D610–D614 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 12 November 2011*
doi:10.1093/nar/gkr883

CoryneRegNet 6.0—Updated database content, new analysis methods and novel features focusing on community demands

Josch Pauling[1,2], Richard Röttger[1], Andreas Tauch[3], Vasco Azevedo[4] and Jan Baumbach[1,2,*]

[1]Computational Systems Biology, Max Planck Institute for Informatics, Campus E1.4, 66123 Saarbrücken, [2]Cluster of Excellence for Multimodal Computing and Interaction, Saarland University, Campus E1.7, 66123 Saarbrücken, [3]Center for Biotechnology, Bielefeld University, D-33594 Bielefeld, Germany and [4]Instituto de Ciencias Biologicas, Universidade Federal de Minas Gerais, Belo Horizonte MG, Brazil

Received August 11, 2011; Revised September 13, 2011; Accepted October 2, 2011

ABSTRACT

Post-genomic analysis techniques such as next-generation sequencing have produced vast amounts of data about micro organisms including genetic sequences, their functional annotations and gene regulatory interactions. The latter are genetic mechanisms that control a cell's characteristics, for instance, pathogenicity as well as survival and reproduction strategies. CoryneRegNet is the reference database and analysis platform for corynebacterial gene regulatory networks. In this article we introduce the updated version 6.0 of CoryneRegNet and describe the updated database content which includes, 6352 corynebacterial regulatory interactions compared with 4928 interactions in release 5.0 and 3235 regulations in release 4.0, respectively. We also demonstrate how we support the community by integrating analysis and visualization features for transiently imported custom data, such as gene regulatory interactions. Furthermore, with release 6.0, we provide easy-to-use functions that allow the user to submit data for persistent storage with the CoryneRegNet database. Thus, it offers important options to its users in terms of community demands. CoryneRegNet is publicly available at http://www.coryneregnet.de.

INTRODUCTION

The utilization of high-throughput next-generation sequencing techniques has made huge amounts of genome data publicly available through numerous life science databases and data warehouses. To understand a cell's response mechanisms to constantly changing environmental influences, it is vital to analyse the data and extract both the expression of genes and the encoded proteins that play a role in the organism's gene regulation [transcription factors, (TFs)]. Putting together these pieces in conjunction with literature-derived knowledge on the regulation of gene expression has made the genome-wide reconstruction of transcriptional regulatory networks possible. Unraveling these networks on a large-scale facilitates a detailed understanding of an organism's adaptation strategies to changing environmental conditions. Publicly available reference databases provide well-structured access to known regulatory interactions as well as integrated analysis features. For prokaryotes, three typical examples are RegulonDB (1–8) for *Escherichia coli*, PRODORIC (9–12) mainly for *Pseudomonas aruginosa* and MycoRegNet (13) for *Mycobacterium tuberculosis*. For corynebacteria, CoryneRegNet is the reference database and analysis platform (14–19). Corybnebacteria are important in biotechnological production processes as well as human and veterinary medicine; see Refs (20–23) for instance. CoryneRegNet is an ontolgy-based data warehouse. It consists of a hierarchically multi-layered concept of transcriptional regulation and utilizes a modular data processing pipeline integrating several software tools that are essential for regulatory network reconstruction such as detection of clusters of homologous proteins (24), binding site motif matching (25), operon prediction (26) and special network visualization and graph analysis algorithms.

In this updated article, we describe further improvements leading to CoryneRegNet version 6.0. The main goal was to implement crucial features fulfilling important community needs and adding user-customized regulatory

*To whom correspondence should be addressed. Tel: +49 681 302 70880; Fax: +49 681 9325 399; Email: jbaumbac@mpi-inf.mpg.de

interactions thus further increasing CoryneRegNet's applicability to typical biological studies. We describe the following updates to CoryneRegNet:

(1) Updated database content including new regulations and all fully sequenced and annotated corynebacteria.
(2) Predicted version (CoryneRegNet v6.0p): updated inter species network transfer pipeline to fully support the new database content.
(3) New options for visualizing transiently imported custom gene expression and gene regulation data with GraphVis that was not deposited with the CoryneRegNet database yet.
(4) Users are now allowed to persistently submit their own data directly to the CoryneRegNet curation team by using new options in the web interface.

In the following, we briefly depict the updated system architecture with a detailed explanation of the above listed items. Subsequently, we discuss the improvements that CoryneRegNet has gained in its current version 6.0. We summarize the updated database content and briefly show both, the transient and the persistent data import facilities. Finally, we sum up and conclude with how CoryneRegNet 6.0 broadens its applicability by extending its functionality based on community demands thus further increasing its contribution to the corynebacteria community.

