W460–W464 *Nucleic Acids Research, 2008, Vol. 36, Web Server issue* *Published online 25 April 2008*
doi:10.1093/nar/gkn217

ProdoNet: identification and visualization of prokaryotic gene regulatory and metabolic networks

Johannes Klein[1], Stefan Leupold[1], Richard Münch[1], Claudia Pommerenke[1,2], Thorsten Johl[2], Uwe Kärst[2], Lothar Jänsch[2], Dieter Jahn[1,*] and Ida Retter[1]

[1]Institute for Microbiology, Technische Universität Braunschweig, Spielmannstr. 7, D-38106 Braunschweig and [2]Department of Cell Biology, Helmholtz Centre for Infection Research, Inhoffenstr. 7, D-38124 Braunschweig, Germany

Received February 1, 2008; Revised April 3, 2008; Accepted April 9, 2008

ABSTRACT

ProdoNet is a web-based application for the mapping of prokaryotic genes and the corresponding proteins to common gene regulatory and metabolic networks. For a given list of genes, the system detects shared operons, identifies co-expressed genes and deduces joint regulators. In addition, the contribution to shared metabolic pathways becomes visible on KEGG maps. Furthermore, the co-occurrence of genes of interest in gene expression profiles can be added to the visualization of the global network. In this way, ProdoNet provides the basis for functional genomics approaches and for the interpretation of transcriptomics and proteomics data. As an example, we present an investigation of an experimental membrane subproteome analysis of *Pseudomonas aeruginosa* with ProdoNet. The ProdoNet dataset on transcriptional regulation is based on the PRODORIC Prokaryotic Database of Gene Regulation and the Virtual Footprint tool. ProdoNet is accessible at http://www.prodonet.tu-bs.de.

INTRODUCTION

The prevalence of high-throughput technologies for the analysis of biological systems causes a strong demand on specialized software for the interpretation of the increasingly complex experimental results. In the case of transcriptome and proteome analyses, the final result is usually a list of proteins or transcripts that show significant differences in their abundance under the compared experimental conditions. The challenge for the scientist is to identify the common properties between co-regulated genes, in order to understand the underlying processes within the analysed cell. Consequently, a computer-aided application is required to explore the common functions of these genes and proteins within the complex cellular network. Such application should map the list of experimentally identified genes and proteins to the known transcriptional and metabolic network and be able to identify new relationships.

A variety of databases supply valuable information on transcription factor binding sites and gene regulation. For prokaryotes, these include databases that focus on a single model organism, as RegulonDB for *Escherichia coli* (1) or DBTBS for *Bacillus subtilis* (2). Others cover a range of species, like RegTransBase (3) or PRODORIC (4). Some databases exclusively present data based on experimental evidence, such as PRODORIC, while other data collections also include data predicted by different algorithms, like Tractor_DB (5), ExtraTrain (6) and SwissRegulon (7). Several databases provide visualization features for the presented data. For example, the database CoryneRegNet presents a visualization of gene regulatory networks from corynebacteria, mycobacteria and *E. coli* (8). Usually, the web interfaces of databases are capable to manage only one item per query and are not prepared to deal with a list of genes and proteins.

However, there are some tools available that allow searching for functional relations within a list of genes. For example, VIS-O-BAC (9) supports the functional exploration of prokaryotic genes and proteins by indicating their genomic positions, matching them onto KEGG pathways and supplying the Gene Ontology annotations. For protein interaction analyses, the STRING website provides a sophisticated platform to indicate functional associations between proteins (10). A major strength of STRING is the visualization of obtained results, where interactions within a set of analysed proteins are displayed as undirected edges.

Apart from the described database-integrated applications, various software tools are available for the visualization of biological networks; an overview was given before (11). Although some of these tools provide advanced network analysis features, most require local

*To whom correspondence should be addressed. Tel: +49 531 391 5801; Fax: +49 531 391 5854; Email: d.jahn@tu-bs.de

installation and do not provide simple access to prokaryotic gene regulatory data.

