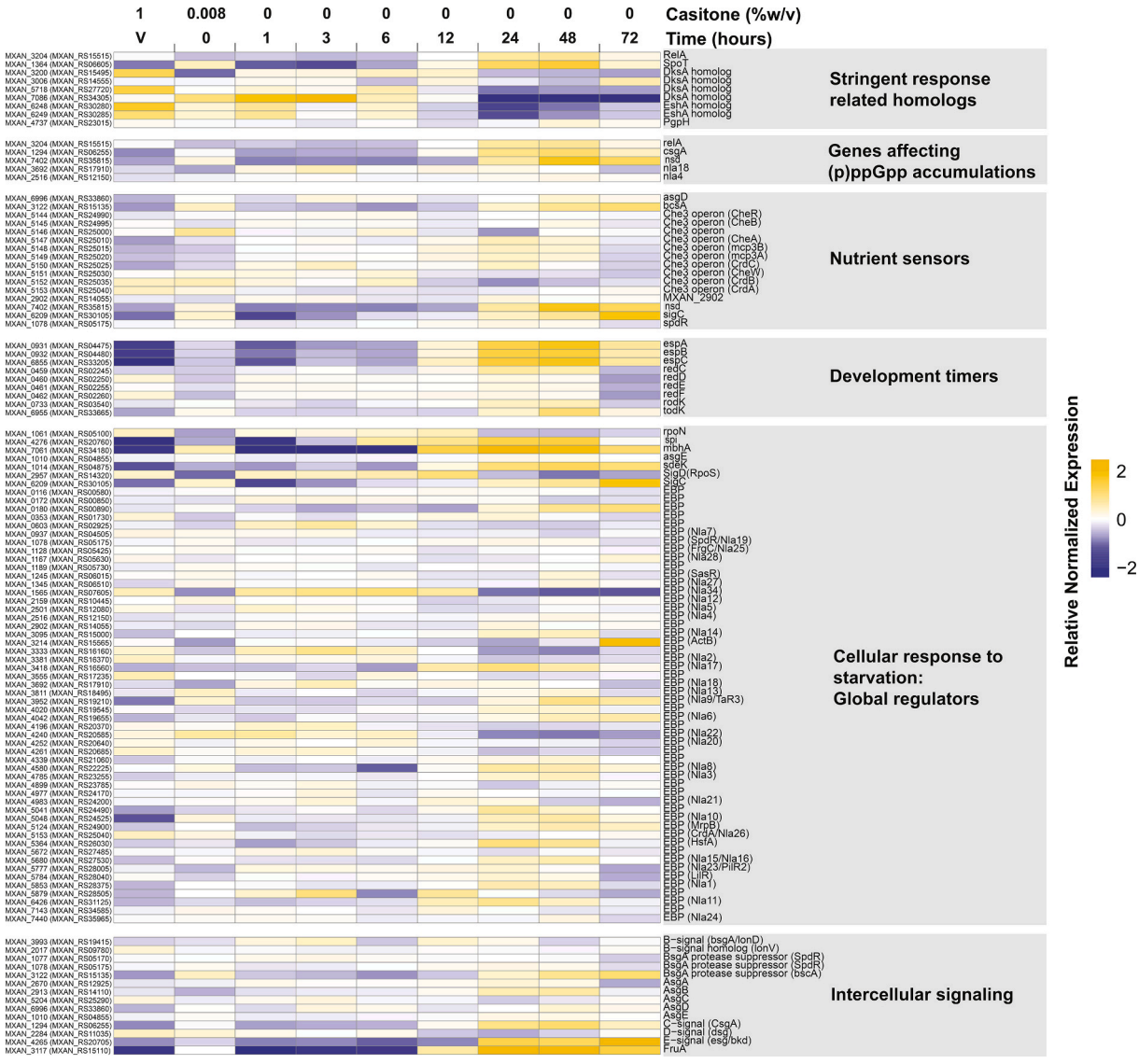


Fig. 5. Transcription expression analysis of genes belonging to six developmental categories during the 72-h developmental time course

whole analysis, and MXAN_7086 has maximum expression during low-nutrient conditions and early development and later decreases. Little is known about growth rate control in *M. xanthus* and this may represent the first data that identifies genes involved in this global regulatory process. Two genes, *eshA* (2 homologs in *M. xanthus;* MXAN_6248 and MXAN_6249) and *pgpH* (MXAN_4737) have been shown to encode proteins that modulate (p)ppGpp levels or (p)ppGpp-dependent gene expression in a variety of bacterial species during vegetative growth [55–57]. Amongst these, *eshA* has been implicated in maintaining (p)ppGpp during the late growth phase [56]. *M. xanthus* has two *eshA* homologs and both have maximum expression during vegetative growth and like two of the *dksA* homologs (MXAN_3200 and MXAN_5718), their expression decreases in low-nutrition conditions and during development. However, the expression of *pgpH* remains relatively constant throughout the time course. *cheB* (from the Che3 cluster) and *MXAN_2902*, a developmentally regulated sigma-54-transcriptional activator shown to affect nutrient sensing [31], exhibit higher levels of expression during growth in rich media and early development as compared to low nutrient media

Based on years of research, many global regulators have been implicated in nutrient-stress control in *M. xanthus*. These include three sigma factors, *rpoN* (MXAN_1061), the RpoS homolog *sigD* (MXAN_2957), and *sigC* (MXAN_6209), known to control the gene expression during sporulation [44,58–60]. Some early developmental genes such as *spi*, *mbhA*, *asgE*, and *sdeK* are transcribed from Sigma-54-like promoters and are maximally expressed during the first few hours of development. Besides these, two of the 52 Enhancer Binding Proteins (EPBs), *nla4,* and *nla18* are required to activate early RpoN-dependent transcription [61,62]. Fig. 5 depicts the relative expression patterns of several known nutrient stress-response genes. Our data shows that *rpoN* is maximally expressed during high nutrient vegetative growth followed by downregulation during the rest of the developmental cycle. The expression of *sigD* shows maximum expression during the vegetative growth and early development followed by a decrease in expression during late development. Transcript abundance is also lower during nutrient-limited growth. This latter result was a bit surprising based on *sigD* described to be the *M. xanthus* stationary phase RpoS homolog [63,64]. Our data suggests that the *M. xanthus rpoS* regulon may not be identical to the *E. coli rpoS* operon and it may have evolved into a variety of unique functions, similar to the divergent regulon *rpoN* in *M. xanthus*. However, contrary to *sigD*, *sigC* is expressed maximally under low nutrient conditions and during the later phases of development. Transcript abundance of *spi*, *mbhA*, *asgE*, and *sdeK* genes is maximal during low nutrient growth and late development and extremely low during the

early portion of development 1–12 h. Transcripts of two EBPs, *nla4,* and *nal18,* show increased abundance once development is initiated, which is consistent with earlier reports [65,66]. One possibility is that the *rpoS* and *rpoC* regulons work in concert to control a variety of post-vegetative and low nutrient growth genes and operons in *M. xanthus*

