CARMEN - Comparative Analysis and *in silico* Reconstruction of organism-specific MEtabolic Networks

J. Schneider[1,2], F.-J. Vorhölter[3], E. Trost[2], J. Blom[1], Y.R. Musa[3], H. Neuweger[1], K. Niehaus[3], S. Schatschneider[3], A. Tauch[2] and A. Goesmann[1]

[1]Computational Genomics Group, Institute for Bioinformatics,
Center for Biotechnology, Bielefeld University, Germany
[2]Systems Biology of Regulatory Networks Group,
Institute for Genome Research and Systems Biology,
Center for Biotechnology, Bielefeld University, Germany
[3]Department of Proteome and Metabolome Research, Faculty of Biology,
Bielefeld University, Germany

Corresponding author: A. Goesmann
E-mail: agoesman@CeBiTec.Uni-Bielefeld.DE

ABSTRACT. New sequencing technologies provide ultra-fast access to novel microbial genome data. For their interpretation, an efficient bioinformatics pipeline that facilitates *in silico* reconstruction of metabolic networks is highly desirable. The software tool CARMEN performs *in silico* reconstruction of metabolic networks to interpret genome data in a functional context. CARMEN supports the visualization of automatically derived metabolic networks based on pathway information from the KEGG database or from user-defined SBML templates; this software also enables comparative genomics. The reconstructed networks are stored in standardized SBML format. We demonstrated the functionality of CARMEN with a major application

example focusing on the reconstruction of glycolysis and related metabolic reactions of *Xanthomonas campestris* pv. campestris B100. The curation of such pathways facilitates enhanced visualization of experimental results, simulations and comparative genomics. A second application of this software was performed on a set of corynebacteria to compare and to visualize their carbohydrate metabolism. In conclusion, using CARMEN, we developed highly automated data analysis software that rapidly converts sequence data into new knowledge, replacing the time-consuming manual reconstruction of metabolic networks. This tool is particularly useful for obtaining an overview of newly sequenced genomes and their metabolic blueprints and for comparative genome analysis. The generated pathways provide automated access to modeling and simulation tools that are compliant with the SBML standard. A user-friendly web interface of CARMEN is available at http://carmen.cebitec.uni-bielefeld.de.

Key words: Systems biology; Metabolic networks; *Xanthomonas*; Genome annotation; *Corynebacterium*

INTRODUCTION

The rapid progress in ultra-fast sequencing technologies leads to a dramatic increase in microbial genome data. After sequencing, the fragmented genome data are assembled to contigs or ideally to a whole genome sequence. Based on these data, an efficient automated analysis is indispensable for integrating the prediction of coding sequences and a first functional classification. This annotation process can be supported by various prokaryotic genome annotation systems such as GenDB (Meyer et al., 2003), AGMIAL (Bryson et al., 2006) and GAMOLA (Altermann and Klaenhammer, 2003). Genes coding for proteins build up a basic metabolic repertoire and facilitate the reconstruction of metabolic networks. These genes can be used to deduce special features of microbes by identifying encoded parts of metabolic pathways. This can be related to the presence or absence of key enzymes that are essential for the formation of biochemical reaction chains (metabolic pathways). Therefore, an extensive metabolic analysis is crucial to understand organism-specific features such as lifestyle, ability for adaptation, pathogenicity, and metabolic phenotype. In addition, reconstructed networks can be used for metabolic modeling (Durot et al., 2009), for instance, to analyze and optimize production strains (Nogales et al., 2008; Puchalka et al., 2008; Kjeldsen and Nielsen, 2009). Furthermore, data from comparative genomics approaches can be integrated to demonstrate differences between related species (Sun et al., 2007; Lee et al., 2009).

Hence, there is a growing demand for the genome-based reconstruction of metabolic networks. As a basis for this kind of analysis, the Systems Biology Markup Language (SBML) (Hucka et al., 2003) was shown to be well suited. SBML is widely accepted as a standard data format for systems biology. It is an open-source XML-based file format that facilitates the description of models and their exchange between various simulation and analysis tools such as the freely available CellDesigner software (Funahashi et al., 2003; Kitano et al., 2005), VANTED (Junker et al., 2006) or Cytoscape (Shannon et al., 2003).

