

Figure 2. Hilton Garden Inn in Bethesda, 7301 Waverly Street, Bethesda, MD 20814, Tel: 1-301-654-8111, Fax 1-301-654-6584, <http://hiltongardeninn.hilton.com>, Walking distance to Bethesda Metro Station. One metro stop south of Medical Center/NIH

Figure 4. Residence Inn by Marriott, 7335 Wisconsin Avenue, Bethesda, MD 20814, Tel: 301-718-0200, <http://www.residenceinnbethesdahotel.com/> or <http://cwp.marriott.com/wasbt/nih/>

Figure 3. Hyatt Regency Bethesda, 1 Bethesda Metro Center, Bethesda, MD 20814, Tel: 657-1234, <http://www.bethesda.hyatt.com>, Walking distance to Metro Station. One metro stop south of Medical Center/NIH

Figure 5. Marriott Pooks Hill, 5151 Pooks Hill Road, Bethesda, Md 20814, Tel: 301-897-9400; 1-800-228-9290, <http://www.marriott.com/hotels/travel/wasbt-bethesda-marriott/>, Shuttle from hotel to Medical Center

M.H. Medema

GSC15p02

MIBiG: Minimal Information about a Biosynthetic Gene Cluster

M.H. Medema^{1,2}, P. Yilmaz³, R. Kottmann³, F.O. Glöckner³, R. Breitling^{2,4}, E. Takano^{1,4}

¹Department of Microbial Physiology

²Groningen Bioinformatics Centre, University of Groningen, Groningen, The Netherlands

³Microbial Genomics and Bioinformatics Group, Max Planck Institute for Marine Microbiology, Bremen, Germany

⁴Faculty of Life Sciences, Manchester Institute of Biotechnology, University of Manchester, United Kingdom

Correspondence: m.h.medema@rug.nl

Keywords: genomic standards, biosynthetic gene cluster, secondary metabolism, natural products

Program session: Implementing Standards in Genomic and Metagenomic Research Projects

Abstract

Bacteria, fungi and plants produce an enormous variety of secondary metabolites with manifold biological activities, e.g., as antibiotics, immunosuppressants, and signaling molecules. The biosynthesis of such molecules is encoded by compact genomic units: biosynthetic gene clusters. Over the past decades, hundreds of biosynthetic gene clusters encoding the biosynthesis of secondary metabolites have been characterized. Although dozens of biosynthetic gene clusters are published and thousands are sequenced annually (with or without their surrounding genome sequence), very little effort has been put into structuring this information. Hence, it is currently very difficult to prioritize gene clusters for experimental characterization, to identify the fundamental architectural principles of biosynthetic gene clusters, to understand which ecological parameters drive their evolution, and to obtain an informative ‘parts registry’ of building blocks for the synthetic biology of secondary metabolite biosynthesis.

Therefore, developing a genomic standard for experimentally characterized biosynthetic gene clusters would be of great value. The standard will build on the MxS standards for ecological and environmental contextualization [1]. Additionally, biochemical, genomic and pharmaceutical metadata will be added in parameters such as enzyme substrate specificities, operon structure, chemical moieties of the product, metabolic precursors and compound bioactivity. Using the already developed computational pipeline for gene cluster analysis antiSMASH [2], information on characterized biosynthetic gene clusters will be linked to the untapped wealth of thousands of unknown gene clusters that have recently been unearthed by massive genome sequencing efforts. Taken together, this has the potential to guide the characterization of new metabolites by allowing to optimize the sampling of diversity at different levels and to identify the biochemical, genomic and ecological parameters that are key predictors of pharmaceutically relevant biological activities. Moreover, it can transform the unordered pile of literature on secondary metabolites into a structured and annotated catalogue of parts that can be used as building blocks to design new biochemical pathways with synthetic biology [3,4].

Yemin Lan

GSC15p03

Identifying functional signatures in microorganism genomes related to polymer decomposition

Yemin Lan¹, Nivedita Clark², Christopher Blackwood² and Gail Rosen¹.

¹Drexel University

²Kent State University

Correspondence: yeminlan@gmail.com

Keywords: polymer decomposition, functional signatures, feature selection

Program session: Development of Resources, Tools or Databases Related to the GSC Mission

Frank Oliver Glöckner

GSC15p05

Micro B3: Marine Microbial Biodiversity, Bioinformatics, Biotechnology

Frank Oliver Glöckner

Jacobs University Bremen, Bremen, Germany

Correspondence: fog@mpi-bremen.de

Keywords: Biodiversity, Bioinformatics, Biotechnology, Standards, Ocean Sampling Day

Program session: Development of Resources, Tools or Databases Related to the GSC Mission, Implementing Standards in Genomic and Metagenomic Research Projects

Abstract

The 32 partner Ocean of Tomorrow Project Micro B3 (Biodiversity, Bioinformatics, Biotechnology, www.microb3.eu) forms teams of experts in bioinformatics, computer science, biology, ecology, oceanography, bioprospecting, biotechnology, ethics and law. The consortium's main aims are to bring together the existing bodies of expertise in ecosystems biology, the processing and interpretation of data, modelling and prediction and the development of intellectual property agreements for the exploitation of high potential commercial applications. At its core Micro B3 aims to develop an innovative, transparent and user friendly open-access system, which will allow for seamless processing, integration, visualisation and accessibility of the huge amount of data collected in ongoing sample campaigns and long-term observations. This will in turn offer new perspectives for the modelling and exploration of marine microbial communities for biotechnological applications.