Currently, web-accessible tools for the mapping of a list of genes and corresponding proteins to the gene regulatory and metabolic network in bacteria with an intuitive visualization are missing. For this purpose, we developed ProdoNet, an application that visualizes the functional relations within a set of prokaryotic genes or proteins with regard to the joined gene regulatory network. ProdoNet uses data derived from the PRODORIC database and displays the hierarchical structure of the underlying network of genes, operons and regulators. Moreover, information is provided on the co-occurrence of analysed genes and proteins in gene expression profiles and metabolic pathways, respectively. To further support the functional exploration of the obtained results, hyperlinks to UniProtKB (12), PRODORIC and the KEGG pathway maps (13) are provided. To complement the PRODORIC dataset that comprises exclusively carefully curated data derived from reliable publications, predictions on operons and regulons are added. For transparency, the ProdoNet visualization allows a clear distinction between experimentally proven and predicted data. The current version of ProdoNet comprises data from the well-characterized model organisms *E. coli*, *B. subtilis* and *Pseudomonas aeruginosa*.

WEB INTERFACE

Input data and query options

ProdoNet identifies and displays the gene regulatory network and metabolic pathways for a user-defined list of genes or proteins. The input list is accepted in most common formats, including comma separated or tab delimited lists. The use of gene or protein symbols within the input list is highly flexible, which means that ProdoNet accepts short names, locus tags and accession numbers from UniProtKB, GenBank, RefSeq and other databases. The explicit usage of locus tags, short names or UniProtKB accession numbers will speed up the query.

In a first step, ProdoNet matches the input data with its own gene dataset and returns a table of recognized genes, comprising their full names. In the case one query name matches more than one gene from the database, corresponding matches are delineated as ambiguous. In the next step, the user can re-select the genes to query and choose the type of analysis to perform. The default selection ('network of operon and regulon' settings) results in the visualization of the operons and regulons that map the selected genes. Optionally, displayed results can include the expression profiles of genes found in DNA array experiments and predicted transcriptional regulations. The query can be limited to the search for matches in the gene expression profile dataset, which generates a table of experimental conditions where candidate genes found in the user list were co-regulated. Alternatively, the user can limit the request to the occurrence of the corresponding proteins in metabolic pathways. In this case, ProdoNet will deduce the enzymes from the input gene set and return a table with hyperlinks to the KEGG pathway maps in which these enzymes are involved. Thereby, enzymes included in the input gene set are marked in red within the pathway maps. Both of these queries can be fused and visualized in the 'network of operons and regulons' query described above by selecting the 'show tables' option. In this case, the tables are depicted below the network. This option will additionally create a table of all involved transcription factors and regulated operons in the corresponding network. All tables can be downloaded as tab-delimited files, allowing for a convenient export and local storage of the requested data.

Network of operons and regulons

In the default mode, the user-defined set of genes is processed and integrated into a directed graph that represents the corresponding gene regulatory network, using the prefuse toolkit for interactive information visualization (14). Hereby, the nodes of the network are regulators, operons or genes while the edges symbolize transcription factor–operon interactions, genes belonging to one operon or the co-occurrence of genes in expression profiles. The result is temporarily stored as GraphML and GML formatted files, which can be downloaded by the user and re-used in other network analysis applications, e.g. in Cytoscape (15).

In the ProdoNet network view (Figures 1 and 3), the network graph is visualized and each gene is depicted as a box, whereas queried genes are highlighted by red letters and transcription factors are marked in yellow. Genes belonging to one operon are shown in one frame. The colour of the connection lines between a transcription factor and its corresponding operon indicates the type of regulation, i.e. green is positive, red is negative, blue represents positive and negative, and black unknown type of regulation. Predicted regulations are indicated in grey. Similarly, the prediction of a gene belonging to an operon is indicated by a grey connection between the gene and the operon node. As an option, genes from the input list that are found co-expressed in a microarray experiment can also be linked by lines, adding an extra layer of information to the network (Figure 1B). The involved genes are marked as ellipses.

For the functional exploration of a particular gene of the network, a gene context menu provides the full name of the corresponding protein and links to the PRODORIC and UniProtKB databases. In case of a gene coding for an enzyme, the menu offers the 'metabolic pathways' option, leading to a table that provides links to the KEGG pathway maps for this enzyme. Similarly, expression profile connections are featured with a clickable dot that displays the name of the experiment, the link to its entry in PRODORIC and a 'select' option to hide all other expression profiles shown in the graph.

