

Laying the Foundation for a Genomic Rosetta Stone: Creating Information Hubs through the Use of Consensus Identifiers

Bart Van Brabant,¹ Tanya Gray,² Bert Verslyppe,^{1,3} Nikos Kyrpides,⁴ Karin Dietrich,^{5,6}
Frank Oliver Glöckner,^{5,6} James Cole,⁷ Ryan Farris,⁷ Lynn M. Schriml,⁸ Paul De Vos,^{1,9}
Bernard De Baets,¹⁰ Dawn Field,² Peter Dawyndt,³ and the Genomic Standards Consortium

Abstract

Given the growing wealth of downstream information, the integration of molecular and non-molecular data on a given organism has become a major challenge. For micro-organisms, this information now includes a growing collection of sequenced genes and complete genomes, and for communities of organisms it includes metagenomes. Integration of the data is facilitated by the existence of authoritative, community-recognized, consensus identifiers that may form the heart of so-called information knuckles. The Genomic Standards Consortium (GSC) is building a mapping of identifiers across a group of federated databases with the aim to improve navigation across these resources and to enable the integration of this information in the near future. In particular, this is possible because of the existence of INSDC Genome Project Identifiers (GPIDs) and accession numbers, and the ability of the community to define new consensus identifiers such as the culture identifiers used in the StrainInfo.net bioportal. Here we outline (1) the general design of the Genomic Rosetta Stone project, (2) introduce example linkages between key databases (that cover information about genomes, 16S rRNA gene sequences, and microbial biological resource centers), and (3) make an open call for participation in this project by providing a vision for its future use.

Introduction

THE CURRENT LANDSCAPE of biology and bioinformatics means that data is spread across many databases. This is commonly accompanied by incompatibilities of formats, semantics, and accessibility of the data. The problem of scattered information is widely acknowledged, and has been aptly discussed by Lincoln Stein in his paper on the “knuckles-and-nodes” concept (Stein, 2003). In particular, he called for a set of resources that unify naturally related sets of other

resources for the sake of achieving global interoperability across bioinformatics domains of specialization.

The concept of knuckles and nodes is highly attractive. Establishing the required mappings in large and distributed information networks, though, requires the implementation of centralized repository activities that keep track of the relations between entities of the same biological type (Van Brabant et al., 2006). The spread of biological information across the Internet, in particular, leads to the fact that local data sets provide information about related objects. For example, cul-

¹Laboratory of Microbiology, Ghent University, Ghent, Belgium.

²Molecular Evolution and Bioinformatics Section, Oxford Centre for Ecology and Hydrology, Oxford, United Kingdom.

³Department of Applied Mathematics and Computer Science, Ghent University, Ghent, Belgium.

⁴Department of Energy, Joint Genome Institute, Walnut Creek, California.

⁵Microbial Genomics Group, Max Planck Institute for Marine Microbiology, Bremen, Germany.

⁶Jacobs University Bremen gGmbH, Bremen, Germany.

⁷Center for Microbial Ecology, Michigan State University, East Lansing, Michigan.

⁸Institute for Genome Sciences and Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, Maryland.

⁹BCCM™/LMG Bacteria Collection, Ghent University, Ghent, Belgium.

¹⁰Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Ghent, Belgium.

tures of the same bacterial strain are spread across a global network of biological resource centers (BRCs) (OECD, 2007), and downstream information on the same genome sequence is reported in autonomous and heterogeneous databases. To build an information network of knuckles and nodes therefore requires “Rosetta Stones” and “resolvers,” where synonymous identifiers are mapped, made persistent, and served to the community as single points of access for certain types of information.

These information hubs should ideally be based upon consensus identifiers. Such identifiers can be issued by authoritative databases (like INSDC GPIDS and accession numbers) or developed by portal projects and communities if no globally unique identifiers were already attributed to certain information objects. An example of the latter is the Strain-Info.net bioportal, which introduces a consensus identifier called a culture identifier (CID) to refer to micro-organisms in a global network of BRCs. These CIDs are automatically assigned to each strain number discovered when indexing the online catalogs of 50 biological resource centers, using an accumulative learning process that is described in detail by Dawyndt et al. (2005). During this process, information about synonymous strain numbers is extracted from the online catalogs of BRCs, and processed to remove redundant and inconsistent information. As a result, information about a given micro-organism may be searched for by whatever strain number assigned to the organism. The StrainInfo.net bioportal may thus be regarded as a knuckle that helps discovering all information that is known about a particular strain.

