BMC Bioinformatics

ORAL PRESENTATION   **Open Access**

Subsystems-based servers for rapid annotation of genomes and metagenomes

Ramy Karam Aziz[1,2]

From UT-ORNL-KBRIN Bioinformatics Summit 2010
Cadiz, KY, USA. 19-21 March 2010

Background

Today, more than 1000 genomes of cellular organisms, mostly microbes, have been completely sequenced and deposited in public databases, in addition to over 2000 viral genomes, and these numbers are expected to sky-rocket in the near future. While sequencing projects remain largely biased towards genomes linked to human interests [1] (e.g., domestic animals and plants, microbial pathogens, and microbes exploited in industry and agriculture), some serious initiatives are being launched for sequencing organisms that represent all branches of the tree of life [2].

Concomitant with the genomic revolution, unprecedented advances in sequencing technology have also led to the emergence of the field of metagenomics, which offers a novel, revolutionary approach for studying (microscopic) life in different environments. Metagenomics allows investigators to assess the biodiversity in a given ecosystem by directly sequencing DNA sampled from that ecosystem [3-5]. As so-called next-generation sequencing technologies evolve, producing tremendous amounts of data [6], the existing tools for sequence annotation are not fast enough to cope with the technological advances. Consequently, manual annotation has almost become impossible; however, automated annotation tools often lead to error propagation and biologically irrelevant ontologies.

Materials and methods

Here, I demonstrate how the use of the subsystems [7] and FIGfams [8,9] technologies, initiated by the Fellowship for Interpretation of Genomes (FIG) and the University of Chicago National Microbial Pathogen Data Resource (NMPDR) project [10], has improved the

accuracy and consistency of genome and metagenome annotation [11]. Using subsystems allows the combination of careful human annotation and the rapid computational propagation of assertions made by human experts through the RAST [8] pipeline for genome annotation, the MG-RAST server for metagenome annotation [12], and Phage-RAST for phage genome annotation (work in progress).

Results and conclusion

Still, although these servers offer relatively rapid annotation, the increasing throughput of sequencing platforms requires even faster pipelines, and annotating a large metagenomic data set can take weeks to months. To address this challenge, researchers at San Diego State University, FIG, and the Argonne National Laboratory are developing a protein family signature-based technology (Robert A. Edwards, Ross Overbeek, et al. submitted) to reduce the annotation speed by an order of magnitude and create a real-time annotation server (URL: http://edwards.sdsu.edu/rtmg). Such server will not only improve speed, but will allow the implementation of annotation pipelines on cell phones (Josh Hoffman et al., unpublished data) and social networks (Daniel Cuevas et al., unpublished data).

Author details
[1]Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo, Egypt. [2]San Diego State University, San Diego, CA 92182, USA.

Published: 23 July 2010

Correspondence: ramy.aziz@salmonella.org
[1]Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo, Egypt

BioMed Central

References

1. Aziz RK: **The case for biocentric microbiology.** *Gut Pathog* 2009, **1**:16.
2. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, Ivanova NN, Kunin V, Goodwin L, Wu M, Tindall BJ, *et al*: **A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea.** *Nature* 2009, **462**:1056-1060.
3. Riesenfeld CS, Schloss PD, Handelsman J: **Metagenomics: genomic analysis of microbial communities.** *Annu Rev Genet* 2004, **38**:525-552.
4. Handelsman J: **Metagenomics: application of genomics to uncultured microorganisms.** *Microbiol Mol Biol Rev* 2004, **68**:669-685.
5. Edwards RA, Rohwer F: **Viral metagenomics.** *Nat Rev Microbiol* 2005, **3**:504-510.
6. Schuster SC: **Next-generation sequencing transforms today's biology.** *Nat Methods* 2008, **5**:16-18.
7. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R, *et al*: **The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.** *Nucleic Acids Res* 2005, **33**:5691-5702.
8. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, *et al*: **The RAST Server: rapid annotations using subsystems technology.** *BMC Genomics* 2008, **9**:75.
9. Meyer F, Overbeek R, Rodriguez A: **FIGfams: yet another set of protein families.** *Nucleic Acids Res* 2009, **37**:6643-6654.
10. McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, Disz T, Edwards RA, Gerdes S, Hwang K, Kubal M, *et al*: **The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation.** *Nucleic Acids Res* 2007, **35**:D347-353.
11. Overbeek R, Bartels D, Vonstein V, Meyer F: **Annotation of bacterial and archaealgenomes: improving accuracy and consistency.** *Chem Rev* 2007, **107**:3431-3447.
12. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, Paczian T, Rodriguez A, Stevens R, Wilke A, *et al*: **The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes.** *BMC Bioinformatics* 2008, **9**:386.