D640–D645 *Nucleic Acids Research, 2016, Vol. 44, Database issue* *Published online 17 October 2015*
doi: 10.1093/nar/gkv1050

CauloBrowser: A systems biology resource for *Caulobacter crescentus*

Keren Lasker[1,*,†], Jared M. Schrader[1,†], Yifei Men[2], Tyler Marshik[1], David L. Dill[2], Harley H. McAdams[1] and Lucy Shapiro[1]

[1]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA and [2]Department of Computer Science, Stanford University, Stanford, CA 94305, USA

Received August 18, 2015; Revised September 29, 2015; Accepted October 01, 2015

ABSTRACT

***Caulobacter crescentus* is a premier model organism for studying the molecular basis of cellular asymmetry. The *Caulobacter* community has generated a wealth of high-throughput spatiotemporal databases including data from gene expression profiling experiments (microarrays, RNA-seq, ChIP-seq, ribosome profiling, LC-ms proteomics), gene essentiality studies (Tn-seq), genome wide protein localization studies, and global chromosome methylation analyses (SMRT sequencing). A major challenge involves the integration of these diverse data sets into one comprehensive community resource. To address this need, we have generated *CauloBrowser* (www.caulobrowser.org), an online resource for *Caulobacter* studies. This site provides a user-friendly interface for quickly searching genes of interest and downloading genome-wide results. Search results about individual genes are displayed as tables, graphs of time resolved expression profiles, and schematics of protein localization throughout the cell cycle. In addition, the site provides a genome viewer that enables customizable visualization of all published high-throughput genomic data. The depth and diversity of data sets collected by the *Caulobacter* community makes *CauloBrowser* a unique and valuable systems biology resource.**

INTRODUCTION

The bacterium *Caulobacter crescentus* is a valuable model system for cell cycle control in which the transcriptional genetic circuitry is interwoven with the 3D deployment of regulatory and structural proteins (1,2). Each cell division in *Caulobacter* is asymmetric and produces daughter cells with distinct morphology and cell fate. *Caulobacter* cell cycle regulation involves far more than transcriptional networks, and includes differential chromosome methylation, transcription factor activation by spatially-restricted phosphosignaling pathways, temporally and spatially-controlled proteolysis, differential translation, post-translational control of gene expression, subcellular localization of regulatory proteins, and the dynamic topology of the cell; all of which are integrated into a robust regulatory network (3–5).

For the past 20 years, the *Caulobacter* community has taken a systems-based approach to study the molecular mechanisms driving the *Caulobacter* cell cycle. These efforts have generated a wealth of high-throughput time-resolved data gathered at multiple times in the cell cycle; including data from 132 microarray experiments, 13 RNAseq experiments, 8 ribosome profiling experiments, 9 LC-MS mass spectrometry experiments, 7 ChIP-seq experiments revealing hundreds of transcription factor binding sites, a Tn-seq experiment delimiting all essential open reading frames (ORFs) and non-coding elements, 438 global protein subcellular localization experiments, and global chromosomal methylation state throughout the cell cycle.

Given this wealth of data, a major challenge involves the integration of these diverse spatiotemporal data sets into one comprehensive official community resource. To address this need, we have generated *CauloBrowser* (www.caulobrowser.org), an online database that serves as an informatics hub composed of all published experimental data derived from global experiments. This site provides a user-friendly interface for quickly searching genes of interest and downloading genome-wide results. Individual genes are displayed with graphs of time resolved expression profiles for transcription at all promoters and translation at every ORF throughout the *Caulobacter* cell cycle. In addition, the site provides a genome viewer that enables customizable visualization of all available high-throughput genomic data.

*To whom correspondence should be addressed. Tel: +1 650 725 7657; Fax: +1 650 725 7739; Email: kerenl@stanford.edu
†These authors contributed equally to the paper as the first author.
Present address: Jared M. Schrader, Department of Biological Sciences, Wayne State University, Detroit, MI, USA.