MATERIALS AND METHODS

Database content and technical changes

Figure 1 shows a timeline of how CoryneRegNet evolved over its previous versions and how this shifted the focus of analysis interests.

During CoryneRegNet's development process, many different applications for this platform's methodology have been tested. For CoryneRegNet version 6.0, as for previous releases, we updated the database to include all fully sequenced and annotated corynebacteria as well as *E. coli* K-12. Table 1 shows the development of the database content from version 1.0 to the current version 6.0. CoryneRegNet 5.0 and 6.0 are available as an 'experimental' and a 'predicted' version; the latter incorporating the results of the inter-species network transfer pipeline that we implemented. CoryneRegNet's strategy for predicting high potential gene regulations is based on the evolutionary conservation of regulatory sites between a target organism and a well-researched model organism, see Refs (21,27) for details. In CoryneRegNet version 6.0, we modified the transfer pipeline to fully support the updated database content. We use the *C. glutamicum* regulatory network as the template for transfers of known regulations to all corynebacterial target organisms.

To date, CoryneRegNet 6.0's database content includes *C. aurimucosum* ATCC 700975, *C. diphtheriae* NCTC 13129, *C. efficiens* YS-314, *C. glutamicum* ATCC 13032, *C. glutamicum* R, *C. jeikeium* K411, *C. kroppenstedtii* DSM 44385, *C. pseudotuberculosis* 1002, *C. pseudotuberculosis* C231, *C. pseudotuberculosis* FRC41, *C. urealyticum* DSM 7109 and *Escherichia coli* K-12 while CoryneRegNet 4.0 included *C. diphtheriae* NCTC 13129, *C. efficiens* YS-314, *C. glutamicum* ATCC 13032, *C. glutamicum* R, *C. jeikeium* K411, *C. urealyticum* DSM 7109, *E. coli* K12, *M. tuberculosis* CDC1551 and *M. tuberculosis* H37Rv. In release 5.0, we utilized the network data from *C. glutamicum* for inter-species transfers to *C. diphtheriae* NCTC 13129, *C. efficiens* YS-314 and *C. jeikeium* K411. With CoryneRegNet 6.0, we now provide transferred networks for all included corynebacteria.

Visualizing transient custom data via GraphVis

One new functionality of CoryneRegNet's web interface is the visualization and analysis of custom user-derived critical data. This comprises regulatory interactions that are not integrated with the database yet (unpublished data) as well as gene expression data. Therefore, we changed CoryneRegNet's visualization tool GraphVis to allow the user to import his own gene expression and gene regulation data. GraphVis runs client-sided as a browser applet thereby ensuring discrete handling of this user's

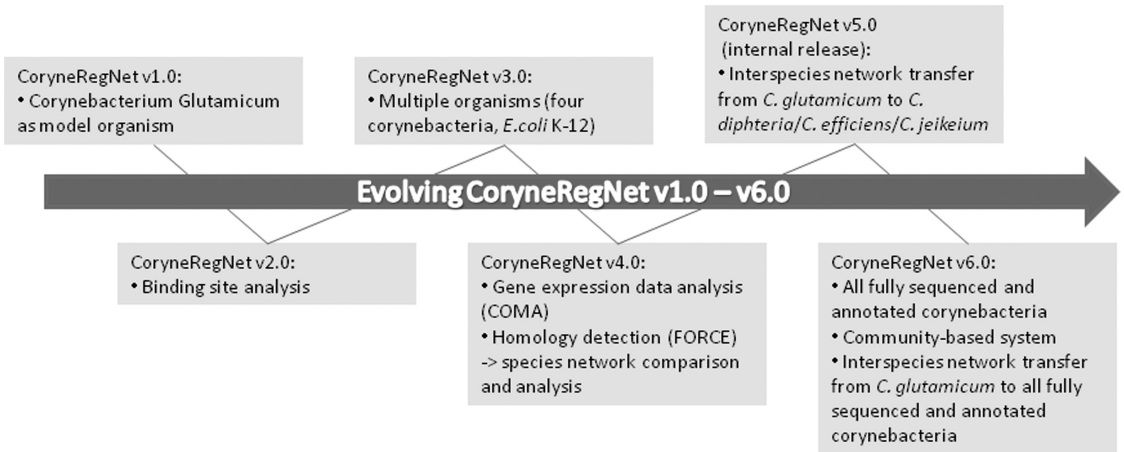

Figure 1. Development stages of CoryneRegNet—from 2006 to 2011.