By default, the outlined network view presents only the part of the complete gene regulatory network that was queried by the selected input genes (Figure 1A). For deeper insights into the network, the view can be interactively expanded by choosing a higher expansion level (Figure 1B). In this case, all transcription factor

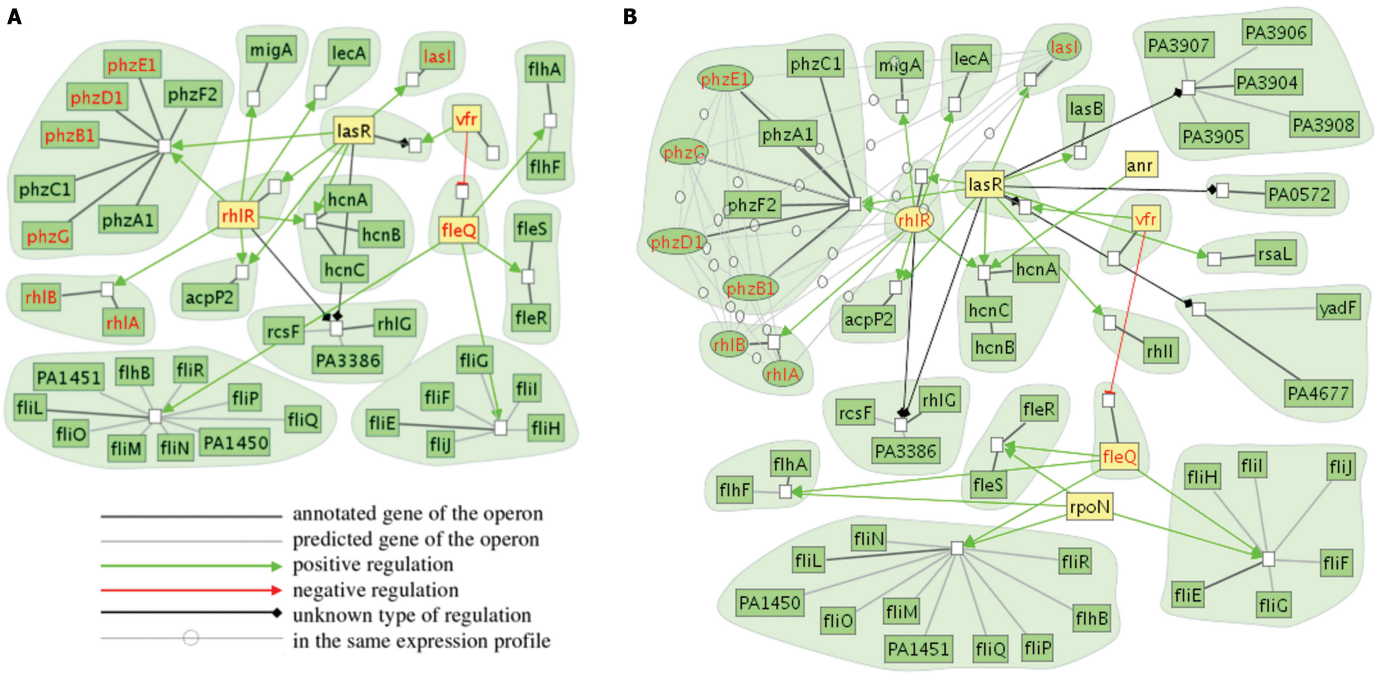
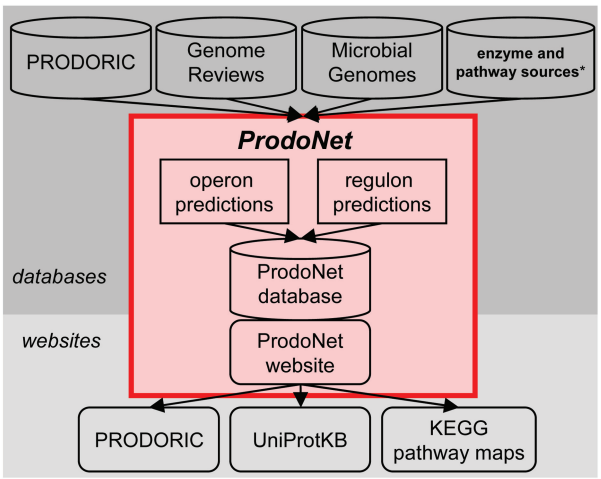


