

Pseudomonas Genome Database: improved comparative analysis and population genomics capability for *Pseudomonas* genomes

Geoffrey L. Winsor¹, David K. W. Lam¹, Leanne Fleming¹, Raymond Lo¹,
Matthew D. Whiteside¹, Nancy Y. Yu¹, Robert E. W. Hancock² and
Fiona S. L. Brinkman^{1,*}

¹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC V5A 1S6 and

²Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

Received August 25, 2010; Accepted September 15, 2010

ABSTRACT

Pseudomonas is a metabolically-diverse genus of bacteria known for its flexibility and leading free living to pathogenic lifestyles in a wide range of hosts. The Pseudomonas Genome Database (<http://www.pseudomonas.com>) integrates completely-sequenced *Pseudomonas* genome sequences and their annotations with genome-scale, high-precision computational predictions and manually curated annotation updates. The latest release implements an ability to view sequence polymorphisms in *P. aeruginosa* PAO1 versus other reference strains, incomplete genomes and single gene sequences. This aids analysis of phenotypic variation between closely related isolates and strains, as well as wider population genomics and evolutionary studies. The wide range of tools for comparing *Pseudomonas* annotations and sequences now includes a strain-specific access point for viewing high precision computational predictions including updated, more accurate, protein subcellular localization and genomic island predictions. Views link to genome-scale experimental data as well as comparative genomics analyses that incorporate robust genera-gated methods for predicting and clustering orthologs. These analyses can be exploited for identifying putative essential and core *Pseudomonas* genes or identifying large-scale evolutionary events. The Pseudomonas Genome Database aims to provide a continually updated, high quality source of genome annotations, specifically tailored for *Pseudomonas* researchers, but using an approach that may be implemented for other genera-level research communities.

INTRODUCTION

Pseudomonas is a genus of bacteria known for its metabolic capacity and ability to occupy a wide range of environments as free-living soil microorganisms or as serious opportunistic pathogens. Since the mosaic-like structure of its members' genomes can influence niche and degree of virulence (1–3), it exhibits the ability to horizontally transfer elements between strains (4) and species can have notable broad-based antimicrobial resistance, the genus is widely studied at the genome level.

The Pseudomonas Community Annotation Project (PseudoCAP) was originally formed to meet the need of providing a conservative, peer-reviewed annotation of the *Pseudomonas aeruginosa* PAO1 genome sequence using an internet-based approach coupled with community-assisted genome annotation. This led to the development of the Pseudomonas Genome Database (<http://www.pseudomonas.com>), which initially was specific for *P. aeruginosa* (5–7). As development of the database continued in parallel with an increase in the number of *Pseudomonas* genomes being sequenced, PseudoCAP recognized the importance of capitalizing on insight obtained through comparative genome analysis (8). By developing comparative analyses for a specific genus like *Pseudomonas*, methods can be developed that are custom made for their level of divergence, generating higher precision comparisons between closely-related species that would otherwise be difficult to achieve when hosting data from more diverse taxonomic sources (8). Databases including the Tuberculosis Database Project (9) and the *Burkholderia* Genome Database (10) focus on providing tools to facilitate comparison with multiple genomes from smaller taxonomic groups and resources more specifically geared for the associated research community interests.

An overview is provided here of the latest updates to the Pseudomonas Genome Database, including a

*To whom correspondence should be addressed. Tel: +778 782 5646; Fax: +778 782 5583; Email: brinkman@sfu.ca

framework for viewing polymorphisms in closely-related *P. aeruginosa* genome sequences, strain-specific portals for accessing whole-genome data (including experimental data) and updated high-precision computational predictions and comparative genome analyses based on robust methods for predicting and clustering orthologs. The *Pseudomonas* Genome Database aims to provide a continually updated, curated source of genome annotations, improved precise computational predictions, genome-scale experimental data and integrated annotations from external databases, tailored specifically for the *Pseudomonas* research community, but using a framework that could be utilized by other similar research communities and groups.

