SubtiWiki—a database for the model organism
Bacillus subtilis that links pathway, interaction and expression information

Raphael H. Michna, Fabian M. Commichau, Dominik Tödter, Christopher P. Zschiedrich and Jörg Stülke*

Department of General Microbiology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Grisebachstrasse 8, D-37077 Göttingen, Germany

Received September 13, 2013; Revised October 3, 2013; Accepted October 4, 2013

ABSTRACT

Genome annotation and access to information from large-scale experimental approaches at the genome level are essential to improve our understanding of living cells and organisms. This is even more the case for model organisms that are the basis to study pathogens and technologically important species. We have generated SubtiWiki, a database for the Gram-positive model bacterium Bacillus subtilis (http://subtiwiki.uni-goettingen.de). In addition to the established companion modules of SubtiWiki, SubtiPathways and SubtiInteract, we have now created SubtiExpress, a third module, to visualize genome scale transcription data that are of unprecedented quality and density. Today, SubtiWiki is one of the most complete collections of knowledge on a living organism in one single resource.

INTRODUCTION

The Gram-positive soil bacterium Bacillus subtilis is one of the best-characterized organisms. It serves as a model organism for important Gram-positive pathogens such as Bacillus anthracis, Listeria monocytogenes or Staphylococcus aureus. Moreover, B. subtilis and its relatives (Bacillus licheniformis and the lactic acid bacteria) are intensively used in biotechnology for commercial production purposes. In the past few years, B. subtilis has been in the focus of several large systems biology projects. These projects, as well as the continuous molecular genetics and biochemical research, have considerably increased our knowledge on B. subtilis. This importance of B. subtilis and the substantial increase of knowledge make the easy and structured access to all information concerning B. subtilis an important task. For a long time, the database SubtiList (1) has been the primary source of rapid information on B. subtilis genes and proteins; unfortunately, SubtiList has not been updated since 2001.

To provide the community with up-to-date information, we created the Wiki-based data source SubtiWiki (1). The initial information listed in SubtiWiki was derived from the SubtiList database as well as from the research as it is published. Therefore, both the scope and information content in SubtiWiki by far exceeded SubtiList already in the first release (2). The Wiki kind of data collection makes it possible to use the collective expertise, as each qualified member of the Bacillus community can contribute to SubtiWiki. The Wiki is accessible to anyone interested in any specific aspect related to B. subtilis and other Gram-positive bacteria.

SubtiWiki was originally a collection of pages providing inter-linked information for each gene of B. subtilis. In the meantime, it has been substantially expanded by adding new modules devoted to the presentation of metabolism and regulation (SubtiPathways) (3) and to the protein–protein interactions (SubtiInteract) (4).

The system-level analyses performed during the recent years provided a wealth of information on gene expression. Therefore, we have developed SubtiWiki, with a specific focus on the presentation of gene expression annotation. For this purpose, we have created a novel module of SubtiWiki, SubtiExpress. In addition, the gene pages have been updated based on new information from the publications.

In this work, we describe the current state of SubtiWiki with emphasis on the new features: (i) SubtiExpress for the intuitive visualization of transcriptome profiling data and (ii) the significant revision of the lists of essential and sporulation genes.

THE PAGES FOR INDIVIDUAL GENES, PROTEINS AND RNAs

The central component of SubtiWiki is the individual pages for each gene that provide all the available...
information on both the gene and its product, usually a protein, sometimes an RNA. Since the last report on SubtiWiki (4), the pages have been continuously updated, and several new features have been introduced on these pages, most strikingly the information on gene expression.

The SubtiWiki gene page for eno (encoding enolase) is shown in Figure 1. To the table at the top of the page, a link to the gene expression module SubtiExpress (see below) has been added. Moreover, an image giving an immediate overview on the expression of the gene under 104 different conditions was inserted. Finally, the links to the DNA sequence (both the gene and the gene with the adjacent region) and the protein sequence have been updated. Below the table, the functional categories and regulons for the gene/protein are shown. This allows immediate access to all related genes or proteins that are members of the same category or regulon. The next sections contain the information about the gene and the protein as well as on gene expression. In the expression part, links to the new expression module SubtiExpress were added. At the bottom of the page, the references for the gene/protein/RNA are listed.