CAULOBROWSER

CauloBrowser start page allows users to search for a gene by its standard name (e.g. *ctr*A) or by its systematic nomenclature (e.g. CCNA_03130 or CC_3035). The user can also search for all genes positioned within a continuous genomic region (e.g. genomic coordinates 1 to 10 000). When specific genes (or a genomic region) have been selected the user is guided to the results page (Figures 1 and 2). The results page is composed of five sections. The first section ('*Overview*') provides general information about the function of a gene, the regulation of its expression and the localization pattern of its protein product (if known). The second section ('*Time resolved gene expression of wild-type cells*') draws from microarray, RNA-seq, ribosome profiling, and proteomics experiments to provide transcriptional and post-transcriptional plots of gene expression across the cell cycle. The third section ('*Gene expression collection*') describes differential expression of genes of interest in a variety of growth conditions (e.g. carbon starvation) and mutant backgrounds (e.g. cell cycle regulatory protein depletions). The fourth section ('*Time resolved* gene expression *collection*') provides plots of gene expression across the cell cycle in different growth conditions (e.g. PYE media) and mutant backgrounds. The fifth section ('*References*') provides a list of relevant papers. In addition, *CauloBrowser* provides an interactive genome visualization tool that includes tracks for all published genome-scale data.

OVERVIEW SECTION

The overview section provides a table composed of four categories listing curated information per gene (Figure 1). The first category ('Gene') includes the following data. 'Name' provides the standard name of the gene, the CCNA locus ID (based on the genome of the NA1000 laboratory strain) (6,7) as well as the CC locus ID (based on the sequence of the closely related CB15 genome) (8). 'Genomic coordinates' denotes the protein coding nucleotides of the gene. The ORF coordinates reflect a recent major reannotation (NCBI accession number CP001340.1) (7). 'Transcriptional control' lists the experimentally identified transcriptional start sites of the gene as well as any known cell cycle-regulated transcription factor binding sites as determined in (9). 'Essentiality' indicates whether the gene was found to be disruptable (i.e. dispensable) or non-disruptable (i.e. essential) for growth in a recent saturating transposon mutagenesis experiment (10). 'Additional information' provides corresponding links to the central databases KEGG (11), BioCyc (12) and NCBI (13).

The second category ('*Product*') includes the following data. '*Function*' describes the function of the gene product based on NCBI prediction. '*Identified by mass spec*' indicates whether the gene product was identified by LC-MS mass spectrometry in normal laboratory growth conditions (7,14). '*Proteolysis*' indicates whether the gene product is a substrate of ClpP associated proteases (15) and whether it is a substrate of tmRNA associated proteases (16). 'Additional information' provides a link to the corresponding UniProt entry (17).

The third category ('*Localization*') summarizes the community knowledge on the subcellular localization of the protein. First, the protein is classified as either cytosolic, cytoplasmic-membrane-associated, extracellular, outer-membrane-associated or periplasmic based on PSORTdb, a subcellular localization database for Bacteria and Archaea (18). Second, we provide a cartoon of the protein subcellular localization across the cell cycle. If the subcellular localization pattern is known, usually by observing a fluorescent protein fusion at multiple points in the cell cycle, we draw a corresponding cartoon and cite the relevant paper. If the subcellular localization pattern is not known we provide a predicted cartoon. When the protein is part of stable protein complex we use a known localization pattern of a member of the complex if available. In all other cases we use the predicted PSORTdb localization. We also provide links to images showing the subcellular localization of the protein tagged with a fluorescent protein. These images are a stored in the *Caulobacter* localizome repository, a comprehensive database of protein localization images from a genome-scale imaging screen involving N- and C-terminal fluorescent protein fusions (19).

The fourth category ('*Actions*') provides two action buttons. Clicking on 'Go to genome viewer' links to our dedicated genome data browser ('*Genome viewer*' section). Clicking on 'Remove gene' removes the gene from the display.

TIME RESOLVED GENE EXPRESSION IN WILD TYPE CELLS SECTION

This section provides up to eight time resolved expression graphs per gene. The graphs display individual promoter activity (9), RNAseq and DNA microarrays (20–23)(Schrader *et al.* in preparation), ribosome profiling (Schrader *et al.* in preparation), and proteomics data (24) (Figure 2). The data sets displayed in the top three graphs have been collected and analyzed using the new annotation of the *Caulobacter* genome (7). The RNA-seq, tiling arrays, and ribosome profiling data sets have been normalized such that levels can be compared between genes. The *CauloBrowser* user interface is suited to comparing expression data between multiple genes. All genes searched by the user are displayed in color-coding on the same graph making it easy to perform gene-by-gene comparisons. The graphs provide a toggle (by simply clicking on the gene in the legend) to hide and show the selected genes.