In this paper we advocate applying a similar approach of mapping consensus identifiers for the integration of genomes and metagenomes, and outline the design of building such a genome knuckle. We also point out the benefits of describing relationships between knuckles that are based upon the use of consensus identifiers.

The Need For A Single List of Genomes and Metagenomes

Molecular biology is shifting from a pregenomic era, characterized by the efforts to sequence parts (e.g., a single gene) of the complete genome of an organism, to a postgenomic era that concentrates on extracting the biological knowledge hidden in a multitude of complete genome sequences. The mission of the Genomic Standards Consortium (GSC) is to build richer descriptions of the complete collection of genomes and metagenomes and work toward improved mechanisms of data exchange and integration. To date, this community-level, consensus-driven effort has been primarily based on producing the “Minimum Information about a Genome Sequence” (MIGS) specification (Field et al., 2008a). Now efforts are shifting to focus on the implementation of the checklist and the capture of data. Both the Genomic Contextual Data Markup Language (GCDML) (Kottmann et al., 2008) and the concept of a Genomic Rosetta Stone (GRS) will be core to this goal, and are essential for the construction of the GSC’s Genome Catalog.

Following the third GSC workshop, both EMBL and NCBI offered their lists of genomes for import into the Genome Catalog. At that time this was not possible, but since then, the INSDC has developed and adopted Genome

Project Identifiers (INSDC GPIDs) that are now suitable for use as community-wide consensus identifiers in this domain.

The GSC’s GRS Project

The GRS project was formally established at the fourth GSC workshop (Field et al., 2008c) and revisited in a session at the fifth GSC workshop (Field et al., 2008b) where it was established as a core GSC project, along with the GCDML (Kottmann et al., 2008). The GRS is a cornerstone for building the long-term vision of integrating the richest possible set of information describing our complete collection of genomes and metagenomes.

Progress has been made toward designing this resource, and an implementation strategy is described below. In summary, the GRS currently involves over 10 databases coming together in a federated approach to build a genome identifier mapping (Table 1). We aim at setting up this resource following a distributed approach to both the responsibility for building and maintaining a mapping of local databases to consensus identifiers and the development of a suitable infrastructure. Although INSDC GPIDs are the consensus identifiers for genomes and metagenomes, identifiers like INSDC accession numbers can also be linked to unify larger “halos” of information that describe BRCs and 16S rRNA gene sequences. The scope and the component parts of the GRS are described below.

Scope of the GRS: a growing list of mapped local identifiers

The GSC’s GRS project already includes participants working on a range of widely used databases holding genomic and metagenomic data (Table 1). It aims at mapping local identifiers used in a multitude of genome and metagenome repositories onto INSDC GPIDs. Furthermore, linking the INSDC GPIDs to identifiers in a variety of other types of databases results in valuable additional information, in particular, where it might be possible to harvest precomputed MIGS/MIMS compliant fields. This includes, for example, culture and specimen voucher collections that provide additional downstream information. For this reason the Straininfo.net bioportal has been an integral collaborator in the GRS project from the start. In addition, several descriptors in MIGS actually belong in other authoritative databases. The MIGS descriptors of “nucleic acid sequence source,” “number of chromosomes,” and “ploidy level,” for example, should ideally come from authoritative databases holding organism (or species) level information. As such databases will be built in the future, one could imagine that MIGS/MIMS compliant data could be “harvested” from external resources instead of being centrally curated.

There is also a range of fields outside the minimum (compact) scope of MIGS/MIMS that are of special interest. Those fields are added to the GCDML as appropriate (Kottmann et al., 2008). Optimal growth temperature, for example, is one of the most widely used ecological parameters in comparative genomic studies and yet is not included in MIGS because it is a descriptor perhaps best curated and maintained within a specialist database. The Prokaryotic Growth Temperature Database (PGTdb) (Huang et al., 2004) is one such database that is now available. Other key data providers will