Several studies within *M. xanthus* have identified genes involved in producing and responding to the four intercellular signaling systems; A, B, C, and *E*-signals (Fig. 5) [58,67–71]. For each of these signaling systems, varieties of genes have been identified that are involved in the production of the signals, and by-pass suppressors of these signals to help identify downstream pathways. Genes of interest for B-signal, the earliest of the extracellular signals, include *bsgA* (*lonD),* a member of the

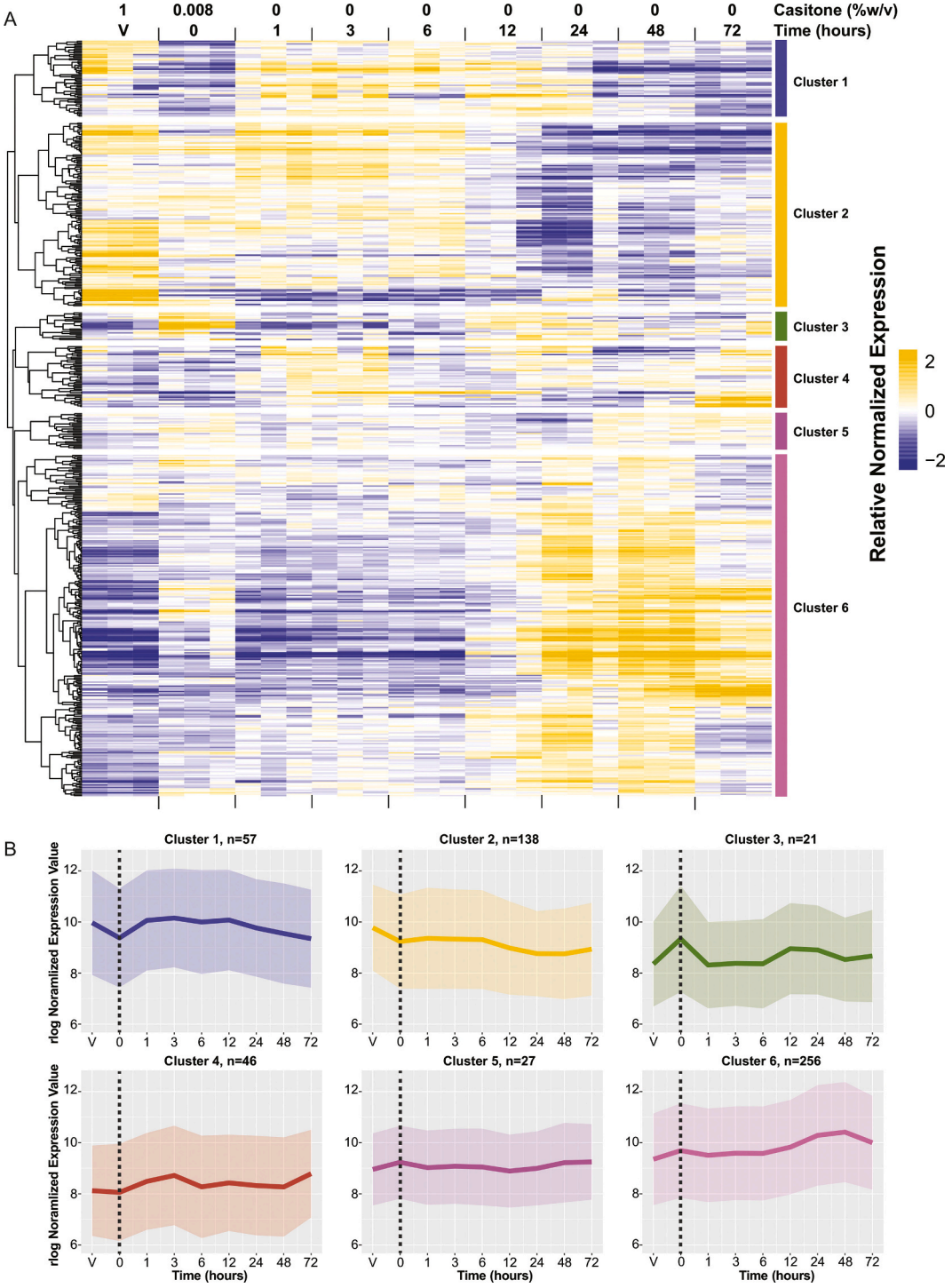


Fig. 6. A. Transcriptional profile of differentially expressed regulatory ($log_2FC > 1$, $p_{adj} < 0.05$) genes in *M. xanthus* DK1622. Casitone concentration (%w/v) and time (hours) are reported on the horizontal axis. 'V' in the time course indicates a vegetative steady state. Here, six clusters representing differentially expressed regulatory genes over the *M. xanthus* DK1622 developmental time course are displayed. Fig. 6B: Trend charts of the six identified clusters representing the differentially expressed genes ($log_2FC > 1$, $p_{adj} < 0.05$) over the *M. xanthus* DK1622 developmental time course reported as normalized expression vs. time (hours). 'V' in the time course indicates a vegetative steady state. The bolded line represents the mean rlog expression of genes in each cluster. Shaded areas represent plus or minus one standard deviation from the mean