The Kyoto Encyclopedia of Genes and Genomes (KEGG) (Ogata et al., 1999) is well established to provide an overview of metabolic pathways. One major component is the KEGG PATHWAY database, which consists of manually drawn pathway maps. The associated computerized information is stored in the KEGG Markup Language (KGML). Unfortunately, KEGG pathways are generalized and static without the ability to add species-specific features or to replace pathways by alternative ones. The tool KEGG2SBML (Funahashi et al., 2004) converts KEGG pathways into SBML networks, but during the conversion process not all relevant information is conserved. For instance, the annotated gene identifiers are lost. The pathway reconstruction is restricted to organisms that are part of the KEGG database and cannot be applied to unpublished genomes. Using the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007), organism-specific pathways can be generated, but may not be curated or edited. Moreover, this tool does not allow comparative pathway reconstruction. An alternative source for metabolic pathway data is the powerful MetaCyc/BioCyc system (Krieger et al., 2004; Caspi et al., 2008), which provides a useful SBML export. However, local software installation and database maintenance are not convenient for users mainly interested in rapidly obtaining SBML-compliant networks for unpublished genomes. With regard to studying cellular metabolism, IdentiCS (Sun and Zeng, 2004) was developed to generate genome annotations and pathway reconstructions. This program combines the identification of coding sequences and the reconstruction, comparison and visualization of the metabolic network directly from unannotated bacterial genome sequences. It allows a fast overview of the metabolic capacities of a newly sequenced microorganism by mapping EC numbers onto KEGG maps. They can be transferred into a Microsoft Excel file for comparison and visualization of various species, but the handling of Excel files is restricted and hinders the illustration of metabolic networks.

To avoid a bottleneck in high-speed data interpretation, CARMEN (Comparative Analysis and Reconstruction of MEtabolic Networks) was developed for the fast and automated reconstruction of metabolic networks. The SBML documents provided by CARMEN are particularly well suited to be subsequently modified with SBML editors. Reconstructed networks can be used to visualize post-genomic experimental data, and they can be utilized as a starting point for model creation in systems biology analyses. Curated SBML models can be used by CARMEN in a recursive manner to extend the network or to include metabolic capabilities of additional organisms. Thus, it enables comparative genome analyses, which is not considered by other SBML generators.

METHODS

To explore the practical value of CARMEN for interdisciplinary applications, a reconstructed and curated metabolic network model of *Xanthomonas campestris* pv. campestris B100 was simultaneously used for a proteomics and a systems biology application.

Proteomics techniques

Protein isolation, two-dimensional (2-D) gel electrophoresis, spot excision, tryptic in-gel digestion, and protein identification by mass spectrometric analysis of peptide mass fingerprints were performed as described earlier (Sidhu et al., 2008). Aliquots representing 2 x 10^{10} *X.*

campestris pv. campestris B100 cells from cultures grown to ODs of 0.8 to 1.2 were harvested and employed for the procedure. For electrophoresis, an Ettan IPGphor (Amersham Biosciences, Freiburg, Germany) was employed in the first dimension, and a PROTEAN II xi Cell (Bio-Rad, Munich, Germany) in the second dimension. Trypsin from Promega (Mannheim, Germany) was employed to digest the proteins of 352 individual spots that were visualized by Coomassie staining. For mass spectrometry and peptide mass fingerprint analysis, an Ultraflex MALDI-TOF (Bruker, Bremen, Germany) was used, combined with the Mascot software (Matrix Science, London, UK) and a database derived from the recently sequenced genome of *X. campestris* pv. campestris B100 (Vorhölter et al., 2008). *X. campestris* pv. campestris enzymes of central metabolism that were identified were highlighted in the SBML model.

Metabolic flux analysis

To predict the metabolic fluxes, metabolic network analysis was used (Schilling and Palsson, 2000; Price et al., 2004; Klamt et al., 2007). This method facilitates a first approximation of fluxes in a metabolic network on the basis of reaction stoichiometry. Version 9.0 of the CellNetAnalyzer software (Klamt et al., 2007) was employed to perform the analysis. CellNetAnalyzer is a plug-in of Matlab (The Mathworks, Natick, MA, USA) for flux optimization analysis, and requires the Optimization Toolbox. CellNetAnalyzer could import the metabolic network from the SBML file provided by CARMEN by means of the sbml2cna script provided with CellNetAnalyzer.

To visualize the network in CellNetAnalyzer for graphical output, a bitmap image is required. The bitmap was obtained from the CARMEN-derived SBML by CellDesigner (Funahashi et al., 2003; Kitano et al., 2005), which facilitated export to a PNG format file. Boxes with digits indicating the predicted fluxes were positioned in CellNetAnalyzer on the top of the respective reaction arrows. For the consumption of ADP, NAD and NADP, regeneration reactions (Krömer et al., 2006) were introduced to the network.

The simulation (flux optimization) was performed for a scenario where 100 units glucose and 10 units 2-dehydro-3-deoxy-D-gluconate were imported from the environment and where the cells produced biomass and xanthan, in this model represented by its precursors glucose-1-phosphate and fructose-1-phosphate, in equimolar proportion. The tentative biomass was derived from the precursors: 1 unit PRPP, 2 units ribose and 7 units pyruvate. The flux distribution is shown as a heat map with a predetermined threshold of 0.