Figure 1. The ProdoNet network view. Exploration of the results from an experimental membrane subproteome analysis of *P. aeruginosa*. Genes included in the query are indicated in red letters. Transcription factors are marked in yellow. (**A**) Result with the default settings. (**B**) Results of a new query with the query genes shown in (A) and the setting 'include gene expression profiles'. The view is expanded by selection of 'level 2' within the Java applet.

nodes are extended both by upstream regulators and further downstream target operons. At a higher expansion level, regulatory cascades or circuits are fully shown and thus provide a broader view on the overall network involved. Furthermore, the network can be expanded by requesting additional genes with the 'add genes' field. Convenient and interactive navigation within the network is ensured by options to search for gene names, zoom in and out, drag and drop genes and operons, move the graph within the screen and re-establish the original view. In addition, nodes and their corresponding edges can be hidden by a right-click of the mouse onto the node. The 'reset visibility' button allows the re-emergence of the hidden nodes.

Data sources and system structure

For the outlined purposes, the ProdoNet web application utilizes several different data sources (Figure 2). The main source is the PRODORIC database, which provides manually curated information on transcription factor binding sites, operon annotation and regulatory interactions between transcription factors and their corresponding binding sites. Further, PRODORIC supplies processed gene expression profiles that were thoroughly evaluated from the literature. In addition to these experimental evidences, predictions for operons and regulons were introduced into the ProdoNet dataset. Operon predictions were included based on the distance between various genes (16). Regulon predictions were performed with the Virtual Footprint algorithm using stringent specificity parameters to decrease the false prediction rate (17).



* Sources for enzyme and metabolic pathway data as described in the text.

Figure 2. ProdoNet structure: data sources and linked websites. The ProdoNet database is automatically created by data extraction from various source databases and addition of the results of operon and regulon predictions. The website provides links to PRODORIC, UniProtKB and KEGG.

The general gene annotation for the involved organisms was extracted from the Genome Review database (18) and the Pseudomonas Genome Database (19). This includes for each gene the name, the unique locus tag and identifiers in other databases, e.g. the UniProtKB accession number. Furthermore, identifiers from GenBank, RefSeq and the NCBI GeneID were gathered from the NCBI Microbial Genomes (20). Although these identifiers are not directly presented on the ProdoNet website, they are used to match

Table 1. Statistics of the ProdoNet Database, release 08.1

Organism	Operons (predicted)	Regulons	Regulatory interactions (predicted)	Expression profiles
B. subtilis (168)	2008 (1590)	84	594 (67)	34
E. coli (K12)	2192 (1640)	78	937 (196)	57
P. aeruginosa (PAO1)	2676 (2518)	31	210 (47)	16

the user's input to the ProdoNet gene dataset. Enzyme and pathway information was extracted from KEGG, BioCyc (21) and ENZYME (22).

The extracted data were imported into an integrated database, which provides the basis system for all offered analyses. This ProdoNet database is freely available for download on our website. The data import is fully automated and regular releases ensure an up-to-date dataset. A summary of the current database statistics is given in Table 1.

APPLICATION OF ProdoNet: ANALYSIS OF THE MEMBRANE SUBPROTEOME OF *Pseudomonas aeruginosa*

To demonstrate the functionality of ProdoNet, we analysed a hitherto unpublished experimental dataset on the membrane subproteome of *P. aeruginosa* using ProdoNet. For this experiment, bacterial cells were grown as a biofilm and broken by a French press passage. Membranes were isolated by sucrose density centrifugation and proteins were digested with trypsin. The resulting peptides were separated by reverse phase chromatography and automatically sequenced by tandem mass spectrometry on a QTOF instrument. In total, 796 unique proteins were identified by searching the peptide fragmentation patterns against a *Pseudomonas* genome database with MASCOT (Matrix Science, London, UK). Further details of the experiment and the names of identified proteins are provided in the Supplementary Material.

The whole list of protein names was entered into ProdoNet and the network of operons and regulons was queried. The resulting network was analysed stepwise by zooming into regions of particular interest. In agreement with published findings, the LasR/RhlR regulon, which is central to the autoinducer-dependent quorum sensing regulatory network involved in biofilm formation, was found completely reconstructed (Figure 1A). It is well known that proteomic experiments usually detect only parts of functional networks. However, with the ProdoNet network view, it is possible to complete partially identified regulatory circuits.