Lon protease family, and a second Lon homolog, *lonV* [72], along with two *bsgA* suppressors- *spdR* and *bcs,* both implicated in nutrient sensing. Five genes (*asgA, asgB, asgC (rpoD), asgD,* and asg*E*) are involved in producing the A-signal. Only one gene has been identified that defines the C-signal (*csgA*), yet several by-pass suppressors have been identified, including *socC* and *socE.* E-signal (*esg/bkd*) is produced by enzyme complexes involved in branched-chain fatty acid and ether lipid biosynthesis [67]. B-signal (*bsgA*) is slightly induced during growth in poor-nutrient conditions compared to rich and increases during early development and fluctuates thereafter. The B-signal homolog (*lonV*), which is not required for development [73] is upregulated during growth in both rich and poor nutrient conditions and then decreases expression during development. This is consistent with previous reports suggesting that LonV is not required for development [72]. Interestingly, all three B-signal suppressors show high expression during growth in poor nutrient conditions and display a strong decrease in expression during the remainder of the development. This suggests that these genes are involved in a process that normally inhibits development during vegetative growth, later degraded by LonD (BsgA) during early development. Therefore, the BsgA suppressor mutants are unable to make these gene products, relieving the requirement of BsgA. Our data shows that these suppressors are not expressed during development which is consistent with this hypothesis. Amongst the genes that encode A-signal proteins, only asgB, classified in cluster 8, has a significant variation during all time points. *asgB* is considerably downregulated during poor nutrient media. C-signal (*csgA*) is expressed highly during growth in nutrient-poor conditions and then again late development; however, *esg* (*E*-signal) is highly expressed only after 24 h of development

3.8. The quest for new developmental regulators

M. xanthus along with other members of order Myxococcales exhibit a complex lifestyle and require a prolific number of regulators [74,75]. Previous studies have extensively identified several putative regulators amongst this organism [32,76–80]. It has been previously mentioned that out of 95 genes known to be involved in *M. xanthus* development, 60 (64%) belong to the signal transduction category (at genome level it is just 7.7%), demonstrating the importance of regulatory proteins in the development [81]. We identified 547 putative and known regulatory genes involved in diverse regulatory functions including nucleotide-binding, two-Component Systems (TCS), ST kinases, sigma factors, MCPs, cyclases, and proteases. We believe that many of these genes might be involved in controlling protein levels or activities in the cell during some phases of growth and development (Table S6). Based on their expression pattern during a developmental time course, these genes can be grouped into six clusters as shown in Figs. 6 and 6B. Of initial interest are those regulatory genes that fall into clusters four and six due to their increase in expression during the developmental time course, implying their role in regulatory behavior. It should be noted that to date, most of the regulators controlling development have been positive regulators. However, we realize that negative regulators may also play an important role in this process. For instance, if developmental genes are repressed during vegetative growth, the removal of the repressor would allow for expression and we would predict that such negative regulators would exhibit a decrease in mRNA levels during development. Such regulators should be found in Cluster 2

Most of the genes in cluster-1 show upregulation during vegetative growth in the rich medium as compared to nutrient-poor medium, while during development these genes are highly expressed during 1, 3, 6, and 12 h and then begin to decrease. Cluster-2 consists of genes whose expression pattern is maximum during the vegetative growth regardless of nutrient levels and then decreases around 6 h and remains low during the remainder of the developmental cycle. Cluster-3 includes only 21 genes and they are expressed mostly during growth at low nutrient concentrations, show a decrease in expression during early development and resume expression at 12 and 24 h, then shut down again. Amongst the 46 genes within Cluster-4, most of them have low expression during vegetative growth, regardless of the nutrient concentration then increase expression within the first hour and maximal expression at 72 h. Genes in this cluster show bi-phasic expression patterns with a high expression both early and late in development. Twenty-seven genes of cluster-5 show low expression during most of the development, with an increase at the late time points. Genes in this cluster also show expression during growth in the low nutrient medium as compared to the rich medium. The genes belonging to the largest cluster, Cluster-6, are expressed maximally during 24, 48, and 72 h, and begin expression at around 12 h. They are poorly expressed during vegetative growth regardless of the nutrient level. This analysis will help the myxobacterial community to consider these regulatory genes one by one for a better understanding of their roles during development

3.9. Variations in central metabolic pathways during development

Several transcriptional changes that we observed upon induction of development are amongst central metabolic processes along with the glycolate pathway and other pathways originated from acetyl-CoA (Table S3, S7, and Fig. 4). One of the earliest response observed is the induction of a large (>70) number of genes encoding degradative enzymes, primarily proteases and glycohydrolases, such as *exo*-alpha-sialidase (MXAN_2203) primarily comprised of genes from clusters 1 and 3. We suggest that these genes represent the low nutrition-induced genes, realigning the cells to possibly scavenge for basic carbon materials. Most of these genes are also up-regulated during growth in low nutrient media as suggested in our results (Fig. 3, Table S3) and in one of our previous reports too [12]. Our analysis revealed the activation of several clusters 7 and 8 genes such as non-ribosomal peptide synthases (MXAN_3935, 4078, 4079, 4414, and 4415), fibronectin (MXAN_2170) amylose (MXAN_0592), and putrescine (MXAN 5481) at around 12 h (Fig. 7). We also identify an increase in the expression of a variety of CRISPER array genes, suggesting that the developmental state may represent a more vulnerable state for infection

Previous studies [82,83] had examined changes in glyoxylate enzyme levels during development. They demonstrated that during glycerol-induced spore formation, there is an increase in the expression of both isocitrate lyase and malate synthase. This is also supported by previously reported transcriptomic analysis of chemical-induced sporulation [47]. This study identified that low nutrition-induced *M. xanthus* development and myxospore formation also show an increased expression of both malate synthase (MXAN_6441) and isocitrate lyase (MXAN_6442). Both genes are in cluster 8, whose expression is highest around 12 h post-starvation (low nutrition). Simultaneously, we find that the expression of several TCA cycle genes categorized in Cluster 3 is high during early development but begins to decrease after 6–12 h into development (Fig. 7). This data suggests that cells begin to shift their carbon flow between 6 and 12 h (around the aggregation to mound stage) to utilize the glyoxylate pathway thereby conserving carbon and the expense of NADH generation. Consistent with this, we identified a cessation at 12 h of transcription for electron transport components such as Quinone biosynthesis, NAD metabolism (Fig. 7D), and Redox buffers such as glutathione. Interestingly, expression of MXAN_5363, which converts NAD^+ to $NADP^+$ increases expression during this time. Between 12 and 24 h, we observe what appears to be a change in the utilization of acetate and acetyl-CoA (Fig. 7E) in addition to the activation of the glyoxylate pathway. Several genes involved in sulfur metabolism (MXAN_6507, 2335, and 2336) are also found activated in the later development. Another major change that we observed was in the genes involved in the biosynthesis of phenylalanine, tyrosine, tryptophan, siderophore, and porphyrin (Fig. 7B). The expressions of several DAHP Synthases, MXAN-3219, and 3526 are down-regulated at 12 h (Cluster 3) whereas MXAN_3642 (3-deoxy-7-phosphoheptulonate synthase) becomes active around 24 h (Cluster 6). As chorismate is being redirected to the formation of Enterochelin, a high-affinity siderophore