TABLE 1. COLLABORATORS OF THE GSCs GENOMIC ROSETTA STONE (GRS) PROJECT

<i>Collaborator</i>	<i>Description</i>	<i>Reference</i>
CMR	The Comprehensive Microbial Resource, information on publicly available, complete prokaryotic genomes.	(Peterson et al., 2001)
Genome Atlas	A database environment for the storage of comparative sequence results for bacterial complete genomes.	(Hallin and Ussery, 2004)
GSC Genome Catalogue	A beta-stage repository for capturing MIGS/MIMS information.	(Field et al., 2008b)
Gemina	An epidemiology metadata database for identification of infectious pathogens and their representative genomic sequences.	(Schriml et al., 2007)
Genome Reviews	The Genome Reviews database provides an up-to-date, standardized and comprehensively annotated view of the genomic sequence of organisms with completely deciphered genomes.	EMBL
GOLD	The Genomes OnLine Database, a resource for comprehensive access to information regarding complete and ongoing genome projects, as well as metagenomes and metadata.	(Liolios et al., 2008)
IMG	The Integrated Microbial Genomes, a resource for comparative annotation of all publicly available genomes.	(Markowitz et al., 2008)
RDP	The Ribosomal Database Project provides ribosome related data, services, and support to the scientific community, including online data analysis and aligned and annotated bacterial small-subunit 16S rRNA sequences.	(Cole et al., 2007)
SEED	A curated genomic database based on a subsystems approach and a repository of open source annotation tools.	(Overbeek et al., 2005)
SILVA	A database of aligned small and large subunit ribosomal RNA sequences.	(Pruesse et al., 2007)
StrainInfo.net	A biportal offering an integrated view on publicly available microbial cultures and their downstream information.	(Dawyndt et al., 2005)

be found in online resources like “mashups” that harvest information from a variety of other sources on the fly (e.g., “iSpecies,” <http://darwin.zoology.gla.ac.uk/~rpage/ispecies/> and Encyclopedia of Life (Wilson, 2003).

Proposed implementation strategy

The final goal of the GRS project is to make available a genome identifier mapping that facilitates discovery of genomic information, comparative genomic studies, and enrichment of databases with hyperlinks and metadata. The GRS project necessarily must be built from several components that need to work together to attain a complete vision. The scope of the mapping therefore needs to be clearly defined. First, the list of candidate databases to be incorporated in the mapping needs to be clearly described. Second, a strategy needs to be developed that describes the way to store and maintain the mapping of IDs and keeping it up to date. A third requirement is a centralized repository that allows users to access the mapping and to make use of it. Finally, information providers must be stimulated to start using the mapping for local use or display. A conceptual scheme for this strategy is depicted in Figure 1.

The GRS builds upon a group of database developers working together to establish the mapping. The collaborators (to join this effort, please contact “curator at ceh.ac.uk.”) identified in Table 1 are already involved in the project. All contributors are expected to map their own local identifiers

to the INSDC GPIDs. Further to a proposal made during the fifth GSC workshop (Field et al., 2008c), the identifier mapping is to be hosted by NCBI and managed via the NCBI LinkOut system (<http://www.ncbi.nlm.nih.gov/projects/>

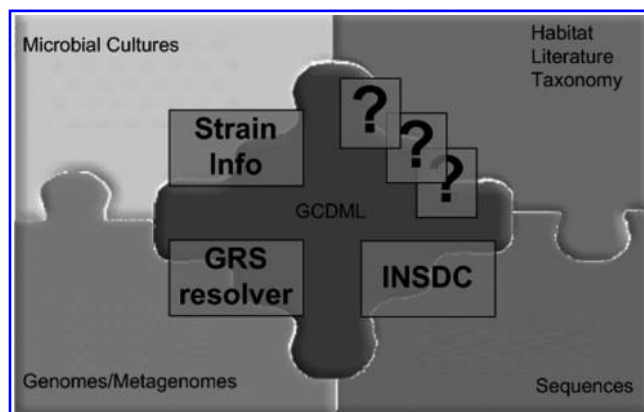


FIG. 1. Knuckles-and-notes conceptual schema for describing genomes and their related information. The jigsaw symbolizes the whole of the information surrounding a genome. Each piece represents all of the nodes of alike data, integrated and connected by the knuckles and discussed in this paper. As depicted, there are still some knuckles missing that might be added in the future. The center piece could be formed by the GCDML that glues it all together.

TABLE 2. CROSS-REFERENCING STATISTICS BETWEEN THE STRAININFO.NET BIOPORTAL, GOLD, AND SILVA

Source	Number of records	With strain number	Linked
GOLD			
<i>Archaea</i>	50	48 (96.0%)	42 (84.0%)
<i>Bacteria</i>	581	578 (99.5%)	312 (53.7%)
<i>Fungi</i>	82	66 (80.5%)	20 (24.4%)
SILVA			
<i>Archaea</i>	2733	1442 (52.8%)	437 (16.0%)
<i>Bacteria</i>	118362	106365 (89.9%)	19027 (16.1%)
<i>Fungi</i>	11017	9672 (87.8%)	1853 (16.8%)