Table 1. Development of the database content

Version	Organisms	TFs	Reg. genes	Regulations	BMs	PWMs	Stimulons	Clusters	Publication
1.0	1	53	331	430	192	23	–	–	(14)
2.0	4	64	499	607	274	29	–	–	(18)
3.0	5	213	1632	2912	1522	130	–	–	(15)
4.0	7	213	1632	2912	1522	130	8	4548	(16)
5.0e	11	245	1986	3712	1759	144	11	5421	–
5.0p	11	350	2888	4928	2553	249	11	5421	–
6.0e	12	245	1986	3712	1759	144	14	3719	This
6.0p	12	482	3946	6352	3429	381	14	3719	This

Note that the database content of CoryneRegNet 5.0e and 6.0e are equal regarding the number of regulations since we gradually kept the database up-to-date, as we will do in the future; TFs, transcription factors; Reg. genes, regulated target genes; BMs, binding motifs; PWMs, position weight matrices.

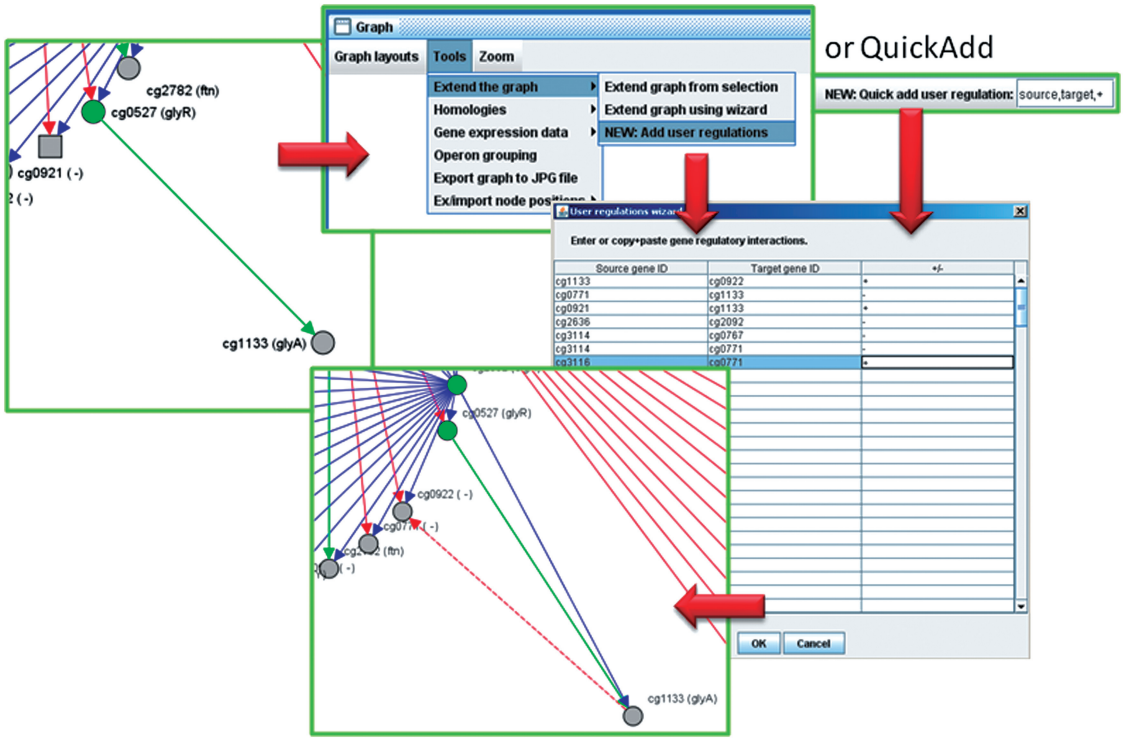

Figure 2. Custom data visualization process. When visualizing an arbitrary known (database) network with the GraphVis feature, the user may utilize several ways to integrate custom data, gene expression data or, as depicted here, transcriptional regulatory interactions. The latter are displayed as dotted lines. This allows for a visual analysis of gene expression data, for instance, together with known (database) and uploaded (user) networks prior to publication.

potentially critical data. As applets run in their own runtime environment, memory is allocated locally and therefore no information about custom data sets is ever transferred over the network. The process of visualizing user regulations together with known gene regulatory networks is easy and intuitive. It is depicted in Figure 2.