RESULTS AND DISCUSSION

For supporting the *in silico* reconstruction of metabolic networks, CARMEN facilitates two major features. First, the KGML-based model generation exploits KEGG's expert knowledge about reactions and their graphical display. It thereby provides a rapid overview of the metabolic repertoire and makes the data obtained available by generating SBML output (SBML Level 2 and latest SBML Level 3 format). Second, the SBML-based model reconstruction uses SBML templates of a reference genome as backbone for pathway reconstruction. This information will be combined with genome annotation data of published or unpublished genomes. They can be extracted from NCBI GenBank files or from a genome annotation system.

Implementation

The software design of CARMEN reflects the two main requirements: specific Perl programs control the workflow of either the KGML- or SBML-based model generation. Both programs employ a core library implemented in Perl that provides functionality for GenBank import, access to KEGG information, the creation of the final SBML file and other reusable functions. Efficient data handling is realized using the general Perl modules XML::DOM and XML::Writer.

KGML-based model generation

The first feature of CARMEN is based on information of the KEGG PATHWAY database (Ogata et al., 1999) to provide a rapid overview of the metabolic properties of an organism. This database contains a collection of pathway maps including computerized information about graphical objects and their relationships stored in KEGG Markup Language (http://www.genome.jp/kegg/docs/xml/). CARMEN extracts the KGML tags "entry", "relation" and "reaction" to gain enzyme, reaction and position data. These data are combined with annotation data such as EC numbers, gene identifiers and gene names of a chosen organism. Beyond the basic reconstruction, command line switches can be used to improve the output of a locally installable version of CARMEN. As an option, cofactors and stoichiometric data can be integrated according to reactions of a pathway based on data derived from KEGG Ligand files (ftp://ftp.genome.jp/pub/kegg/ligand/). To optimize the reconstructed network, CARMEN can merge identical enzyme symbols that were split in KEGG due to their participation in similar reactions. In addition, the compound nomenclature used within the reconstructed pathways can be chosen. Currently, KEGG compound IDs, KEGG compound names and user-defined abbreviations stored in a tab-separated list are supported.

SBML-based model generation

The second feature of CARMEN is the generation of metabolic networks based on pre-existing SBML templates. This facilitates the precise and rapid reconstruction of the metabolism of closely related genomes, for instance, of various strains belonging to the same species. The input is an SBML file, which provides layout and gene-specific information of a metabolic pathway. Genes of a new organism are mapped onto the SBML template using bidirectional best-BLAST hits (BBHs). BBHs are widely used for homology-based orthology estimation and can be generated in comparably short time. Recently, Altenhoff and Dessimoz (2009) concluded that BBHs give results comparable to the more sophisticated methods for orthology estimation. Thus, in CARMEN, BBHs are used based on BLASTP results, using the default scoring matrix BLOSUM62 and a user-defined E-value cutoff (default: $1e^{-10}$). Optionally, CARMEN proposes pathway extensions to find alternative or continuative pathways extending the template model based on KEGG Ligand files. Affiliated reactions have to satisfy two criteria. First, the reactions need to be adjacent to reactions of the SBML template. Second, genes encoding proteins catalyzing a certain reaction are annotated in the newly sequenced and annotated genome. Thus, it is possible to automatically identify direct connections of related pathways, which will be stored in a newly generated SBML file.

Moreover, KGML-based models can be used for SBML-based model generation. This strategy offers the mapping of exclusively (manually) drawn metabolic pathways to build up a library of verified metabolic networks for different organisms.

Web application of CARMEN

To provide convenient access to the implemented CARMEN functionality, a platform-independent web interface has been developed. An Apache web server (http://www.apache.org/), using CGI scripts (http://search.cpan.org/dist/CGI.pm/), generates the user interface that provides the functionality of the KGML-based model generation. A continuously updated NCBI genome database together with the KEGG PATHWAY map data is employed. The user may choose one organism and a multitude of KEGG pathways and perform the reconstruction analysis. After a CARMEN analysis run has finished, the user can directly download the resulting SBML files, which can be visualized by SBML-compliant tools, e.g., the CellDesigner software. Additionally, a Scalable Vector Graphics (SVG) representation of the metabolic reconstruction is rendered on the fly, displaying the selected metabolic pathways. The SVG image can be downloaded or reviewed in most recent browsers directly or via the freely available Adobe SVG plugin. Researchers have access to the Perl source code; detailed documentation is supplied on the CARMEN homepage.

Applications

CARMEN can be used as a stand-alone tool or via the web interface. Furthermore, it is embedded in the genome analysis pipeline of the Center for Biotechnology at Bielefeld University. Here, the tool is connected to the genome annotation system GenDB (Meyer et al., 2003), the comparative genome analysis tool EDGAR (Blom et al., 2009) and the multi-omics viewer ProMeTra (Neuweger et al., 2009). GenDB can be used to obtain genome annotation data and BLAST hits between predicted proteins of two unpublished species. EDGAR identifies orthologous genes in different genomes and classifies genes as core genes or singletons. Gene sets of an organism group can be exported and mapped by CARMEN onto a graphical metabolic representation. ProMeTra uses SBML output of CARMEN for the mapping of quantitative omics data such as microarray data.