GENE EXPRESSION COLLECTION SECTION

This section lists a collection of 35 microarrays and 4 proteomics experiments from various different growth conditions and/or strains containing mutations (Figure 2). The table reports the $\log_2$ ratio between the mRNA or protein levels between the indicated conditions. Two of the data sets show proteins identified as substrates to ClpP protease or substrates of the tmRNA ribosomal rescue system.

CAULOBACTER RESEARCH | Home | CauloBrowser | *Caulobacter* Labs | Help | admin

- Click on gene names to view more info on the gene from KEGG.
- Hover over a point on one of the graphs with the cursor to view more info.
- Click on a gene chart legend to toggle its visibility.
- Zoom in on a graph by scrolling when the cursor is over the graph. Right click on the graph to return it to the default view.

Search Additional Gene: Add Gene

Download csv

1. Overview

Help

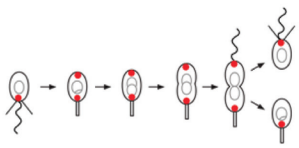
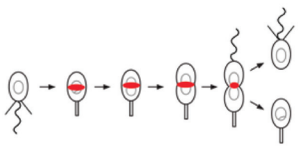
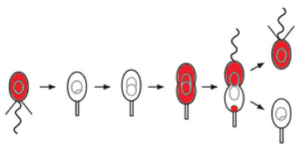
		popZ	*ftsZ*	*ctrA*
Gene	**Name**	*popZ*, CCNA_01380, CC_1319	*ftsZ*, CCNA_02623, CC_2540	*ctrA*, CCNA_03130, CC_3035
	Genomic coordinates	1494812-1495345, forward	2771579-2773105, reverse	3278952-3279647, reverse
	Transcriptional control	P1: RpoD, +1494759	P1: DnaA, CcrM, GcrA, -2773216	P1: RpoD, CtrA_half_repressor, SciP_half_motif, CcrM, GcrA, -3279769
		P2: +1494784		P2: CtrA_full_motif, GcrA, -3279712
		P3: +1494793		P3: CtrA_full_motif, SciP_half_motif, GcrA, -3279737
	Essentially	Essential	Essential	Essential
	Additional information	KEGG, BioCyc, NCBI	KEGG, BioCyc, NCBI	KEGG, BioCyc, NCBI
Product	**Function**	pole-organizing protein popZ	cell division protein FtsZ	cell cycle response regulator ctrA
	Identified by mass spec	Yes	Yes	Yes
	Proteolysis	Not a substrate of ClpP associated proteases Not a substrate of tmRNA associated proteases	A substrate of ClpP associated proteases Not a substrate of tmRNA associated proteases	A substrate of ClpP associated proteases Not a substrate of tmRNA associated proteases
	Additional information	UniProt	UniProt	UniProt
Localization	**PSORTdb prediction**	Cytoplasmic protein	Cytoplasmic protein	Cytoplasmic protein
	Cell Cycle			
		Bowman et al Cell 2008, Ebersbach et al Cell 2008	Goley et al Mol Micro 2011	Ryan et al PNAS 2004
	Localisome			N-terminal fusion
Actions		Go to genome viewer	Go to genome viewer	Go to genome viewer
		Remove gene	Remove gene	Remove gene

Figure 1. Screenshot of *CauloBrowser* overview section. The figure shows the overview section search results for genes *ctrA, ftsZ and popZ*. The results are organized as a table with a row per data category and a column per gene. Links to central databases such as KEGG, BioCyc, NCBI, and UniProt are colored in light blue. The 'Actions' row provides two buttons. 'Go to genome viewer' links to our *Caulobacter* genome viewer page zoomed to the region of the corresponding ORF. 'Remove gene' removes the gene from the display.