The regular updates of SubtiWiki had significant consequences: 130 genes previously assigned to the category ‘Protein of unknown function’ could be deleted from the list since the last update. Moreover, 13 newly discovered regulons have been added, and for one previously suspected regulator (LytR), experimental research has revealed that the protein is not a regulator but is involved in the attachment of anionic polymers to peptidoglycan, and the protein was accordingly renamed TagU (5). In total, 62 proteins were given new designations in the past 2 years; accordingly, the names of the corresponding pages were changed in a way that both the old and the new names guide the user to the most

Figure 1. The eno gene page in SubtiWiki. Pages for each gene of B. subtilis exist in SubtiWiki. The table at the top of the page provides the most important information including links to the modules SubtiExpress, SubtiInteract and SubtiPathways. The image with the expression profile is clickable and leads the user to the SubtiExpress page of the gene (see Figures 3 and 4).
updated page (to track changed designations, the move log of SubtiWiki can be consulted: http://subtiwiki.uni-goettingen.de/wiki/index.php?title=Special%3ALog&type=move).

In the previous update, we have introduced the functional categories and regulon lists. In the meantime, these lists have proven to be valuable instruments to support experimental research (6).

ESSENTIAL GENES

Recently, biology has developed a major interest in determining the set of genes that is required to sustain a living cell. The basis of this set is the essential genes of an organism. However, the essential genes are not sufficient to provide all necessary biochemical activities, as many essential metabolites can be obtained in different ways; e.g. the availability of all proteinogenic amino acids is essential for any living cell but only two genes involved in amino acid metabolism (glyA and merK) are essential in B. subtilis. This is due to the fact that amino acids can be obtained in two different ways: (i) by biosynthesis or (ii) by uptake from the medium. Thus, either biosynthetic enzymes or amino acid transporters have to be present in a minimal cell, although none of the individual genes is essential. Nonetheless, the essential genes are the basis for all projects to construct minimal organisms. For B. subtilis, the set of essential genes was first studied in 2003 (7), and the functional category ‘Essential genes’ had been introduced in 2010 in SubtiWiki to provide an easy overview on these important genes. In the meantime, we have re-analyzed the set of essential B. subtilis genes, both by experimental studies and by data mining from the literature. Interestingly, it turned out that many ribosomal proteins are not essential in B. subtilis (8). Similarly, and in contrast to the previous report, several glycolytic enzymes were found to be dispensable (9). This finding is in good agreement with metabolic models that do not see any need for glycolytic enzymes as long as intermediates from both ends of the pathway are available (such as glucose and malate as an intermediate of the citric acid cycle) (10). Recently, the odhB gene encoding the E2 subunit of the 2-oxoglutarate dehydrogenase, as well as the gene pairs ydhT and yomL, and yloU and yphY were found to be dispensable as well (our unpublished results).

The dynamics of the list of essential genes is not only due to the fact that 36 genes previously said to be essential were shown to be actually dispensable; in addition, 20 novel essential genes have been discovered in the past 10 years. All this information is now covered in the updated SubtiWiki page for the category ‘Essential genes’ (http://subtiwiki.uni-goettingen.de/wiki/index.php/Essential_genes, Figure 2). As of September 2013, 254 protein-coding genes of B. subtilis are regarded as being essential. The largest group of these proteins is required for protein synthesis, secretion and quality control; large sets are also involved in metabolism and cell division. Owing to the intensive research on B. subtilis, only two essential proteins are of unknown function. The regularly updated list of essential proteins will be an important foundation for efforts to create a minimal cell based on B. subtilis.

SubtiPathways and SubtiInteract–TWO RESOURCES THAT COMPLEMENT SubtiWiki

To enhance the properties of SubtiWiki, we have created accompanying Web sites that provide additional sets of information. Two of these modules, SubtiPathways and SubtiInteract, are devoted to the presentation of metabolic and regulatory pathways and of protein–protein interactions, respectively.

SubtiPathways is based on published models and experimental data on the metabolism and gene regulation in B. subtilis. As of autumn 2013, SubtiPathways encompasses 35 diagrams covering many aspects of B. subtilis physiology, development and regulation. The diagrams are extensively linked to the SubtiWiki pages of the relevant genes and, on the other hand, each gene page contains a link to the specific diagram, if available.

The protein–protein interactions in B. subtilis are displayed in the module SubtiInteract. There are two complementary presentations: a genome-scale model was created using Cytoscape and converted to a zoomable and clickable map using the Google maps API and the program CellPublisher (11). In addition, specific pages for each protein show the respective interactions at the first (primary interactions) and second levels (interactions of the primary partners), up to the fourth level. This allows getting a picture of the whole interaction network of any given protein. As described above for SubtiPathways, all proteins shown in SubtiInteract are labeled with clickable markers to link them to their SubtiWiki pages. Moreover, a click on any protein of an interaction network will re-center the presentation around this protein. By September 2013, SubtiInteract contains 2055 interactions involving 917 different proteins and 5 RNAs. This corresponds to 225 novel interactions and 116 novel proteins participating in interactions since the last report on SubtiWiki. The protein-specific interaction networks are directly accessible from the table on the top of the gene pages (Figure 1).