The *in silico* pathway reconstruction of CARMEN was used to obtain insights into the lifestyle of microorganisms, which have roles in both industrial biotechnology and pathogenicity, such as *X. campestris* pv. campestris B100 and corynebacteria. The first major example is the KGML-based reconstruction of the glycolysis pathway in *X. campestris* pv. campestris B100. Subsequently, further reconstructed networks have been curated, extended and used for various applications such as the mapping of proteome data or metabolic flux analysis. Examples illustrate these applications and demonstrate the usefulness of CARMEN for postgenomics experiments and systems biology. Another example arises from comparative genomics in combination with the SBML-based model reconstruction. As a use case, two recently sequenced pathogenic corynebacteria from human clinical sources named *Corynebacterium kroppenstedtii* and *C. aurimucosum* were selected. A metabolic network of the central carbohydrate metabolism of *C. kroppenstedtii* reconstructed in SBML (Tauch et al., 2008) was used to obtain the corresponding pathways of *C. aurimucosum* in less than a minute using CARMEN.

KGML-based models of *Xanthomonas campestris*

Gram-negative bacteria of the genus *Xanthomonas* are plant-associated and usually plant-pathogenic proteobacteria that cause substantial harm to plants (Swings, 1993). At the same time they produce a polysaccharide called xanthan, which has many applications in industry, making it the most important biotechnological polysaccharide synthesized by bacteria (Seisun, 2006). While previous genome analyses focused on plant pathogenicity (Da Silva et al., 2002; Qian et al., 2005), the recent genome analysis of *X. campestris* pv. campestris B100 (Vorhölter et al., 2008) compiled the genetic blueprints for all metabolic processes involved in xanthan biosynthesis, ranging from uptake of sugars via their processing in the central metabolism to the polymerization and export of xanthan. Based on this annotated genome, glycolysis was reconstructed by CARMEN (Figure 1). Biochemical analyses had shown that *X. campestris* pv. campestris metabolizes imported glucose via the pentose-phosphate and Entner-Doudoroff pathways (Whitfield et al., 1982; Pielken et al., 1988), while no activity could be shown for the phosphofructokinase reaction, which is an essential key enzyme of the Embden-Meyerhof pathway (glycolysis). This resulted in the assumption that there is no glycolysis in *X. campestris*, but recent genome analyses revealed a gene for phosphofructokinase, *pfkA* (Vorhölter et al., 2008). Accordingly, the automated reconstruction with CARMEN of *X. campestris* pv. campestris B100 provided a metabolic pathway that comprised all key reactions of glycolysis, illustrating the fundamental capacity to utilize this catabolic pathway.
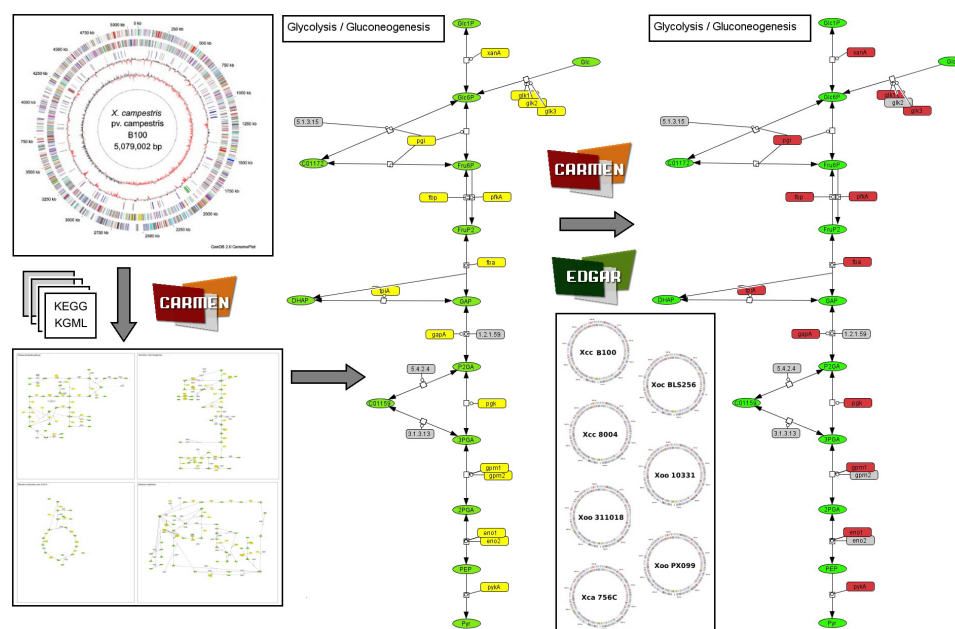

Figure 1. KGML-based model reconstruction of *Xanthomonas campestris* pv. campestris B100 and its curation. The reconstruction and manual curation was performed based on the KGML file of the glycolysis. Gray symbols displaying the EC number of the encoded protein represent missing genes; otherwise, they are colored yellow. Pathway intermediates are displayed green. This pathway was used as a template for a core genome mapping of seven *Xanthomonas* species. The core genome was computed by EDGAR.