NEW FEATURES

Ability to view polymorphisms/SNPs in closely-related *Pseudomonas* genomes

Next-generation sequencing technologies have already demonstrated that *P. aeruginosa* reference strain PAO1 isolates are undergoing microevolution contributing to multiple phenotypic differences (11). With the release of incompletely sequenced genomes outpacing completed genomes it has become important to apply a comprehensive, yet conservative approach to comparing reference genomes against these kinds of sequences. As a solution, we developed a framework for aligning partial genome sequences or single genes against the PAO1 reference sequence and indexing positions where polymorphisms are located. The MUMmer 3.22 (12) program NUCmer was used to align genomic DNA or cDNA sequences from incomplete *P. aeruginosa* genomes or individual polymorphic sequences available at the National Center for Biotechnology Information (NCBI) whereas the MUMmer program show that SNPs were used to identify SNPs and indels in non-repetitive regions. Once identified at the genomic level, polymorphisms were classified according to type of mutation (e.g. missense, silent, etc.) and its location within the gene's nucleotide and protein sequences, along with the downstream effect on the amino acid properties being recorded. Polymorphisms in genomic DNA, cDNA and amino acid sequences were also assigned a standardized description following recommendations made by the Human Genome Variation Society (13), a standard adopted by other SNP databases including dbSNP (14) and Ensembl (15). Access to this data is currently under the 'polymorphisms' tab of any *P. aeruginosa* PAO1 gene (or 'gene card') page where an overview of point mutations is presented along with an image of the specific sequence containing hyperlinked mutations at each site. The nucleotides at these sites are colored in order to distinguish synonymous and non-synonymous mutations from insertions and deletions, whereas a single strain can be selected from an adjacent list in order to apply a filter that hides all other strains. Links are provided to more detailed point mutation descriptions, experimental details, sequence or alignment downloads and effects that a mutation would have on amino acid properties. Links to a GBrowse genome viewer (16) representation of the

local genomic landscape is provided, whereby all known SNPs/indels are annotated and colored by type. This presentation can be quite useful for visualizing regions undergoing a putatively higher degree of selection, as represented by a clustering of non-synonymous mutations (for example, see OprD at <http://www.pseudomonas.com/getAnnotation.do?locusID=PA0958>). Finally, we offer additional SNP information including the origin of the SNP call (e.g. experimental method such as restriction digest or computational prediction based on sequence alignment) to help provide users with a level of confidence regarding a given result. This can be easily expanded in the future to handle a wider range of *Pseudomonas* species and incorporate additional information for SNP validation and quality scores.

New search options including a portal to strain-specific analyses and experimental data, plus expanded comparative genomic analysis capabilities

All annotations may be searched using either the simple or advanced Boolean-based search tools which have recently been enhanced with an option to filter the list of genes returned to a specified size range or protein subcellular localization of the gene products. Other filters limit for properties such as genes encoding known drug targets (based on manually curated data) or also putatively essential genes as determined by saturation transposon mutagenesis [currently *P. aeruginosa* only (17–19)].

We have also recently introduced a higher-level portal to strain-specific, whole-genome analyses and experimental data. By linking from the home page, one can go to an overview of any reference strain's chromosome or plasmids with a summary of their genes, broken down by feature type or its product's primary subcellular localization. It also serves as a starting point for other strain-related searches including browsing by function, localization, virulence factors, drug targets and genes located in genomic islands or by performing sequence searches. Under the 'Tools' tab, it facilitates an IUPAC-formatted DNA motif search against other *Pseudomonas* reference genomes to return a list of genomic features containing that motif which can be applicable to finding putative transcription factor binding sites and other interesting motifs. The 'Experimental data' tab provides a starting point for identifying expression data studies and linking to respective data at NCBI's Gene Expression Omnibus (20) or ArrayExpress (21) or to sequence read data at NCBI's Trace Archive or Short Reads Archive (22). Finally, the strain-specific access page provides an entry point to whole-genome comparative analyses including a high-precision set of putative orthologs and sets of human homologs which are of particular relevance to researchers investigating *Pseudomonas*-specific drug targets. The orthologous genes set consists of *Pseudomonas* genes mapped to their respective orthologs in other *Pseudomonas* species using our *Pseudomonas* Orthologous Genes (POGs) classification (8) with further assessment by a high-precision method called Ortholuge which examines phylogenetic distance ratios between two

comparison species and an outgroup species (23). This Orthologue method was also recently improved (24).

We have recently added a search feature that enables users to perform their own comparative analysis of putatively orthologous genes. This should help address questions including what orthologous genes are present in one set of *Pseudomonas* genomes that are not present in another or find genes in one strain that have no orthologs in the other strains. In order to perform this analysis, users simply indicate which species they would like to have orthologs returned for by selecting from a list of genomes. They can optionally make a selection that limits results to putatively essential genes or genes with/without human homologs and even specify results matching multiple keywords specified in a Boolean search form.