The table shows the total number of archaeal, bacterial, and fungal records in each dataset, the number of entries for which a strain number was available, and the amount of strain numbers that could be linked to a culture identifier. For the GOLD dataset, strain numbers were retrieved from the original publication or from the genome sequencing project Web site. As the extraction was performed manually, strain numbers could be retrieved for nearly all genome sequences. Mapping the strain numbers to their corresponding culture identifiers was initially performed in an automated way, linking only those strain numbers that were considered to be unique within a global namespace by the StrainInfo.net bioportal. The remaining ambiguous strain numbers were then subjected to further manual resolution. For the SILVA dataset, strain numbers were automatically parsed out of the corresponding EMBL sequence records using regular expressions. The resolution of these strain numbers to their corresponding culture identifiers was performed automatically.

linkout/doc/liblinkout.html). Each GRS data provider is to register separately with NCBI LinkOut as an external data provider. The data provider will provide a file describing the data resource, and a resource file that includes a mapping of the data provider resources to identifiers in the NCBI genome project database. The data provider is responsible for maintaining these files and publishing updates to a dedicated ftp directory provided by NCBI. The information contained in the data provider resource file will be publically available via the NCBI Genome Project Web site, and can also be queried via the LinkOut web service. As of April

2008, the RDP and Gemina databases have registered as external data providers with NCBI LinkOut and included identifier mappings in their resource files.

The GRS resolver, a tool that brokers the exchange of genome project identifiers, enables INSDC GPIDs to play a role as consensus identifiers for genomes and metagenome projects. To make best use of the genome identifier mapping, the GSC aims at engineering an online resolution service that produces lists of links to all databases in which an instance of a particular genome occurs. The resolver will function much like a currency converter. Entering an identifier will return synonymous identifiers for a particular set of genomes or metagenomes. The mappings and source code will be made freely available. James Cole and colleagues at Michigan State University have developed a client tool that consumes data directly from the GRS resolver. James presented this during the fifth GSC workshop (Field et al., 2008c). He is planning to make this tool distributable for use by other database and Web sites.

Linking Knuckles

Further development envisages a working GRS that is one of many knuckles that can, in turn, be traversed and mined. This goal will be achieved by the strategy that is depicted in Figure 2. The use of consensus identifiers discussed in this paper allows the development of resolvers, each acting as knuckles on the nodes in which biological information of a certain type is stored, for example, the multitude of biological resource centers containing information on well-characterized micro-organisms. By linking between knuckles rather than between individual nodes, redundant relationships can be greatly reduced while users get a more complete view on the information. Preliminary work toward this goal is now underway. The StrainInfo.net bioportal, for example, has enriched its organismal information with a series of deep links to the genome projects in the Genomes OnLine Database (GOLD) (Liolios et al., 2008) and to the curated rRNA gene sequence information in the SILVA database (Pruesse et al., 2007) (Table 2). Furthermore, a link mechanism is provided to allow other genome information providers to easily hook

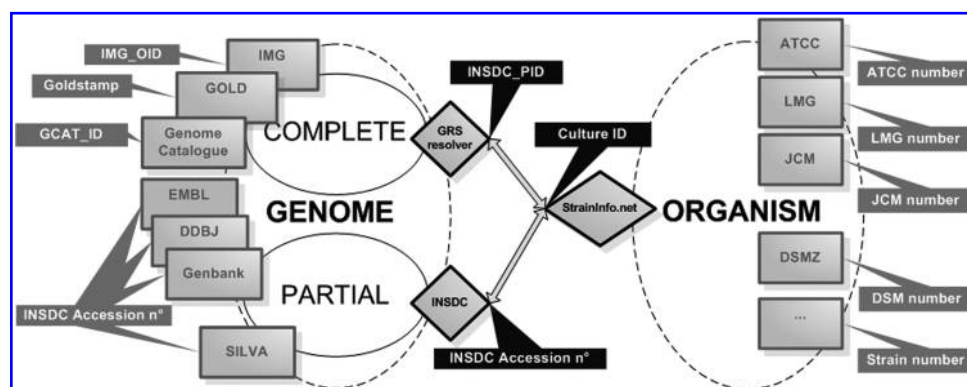


FIG. 2. Schematic representation of the extended information network used to implement some example mappings within the scope of this paper. The existing information network is expanded with information knuckles that are involved in the process of establishing the mappings between the genome and organismal information. Knuckles act as brokers for the various identifiers that are in use for a given information type, and provide consensus identifiers for all information objects (genome projects, sequences records, or microbial cultures) within their scope. As such, the mapping of biological information becomes far more complete and integrated than without the use of these resolvers.

up their resources to the organismal information contained in the StrainInfo.net bioportal, thereby allowing them to enrich their resources with fully integrated microbial information. Detailed information on how these links have been implemented is provided on the GRS wiki (http://gensc.org/gc_wiki/index.php/Genomic_Rosetta_Stone).