In a second step, CARMEN was employed for metabolic analyses based on comparative genomics. The genome annotation of *X. campestris* pv. campestris B100 had revealed several isogenes that encode multiple variants of central metabolism enzymes, including glycolysis enzymes such as phosphoglycerate mutase (EC 5.4.2.1, *gpm*), phosphopyruvate hydratase (EC 4.2.11, *eno*) or glucokinase (EC 2.7.1.2, *glk*). The specific role of the individual isogenes remained unclear. To shed light on this phenomenon, the distribution of these genes was checked in other *Xanthomonas* genomes that were publicly available. The curated SBML file of *X. campestris* pv. campestris B100 glycolysis, which had been generated by CARMEN, was re-used as a reference for this purpose. The core genome of seven *Xanthomonas* strains representing *X. campestris* pv. campestris (Xcc), *X. oryzae* pv. oryzicola (Xoc), *X. oryzae* pv. oryzae (Xoo), and *X. campestris* pv. amoraciae (Xca) was computed by EDGAR and stored in a tab-separated file. Subsequently, this information was mapped by CARMEN onto the SBML template. The result demonstrates that all proteins essential for glycolysis including phosphofructokinase are conserved between the *Xanthomonas* strains (Figure 1). At the same time, it became obvious that from each group of glycolysis enzyme isogenes, one representative was not included in the core genome, indicating that it might have a less essential role than the well-preserved isogenes.

Curation and application of metabolic networks

Besides visualizing metabolic pathways based on data from genomics or comparative genomics, CARMEN is designed to be a valuable tool for the manual curation of genome annotation data related to metabolism. For this purpose, CARMEN can be used in combination with an easy-to-use SBML editor such as CellDesigner. In such a scenario, CARMEN provides SBML-output at a semi-finished annotation state (e.g., after an automated functional annotation step). SBML editors exploit information generated by CARMEN such as EC numbers and KEGG reaction identifiers to facilitate database lookup by a single mouse-click, with the KEGG database system as a portal to cross-linked metabolic databases. Further on, users conveniently get access to integrated data such as gene identifiers. The user-friendly database access enables the rapid curation of annotation data. Metabolic reactions may be deleted, added or modified to complete the annotation. In parallel, the visual display of the analyzed organism's metabolism can be further enhanced. The metabolic maps obtained can be used as templates when further functional data need to be displayed. An example of this feature is again the central metabolism of *X. campestris* pv. campestris B100, where glycolysis was curated and interconnected with the Entner-Doudoroff pathway and the pentose-phosphate pathway to clearly illustrate this complex metabolic network (Figure 2).

This metabolic network reconstructed by CARMEN was utilized to visualize experimental results and predictions from systems biology simulations. The example indicates the presence of network nodes such as metabolites or enzymes that are identified in laboratory experiments. A use case is given in Figure 2A, where those enzymes of the *X. campestris* pv. campestris B100 central carbohydrate metabolism that were identified in the laboratory are highlighted. Not all proteins involved in the central metabolism could be identified experimentally due to biological and technical reasons (Stults and Arnott, 2005). For *Xanthomonas*, basic proteomics techniques are well established. Recently, proteomics was employed to in-

vestigate the outer membrane vesicles of *X. campestris* pv. campestris B100, while the 30 most abundant cytoplasmic proteins were already identified in 2003 (Sidhu et al., 2008). For this use case, 215 proteins of *X. campestris* pv. campestris B100 were identified by 2-D gel chromatography (Figure 2B) and subsequent analysis of their peptide mass fingerprints. The gene highlighting option of CARMEN's SBML-based model generation was used to pinpoint the experimentally identified proteins within the reconstructed metabolic network based on a list of gene identifiers of the detected proteins.