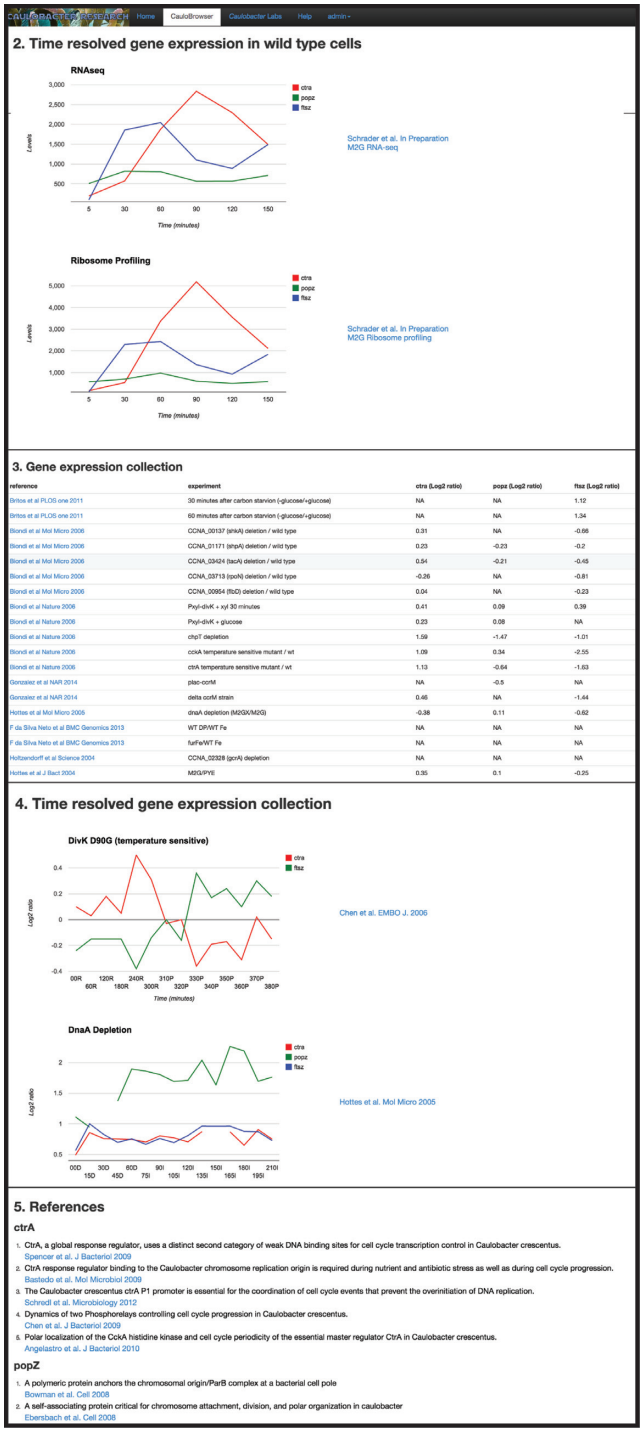


Figure 2. Screenshot of *CauloBrowser sections 2–5.* The figure shows sections 2–5 search results for genes *ctrA* (red), *ftsZ* (blue), and *popZ* (green). Time resolved expression data (sections 2,4) is visualized using plots of gene expression across the cell cycle in wild type or different growth conditions and mutant background. *Gene expression collection data (section 3)* is organized as a table with a row per experiment and a column per gene. Relevant PubMed references (section 5) are listed with links to corresponding PubMed entries.

TIME RESOLVED GENE EXPRESSION COLLECTION SECTION

This section provides graphs for a collection of additional cell cycle microarrays performed on mutant strains and/or in various different growth conditions. These experiments track mRNA levels at the indicated time points across the cell cycle (Figure 2).

REFERENCES SECTION

This final section lists per gene of interest relevant papers (Figure 2). The list is gathered from the gene-centered information resource at NCBI (13). For some genes, e.g. *popZ*, the list of paper generated by NCBI is incomplete and missing key papers. In such cases we manually expanded the list of papers displayed in *CauloBrowser*. We are constantly updating the list of relevant papers per gene and invite the community to participate in this effort.