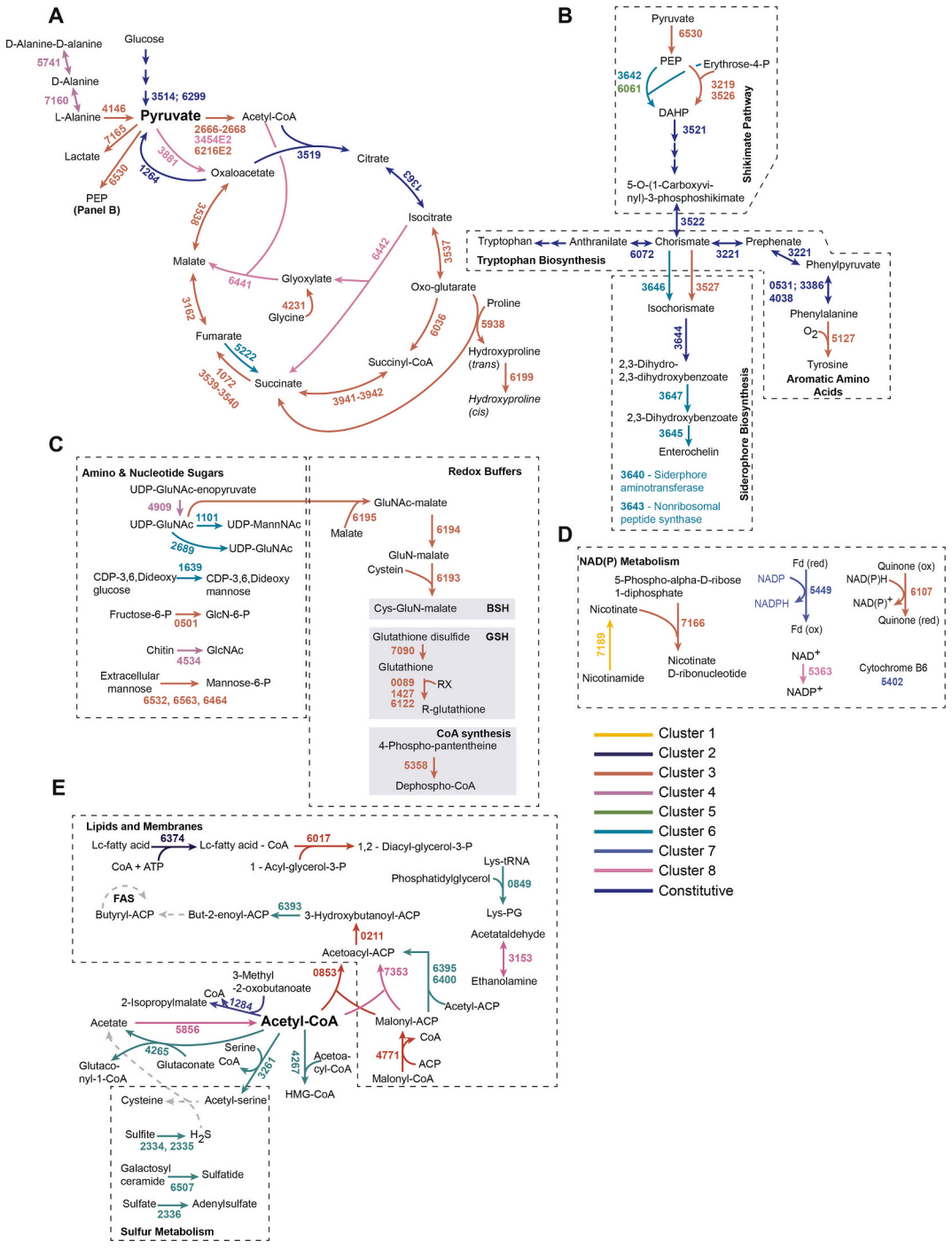


Fig. 7. The network of metabolic pathways implicated in the progression of development in *M. xanthus*. MXAN gene names encoding the enzymes catalyzing steps shown in the figure are indicated by their respective four-digit MXAN gene number. Arrows indicating enzyme-catalyzed steps and the associated gene number are color-coded as per the corresponding gene's membership in the clusters identified in Figure 4AB. Several Pathways 7A. The pathways of central carbon metabolism as centered on pyruvate. The first 9 steps of glycolysis found to be constitutively expressed are represented with multiple arrows in series. 7B. Metabolic pathways linking the shikimate pathway, aromatic amino acid biosynthesis, tryptophan biosynthesis, and siderophore biosynthesis are depicted. Multiple steps of the shikimate pathway and tryptophan biosynthesis found to be constitutively expressed are indicated by multiple arrows in series. 7C. Pathways associated with the synthesis of redox buffers and amino and nucleotide sugars; Gray boxes highlight key redox buffers. 7D. Pathways involved in NAD(P) metabolism. 7E. Metabolic pathways centered around acetyl-CoA. Steps in pathways illustrated with a dashed gray line indicate connections between genes that have not been explicitly found to meet the criteria for colored highlighting, however, for which observing the close connection may be informative. This includes the biosynthesis of cysteine and acetate from hydrogen sulfide and acetyl-serine and enzyme activity by the fatty acid synthase

during these late development hours, it might suggest that cells are scavenging iron