Conclusions

The information content of biological databases may be vastly improved if users and software agents can easily navigate between them. In this paper we have discussed the use of a knuckles-and-nodes strategy to establish these links. When services of individual nodes are bundled into common knuckles, relationships between biological information may be expressed using whatever identifier that is serviced by the knuckles. After all, the knuckles ensure straightforward cross-referencing. This setup should not be restricted to a narrow genomic scope. Moreover, the existence of knuckles allows new information providers (nodes) to easily hook up onto the network, enriching its information content, while the network itself stays highly maintainable.

Besides linking genomes to the organisms from which they were derived and vice versa, the network can be expanded with informational links on isolation environment, relevant publications, etc. Doing so, different domains of information become interconnected in a way that allows scientists active in the domains of molecular biology, microbiology, microbial ecology, and medicine to easily retrieve all the relevant data, and even more so, if it were possible in the future to harvest integrated data from all of the associated resources with GCDML. Achieving this goal requires widespread participation in the GRS and adoption of GCDML. As such, the GRS described in this work could provide a framework for the creation of fully integrated and flexible genome information networks.

Acknowledgment

B.V.B., B.V., B.D.B., P.D.V., and P.D. acknowledge the financial support of the Belgian Science Policy IAP VI program, no. P6UCL06.

Author Disclosure Statement

The authors declare that no competing financial interests exist.

References

Cole, J.R., Chai, B., Farris, R.J., Wang, Q., et al. (2007). The ribosomal database project (RDP-II): Introducing myRDP space and quality controlled public data. *Nucleic Acids Res* 35 (Database Issue): D169-D172; doi: 10.1093/nar/gkl889.

Dawyndt, P., Vancanneyt, M., De Meyer, H., and Swings, J. (2005). Knowledge accumulation and resolution of data inconsistencies during the integration of microbial information sources. *IEEE Trans Knowledge Data Eng* 17, 1111-1126.

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., et al. (2008a). Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum In-

formation about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 26, 541-547.

Field, D., Garrity, G.M., Sansone, S.A., Sterk, P., Gray, T., and Glöckner, F.O. (2008b). Meeting report: The Fifth Genomic Standards Consortium Workshop. *OMICS* (this issue).

Field, D., Glöckner, F.O., Garrity, G.M., Gray, T., Sterk, P., Cochrane, G., et al. (2008c). Meeting report: The Fourth Genomic Standards Consortium (GSC) workshop. *OMICS* (this issue).

Hallin, P.F., and Ussery, D.W. (2004). CBS Genome Atlas Database: a dynamic storage for bioinformatic results and sequence data. *Bioinformatics* 20, 3682-3686.

Huang, S.-L., Wu, L.-C., Liang, H.-K., Pan, K.-T., Horng, J.-T., and Ko, M.-T. (2004). PGTdb: a database providing growth temperatures of prokaryotes. *Bioinformatics* 20, 276-278.

Kottmann, R., Gray, T., Murphy, S., Kagan, L., Kravitz, S., Lombardot, T., et al. (2008). A standard MIGS/MIMS compliant XML Schema: Toward the development of the Genomic Contextual Data Markup Language (GCDML). *OMICS* (this issue).

Liolios, K., Mavromatis, K., Tavernarakis, N., and Kyrpides, N.C. (2008). The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 36(Suppl 1), D475-D479.

Markowitz, V.M., Szeto, E., Palaniappan, K., Grechkin, Y., Chu, K., Chen, I.M.A., et al. (2008). The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res* 36, D528-D533.

OECD. (2007). *OECD Best Practices Guidelines for Biological Resource Centres*. Paris: Organisation for Economic Co-operation and Development.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33, 5691-5702.

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.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., et al. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 35, 7188-7196.

Schriml, L.M., Gussman, A., Phillipy, K., Angiuoli, S., et al. (2007). Gemina: A web-based epidemiology and genomic metadata system designed to identify infectious pathogens. D Zeng et al. (eds.). *BioSurveillance*, LNCS 4506, pp. 228-229.

Stein, L.D. (2003). Integrating biological databases. *Nat Rev Genet* 4, 337-345.