NON-CODING RNA

The user can search for a non-coding RNA by its systematic nomenclature (e.g. CCNA_R0001 or CC_t01). The results page is composed of five sections as described above. However, many of the result fields are empty either because they are not relevant or because the corresponding data were not collected. The data displayed for a non-coding RNA includes fields in the overview table and time resolved RNAseq data.

GENOME VIEWER

CauloBrowser contains a genomic data viewer to visualize multiple genome wide data sets (tracks) on the *Caulobacter* chromosome. When loading the genome viewer two default tracks appear. The Genome track displays all ORFs and the NA1000 operons track displays the *Caulobacter* operon structures based upon 5′ Global RACE and RNA-seq data sets (7,9). The user can toggle the visible data sets using the + button. Additional tracks include data gathered from ChIP-seq (25–27), RNA-seq (7,9,20), ribosome profiling (7), and Tn-seq experiments (10) (Figure 3). The user can configure the track display options using the track button, export an image of the data using the printer button, change browsing options using the gear button, and find help using the question mark button. Users can also upload their own data sets using the + button.

IMPLEMENTATION

CauloBrowser (www.caulobrowser.org) web interface was developed using python, HTML, CSS, JavaScript, Google chart tools, and runs on an Apache HTTP server (version 2.4.7) hosted on an Ubuntu Linux server (version 14.04.2 LTS). The *CauloBrowser* database is stored on a MySQL server (version 5.3.43). The Genome viewer was developed based on Dalliance (28).

OUTLOOK

CauloBrowser is a comprehensive resource for the study of *Caulobacter crescentus* cell cycle as a paradigm for asymmetric cell division. *CauloBrowser* allows the user to quickly