Persistent submission of user data to the CoryneRegNet database

Being the reference database and analysis platform for corynebacterial gene regulatory networks, it is important to constantly expand the database content with user-derived data. As of yet, we exclusively relied on our data curation team. CoryneRegNet version 6.0 now also allows its users to directly submit their own data persistently to the CoryneRegNet database. By utilizing an intuitive web front-end, integrated in the CoryneRegNet platform, users may send their data sets to the database curators. We decide whether or not the data meets the CoryneRegNet's quality standards and if so, accept the new data set. Once accepted the data is directly added to the CoryneRegNet database persistently. This allows for a well-structured and faster provision of new data without the error-prone detour via literature mining. This assures CoryneRegNet's applicability to the most recent data sets on ongoing matters. With this feature, CoryneRegNet eases the information exchange and further improves its contribution to the (corynebacteria) community.

RESULTS AND DISCUSSION

With CoryneRegNet version 6.0, we updated the database content by adding all recently sequenced corynebacteria to the database. As we did in the past, the database will be kept up-to-date as more regulations are deciphered in the wet lab and more corynebacterial whole-genome annotations become available. To date, CoryneRegNet is the largest available knowledge base for corynebacterial gene regulatory interactions world wide. Moreover, our inter-species network transfer integration ('predicted' version), roughly doubles the database content of the 'experimental' version of CoryneRegNet.

In addition to the growing database content, we also extended our network tool GraphVis to allow its users to import, visualize and analyse their own gene expression and gene regulation data with full functionality. This mainly allows the biologists to visualize and analyse their own, yet unpublished, findings together with known data, such as regulatory interactions, from the database. It is ensured that the data are kept transiently on the client only during runtime and no information is exchanged over the network. Additionally, users can submit their own data to the CoryneRegNet database for persistent storage. Here, we provide the community with a direct, self-explanatory interface and thereby limit the potential for typing mistakes. This allows for a faster handling of evolving new data in a well-structured way.

CONCLUSION

CoryneRegNet 6.0 is a versatile systems biology platform to support the efficient and large-scale analysis of transcriptional regulation of gene expression in corynebacterial microorganisms. With the aforementioned changes, we extended the CoryneRegNet platform to be a community-based system. In conclusion, CoryneRegNet remains the reference database and analysis platform for corynebacterial transcriptional gene regulatory networks.

FUNDING

DFG Cluster of Excellence for Multimodel Computing and Interaction (MMCI) and International Max Planck Research School (IMPRS). Funding for open access charge: German Research Foundation, Max-Planck-Society.

Conflict of interest statement. None declared.

REFERENCES

1. Gama-Castro,S., Salgado,H., Peralta-Gil,M., Santos-Zavaleta,A., Muñiz-Rascado,L., Solano-Lira,H., Jimenez-Jacinto,V., Weiss,V., García-Sotelo,J.S., López-Fuentes,A. *et al.* (2011) RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Res.*, **39**, D98–D105.
2. Huerta,A.M., Salgado,H., Thieffry,D. and Collado-Vides,J. (1998) RegulonDB: a database on transcriptional regulation in Escherichia coli. *Nucleic Acids Res.*, **26**, 55–59.
3. Salgado,H., Santos,A., Garza-Ramos,U., van Helden,J., Díaz,E. and Collado-Vides,J. (1999) RegulonDB (version 2.0): a database on transcriptional regulation in Escherichia coli. *Nucleic Acids Res.*, **27**, 59–60.
4. Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millán-Zárate,D., Blattner,F.R. and Collado-Vides,J. (2000) RegulonDB (version 3.0): transcriptional regulation and operon organization in Escherichia coli K-12. *Nucleic Acids Res.*, **28**, 65–67.
5. Salgado,H., Santos-Zavaleta,A., Gama-Castro,S., Millán-Zárate,D., Díaz-Peredo,E., Sánchez-Solano,F., Pérez-Rueda,E., Bonavides-Martínez,C. and Collado-Vides,J. (2001) RegulonDB (version 3.2): transcriptional regulation and operon organization in Escherichia coli K-12. *Nucleic Acids Res.*, **29**, 72–74.
6. Salgado,H., Gama-Castro,S., Martínez-Antonio,A., Díaz-Peredo,E., Sánchez-Solano,F., Peralta-Gil,M., Garcia-Alonso,D., Jiménez-Jacinto,V., Santos-Zavaleta,A., Bonavides-Martínez,C. *et al.* (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in Escherichia coli K-12. *Nucleic Acids Res.*, **32**, D303–D306.
7. Salgado,H., Gama-Castro,S., Peralta-Gil,M., Díaz-Peredo,E., Sánchez-Solano,F., Santos-Zavaleta,A., Martínez-Flores,I., Jiménez-Jacinto,V., Bonavides-Martínez,C., Segura-Salazar,J. *et al.* (2006) RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.*, **34**, D394–D397.
8. 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**, D120–D124.
9. 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.
10. 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.
11. Grote,A., Klein,J., Retter,I., Haddad,I., Behling,S., Bunk,B., Biegler,I., Yarmolinetz,S., Jahn,D. and Münch,R. (2009) PRODORIC (release 2009): a database and tool platform for the analysis of gene regulation in prokaryotes. *Nucleic Acids Res.*, **37**, D61–D65.
12. 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.
13. Krawczyk,J., Kohl,T.A., Goesmann,A., Kalinowski,J. and Baumbach,J. (2009) From Corynebacterium glutamicum to Mycobacterium tuberculosis–towards transfers of gene regulatory networks and integrated data analyses with MycoRegNet. *Nucleic Acids Res.*, **37**, e97.
14. Baumbach,J., Brinkrolf,K., Czaja,L.F., Rahmann,S. and Tauch,A. (2006) CoryneRegNet: an ontology-based data warehouse of corynebacterial transcription factors and regulatory networks. *BMC Genomics*, **7**, 24.
15. Baumbach,J., Wittkop,T., Rademacher,K., Rahmann,S., Brinkrolf,K. and Tauch,A. (2007) CoryneRegNet 3.0–an interactive systems biology platform for the analysis of gene regulatory networks in corynebacteria and Escherichia coli. *J. Biotechnol.*, **129**, 279–289.
16. Baumbach,J. (2007) CoryneRegNet 4.0 - A reference database for corynebacterial gene regulatory networks. *BMC Bioinformatics*, **8**, 429.
17. Neuweger,H., Baumbach,J., Albaum,S., Bekel,T., Dondrup,M., Hüser,A.T., Kalinowski,J., Oehm,S., Pühler,A., Rahmann,S. *et al.* (2007) CoryneCenter - an online resource for the integrated analysis of corynebacterial genome and transcriptome data. *BMC Syst. Biol.*, **1**, 55.
18. Baumbach,J., Brinkrolf,K., Wittkop,T., Tauch,A. and Rahmann,S. (2006) CoryneRegNet 2: an integrative bioinformatics approach for reconstruction and comparison of transcriptional regulatory networks in prokaryotes. *J. Int. Bioinformatics*, **3**, 24.
19. Baumbach,J., Wittkop,T., Kleindt,C.K. and Tauch,A. (2009) Integrated analysis and reconstruction of microbial transcriptional