Interestingly, the ArgR regulon revealed both positive and negative regulatory effects (Figure 3). Since ArgR can serve as either a repressor or an activator (23) and the proteome data clearly showed induced and repressed genes of the ArgR regulon, we assume that additional regulatory proteins are involved in the ArgR response. Similar observations were made for Fur, PrsA and AlgR regulons (data not shown). Consequently, ProdoNet helps to formulate new hypotheses for experimental testing.

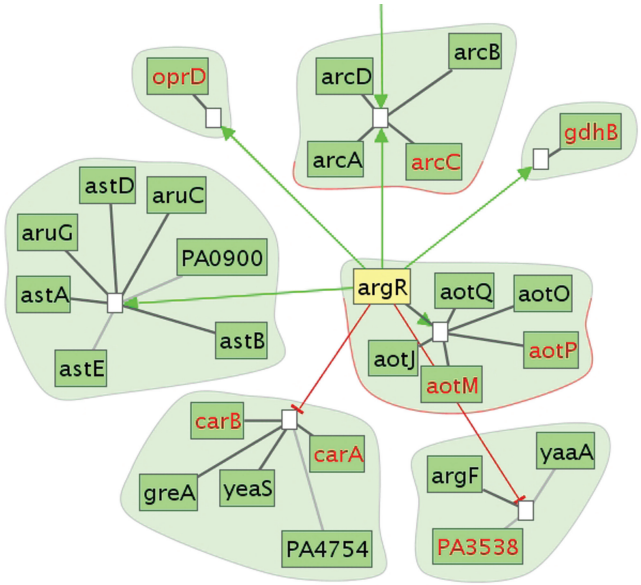


Figure 3. The ArgR regulon. The ProdoNet network view shows genes encoding for proteins that were found in the membrane subproteome of *P. aeruginosa* in the context of the ArgR regulon. These include genes from both activated and repressed operons.

IMPLEMENTATION

The ProdoNet website is generated by PHP scripts that operate on an Apache 2.2 web server. The underlying database is managed by a PostgreSQL relational database management system. The network view is created by use of a Java applet implemented with the prefuse library (14). For this reason, the website requires a browser with activated Java plugin (version 6 or higher), and JavaScript activation is necessary. The KEGG maps are generated by using the KEGG SOAP API (24).