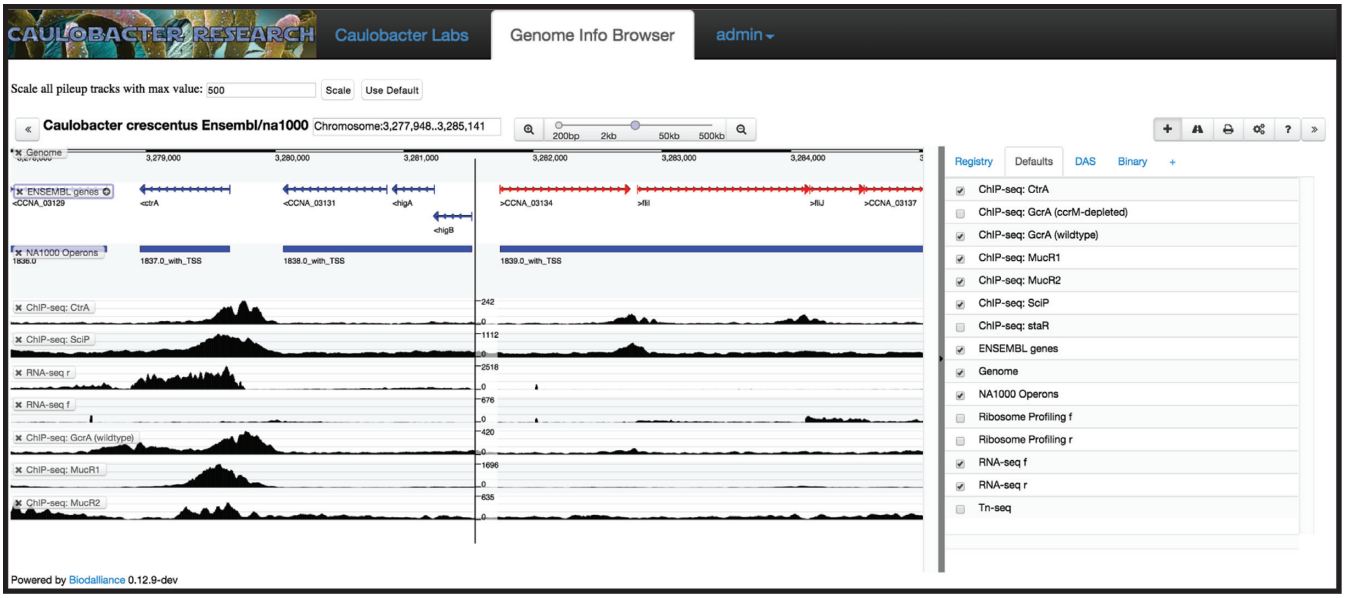


Figure 3. Screenshot of the genome viewer. The figure shows our genomic viewer zoomed to genomic coordinates 3 277 948–3 285 141. Nine tracks are open demonstrating the power of simultaneously visualizing data from different experiments. The top ENSEMBL track shows NA1000 genome annotation with forward genes in red and reverse genes in blue. The second track shows the *Caulobacter* operon structure. Seven additional tracks are shown visualizing multiple Chip-seq and RNS-seq experiments. A list of all available data sets is shown to the right.

search genes of interest and browse all published high throughput experiments. We encourage members of the community to take an active role in the development of the portal by providing feedback, suggesting new functionalities as well as additional data sets to visualize. We plan to constantly incorporate new experimental data and maintain the database for at least 5 years. *CauloBrowser* is designed to provide easy access to the multitude of diverse *Caulobacter* data sets. This resource is a valuable tool for the systems level understanding of the spatiotemporal program that controls the *Caulobacter* cell cycle.

ACKNOWLEDGEMENTS

We thank Dante Ricci for critical comments. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

FUNDING

Gordon and Betty Moore Foundation [GBMF 2550.03 to Life Sciences Research Foundation to K.L.] and the Weizmann Institute of Science National Postdoctoral Award Program for Advancing Women in Science [to K.L.]; NIH postdoctoral fellowship [F32 GM100732 to J.M.S.]; NIH [R01 GM51426 and R01 GM32506 to L.S.]; NSF [CCF 1344284 to H.H.M. and D.L.D.] Funding for open access charge: NIH [R01 GM51426 and R01 GM32506].
Conflict of interest statement. None declared.

REFERENCES

1. Laub,M.T., Shapiro,L. and McAdams,H.H. (2007) Systems biology of Caulobacter. *Annu. Rev. Genet.*, **41**, 429–441.
2. McAdams,H.H. and Shapiro,L. (2009) System-level design of bacterial cell cycle control. *FEBS Lett.*, **583**, 3984–3991.
3. Curtis,P.D. and Brun,Y.V. (2010) Getting in the loop: regulation of development in Caulobacter crescentus. *Microbiol. Mol. Biol. Rev.*, **74**, 13–41.
4. McAdams,H.H. and Shapiro,L. (2011) The architecture and conservation pattern of whole-cell control circuitry. *J. Mol. Biol.*, **409**, 28–35.
5. Tsokos,C.G. and Laub,M.T. (2012) Polarity and cell fate asymmetry in *Caulobacter crescentus*. *Curr. Opin. Microbiol.*, **15**, 744–750.
6. Marks,M.E., Castro-Rojas,C.M., Teiling,C., Du,L., Kapatral,V., Walunas,T.L. and Crosson,S. (2010) The genetic basis of laboratory adaptation in *Caulobacter crescentus*. *J. Bacteriol.*, **192**, 3678–3688.
7. Schrader,J.M., Zhou,B., Li,G.W., Lasker,K., Childers,W.S., Williams,B., Long,T., Crosson,S., McAdams,H.H., Weissman,J.S. *et al.* (2014) The coding and noncoding architecture of the Caulobacter crescentus genome. *PLoS Genet.*, **10**, e1004463.
8. Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J.A., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R. *et al.* (2001) Complete genome sequence of *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4136–4141.
9. Zhou,B., Schrader,J.M., Kalogeraki,V.S., Abeliuk,E., Dinh,C.B., Pham,J.Q., Cui,Z.Z., Dill,D.L., McAdams,H.H. and Shapiro,L. (2015) The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLoS Genet.*, **11**, e1004831.
10. Christen,B., Abeliuk,E., Collier,J.M., Kalogeraki,V.S., Passarelli,B., Coller,J.A., Fero,M.J., McAdams,H.H. and Shapiro,L. (2011) The essential genome of a bacterium. *Mol. Syst. Biol.*, **7**, 528.
11. Kanehisa,M., Goto,S., Sato,Y., Kawashima,M., Furumichi,M. and Tanabe,M. (2014) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.*, **42**, D199–D205.
12. Caspi,R., Altman,T., Billington,R., Dreher,K., Foerster,H., Fulcher,C.A., Holland,T.A., Keseler,I.M., Kothari,A., Kubo,A. *et al.* (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, **42**, D459–D471.
13. Brown,G.R., Hem,V., Katz,K.S., Ovetsky,M., Wallin,C., Ermolaeva,O., Tolstoy,I., Tatusova,T., Pruitt,K.D., Maglott,D.R. *et al.* (2015) Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.*, **43**, D36–D42.
14. Britos,L., Abeliuk,E., Taverner,T., Lipton,M., McAdams,H. and Shapiro,L. (2011) Regulatory response to carbon starvation in Caulobacter crescentus. *PLoS One*, **6**, e18179.