In a second use case, the same CARMEN-derived SBML model of the *X. campestris* pv. campestris B100 central carbohydrate metabolism was employed for the analysis of metabolic fluxes (Figure 2C). Based on the reaction stoichiometries and a linear optimization algorithm, the reaction rates along the reconstructed pathways were simulated by this method (Klamt et al., 2007), which is often applied in systems biology. A comparison of both use case results indicates their frequent conformity. This seems to be plausible because experimentally detected enzymes are likely to be highly abundant (Stults and Arnott, 2005), while high enzyme abundance may lead to substantial turnover rates for those metabolic pathways that are constituted by the abundant enzymes.
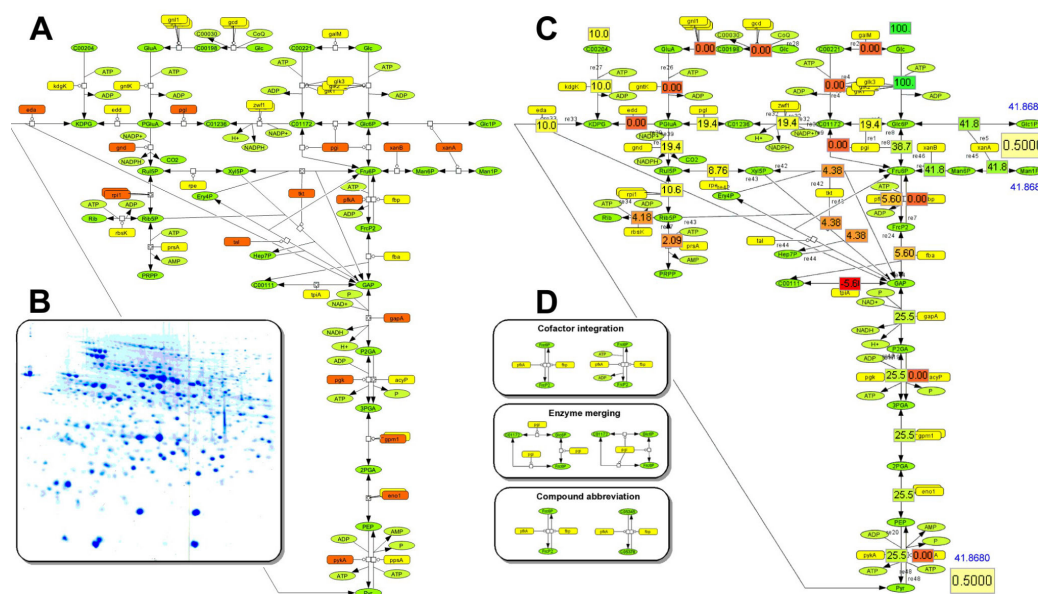

Figure 2. Applying models reconstructed by CARMEN to enhance visualization of network-related experimental results. The central metabolism of *Xanthomonas campestris* pv. campestris B100 comprises the Entner-Doudoroff pathway, the pentose phosphate pathway and glycolysis/gluconeogenesis. **A.** The map can be used to visualize proteomics results derived from 2-D gel electrophoresis **B.** Proteins that were detected are highlighted in orange. **C.** The curated SBML model was employed to visualize metabolic fluxes predicted by stoichiometry-based metabolic network analysis. Ideal routes were predicted to convert imported glucose either into biomass or the exopolysaccharide xanthan. High predicted fluxes through the individual reactions lead to green coloration of boxes, low fluxes to red coloration. Numbers within the boxes represent the predicted reaction rates. Negative numbers indicate reactions in the reverse direction. The *in silico* reconstruction was improved by the integration of cofactors, joining of closely located equal enzymes and the use of metabolite abbreviations (**D**).

Employing an SBML-based metabolic network for comparative analysis

The genus *Corynebacterium* covers Gram-positive bacteria, which are non-motile and non-spore forming with a high G+C-content DNA. Two human-pathogenic members of this genus are *C. aurimucosum* and *C. kroppenstedtii*. *C. kroppenstedtii*, causing inflammatory breast disease (Taylor et al., 2003), is a lipophilic and facultative anaerobic bacterium that had been described taxonomically in 1998 (Collins et al., 1998) while its complete genome was sequenced in 2008 (Tauch et al., 2008). After sequencing and assembly, in the course of genome annotation, a metabolic network for *C. kroppenstedtii* was developed, visualized and stored in SBML. This SBML file was subsequently used as a basis for the SBML-based model reconstruction.

