High-Throughput Next-Generation Sequencing Technologies Foster New Cutting-Edge Computing Techniques in Bioinformatics

Mary Qu Yang
National Human Genome Research Institute, yangma@mail.nih.gov

Brian D. Athey
National Center for Integrated Biomedical Informatics

Hamid R. Arabnia
University of Georgia, hra@cs.uga.edu

Andrew H. Sung
New Mexico Institute of Mining and Technology, andrew.sung@usm.edu

Qingzhong Liu
New Mexico Institute of Mining and Technology, liu@shsu.edu

See next page for additional authors

Follow this and additional works at: https://aquila.usm.edu/fac_pubs

Part of the Genetics and Genomics Commons

Recommended Citation
Available at: https://aquila.usm.edu/fac_pubs/8415

This Article is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
Introduction

High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics

Mary Qu Yang1, Brian D Athey2, Hamid R Arabnia3, Andrew H Sung4, Qingzhong Liu4, Jack Y Yang5, Jinghe Mao6 and Youping Deng*7

Address: 1National Human Genome Research Institute, National Institutes of Health (NIH), U.S. Department of Health and Human Services, Bethesda, MD 20892, USA, 2National Center for Integrated Biomedical Informatics (NCIIBI) and Center for Computational Medicine and Biology, University of Michigan, Ann Arbor, MI 48109, USA, 3Department of Computer Science, University of Georgia, Athens, GA 30602-7404, USA, 4Department of Computer Science, New Mexico Tech, Socorro, NM 87801, USA, 5Harvard University, P.O. Box 400888, Cambridge, MA 02140-0888, USA, 6Department of Biology, Tougaloo College, Tougaloo, MS 39174, USA and 7Department of Biological Sciences, University of Southern Mississippi, Hattiesburg, MS 39406, USA

Email: Mary Qu Yang - yangma@mail.nih.gov; Brian D Athey - bleu@umich.edu; Hamid R Arabnia - hra@cs.uga.edu; Andrew H Sung - sung@cs.nmt.edu; Qingzhong Liu - sung@cs.nmt.edu; Jack Y Yang - Dr.Yang@JHU.edu; Jinghe Mao - jmao@tougaloo.edu; Youping Deng* - Youping.deng@usm.edu

* Corresponding author

Abstract

The advent of high-throughput next generation sequencing technologies have fostered enormous potential applications of supercomputing techniques in genome sequencing, epi-genetics, metagenomics, personalized medicine, discovery of non-coding RNAs and protein-binding sites. To this end, the 2008 International Conference on Bioinformatics and Computational Biology (BioCOMP) – 2008 World Congress on Computer Science, Computer Engineering and Applied Computing (Worldcomp) was designed to promote synergistic inter/multidisciplinary research and education in response to the current research trends and advances. The conference attracted more than two thousand scientists, medical doctors, engineers, professors and students gathered at Las Vegas, Nevada, USA during July 14–17 and received great success. Supported by International Society of Intelligent Biological Medicine (ISIBM), International Journal of Computational Biology and Drug Design (IJCBDD), International Journal of Functional Informatics and Personalized Medicine (IJFIPM) and the leading research laboratories from Harvard, M.I.T., Purdue, UIUC, UCLA, Georgia Tech, UT Austin, U. of Minnesota, U. of Iowa etc, the conference received thousands of research papers. Each submitted paper was reviewed by at least three reviewers and accepted papers were required to satisfy reviewers’ comments. Finally, the review board and the committee decided to select only 19 high-quality research papers for inclusion in this supplement to BMC Genomics based on the peer reviews only. The conference committee was very grateful for the Plenary Keynote Lectures given by: Dr. Brian D. Athey (University of Michigan Medical School), Dr. Vladimir N. Uversky (Indiana University School of Medicine), Dr. David A. Patterson (Member of United States National Academy of Sciences and National Academy of Engineering, University of California at Berkeley) and Anousheh Ansari (Prodea Systems, Space Ambassador). The theme of the conference to promote synergistic research and education has been achieved successfully.


This article is available from: http://www.biomedcentral.com/1471-2164/10/S1/I1

© 2009 Yang et al; licensee BioMed Central Ltd.
This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Introduction
With the advent of high-throughput next generation sequencing technologies, gigabases of sequence information can be obtained in just a few days. The technologies offer drastically faster and cost-effective sequence throughput and are vastly superior to shotgun sequencing due to the high volume of data and the drastically short time to sequence a whole genome or disease genome, but genome assembly is much more computationally expensive. Therefore, the next generation sequencing technologies will foster enormous potential applications of supercomputing techniques in genome sequencing, epigenetics, metagenomics, personalized medicine, discovery of non-coding RNAs and protein-binding sites. Furthermore, Next-generation sequencing will substitute microarray, the mostly used technology in genomics and bioinformatics. Like next-generation sequencing, microarrays can be used to examine thousands of genes in one experiment and can be used to obtain gene profiles, but the drawback of Microarrays are based on hybridization. Gene expression levels are measured by fluoresce from hybridizes but quantification of the fluorescence of vast amount of spots on a chip are often unreliable from experiment to experiment. Furthermore many DNA samples can hybridize to more than one spot, thus, generating misleading results. Next-generation sequencing overcomes problems of Microarrays by generating actual sequence reads and is ideally detect genetic mutations. To measure the gene expression level reflected by the amount of a particular RNA molecule, simply tally up the number of sequence reads corresponding to that RNA molecule rather measuring an inaccurate fluorescent spot and trying to control for all sorts of experimental variation. Gene expression can be actually more accurately obtained by counting sequence reads. Next generation sequencing has distinct advantages to obtain regulatory markings in chromatin, and to identify neural regulatory protein binds in the genome, as well as to investigate differences between stem cells and differentiated cells such as cancerous cells, and to determine how gene regulatory network can be altered by an activated external signal. Next generation sequencing is cheaper, fast, and less time consuming but computational expensive. Given situations, the development of computational techniques is important for future bioinformatics data mining. To this end, the International Society of Intelligent Biological Medicine [http://www.ISIBM.org] works with academic conferences to promote the cutting edge research.

Research presentations
ease study and biology database. The authors presented a variety of approaches and tools that can be used in analyzing next generation sequencing data. Meanwhile the researchers can greatly benefit from quality and quantity of data generating by next generation sequencing technology.

**Future meeting**
The next annual conference will be held in same location in Las Vegas, Nevada on July 13–16, 2009. The web site, [http://www.world-academy-of-science.org](http://www.world-academy-of-science.org), contains further information on future meetings. The meeting is a large international conference with more than two thousand attendees from more than 80 countries annually. The conference continuously aims at promoting computational science research and education in biomedical sciences. For articles from the 2008 conference, please see [http://www.biomedcentral.com/1471-2164/9?issue=S1](http://www.biomedcentral.com/1471-2164/9?issue=S1), and also [http://www.biomedcentral.com/1471-2164/9?issue=S2](http://www.biomedcentral.com/1471-2164/9?issue=S2) for further reading.

**Competing interests**
The authors declare that they have no competing interests.

**Authors’ contributions**
All authors provided professional services to the conferences and contributed in writing this introductory article. All authors reviewed and agreed on the content of this introductory article.

**Acknowledgements**
This article has been published as part of *BMC Genomics* Volume 10 Supplement 1, 2009: The 2008 International Conference on Bioinformatics & Computational Biology (BIOCOMP’08). The full contents of the supplement are available online at [http://www.biomedcentral.com/1471-2164/10/S1](http://www.biomedcentral.com/1471-2164/10/S1).

**References**