3.10. How multicellular development might have evolved amidst myxobacteria

Order Myxococcales is one of nine orders within the class

Deltaproteobacteria, well-known for their large (most of them between 9 and 16 Mb) genome, aerobic respiration, and complex social behaviors (development, sporulation, gliding motility, etc.) with a few exceptions such as *Pajaroellobacter* (~2 Mb genome size, pathogenic, and anaerobic), *Anaeromyxobacter* (~5 Mb genome size and anaerobic), *Vulgatibacter* (~4 Mb genome size and aerobic). Myxobacteria have been phylogenetically classified within the Deltaproteobacteria based on their 16S RNA homology, however, phenotypically, genetically, and genome-wise, they are quite diverse from the latter. Therefore, this study allows us to address two additional questions i.e. "How unique these developmentally regulated genes (taxonomically and homology-wise) are" and "from where these genes might have originated from"

While looking at the 591 upregulated expressed genes from T01 to T72 (Table S5), ~45% of those genes are poorly characterized (COG category) out of which ~50% are unique to order Myxococcales. However, only ~10% of the COG categorized genes are unique to the Myxococcales. Amongst 429-downregulated genes, we found ~30% poorly characterized proteins out of which 35% are only present in myxobacteria and only 6 of the COG categorized proteins are unique amongst myxobacteria. Conclusively, we can argue that while many, if not the majority, of the genes involved in the developmental program, are evolutionary modifications of genes that encode gene products of known functions, a large fraction, almost ~20%, represent genes of unknown function that are only found in Myxococcales. Interestingly, of the 591 upregulated and 429 downregulated genes, a mere 5% of genes show closeness with non-Myxococcales Deltaproteobacteria, which suggests that either this myxobacterium has procured these genes/proteins from diverse bacterial groups or they might have evolved independently. For the remainder genes, Gammaproteobacteria (~16%), Firmicutes (~13%), Actinobacteria (~10%), Betaproteobacteria (~7%) are the closest lineage. This propensity of the Myxococcales to incorporate genes from other organisms and assimilate them into their genome has been previously reported in *M. xanthus* DK1622 [84]. What is of interest here is how these genes got assimilated into the developmental program. As many of these genes are of unknown function, therefore, they represent a rich new group of genes and gene products for study. This study brings out the first basic understanding of their function i.e. they are associated with development. The question of future interest is to identify what role these new and novel genes play in the development followed up by their molecular and biochemical function characterization

4. Conclusion

Understanding the transcriptional regulation of complex biological phenotypes, such as development, is now possible due to the recent advances of new generation sequencing and RNA analyses. This study focuses on the identification of genes involved in the *M. xanthus* developmental program, first during nutrient stress, and later during the developmental time course. Here we implement a simple cost-effective strategy to remove >90% of the rRNAs from isolated transcriptome using RNaseH, which provided us higher coverage by using multiplexing-based sequencing for vegetative, low nutrition, and 72-h development time points in *M. xanthus*. This work has identified several unique and unexpected changes in developmental gene expression, including the activation of enterochelin and other siderophores synthesis and the involvement of genes encoding enzymes of key central metabolic pathways in mediating this development. We have also compared our results with a recently published analysis and found that only ~30% of the up/down-regulated genes are shared between two studies. Most of the up/down-regulated genes show more closeness with non-Myxococcales Deltaproteobacteria, suggesting that these putatively horizontally transferred genes might have assimilated into the developmental program during evolution. Our dataset will serve as a resource for other myxobacterial researchers seeking to dissect the genetic basis of development in *M. xanthus*. Finally, we expect that these data and protocols should also facilitate similar studies on closely related species, thereby enabling cross-species comparative studies that will permit a better understanding of the evolution of social behavior in bacteria

Funding

This work was supported by the National Science Foundation grant: IOS-1354562 to M.S

Authors' contributions

GTS, JL, and ML did the experimental work. GS, AIY, GTS, MTF, and MS analyzed the data. AIY and GS generated the figures/tables. GS and GTS wrote the manuscript. MS provided funding and intellectual contributions. All authors read and approved the manuscript

Declaration of Competing Interest

Authors declare that they have no conflicts of interest

Acknowledgments

The authors would like to thank Blythe Durbin-Johnson at the UC Davis bioinformatics core facility for statistical assistance, Ivy Jose for technical assistance, and Scott Dawson for critical discussions. This work was supported by the National Science Foundation grant: IOS-1354562 to M.S. GS is his present position is supported by the DST-INSPIRE Faculty Award from the Government of India

Appendix A. Supplementary data

Raw reads for all RNAseq experiments can be accessed at the NIH Sequence Read Archive database (http://www.ncbi.nlm.nih.gov/sra) with SRA number: SRP181226, BioProject accession number: PRJNA516475. SRA accession numbers for the corresponding development phase are SRR8474501 (vegetative), SRR8474502 (low nutrient), SRR8474504 (1 h.), SRR8474503 (3 h.), SRR8474506 (6 h.), SRR8474505 (12 h.), SRR8474508 (24 h.), SRR8474507 (48 h.), and SRR8474509 (72 h.). The complete transcriptional profile of all *M. xanthus* DK1622 encoded genes is also presented in Table S7. Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2020.11.030.

References

[1] M. Dworkin, D. Kaiser, Cell interactions in myxobacterial growth and development, Science 230 (1985) 18–24.
[2] D. Kaiser, Signaling in myxobacteria, Annu. Rev. Microbiol. 58 (2004) 75–98.
[3] D.T. Pathak, X. Wei, D. Wall, Myxobacterial tools for social interactions, Res. Microbiol. 163 (2012) 579–591.
[4] L.J. Shimkets, Control of morphogenesis in myxobacteria, Crit. Rev. Microbiol. 14 (1987) 195–227.
[5] Z. Vaksman, H.B. Kaplan, Myxococcus xanthus Growth, Development, and Isolation, Current protocols in microbiology 39 (2015), 7a.1.1-7a.1.21.
[6] L. Kroos, A. Kuspa, D. Kaiser, A global analysis of developmentally regulated genes in *Myxococcus xanthus*, Dev. Biol. 117 (1986) 252–266.
[7] D.R. Yoder-Himes, L. Kroos, Regulation of the *Myxococcus xanthus* C-signal-dependent Omega4400 promoter by the essential developmental protein FruA, J. Bacteriol. 188 (2006) 5167–5176.
[8] L. Kroos, A. Kuspa, D. Kaiser, Defects in fruiting body development caused by Tn5 lac insertions in *Myxococcus xanthus*, J. Bacteriol. 172 (1990) 484–487.
[9] D. Kaiser, L. Kroos, Intercellular Signaling. Chapter 12, In "Myxobacteria II", ASM Press, Washington D.C, 1993. American Society for Microbiology, Washington, D. C.
[10] L. Kroos, Highly signal-responsive gene regulatory network governing Myxococcus development, Trends Genet. 33 (2017) 3–15.
[11] J. Muñoz-Dorado, F.J. Marcos-Torres, E. García-Bravo, A. Moraleda-Muñoz, J. Pérez, Myxobacteria: Moving, Killing, Feeding, and Surviving Together, Frontiers in Microbiology, 7, 2016.
[12] D.L. Whitfield, G. Sharma, G.T. Smaldone, M. Singer, Peripheral rods: a specialized developmental cell type in *Myxococcus xanthus*, Genomics 112 (2020) 1588–1597.