15. Bhat,N.H., Vass,R.H., Stoddard,P.R., Shin,D.K. and Chien,P. (2013) Identification of ClpP substrates in Caulobacter crescentus reveals a role for regulated proteolysis in bacterial development. *Mol Microbiol*, **88**, 1083–1092.
16. Hong,S.J., Lessner,F.H., Mahen,E.M. and Keiler,K.C. (2007) Proteomic identification of tmRNA substrates. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 17128–17133.
17. The UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
18. Yu,N.Y., Laird,M.R., Spencer,C. and Brinkman,F.S. (2011) PSORTdb–an expanded, auto-updated, user-friendly protein subcellular localization database for Bacteria and Archaea. *Nucleic Acids Res.*, **39**, D241–D244.
19. Werner,J.N., Chen,E.Y., Guberman,J.M., Zippilli,A.R., Irgon,J.J. and Gitai,Z. (2009) Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 7858–7863.
20. Fang,G., Passalacqua,K.D., Hocking,J., Llopis,P.M., Gerstein,M., Bergman,N.H. and Jacobs-Wagner,C. (2013) Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. *BMC Genomics*, **14**, 450.
21. Laub,M.T., McAdams,H.H., Feldblyum,T., Fraser,C.M. and Shapiro,L. (2000) Global analysis of the genetic network controlling a bacterial cell cycle. *Science*, **290**, 2144–2148.
22. McGrath,P.T., Lee,H., Zhang,L., Iniesta,A.A., Hottes,A.K., Tan,M.H., Hillson,N.J., Hu,P., Shapiro,L. and McAdams,H.H. (2007) High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. *Nat. Biotechnol.*, **25**, 584–592.
23. Hottes,A.K., Shapiro,L. and McAdams,H.H. (2005) DnaA coordinates replication initiation and cell cycle transcription in Caulobacter crescentus. *Mol. Microbiol.*, **58**, 1340–1353.
24. Grunenfelder,B., Rummel,G., Vohradsky,J., Roder,D., Langen,H. and Jenal,U. (2001) Proteomic analysis of the bacterial cell cycle. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 4681–4686.
25. Fiebig,A., Herrou,J., Fumeaux,C., Radhakrishnan,S.K., Viollier,P.H. and Crosson,S. (2014) A cell cycle and nutritional checkpoint controlling bacterial surface adhesion. *PLoS Genet.*, **10**, e1004101.
26. Fioravanti,A., Fumeaux,C., Mohapatra,S.S., Bompard,C., Brilli,M., Frandi,A., Castric,V., Villeret,V., Viollier,P.H. and Biondi,E.G. (2013) DNA binding of the cell cycle transcriptional regulator GcrA depends on N6-adenosine methylation in *Caulobacter crescentus* and other Alphaproteobacteria. *PLoS Genet.*, **9**, e1003541.
27. Fumeaux,C., Radhakrishnan,S.K., Ardissone,S., Theraulaz,L., Frandi,A., Martins,D., Nesper,J., Abel,S., Jenal,U. and Viollier,P.H. (2014) Cell cycle transition from S-phase to G1 in *Caulobacter* is mediated by ancestral virulence regulators. *Nat. Commun.*, **5**, 4081.
28. Down,T.A., Piipari,M. and Hubbard,T.J. (2011) Dalliance: interactive genome viewing on the web. *Bioinformatics*, **27**, 889–890.