SubtiExpress–A PRESENTATION OF GENE EXPRESSION IN B. subtilis

As most other bacteria, B. subtilis is capable of adapting to a variety of changing environments. This is the case for all nutrients including carbon and nitrogen sources and also for different stress conditions such as temperature or osmotic stress. The main mechanism of this adaptation is the changes in gene expression to ensure that under any specific condition precisely those genes are expressed that allow most rapid growth under the given circumstances. Thus, knowledge of gene expression is central to our understanding of the biology of an organism. Recently, two European consortia have investigated the transcriptome of B. subtilis under 104 different conditions using tiling arrays (6). Moreover, data from a chronotranscriptome analysis that cover gene expression during
growth in a complex medium have become available (12). This depth of information on gene expression is unprecedented and has no match for any other organism. Unfortunately, there was no easy and intuitive access to the data of these studies. Therefore, we created a third module to accompany SubtiWiki, SubtiExpress. SubtiExpress provides pages for each individual gene that can be accessed from the respective SubtiWiki pages. Moreover, an overview on the gene expression profile was integrated on the SubtiWiki gene pages (Figure 1).

The SubtiExpress pages contain three important sets of information. First, they present the expression of the gene under the 104 different conditions. To make the images more easily accessible, we have defined a frame that allows to mouse over the image and to see the specific experimental condition once the mouse is situated at a data point. Additionally, clicking on the data point will give access to a new page that shows the precise experimental conditions used to study the expression levels (Figure 3). In addition, these pages for each condition also contain a list with all the genes that are highly expressed under the given condition. As usual, the genes in this list are linked to their respective SubtiWiki pages. The second part of the SubtiExpress pages shows the results from the chronotranscriptome study, i.e. during growth in a rich medium at 10-min intervals. Finally, the bottom part of the pages presents the genomic organization of the gene and the transcriptional landscape as deduced from the high-density tiling array experiment (transcriptional intensities, start and stop signals) for the two DNA strands (Figure 4). At the top of the gene-specific expression pages, there are links to the corresponding SubtiWiki and SubtiInteract pages as well as to the page of the original expression browser from which the information was imported to SubtiExpress (Figure 3).

### NOVEL SPORULATION GENES

In the past few years, several genomic and transcription factor-specific transcriptome analyses of the genes involved in sporulation have been published (13).
Importantly, Galperin et al. (14) have studied the minimal set of sporulation-specific genes in bacilli and clostridia. The recent transcriptome analysis under 104 defined conditions confirmed sporulation genes identified in earlier transcriptome studies that were, however, not included in the regulation database DBTBS (15). Moreover, this study identified many novel genes that are specifically expressed under sporulation conditions (6). For most of these new sporulation genes, no function had been known before. However, some metabolic proteins, such as a citrate transporter and the alternative citrate synthase CitA, are specifically expressed under sporulation conditions, suggesting a role for citrate metabolism during sporulation. Based on transcription profiling, 145 new proteins were found to be functionally related to sporulation. These new functional assignments demonstrate how the integration of data in one central tool helps to accomplish the central claim of functional genomics, i.e. to provide indications for the functions of so far unknown proteins. For several of the newly discovered sporulation proteins, their implication in sporulation has recently been experimentally demonstrated (16–18). The annotation of sporulation genes has recently been significantly enhanced by the creation of SporeWeb, an interactive knowledge platform about the sporulation cycle of *B. subtilis* (12). The information in SporeWeb is linked to gene-specific pages of SubtiWiki to facilitate the interaction between the two databases.

**PERSPECTIVES**

SubtiWiki (with SubtiPathways, SubtiInteract and SubtiExpress) has become one of the most complete inventories of knowledge on a living organism in one single resource. The continuous updates and the novel features both contribute to its popularity, which is reflected by >2.5 million page visits during the past 12 months. This corresponds to a 250% increase in the past 2 years.

In the future, keeping up-to-date with the forefront of research will remain a key task for the development of SubtiWiki. In addition, we will develop SubtiWiki to become a major foundation for genome minimization projects, and we will include broader information on protein localization.
ACKNOWLEDGEMENTS

The authors are grateful to Katrin Gunka, Victoria Keidel, Anika Klewing and Felix Mehne for their help with the characterization of essential genes.

FUNDING

The Federal Ministry of Education and Research SYSMO network and SFB860 (to J.S.) and the DFG [CO1139/1-1 to F.M.C.]. Funding for open access charge: DFG [SFB860 to J.S.].

Conflict of interest statement. None declared.