[13] J. Muñoz-Dorado, A. Moraleda-Muñoz, F.J. Marcos-Torres, F.J. Contreras-Moreno, A.B. Martin-Cuadrado, J.M. Schrader, P.I. Higgs, J. Pérez, Transcriptome dynamics of the Myxococcus xanthus multicellular developmental program, eLife 8 (2019), e50374.

[14] G.T. Smaldone, Y. Jin, D.L. Whitfield, A.Y. Mu, E.C. Wong, S. Wuertz, M. Singer, Growth of *Myxococcus xanthus* in continuous-flow-cell bioreactors as a method for studying development, Appl. Environ. Microbiol. 80 (2014) 2461–2467.

[15] D. Kaiser, Social gliding is correlated with the presence of pili in *Myxococcus xanthus*, Proc. Natl. Acad. Sci. U. S. A. 76 (1979) 5952–5956.

[16] J.D. Morlan, K. Qu, D.V. Sinicropi, Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue, PLoS One 7 (2012), e42882.

[17] S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2014.

[18] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.

[19] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.

[20] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169.

[21] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.

[22] R. Kolde, M.R. Kolde, Package 'Pheatmap', R Package, 1, 2015.

[23] P. Mestdagh, N. Hartmann, L. Baeriswyl, D. Andreasen, N. Bernard, C. Chen, D. Cheo, P. D'Andrade, M. DeMayo, L. Dennis, S. Derveaux, Y. Feng, S. Fulmer-Smentek, B. Gerstmayer, J. Gouffon, C. Grimley, E. Lader, K.Y. Lee, S. Luo, P. Mouritzen, A. Narayanan, S. Patel, S. Peiffer, S. Ruberg, G. Schroth, D. Schuster, J.M. Shaffer, E.J. Shelton, S. Silveria, U. Ulmanella, V. Veeramachaneni, F. Staedtler, T. Peters, T. Guettouche, L. Wong, J. Vandesompele, Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study, Nature methods 11 (2014) 809–815.

[24] C. Everaert, M. Luypaert, J.L.V. Maag, Q.X. Cheng, M.E. Dinger, J. Hellemans, P. Mestdagh, Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data, Sci. Rep. 7 (2017) 1559.

[25] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, A. Mortazavi, A survey of best practices for RNA-seq data analysis, Genome Biol. 17 (2016) 13.

[26] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden, BLAST+: architecture and applications, BMC Bioinformatics 10 (2009) 421.

[27] H. Yi, Y.J. Cho, S. Won, J.E. Lee, H. Jin Yu, S. Kim, G.P. Schroth, S. Luo, J. Chun, Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq, Nucleic Acids Res. 39 (2011), e140.

[28] X. Adiconis, D. Borges-Rivera, R. Satija, D.S. DeLuca, M.A. Busby, A.M. Berlin, A. Sivachenko, D.A. Thompson, A. Wysoker, T. Fennell, A. Gnirke, N. Pochet, A. Regev, J.Z. Levin, Comparative analysis of RNA sequencing methods for degraded or low-input samples, Nat. Methods 10 (2013) 623–629.

[29] H.B. Bode, M.W. Ring, G. Schwar, M.O. Altmeyer, C. Kegler, I.R. Jose, M. Singer, R. Muller, Identification of additional players in the alternative biosynthesis pathway to isovaleryl-CoA in the myxobacterium *Myxococcus xanthus*, Chembiochem : a European journal of chemical biology 10 (2009) 128–140.

[30] V.D. Pham, C.W. Shebelut, I.R. Jose, D.A. Hodgson, D.E. Whitworth, M. Singer, The Response Regulator PhoP4 Is Required for Late Developmental Events in *Myxococcus xanthus*, Microbiology (Reading, England), 152 (2006), 1609-1620.

[31] J.S. Jakobsen, L. Jelsbak, L. Jelsbak, R.D. Welch, C. Cummings, B. Goldman, E. Stark, S. Slater, D. Kaiser, Sigma54 enhancer binding proteins and *Myxococcus xanthus* fruiting body development, J. Bacteriol. 186 (2004) 4361–4368.

[32] X. Shi, S. Wegener-Feldbrugge, S. Huntley, N. Hamann, R. Hedderich, L. Sogaard-Andersen, Bioinformatics and experimental analysis of proteins of two-component systems in *Myxococcus xanthus*, J. Bacteriol. 190 (2008) 613–624.

[33] C.J. Rosario, M. Singer, Developmental expression of dnaA is required for sporulation and timing of fruiting body formation in *Myxococcus xanthus*, Mol. Microbiol. 76 (2010) 1322–1333.

[34] M. Ogawa, S. Fujitani, X. Mao, S. Inouye, T. Komano, FruA, a putative transcription factor essential for the development of *Myxococcus xanthus*, Mol. Microbiol. 22 (1996) 757–767.

[35] H. Sun, W. Shi, Analyses of mrp genes during *Myxococcus xanthus* development, J. Bacteriol. 183 (2001) 6733–6739.

[36] A.G. Garza, J.S. Pollack, B.Z. Harris, A. Lee, I.M. Keseler, E.F. Licking, M. Singer, SdeK is required for early fruiting body development in *Myxococcus xanthus*, J. Bacteriol. 180 (1998) 4628–4637.

[37] M. Cumsky, D.R. Zusman, Myxobacterial hemagglutinin: a development-specific lectin of *Myxococcus xanthus*, Proc. Natl. Acad. Sci. U. S. A. 76 (1979) 5505–5509.