This reconstruction was applied to create the central carbohydrate metabolism of *C. aurimucosum*, a bacterial species that was isolated from human vaginal infections. The complete SBML model was automatically generated based on the manually curated metabolic network of *C. kroppenstedtii* in less than one minute (Figure 3A). Results of reciprocal best-
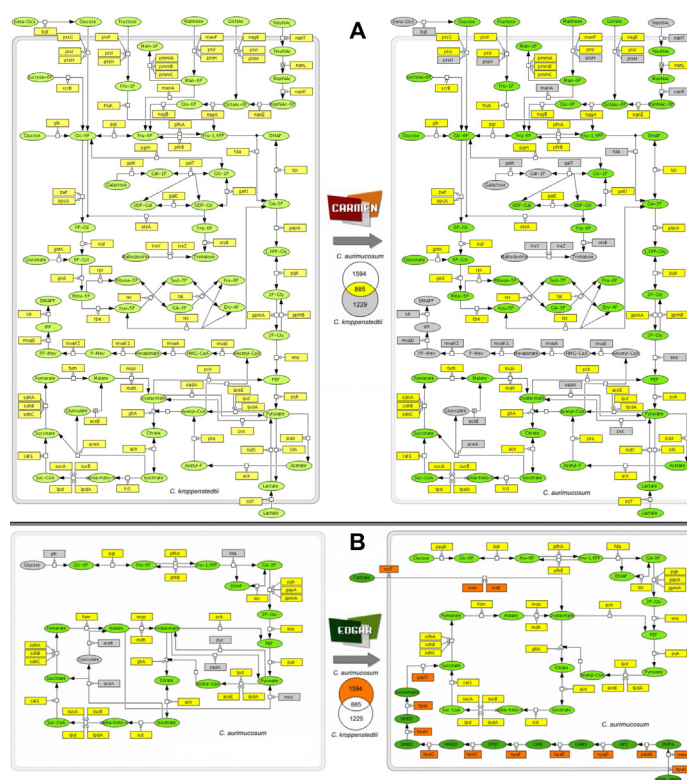

Figure 3. SBML-based model reconstruction of *Corynebacterium kroppenstedtii* and *C. aurimucosum*. **A.** The manual reconstruction of the central carbohydrate metabolism of *C. kroppenstedtii* was used as a draft template for *C. aurimucosum* displaying orthologous genes. Proteins given by gene names are colored yellow and simple molecules are green. Missing genes and associated molecules are colored gray. Venn diagrams were computed by EDGAR. **B.** Singleton genes encoding unique proteins of *C. aurimucosum* were manually added to the metabolic pathway and colored orange. The CellDesigner software was used for visualization and editing.

BLAST hits between all genes of both species were employed. A detailed analysis of this network with CARMEN revealed that the majority of genes involved in glycolysis, pentose phoshate pathway and TCA cycle are conserved. Furthermore, the resulting network could be directly used to infer the potential natural habitat from the annotated genome of *C. aurimucosum* (Figure 3B): the presence of genes coding for L-tartrate degradation and the HPC meta-cleavage pathway indicate that *C. aurimucosum* is a resident of the human gut and as a contaminant of the vaginal tract could cause fatal abortion during pregnancy (Trost et al., 2010).

CONCLUSIONS

We have implemented an automated and fast *in silico* metabolic reconstruction tool named CARMEN, which employs the approved standardized SBML format. Using this format enables the exchange between different visualization and simulation tools and the modeling of metabolic pathways, as well as the manual editing and verification of different networks. CARMEN is applicable for published or newly sequenced genomes and provides the KGML-based model reconstruction of metabolic networks as one of its major features to get a rapid overview of the metabolic repertoire of a microbe. The second feature is the creation of SBML-based models using already existing SBML template files of a reference genome, which can be used to focus on the reconstruction of specific pathways, for example manually verified ones. The possibility of the continuative use of the reconstructed or manually drawn metabolic networks is an important benefit of this software in comparison to tools that are tailored to KEGG-dependent pathway reconstruction. In summary, CARMEN is particularly useful for rapidly obtaining an overview of newly sequenced genomes and their metabolic blueprints, for visualizing post-genomic experimental or simulation data, and for comparative genome analysis.

ACKNOWLEDGMENTS

REFERENCES

Altenhoff AM and Dessimoz C (2009). Phylogenetic and functional assessment of orthologs inference projects and methods. *PLoS. Comput. Biol.* 5: e1000262.

Altermann E and Klaenhammer TR (2003). GAMOLA: a new local solution for sequence annotation and analyzing draft and finished prokaryotic genomes. *OMICS* 7: 161-169.

Blom J, Albaum SP, Doppmeier D, Puhler A, et al. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10: 154.

Bryson K, Loux V, Bossy R, Nicolas P, et al. (2006). AGMIAL: implementing an annotation strategy for prokaryote genomes as a distributed system. *Nucleic Acids Res.* 34: 3533-3545.

Caspi R, Foerster H, Fulcher CA, Kaipa P, et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* 36: D623-D631.

Collins MD, Falsen E, Akervall E, Sjoden B, et al. (1998). *Corynebacterium kroppenstedtii* sp. nov., a novel *corynebacterium* that does not contain mycolic acids. *Int. J. Syst. Bacteriol.* 48 (Pt 4): 1449-1454.

Da Silva AC, Ferro JA, Reinach FC, Farah CS, et al. (2002). Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. *Nature* 417: 459-463.

Durot M, Bourguignon PY and Schachter V (2009). Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol. Rev.* 33: 164-190.

Funahashi A, Tanimura N, Morohashi M and Kitano H (2003). CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 1: 159-162.

Funahashi A, Jouraku A and Kitano H (2004). Converting KEGG pathway database to SBML. Proceedings of the 8th Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004), San Diego.