Precise computational predictions added: improved protein subcellular localization prediction and identification of genomic islands

Our desire to make high quality *Pseudomonas* annotations available has led to the integration of computational predictions based on high-precision methods. We have recently updated our computational predictions of protein subcellular localization data based on PSORTb version 3.0 software (25) containing enhancements over our very successful PSORTb version 2.0 (26). PSORTb 3.0 adds several new sub-category localizations related to bacterial organelles while differentiating proteins targeted to a host cell. The new version also exhibits higher sensitivity and genome prediction coverage compared with the previous version, with an average of 15% genome coverage increase for Gram-negative species over PSORTb 2.0. For *P. aeruginosa* annotations, additional experimentally demonstrated localizations in *P. aeruginosa* or highly similar proteins from closely related species are made available in place of PSORTb predictions and are assigned a value based on the degree of confidence in the localization.

Pseudomonas genomes are a mosaic of horizontally-transferred genes (e.g. arising from genomic islands) which play an important role in its species' adaptation to environmental niches. To identify genomic islands (GIs), our lab developed IslandViewer (27), a computational tool integrating two sequence composition GI prediction methods, SIGI-HMM (28) and IslandPath-DIMOB (29), with a comparative GI prediction method, IslandPick (30). Views of the data in IslandViewer have been integrated into the *Pseudomonas* Genome Database from a 'browse genomic island' page and from individual strain summary pages. These sections link to IslandViewer website pages containing circular chromosome images overlaid with details of genomic islands identified by the various methods with links to download lists of more detailed results about genes within islands, for further analysis.

Integration of analyses from external sources

For all genomes, we now provide precise operon predictions based on the Database of Prokaryotic Operons (DOOR), rated one of the best programs for

operon prediction (31,32). We also incorporated updated Rho-independent transcription terminator predictions using TransTermHP (33) and computationally-identified inverted repeats (palindromes) using EMBOSS's Palindrome software (34). The database also hosts data from three transposon mutant libraries based on *P. aeruginosa* PAO1 (17,18) and *P. aeruginosa* PA14 (19) and will continue to add relevant data from these and other strains as it arises in order to form a foundation for identifying putative essential genes.

Curated updates to genome annotations

Our continual updates to genome annotations come from a variety of sources including in-house curation, submission from members of the *Pseudomonas* research community, curators belonging to other *Pseudomonas* sequencing projects and directly from NCBI, where many genome centers directly submit their annotation updates. Since 2001, more than 2000 annotation updates have been made to the *P. aeruginosa* PAO1 annotation through a variety of curation methods described earlier (7). Annotation updates are also made to other *Pseudomonas* genomes in the database through the process of contacting curators at other genome centers including those responsible for the ongoing annotation of the *P. syringae* DC3000 (35), *P. aeruginosa* PA14 (2) and *P. aeruginosa* LESB58 (36) reference strains. The quality of annotations available and flexibility of ways to view results is widely recognized. Large annotation databases, including the Comprehensive Microbial Resource (J. Craig Venter Institute) (37), UniProt (38) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) (39), as well as *Pseudomonas*-specific resources including the *P. syringae* project and Systemomonas (40) link to views of data on our website or have requested our curated data sets of updated gene annotation information.

FUTURE DEVELOPMENT

In addition to continually updating genome annotations with a focus on *P. aeruginosa* curation, we will continue to add new strains and update other *Pseudomonas* species annotations from publicly available repositories and *Pseudomonas* genome sequencing centers. We plan to extend the scope of the database to include a system for viewing and interrogating microarray expression data and RNA-seq-based transcriptome data in a more sophisticated manner. We are also developing a parallel interaction database using the InnateDB framework we previously developed for human and mouse innate immunity (and larger proteome) protein-protein interactions (41). Throughout these and other future efforts, we aim to continue to provide a high quality, user-friendly and powerful resource for the *Pseudomonas* research community.

AVAILABILITY

All features of this database are fully accessible to the public. The source code is freely available under the GNU GPL license.

ACKNOWLEDGEMENTS

We thank Dr Jens Klockgether (Hannover Medical School, Germany) and Dr Shannan Ho Sui (SFU, Canada) for their feedback on the website presentation of SNP data. We also thank all 150 community annotation update participants (listed at <http://www.pseudomonas.com/researchList.jsp>) for their valuable contributions and all the *Pseudomonas* genome projects, without which this database would not be possible.

FUNDING

The Cystic Fibrosis Foundation (Cystic Fibrosis Foundation Therapeutics) with additional support for some tool development by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the SFU Community Trust; Michael Smith Foundation for Health Research (MSFHR), Junior Graduate Studentship Award (to M.D.W.); F.S.L.B. is a MSFHR Senior Scholar and R.E.W.H. holds a Canada Research Chair. Funding for open access charge: Cystic Fibrosis Foundation and SFU Community Trust Endowment Fund.

Conflict of interest statement. None declared.