[38] M. Robinson, B. Son, D. Kroos, L. Kroos, Transcription factor MrpC binds to promoter regions of hundreds of developmentally-regulated genes in *Myxococcus xanthus*, BMC Genomics 15 (2014) 1123.

[39] P.T. McLaughlin, V. Bhardwaj, B.E. Feeley, P.I. Higgs, MrpC, a CRP/Fnr homolog, functions as a negative autoregulator during the Myxococcus xanthus multicellular developmental program, Mol. Microbiol. 109 (2018) 245–261.

[40] H. Sun, W. Shi, Genetic Studies of mrp, a Locus Essential for Cellular Aggregation and Sporulation of Myxococcus xanthus, Journal of Bacteriology 183 (2001) 4786–4795.

[41] J.S. Downard, D.R. Zusman, Differential expression of protein S genes during *Myxococcus xanthus* development, J. Bacteriol. 161 (1985) 1146–1155.

[42] S. Inouye, H. Nariya, J. Muñoz-Dorado, 11 Protein Ser/Thr Kinases and Phosphatases in Myxococcus Xanthus, in: Myxobacteria, American Society of Microbiology, 2008.

[43] P.G. Livingstone, A.D. Millard, M.T. Swain, D.E. Whitworth, Transcriptional changes when Myxococcus xanthus preys on Escherichia coli suggest myxobacterial predators are constitutively toxic but regulate their feeding, Microbial genomics 4 (2018), e000152.

[44] M.E. Diodati, R. Gill, L. Plamann, M. Singer, in: D. Whitworth (Ed.), Development I: Initiation and Early developmental events. Chapter 3 in 'Myxobacteria. Multicellularity and Differentiation', ASM Press, Washington D.C, 2008, pp. 479–489.

[45] M.Y. Galperin, K.S. Makarova, Y.I. Wolf, E.V. Koonin, Expanded microbial genome coverage and improved protein family annotation in the COG database, Nucleic Acids Res. 43 (2015) D261–D269.

[46] R.P. Burchard, M. Dworkin, A bacteriophage for Myxococcus xanthus: isolation, characterization and relation of infectivity to host morphogenesis, J. Bacteriol. 91 (1966) 1305–1313.

[47] F.-D. Müller, A. Treuner-Lange, J. Heider, S.M. Huntley, P.I. Higgs, Global transcriptome analysis of spore formation in *Myxococcus xanthus* reveals a locus necessary for cell differentiation, BMC Genomics 11 (2010) 264.

[48] J. Munoz-Dorado, S. Inouye, M. Inouye, A gene encoding a protein serine/threonine kinase is required for normal development of *M. xanthus*, a gram-negative bacterium, Cell 67 (1991) 995–1006.

[49] L. Thony-Meyer, D. Kaiser, devRS, an autoregulated and essential genetic locus for fruiting body development in *Myxococcus xanthus*, J Bacteriol 175 (1993) 7450–7462.

[50] R. Rajagopalan, L. Kroos, The dev operon regulates the timing of sporulation during Myxococcus xanthus development, J. Bacteriol. 199 (2017) e00788-00716.

[51] J.R. Kirby, D.R. Zusman, Chemosensory regulation of developmental gene expression in *Myxococcus xanthus*, Proc. Natl. Acad. Sci. 100 (2003) 2008–2013.

[52] M. Cumsky, D.R. Zusman, Myxobacterial hemagglutinin: a development-specific lectin of *Myxococcus xanthus*, Proc. Natl. Acad. Sci. 76 (1979) 5505–5509.

[53] B.Z. Harris, D. Kaiser, M. Singer, The guanosine nucleotide (p)ppGpp initiates development and A-factor production in *Myxococcus xanthus*, Genes Dev. 12 (1998) 1022–1035.

[54] M. Singer, D. Kaiser, Ectopic production of guanosine penta- and tetraphosphate can initiate early developmental gene expression in *Myxococcus xanthus*, Genes Dev. 9 (1995) 1633–1644.

[55] S. Masuda, C.E. Bauer, Null mutation of HvrA compensates for loss of an essential relA/spoT-like gene in *Rhodobacter capsulatus*, J. Bacteriol. 186 (2004) 235–239.

[56] N. Saito, J. Xu, T. Hosaka, S. Okamoto, H. Aoki, M.J. Bibb, K. Ochi, EshA accentuates ppGpp accumulation and is conditionally required for antibiotic production in *Streptomyces coelicolor* A3(2), J. Bacteriol. 188 (2006) 4952–4961.

[57] S. Liu, D.O. Bayles, T.M. Mason, B.J. Wilkinson, A cold-sensitive *Listeria monocytogenes* mutant has a transposon insertion in a gene encoding a putative membrane protein and shows altered (p)ppGpp levels, Appl. Environ. Microbiol. 72 (2006) 3955–3959.

[58] I.M. Keseler, D. Kaiser, An early A-signal-dependent gene in *Myxococcus xanthus* has a sigma 54-like promoter, J. Bacteriol. 177 (1995) 4638–4644.

[59] T. Ueki, S. Inouye, A new sigma factor, SigD, essential for stationary phase is also required for multicellular differentiation in *Myxococcus xanthus*, Genes to cells : devoted to molecular & cellular mechanisms 3, 1998, pp. 371–385.

[60] D. Apelian, S. Inouye, A new putative sigma factor of *Myxococcus xanthus*, J. Bacteriol. 175 (1993) 3335–3342.

[61] K.M. Giglio, N. Caberoy, G. Suen, D. Kaiser, A.G. Garza, A cascade of coregulating enhancer binding proteins initiates and propagates a multicellular developmental program, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) E431–E439.

[62] K.M. Giglio, J. Eisenstatt, A.G. Garza, Identification of enhancer binding proteins important for *Myxococcus xanthus* development, J. Bacteriol. 192 (2010) 360–364.

[63] T. Ueki, S. Inouye, A new sigma factor SigD, essential for stationary phase is also required for multicellular differentiation in Myxococcus xanthus, Genes to Cells 3 (1998) 371–385.

[64] P. Viswanathan, M. Singer, L. Kroos, Role of sigmaD in regulating genes and signals during Myxococcus xanthus development, J. Bacteriol. 188 (2006) 3246–3256.