Hucka M, Finney A, Sauro HM, Bolouri H, et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19: 524-531.

Junker BH, Klukas C and Schreiber F (2006). VANTED: a system for advanced data analysis and visualization in the context of biological networks. *BMC Bioinformatics* 7: 109.

Kitano H, Funahashi A, Matsuoka Y and Oda K (2005). Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* 23: 961-966.

Kjeldsen KR and Nielsen J (2009). *In silico* genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnol. Bioeng.* 102: 583-597.

Klamt S, Saez-Rodriguez J and Gilles ED (2007). Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.* 1: 2.

Krieger CJ, Zhang P, Mueller LA, Wang A, et al. (2004). MetaCyc: a multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* 32: D438-D442.

Krömer JO, Wittmann C, Schroder H and Heinzle E (2006). Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. *Metab. Eng.* 8: 353-369.

Lee DS, Burd H, Liu J, Almaas E, et al. (2009). Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. *J. Bacteriol.* 191: 4015-4024.

Meyer F, Goesmann A, McHardy AC, Bartels D, et al. (2003). GenDB - an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31: 2187-2195.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, et al. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35: W182-W185.

Neuweger H, Persicke M, Albaum SP, Bekel T, et al. (2009). Visualizing post genomics data-sets on customized pathway maps by ProMeTra-aeration-dependent gene expression and metabolism of *Corynebacterium glutamicum* as an example. *BMC Syst. Biol.* 3: 82.

Nogales J, Palsson BO and Thiele I (2008). A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* 2: 79.

Ogata H, Goto S, Sato K, Fujibuchi W, et al. (1999). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* 27: 29-34.

Pielken P, Schimz KL, Eggeling L and Sahm H (1988). Glucose metabolism in *Xanthomonas campestris* and influence of methionine on the carbon flow. *Can. J. Microbiol.* 34: 1333-1337.

Price ND, Reed JL and Palsson BO (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat. Rev. Microbiol.* 2: 886-897.

Puchalka J, Oberhardt MA, Godinho M, Bielecka A, et al. (2008). Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS. Comput. Biol.* 4: e1000210.

Qian W, Jia Y, Ren SX, He YQ, et al. (2005). Comparative and functional genomic analyses of the pathogenicity of phytopathogen *Xanthomonas campestris* pv. campestris. *Genome Res.* 15: 757-767.

Schilling CH and Palsson BO (2000). Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J. Theor. Biol.* 203: 249-283.

Seisun D (2006). The market of hydrocolloids and other uses of seaweeds, Aqua Sur. Proceedings of the 3rd International Conference on Aquaculture, Puerto Varas, Chile.

Shannon P, Markiel A, Ozier O, Baliga NS, et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13: 2498-2504.

Sidhu VK, Vorholter FJ, Niehaus K and Watt SA (2008). Analysis of outer membrane vesicle associated proteins isolated from the plant pathogenic bacterium *Xanthomonas campestris* pv. campestris. *BMC Microbiol.* 8: 87.

Stults JT and Arnott D (2005). Proteomics. *Methods Enzymol.* 402: 245-289.

Sun J and Zeng AP (2004). Identics-identification of coding sequence and *in silico* reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence. *BMC Bioinformatics* 5: 112.

Sun J, Lu X, Rinas U and Zeng AP (2007). Metabolic peculiarities of *Aspergillus niger* disclosed by comparative metabolic genomics. *Genome Biol.* 8: R182.

Swings JG (1993). Xanthomonas. 1st edn. Chapman and Hall, London.

Tauch A, Schneider J, Szczepanowski R, Tilker A, et al. (2008). Ultrafast pyrosequencing of *Corynebacterium kroppenstedtii* DSM44385 revealed insights into the physiology of a lipophilic *Corynebacterium* that lacks mycolic acids. *J. Biotechnol.* 136: 22-30.

Taylor GB, Paviour SD, Musaad S, Jones WO, et al. (2003). A clinicopathological review of 34 cases of inflammatory breast disease showing an association between corynebacteria infection and granulomatous mastitis. *Pathology* 35: 109-119.

Trost E, Gotker S, Schneider J, Schneiker-Bekel S, et al. (2010). Complete genome sequence and lifestyle of black-pigmented *Corynebacterium aurimucosum* ATCC 700975 (formerly *C. nigricans* CN-1) isolated from a vaginal swab of a woman with spontaneous abortion. *BMC Genomics* 11: 91.

Vorhölter FJ, Schneiker S, Goesmann A, Krause L, et al. (2008). The genome of *Xanthomonas campestris* pv. campestris B100 and its use for the reconstruction of metabolic pathways involved in xanthan biosynthesis. *J. Biotechnol.* 134: 33-45.

Whitfield C, Sutherland IW and Cripps RE (1982). Glucose metabolism of *Xanthomonas campestris*. *J. Gen. Microbiol.* 128: 981-985.