REFERENCES

- Mathee,K., Narasimhan,G., Valdes,C., Qiu,X., Matewish,J.M., Koehrsen,M., Rokas,A., Yandava,C.N., Engels,R., Zeng,E. *et al.* (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proc. Natl Acad. Sci. USA*, **105**, 3100–3105.
- Lee,D.G., Urbach,J.M., Wu,G., Liberati,N.T., Feinbaum,R.L., Miyata,S., Diggins,L.T., He,J., Saucier,M., Deziel,E. *et al.* (2006) Genomic analysis reveals that *Pseudomonas aeruginosa* virulence is combinatorial. *Genome Biol.*, **7**, R90.
- Joardar,V., Lindeberg,M., Jackson,R.W., Selengut,J., Dodson,R., Brinkac,L.M., Daugherty,S.C., Deboy,R., Durkin,A.S., Giglio,M.G. *et al.* (2005) Whole-genome sequence analysis of *Pseudomonas syringae* pv. *phaseolicola* 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.*, **187**, 6488–6498.
- Qiu,X., Gurkar,A.U. and Lory,S. (2006) Interstrain transfer of the large pathogenicity island (PAPI-1) of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA*, **103**, 19830–19835.
- Brinkman,F.S., Hancock,R.E. and Stover,C.K. (2000) Sequencing solution: use volunteer annotators organized via internet. *Nature*, **406**, 933.
- Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M. *et al.* (2000) Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature*, **406**, 959–964.
- 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**, D338–D343.
- Winsor,G.L., Van Rossum,T., Lo,R., Khaira,B., Whiteside,M.D., Hancock,R.E. and Brinkman,F.S. (2009) *Pseudomonas* genome database: facilitating user-friendly, comprehensive comparisons of microbial genomes. *Nucleic Acids Res.*, **37**, D483–D488.
- Reddy,T.B., Riley,R., Wymore,F., Montgomery,P., DeCaprio,D., Engels,R., Gellesch,M., Hubble,J., Jen,D., Jin,H. *et al.* (2009) TB database: an integrated platform for tuberculosis research. *Nucleic Acids Res.*, **37**, D499–D508.
- Winsor,G.L., Khaira,B., Van Rossum,T., Lo,R., Whiteside,M.D. and Brinkman,F.S. (2008) The Burkholderia genome database: facilitating flexible queries and comparative analyses. *Bioinformatics*, **24**, 2803–2804.
- Klockgether,J., Munder,A., Neugebauer,J., Davenport,C.F., Stanke,F., Larbig,K.D., Heeb,S., Schock,U., Pohl,T.M., Wiehmann,L. *et al.* (2010) Genome diversity of *Pseudomonas aeruginosa* PA01 laboratory strains. *J. Bacteriol.*, **192**, 1113–1121.
- Kurtz,S., Phillippy,A., Delcher,A.L., Smoot,M., Shumway,M., Antonescu,C. and Salzberg,S.L. (2004) Versatile and open software for comparing large genomes. *Genome Biol.*, **5**, R12.
- Ogino,S., Gulley,M.L., den Dunnen,J.T., Wilson,R.B. and Association for Molecular Pathology Training and Education Committee (2007) Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. *J. Mol. Diagn.*, **9**, 1–6.
- Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.
- Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T. *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.
- Stein,L.D., Mungall,C., Shu,S., Caudy,M., Mangone,M., Day,A., Nickerson,E., Stajich,J.E., Harris,T.W., Arva,A. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–610.
- Jacobs,M.A., Alwood,A., Thaipisuttikul,I., Spencer,D., Haugen,E., Ernst,S., Will,O., Kaul,R., Raymond,C., Levy,R. *et al.* (2003) Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl Acad. Sci. USA*, **100**, 14339–14344.
- Lewenza,S., Falsafi,R.K., Winsor,G., Gooderham,W.J., McPhee,J.B., Brinkman,F.S. and Hancock,R.E. (2005) Construction of a mini-Tn5-luxCDABE mutant library in *Pseudomonas aeruginosa* PA01: a tool for identifying differentially regulated genes. *Genome Res.*, **15**, 583–589.
- Liberati,N.T., Urbach,J.M., Miyata,S., Lee,D.G., Drenkard,E., Wu,G., Villanueva,J., Wei,T. and Ausubel,F.M. (2006) An ordered, nonredundant library of *Pseudomonas aeruginosa* strain PA14 transposon insertion mutants. *Proc. Natl Acad. Sci. USA*, **103**, 2833–2838.
- Barrett,T., Troup,D.B., Wilhite,S.E., Ledoux,P., Rudnev,D., Evangelista,C., Kim,I.F., Soboleva,A., Tomashevsky,M., Marshall,K.A. *et al.* (2009) NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Res.*, **37**, D885–D890.
- Parkinson,H., Kapushesky,M., Kolesnikov,N., Rustici,G., Shojatalab,M., Abeygunawardena,N., Berube,H., Dylag,M., Emam,I., Farne,A. *et al.* (2009) ArrayExpress update—from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.*, **37**, D868–D872.
- Sayers,E.W., Barrett,T., Benson,D.A., Bolton,E., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., Dicuccio,M., Federhen,S. *et al.* (2010) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **38**, D5–D16.
- Fulton,D.L., Li,Y.Y., Laird,M.R., Horsman,B.G., Roche,F.M. and Brinkman,F.S. (2006) Improving the specificity of high-throughput ortholog prediction. *BMC Bioinformatics*, **7**, 270.
- Min,J.E., Whiteside,M.D., Brinkman,F.S.L., McNeney,B. and Graham,J. (2011) A statistical approach to high-throughput screening of predicted orthologs. *Comput. Stat. Data Anal.*, **55**, 935–943.
- Yu,N.Y., Wagner,J.R., Laird,M.R., Melli,G., Rey,S., Lo,R., Dao,P., Sahinalp,S.C., Ester,M., Foster,L.J. *et al.* (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*, **26**, 1608–1615.
- Gardy,J.L., Laird,M., Chen,F., Rey,S., Walsh,C.J., Tusnády,G.E., Ester,M. and Brinkman,F.S.L. PSORT-B v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, **21**, 617–623.
- Langille,M.G. and Brinkman,F.S. (2009) IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics*, **25**, 664–665.