A key boost to the work will be provided by the Ocean Sampling Day (OSD, www.oceansamplingday.org), scheduled to take place on summer solstice - 21 June 2014. OSD will take place worldwide, with pilots conducted in 2012 and 13 to establish standardized sampling techniques. Adhering to the Minimum information

checklists (MIxS) standard for describing molecular samples as outlined by the Genomic Standards Consortium will be essential for OSD. The event will generate a massive amount of useful marine microbial data to be included in the project's integrated MB3-Information System, providing the members of the biotechnology team with information to generate hypotheses for more cost- and time-efficient biotechnological testing and applications.

In summary Micro B3 is set to revolutionise Europe's capacity for bioinformatics and marine microbial data integration, to the benefit of a variety of disciplines in bioscience, technology, computing, standardisation and law.

Micro B3 is financially supported by the 7FP Ocean of Tomorrow Grant #287589

Tonia Korves

GSC15p06

Applying Publicly Available Genomic Metadata to Disease Outbreak Investigation

Tonia Korves, Matthew Peterson, Wenling Chang and Lynette Hirschman

MITRE, Bedford, MA, USA and McLean, VA, USA

Correspondence: tkorves@mitre.org

Keywords: disease outbreaks, data integration, public databases, applying genomic metadata

Program session: Development of Resources, Tools or Databases Related to the GSC Mission

Abstract

An important application of genomic metadata is in the investigation of disease outbreaks. The ability to identify the sources of disease outbreaks can prevent repeat events and potentially curtail outbreaks, saving lives, preventing societal disruption, and reducing economic costs. One way to discover the source of a pathogen is to identify other strains with shared biological properties, and then

becomes the center of the ring chart and the four levels below it are displayed. By clicking on the center of a ring chart, users can zoom out one level. By hovering the mouse over a slice, a tooltip is displayed that gives more information about the taxonomic node associated with the slice. The tooltip shows the number of peptides

that have the taxon as their LCA, and the number of peptides whose LCA is the taxon or one of its descendants in the NCBI taxonomy. These visualizations make Unipept an essential tool for gaining novel insights into the biodiversity of complex metaproteome samples [7-9].

Figure 6. Ring chart visualization of treemap data.

Abstract

Standards in Genomics Sciences (SIGS; www.standardsingenomics.org) was founded as an experimental open access publication to promote the data standardization efforts of the Genomic Standards Consortium (GSC). Whereas peer-reviewed publication of genome sequences was commonplace at the outset of the genomics era, many of the established journals in the life sciences abandoned such papers as the number of sequenced genomes increased sharply, leading to a loss of contextual information needed to correctly analyze and interpret genome sequence data. SIGS uses highly structured, easy to read and edit reports of genome and metagenome sequences, standard operating procedures, meeting reports, white papers and other articles that are relevant to a growing readership. Highlights for 2012 included SIGS receiving its first Impact Factor (1.63) and effective publication of a number of new species and genus descriptions that incorporate genome sequence data into the descriptions. At the time of writing, SIGS had published 280 articles, including 234 short genome reports. Our readership continues to grow, topping 64,000 readers in 171 countries, representing a 71% increase in audience in 2012 during 2012. SIGS is listed in CrossRef, PubMedCentral, Scopus, the Web of Science, ChemAbstracts, EBSCO and the Directory of Open Access Journals (DOAJ) and has become one of top three journals publishing papers on new genome sequences.

Ilene Mizrachi

GSC15p35

Standards in the INSDC

Ilene Mizrachi

National Center for Biotechnology Information, U.S. National Library of Medicine, Bethesda MD, USA

Abstract

The INSDC is committed to archiving and distributing high quality sequences and annotation along

with rich metadata to describe the project and biological sample. BioProject and BioSample databases have been developed to facilitate the capture of structured metadata for diverse biological research projects and samples represented in NCBI's archival databases. NCBI and its INSDC partners promote the use of standards for describing samples, experimental methodology and annotation. The adoption of common standards by the databases and scientific community will simplify the exchange of knowledge between different resources and improve the utility of the data for making scientific discoveries. Through a newly developed Submission Portal (<http://submit.ncbi.nlm.nih.gov>) for the deposition of experimental data and associated metadata, NCBI can enforce rules to ensure that annotation and metadata standards are being met for as a condition of submission.

Petra ten Hoopen

GSC15p36

ENA tools for sample metadata compliance

Petra ten Hoopen, Blaise Alako, Clara Amid, Lawrence Bower, Ana Cerdeño-Tárraga, Iain Cleland, Richard Gibson, Neil Goodgame, Mikyung Jang, Simon Kay, Rasko Leinonen, Xiu Lin, Arnaud Oisel, Nima Pakseresht, Swapna Pallreddy, Sheila Plaister, Rajesh Radhakrishnan, Stéphane Rivière, Marc Rossello, Alexander Senf, Nicole Silvester, Dmitriy Smirnov, Ana Toribio, Daniel Vaughan, Vadim Zalunin and Guy Cochrane

European Nucleotide Archive, European Molecular Biology Laboratory-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK

Keywords: metadata, standards compliance, ENA, INSDC, MiXs, MicroB3

Program session: Development of Resources, Tools or Databases Related to the GSC Mission

Abstract

Harmonization of (meta)data collection becomes an essential effort in the age when data generation is often easier and more affordable than