Van Brabant, B., Dawyndt, P., De Baets, B., and De Vos, P. (2006). A knuckles-and-nodes approach to the integration of microbiological resource data. *Lecture Notes Comput Sci* 4277, 740-750.

Wilson, E.O. (2003). The encyclopedia of life. *Trends Ecol Evol* 18, 77-80.

Address reprint requests to:

Prof. Peter Dawyndt

Department of Applied Mathematics and Computer Science

Ghent University

Krijgslaan 281 S9, B-9000 Ghent, Belgium

E-mail: Peter.Dawyndt@UGent.be

This article has been cited by:

1. Jack A. Gilbert, Yiming Bao, Hui Wang, Susanna-Assunta Sansone, Scott C. Edmunds, Norman Morrison, Folker Meyer, Lynn M. Schriml, Neil Davies, Peter Sterk, Jared Wilkening, George M. Garrity, Dawn Field, Robert Robbins, Daniel P. Smith, Ilene Mizrahi, Corrie Moreau. 2012. Report of the 13th Genomic Standards Consortium Meeting, Shenzhen, China, March 4–7, 2012. *Standards in Genomic Sciences* 6:2, 276-286. [[Crossref](#)]
2. Bert Verslyppe, Renzo Kottmann, Wim De Smet, Bernard De Baets, Paul De Vos, Peter Dawyndt. 2010. Microbiological Common Language (MCL): a standard for electronic information exchange in the Microbial Commons. *Research in Microbiology* 161:6, 439-445. [[Crossref](#)]
3. Stefano Mocali, Anna Benedetti. 2010. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Research in Microbiology* 161:6, 497-505. [[Crossref](#)]
4. Jack A. Gilbert, Folker Meyer, Rob Knight, Dawn Field, Nikos Kyrpides, Pelin Yilmaz, John Wooley. 2010. Meeting report: GSC M5 roundtable at the 13th International Society for Microbial Ecology meeting in Seattle, WA, USA August 22-27, 2010. *Standards in Genomic Sciences* 3:3, 235-239. [[Crossref](#)]
5. Tanja Davidsen, Ramana Madupu, Peter Sterk, Dawn Field, George Garrity, Jack Gilbert, Frank Oliver Glöckner, Lynette Hirschman, Eugene Kolker, Renzo Kottmann, Nikos Kyrpides, Folker Meyer, Norman Morrison, Lynn Schriml, Tatiana Tatusova, John Wooley. 2010. Meeting Report from the Genomic Standards Consortium (GSC) Workshop 9. *Standards in Genomic Sciences* 3:3, 216-224. [[Crossref](#)]
6. Elizabeth Glass, Folker Meyer, Jack A Gilbert, Dawn Field, Sarah Hunter, Renzo Kottmann, Nikos Kyrpides, Susanna Sansone, Lynn Schriml, Peter Sterk, Owen White, John Wooley. 2010. Meeting Report from the Genomic Standards Consortium (GSC) Workshop 10. *Standards in Genomic Sciences* 3:3, 225-231. [[Crossref](#)]
7. Dawn Field, Iddo Friedberg, Peter Sterk, Renzo Kottmann, Frank Oliver Glöckner, Lynette Hirschman, George M. Garrity, Guy Cochrane, John Wooley, Jack Gilbert. 2009. Meeting Report: Metagenomics, Metadata and Meta-analysis; (M3) Special Interest Group at ISMB 2009. *Standards in Genomic Sciences* 1:3, 278-282. [[Crossref](#)]
8. Dawn Field, Peter Sterk, Nikos Kyrpides, Renzo Kottmann, Frank Oliver Glöckner, Lynette Hirschman, George M. Garrity, John Wooley, Paul Gilna. 2009. Meeting Report from the Genomic Standards Consortium (GSC) Workshops 6 and 7. *Standards in Genomic Sciences* 1:1, 68-71. [[Crossref](#)]
9. John C. Wooley, Dawn Field, Frank-Oliver Glöckner. 2009. Extending Standards for Genomics and Metagenomics Data: A Research Coordination Network for the Genomic Standards Consortium (RCN4GSC). *Standards in Genomic Sciences* 1:1, 85-90. [[Crossref](#)]
10. Bert Verslyppe, Bram Slabbinck, Wim De Smet, Paul De Vos, Bernard De Baets, Peter Dawyndt. StrainInfo.net Web Services: Enabling Microbiologic Workflows Such as Phylogenetic Tree Building and Biomarker Comparison 603-607. [[Crossref](#)]