CONCLUSION

ProdoNet constitutes a powerful tool in the functional exploration for lists of prokaryotic genes and proteins. Information on gene regulatory networks is integrated with gene expression profiles and information on metabolic pathways. In this way, ProdoNet provides a convenient and intuitive data analysis method that includes detailed information on the genes and proteins of interest. The current dataset comprises the well-established model organisms *E. coli*, *B. subtilis* and *P. aeruginosa*. In future versions, data on more bacterial species will be added.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Hedwig Schrader for excellent technical assistance and Dr Max Schobert for fruitful discussions on regulatory networks in *P. aeruginosa*. Special thanks go to Isam Haddad for sharing his outstanding expertise on Java applets and libraries. This study was funded by German Bundesministerium für Bildung und Forschung (ERA-NET grant no 0313936C to J.K., NGFN2-EP grant no. 0313398A to C.P) and the Volkswagen Foundation (to S.L. and R.M.). Funding to pay the Open Access publication charges for this article was provided by the Volkswagen Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Gama-Castro,S., Jiménez-Jacinto,V., Peralta-Gil,M., Santos-Zavaleta,A., Peñaloza-Spinola,M.I., Contreras-Moreira,B., Segura-Salazar,J., Muñiz-Rascado,L., Martínez-Flores,I., Salgado,H. *et al.* (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36** (Database issue), D120–D124.
2. Sierro,N., Makita,Y., de Hoon,M.J.L. and Nakai,K. (2008) DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.*, **36** (Database issue), D93–D96.
3. Kazakov,A.E., Cipriano,M.J., Novichkov,P.S., Minovitsky,S., Vinogradov,D.V., Arkin,A., Mironov,A.A., Gelfand,M.S. and Dubchak,I. (2007) RegTransBase – a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. *Nucleic Acids Res.*, **35** (Database issue), D407–D412.
4. Münch,R., Hiller,K., Barg,H., Heldt,D., Linz,S., Wingender,E. and Jahn,D. (2003) PRODORIC: prokaryotic database of gene regulation. *Nucleic Acids Res.*, **31**, 266–269.
5. Pérez,A.G., Angarica,V.E., Vasconcelos,A.T. and Collado-Vides,J. (2007) Tractor_DB (version 2.0): a database of regulatory interactions in gamma-proteobacterial genomes. *Nucleic Acids Res.*, **35** (Database issue), D132–D136.
6. Pareja,E., Pareja-Tobes,P., Manrique,M., Pareja-Tobes,E., Bonal,J. and Tobes,R. (2006) ExtraTrain: a database of extragenic regions and transcriptional information in prokaryotic organisms. *BMC Microbiol.*, **6**, 29.
7. Pachkov,M., Erb,I., Molina,N. and van Nimwegen,E. (2007) SwissRegulon: a database of genome-wide annotations of regulatory sites. *Nucleic Acids Res.*, **35** (Database issue), D127–D131.
8. Baumbach,J. (2007) CoryneRegNet 4.0 – a reference database for corynebacterial gene regulatory networks. *BMC Bioinform.*, **8**, 429.
9. Dieterich,G., Kärst,U., Wehland,J. and Jänsch,L. (2006) VIS-O-BAC: exploratory visualization of functional genome studies from bacteria. *Bioinformatics*, **22**, 630–631.
10. von Mering,C., Jensen,L.J., Kuhn,M., Chaffron,S., Doerks,T., Krüger,B., Snel,B. and Bork,P. (2007) STRING 7 – recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res.*, **35** (Database issue), D358–D362.
11. Suderman,M. and Hallett,M. (2007) Tools for visually exploring biological networks. *Bioinformatics*, **23**, 2651–2659.
12. UniProt Consortium (2008) The universal protein resource (UniProt). *Nucleic Acids Res.*, **36** (Database issue), D190–D195.
13. Kanehisa,M., Araki,M., Goto,S., Hattori,M., Hirakawa,M., Itoh,M., Katayama,T., Kawashima,S., Okuda,S., Tokimatsu,T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36** (Database issue), D480–D484.
14. Heer,J., Card,S.K. and Landay,J.A. (2005) Prefuse: a toolkit for interactive information visualization. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 421–430.
15. Cline,M.S., Smoot,M., Cerami,E., Kuchinsky,A., Landys,N., Workman,C., Christmas,R., Avila-Campilo,I., Creech,M., Gross,B. *et al.* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.
16. Moreno-Hagelsieb,G. and Collado-Vides,J. (2002) A powerful non-homology method for the prediction of operons in prokaryotes. *Bioinformatics*, **18** (Suppl 1), S329–S336.
17. Münch,R., Hiller,K., Grote,A., Scheer,M., Klein,J., Schobert,M. and Jahn,D. (2005) Virtual Footprint and PRODORIC: an integrative framework for regulon prediction in prokaryotes. *Bioinformatics*, **21**, 4187–4189.
18. Sterk,P., Kersey,P.J. and Apweiler,R. (2006) Genome reviews: standardizing content and representation of information about complete genomes. *OMICS*, **10**, 114–118.
19. Winsor,G.L., Lo,R., Sui,S.J., Ung,K.S., Huang,S., Cheng,D., Ching,W.K., Hancock,R.E. and Brinkman,F.S. (2005) *Pseudomonas aeruginosa* Genome Database and PseudoCAP: facilitating community-based, continually updated, genome annotation. *Nucleic Acids Res.*, **33** (Database issue), D338–D343.
20. Wheeler,D.L., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Edgar,R., Federhen,S. *et al.* (2008) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **36** (Database issue), D13–D21.
21. Caspi,R., Foerster,H., Fulcher,C.A., Kaipa,P., Krummenacker,M., Latendresse,M., Paley,S., Rhee,S.Y., Shearer,A.G., Tissier,C. *et al.* (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.*, **36** (Database issue), D623–D631.
22. Bairoch,A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.*, **28**, 304–305.
23. Lu,C.D., Yang,Z. and Li,W. (2004) Transcriptome analysis of the ArgR regulon in Pseudomonas aeruginosa. *J. Bacteriol.*, **186**, 3855–3861.
24. Kanehisa,M., Goto,S., Hattori,M., Aoki-Kinoshita,K.F., Itoh,M., Kawashima,S., Katayama,T., Araki,M. and Hirakawa,M. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.*, **34** (Database issue), D354–D357.