REFERENCES

1. Moszer,I., Jones,L.M., Moreira,S., Fabry,C. and Danchin,A. (2002) SubtiList: the reference database for the Bacillus subtilis genome. Nucleic Acids Res., 30, 62–65.
2. Flórez,L.A., Roppel,S.F., Schmeisky,A.G., Lammers,C.R. and Stülke,J. (2009) A community-curated consensual annotation that is continuously updated: the Bacillus subtilis centred wiki SubtWiki. Database, 2009, hapt012.
3. Lammers,C.R., Flórez,L.A., Schmeisky,A.G., Roppel,S.F., Mäder,U., Hamoen,L. and Stülke,J. (2010) Connecting parts with processes: SubtiWiki and SubtiPathways integrate gene and pathway annotation for Bacillus subtilis. Microbiology, 156, 849–859.
4. Mäder,U., Schmeisky,A.G., Flórez,L.A. and Stülke,J. (2012) SubtiWiki – a comprehensive community resource for the model organism Bacillus subtilis. Nucleic Acids Res., 40, D1278–D1287.
5. Kawai,Y., Marles-Wright,J., Cleverley,R.M., Emmins,R., Ishikawa,S., Kuvano,M., Heinz,N., Bui,N.K., Hoyland,C.N., Ogasawara,N. et al. (2011) A widespread family of bacterial cell wall assembly proteins. EMBO J., 30, 4931–4941.
6. Nicolas,P., Mäder,U., Dervyn,E., Rochat,T., Leduc,A., Pigeonneau,N., Bidenko,E., Marchadier,E., Hoebeke,M., Aymerich,S. et al. (2012) The condition-dependent whole-transcriptome reveals high-level regulatory architecture in bacteria. Science, 335, 1103–1106.
7. Kobayashi,K., Ehrlich,S.D., Albertini,A., Amati,G., Andersen,K.K., Arnaud,M., Asai,K., Ashikaga,S., Aymerich,S., Bessieres,P. et al. (2003) Essential Bacillus subtilis genes. Proc. Natl Acad. Sci. USA, 100, 4678–4683.
8. Akanuma,G., Namamiya,H., Natori,Y., Yano,K., Suzuki,S., Omata,S., Ishizuka,M., Sekine,Y. and Kawamura,F. (2012) Inactivation of ribosomal protein genes in Bacillus subtilis reveals
importance of each ribosomal protein for cell proliferation and cell differentiation. J. Bacteriol., 194, 6282–6291.
9. Commichau, F.M., Pietack, N. and Stülke, J. (2013) Essential genes in Bacillus subtilis: a re-evaluation after ten years. Mol. Biosyst., 9, 1068–1075.
10. Florez, L.A., Gunka, K., Polania, R., Tholen, S. and Stülke, J. (2011) SPABBATS: A pathway-discovery method based on Boolean satisfiability that facilitates the characterization of suppressor mutants. BMC Systems Biol., 5, 5.
11. Florez, L.A., Lammers, C.R., Michna, R., and Stülke, J. (2010) CellPublisher: a web platform for the intuitive visualization and sharing of metabolic, signaling and regulatory pathways. Bioinformatics, 26, 2997–2999.
12. Blom, E.J., Ridder, A.N., Lulko, A.T., Roerdink, J.B. and Kuipers, O.P. (2011) Time-resolved transcriptomics and bioinformatic analyses reveal intrinsic stress responses during batch culture of Bacillus subtilis. PLoS One, 6, e27160.
13. Eijlander, R.T., de Jong, A., Krawczyk, A.O., Holsapple, S. and Kuipers, O.P. (2014) SporeWeb: an interactive journey through the complete sporulation cycle of Bacillus subtilis. Nucleic Acids Res. (epub ahead of print).
14. Galperin, M.Y., Mekhekov, S.L., Puigbo, P., Smirnov, S., Wolf, Y.I. and Ridgen, D.J. (2012) Genomic determinants of sporulation in Bacilli and Clostridia: towards the minimal set of sporulation-specific genes. Environ. Microbiol., 14, 2870–2890.
15. Sierro, N., Makita, Y., de Hoon, M. and Nakai, K. (2008) DBTBS: a database of transcriptional regulation in Bacillus subtilis containing upstream intergenic conservation information. Nucleic Acids Res., 36, D93–D96.
16. Abecasis, A.B., Serrano, M., Alves, R., Quintais, L., Pereira-Leal, J.B. and Henriques, A.O. (2013) A genomic signature and the identification of new sporulation genes. J. Bacteriol., 195, 2101–2115.
17. Traag, B.A., Ramirez-Peralta, A., Wang Erickson, A.F., Setlow, P. and Losick, R. (2013) A novel RNA polymerase-binding protein controlling genes involved in spore germination in Bacillus subtilis. Mol. Microbiol., 89, 113–122.
18. Traag, B.A., Pugliese, A., Eisen, J.A. and Losick, R. (2013) Gene conservation among endospore-forming bacteria reveals additional sporulation genes in Bacillus subtilis. J. Bacteriol., 195, 253–260.