gene regulatory networks using CoryneRegNet. *Nat. Protoc.*, **4**, 992–1005.
20. Schröder,J. and Tauch,A. (2010) Transcriptional regulation of gene expression in Corynebacterium glutamicum: the role of global, master and local regulators in the modular and hierarchical gene regulatory network. *FEMS Microbiol. Rev.*, **34**, 685–737.
21. Baumbach,J. (2010) On the power and limits of evolutionary conservation–unraveling bacterial gene regulatory networks. *Nucleic Acids Res.*, **38**, 7877–7884.
22. Ruiz,J.C., D'Afonseca,V., Silva,A., Ali,A., Pinto,A.C., Santos,A.R., Rocha,A.A.M.C., Lopes,D.O., Dorella,F.A., Pacheco,L.G.C. *et al.* (2011) Evidence for reductive genome evolution and lateral acquisition of virulence functions in two Corynebacterium pseudotuberculosis strains. *PLoS ONE*, **6**, e18551.
23. Cerdeira,L.T., Carneiro,A.R., Ramos,R.T.J., de Almeida,S.S., Afonseca,V.D., Schneider,M.P.C., Baumbach,J., Tauch,A., McCulloch,J.A., Azevedo,V.A.C. *et al.* (2011) Rapid hybrid de novo assembly of a microbial genome using only short reads: Corynebacterium pseudotuberculosis I19 as a case study. *J. Microbiol. Methods*, **86**, 218–223.
24. Wittkop,T., Emig,D., Truss,A., Albrecht,M., Böcker,S. and Baumbach,J. (2011) Comprehensive cluster analysis with Transitivity Clustering. *Nat. Protoc.*, **6**, 285–295.
25. Beckstette,M., Homann,R., Giegerich,R. and Kurtz,S. (2006) Fast index based algorithms and software for matching position specific scoring matrices. *BMC Bioinformatics*, **7**, 389.
26. Dehal,P.S., Joachimiak,M.P., Price,M.N., Bates,J.T., Baumohl,J.K., Chivian,D., Friedland,G.D., Huang,K.H., Keller,K., Novichkov,P.S. *et al.* (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.*, **38**, D396–D400.
27. Baumbach,J., Rahmann,S. and Tauch,A. (2009) Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms. *BMC Syst. Biol.*, **3**, 8.