[65] F. Ossa, M.E. Diodati, N.B. Caberoy, K.M. Giglio, M. Edmonds, M. Singer, A. G. Garza, The *Myxococcus xanthus* Nla4 protein is important for expression of stringent response-associated genes, ppGpp accumulation, and fruiting body development, J. Bacteriol. 189 (2007) 8474–8483.

[66] M.E. Diodati, F. Ossa, N.B. Caberoy, I.R. Jose, W. Hiraiwa, M.M. Igo, M. Singer, A. G. Garza, Nla18, a key regulatory protein required for Normal growth and development of *Myxococcus xanthus*, J. Bacteriol. 188 (2006) 1733–1743.

[67] T. Ahrendt, C. Dauth, H.B. Bode, An iso-15 : 0 O-alkylglycerol moiety is the key structure of the E-signal in *Myxococcus xanthus*, Microbiology (Reading, England) 162 (2016) 138–144.

[68] M.G. Bowden, H.B. Kaplan, The *Myxococcus xanthus* developmentally expressed asgB-dependent genes can be targets of the a signal-generating or a signal-responding pathway, J. Bacteriol. 178 (1996) 6628–6631.

[69] E. Ellehauge, M. Norregaard-Madsen, L. Sogaard-Andersen, The FruA signal transduction protein provides a checkpoint for the temporal co-ordination of intercellular signals in *Myxococcus xanthus* development, Mol. Microbiol. 30 (1998) 807–817.

[70] J. Downard, S.V. Ramaswamy, K.S. Kil, Identification of esg, a genetic locus involved in cell-cell signaling during *Myxococcus xanthus* development, J. Bacteriol. 175 (1993) 7762–7770.

[71] D.C. Hagen, A.P. Bretscher, D. Kaiser, Synergism between morphogenetic mutants of *Myxococcus xanthus*, Dev. Biol. 64 (1978) 284–296.
[72] N. Tojo, S. Inouye, T. Komano, The lonD gene is homologous to the lon gene encoding an ATP-dependent protease and is essential for the development of Myxococcus xanthus, J. Bacteriol. 175 (1993) 4545–4549.
[73] N. Tojo, S. Inouye, T. Komano, Cloning and nucleotide sequence of the Myxococcus xanthus lon gene: indispensability of lon for vegetative growth, J. Bacteriol. 175 (1993) 2271–2277.
[74] D.R. Zusman, A.E. Scott, Z. Yang, J.R. Kirby, Chemosensory pathways, motility and development in Myxococcus xanthus, Nat. Rev. Microbiol. 5 (2007) 862–872.
[75] S. Schneiker, O. Perlova, O. Kaiser, K. Gerth, A. Alici, M.O. Altmeyer, D. Bartels, T. Bekel, S. Beyer, E. Bode, H.B. Bode, C.J. Bolten, J.V. Choudhuri, S. Doss, Y. A. Elnakady, B. Frank, L. Gaigalat, A. Goesmann, C. Groeger, F. Gross, L. Jelsbak, L. Jelsbak, J. Kalinowski, C. Kegler, T. Knauber, S. Konietzny, M. Kopp, L. Krause, D. Krug, B. Linke, T. Mahmud, R. Martinez-Arias, A.C. McHardy, M. Merai, F. Meyer, S. Mormann, J. Munoz-Dorado, J. Perez, S. Pradella, S. Rachid, G. Raddatz, F. Rosenau, C. Ruckert, F. Sasse, M. Scharfe, S.C. Schuster, G. Suen, A. Treuner-Lange, G.J. Velicer, F.J. Vorholter, K.J. Weissman, R.D. Welch, S. C. Wenzel, D.E. Whitworth, S. Wilhelm, C. Wittmann, H. Blocker, A. Puhler, R. Muller, Complete genome sequence of the myxobacterium *Sorangium cellulosum*, Nature biotechnology 25 (2007) 1281–1289.
[76] G. Sharma, I. Khatri, S. Subramanian, Comparative Genomics of Myxobacterial Chemosensory Systems, Journal of Bacteriology, 200, 2018.
[77] D.E. Whitworth, P.J. Cock, Two-component systems of the myxobacteria: structure, diversity and evolutionary relationships, Microbiology (Reading, England), 154 (2008) 360–372.
[78] D.E. Whitworth, Genome-wide analysis of myxobacterial two-component systems: genome relatedness and evolutionary changes, BMC Genomics 16 (2015) 780.
[79] Z. Sarwar, A.G. Garza, Two-component signal transduction systems that regulate the temporal and spatial expression of Myxococcus xanthus sporulation genes, J. Bacteriol. 198 (2016) 377–385.
[80] L. Kroos, The Bacillus and Myxococcus developmental networks and their transcriptional regulators, Annu. Rev. Genet. 41 (2007) 13–39.
[81] S. Huntley, N. Hamann, S. Wegener-Feldbrugge, A. Treuner-Lange, M. Kube, R. Reinhardt, S. Klages, R. Muller, C.M. Ronning, W.C. Nierman, L. Sogaard-Andersen, Comparative genomic analysis of fruiting body formation in Myxococcales, Mol. Biol. Evol. 28 (2011) 1083–1097.
[82] M. Orlowski, P. Martin, D. White, M.C. Wong, Changes in activity of glyoxylate cycle enzymes during myxospore development in Myxococcus xanthus, J. Bacteriol. 111 (1972) 784–790.
[83] J. Bland, W.K. Yeh, D. White, A. Hendricks, Increase in glyoxylate shunt enzymes during cellular morphogenesis in Myxococcus xanthus, Can. J. Microbiol. 17 (1971) 209–211.
[84] B.S. Goldman, W.C. Nierman, D. Kaiser, S.C. Slater, A.S. Durkin, J.A. Eisen, C. M. Ronning, W.B. Barbazuk, M. Blanchard, C. Field, C. Halling, G. Hinkle, O. Iartchuk, H.S. Kim, C. Mackenzie, R. Madupu, N. Miller, A. Shvartsbeyn, S. A. Sullivan, M. Vaudin, R. Wiegand, H.B. Kaplan, Evolution of sensory complexity recorded in a myxobacterial genome, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 15200–15205.