28. Waack,S., Keller,O., Asper,R., Brodag,T., Damm,C., Fricke,W.F., Surovcik,K., Meinicke,P. and Merkl,R. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden markov models. *BMC Bioinformatics*, **7**, 142.
29. Hsiao,W., Wan,I., Jones,S.J. and Brinkman,F.S. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics*, **19**, 418–420.
30. Langille,M.G., Hsiao,W.W. and Brinkman,F.S. (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics*, **9**, 329.
31. Dam,P., Olman,V., Harris,K., Su,Z. and Xu,Y. (2007) Operon prediction using both genome-specific and general genomic information. *Nucleic Acids Res.*, **35**, 288–298.
32. Brouwer,R.W., Kuipers,O.P. and van Hijum,S.A. (2008) The relative value of operon predictions. *Brief Bioinform*, **9**, 367–375.
33. Kingsford,C.L., Ayanbule,K. and Salzberg,S.L. (2007) Rapid, accurate, computational discovery of rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol.*, **8**, R22.
34. Rice,P., Longden,I. and Bleasby,A. (2000) EMBOSS: the European molecular biology open software suite. *Trends Genet.*, **16**, 276–277.
35. Buell,C.R., Joardar,V., Lindeberg,M., Selengut,J., Paulsen,I.T., Gwinn,M.L., Dodson,R.J., Deboy,R.T., Durkin,A.S., Kolonay,J.F. *et al.* (2003) The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. *tomato* DC3000. *Proc. Natl Acad. Sci. USA*, **100**, 10181–10186.
36. Winstanley,C., Langille,M.G., Fothergill,J.L., Kukavica-Ibrulj,I., Paradis-Bleau,C., Sanschagrin,F., Thomson,N.R., Winsor,G.L., Quail,M.A., Lennard,N. *et al.* (2009) Newly introduced genomic prophage islands are critical determinants of in vivo competitiveness in the liverpool epidemic strain of *Pseudomonas aeruginosa*. *Genome Res.*, **19**, 12–23.
37. Peterson,J.D., Umayam,L.A., Dickinson,T., Hickey,E.K. and White,O. (2001) The comprehensive microbial resource. *Nucleic Acids Res.*, **29**, 123–125.
38. UniProt Consortium. (2009) The universal protein resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
39. 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**, D480–D484.
40. Choi,C., Munch,R., Leupold,S., Klein,J., Siegel,I., Thielen,B., Benkert,B., Kucklick,M., Schobert,M., Barthelmes,J. *et al.* (2007) SYSTOMONAS—an integrated database for systems biology analysis of *Pseudomonas*. *Nucleic Acids Res.*, **35**, D533–D537.
41. Lynn,D.J., Winsor,G.L., Chan,C., Richard,N., Laird,M.R., Barsky,A., Gardy,J.L., Roche,F.M., Chan,T.H., Shah,N. *et al.* (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.*, **4**, 218.