HKU Centre for Genomic Sciences and HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory Joint Workshop

September 5, 2012

Lecture Theatre 2
Cheung Kung Hai Conference Centre
Faculty of Medicine Building, HKU

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF
COMPUTER SCIENCE

CENTRE FOR GENOMIC SCIENCES
LI KA SHING FACULTY OF MEDICINE
THE UNIVERSITY OF HONG KONG

Contents

Introduction

This one-day workshop, co-organized by HKU Centre for Genomic Sciences (CGS) and HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory (BAL), provides a forum for local researchers in Bioinformatics to share research ideas and discuss research works related to the Next Generation Sequencing technology. It will feature one keynote speech and seven talks by researchers from CGS, BAL, BGI-Shenzhen, HKU Department of Computer Science and School of Biological Science. Recent results on meta-genomics, SNP detections, short read alignment, disease gene identification and systems biology will be presented.

The workshop will be preceded by the inauguration ceremony of the HKU-BGI Bioinformatics Algorithms and Core Technology Research Laboratory (BAL). The lab is a research collaboration between HKU Computer Science department and BGI-Shenzhen. Its main mission is to foster the research and development in computing technologies for high-throughput analysis of sequence data. Prof. Jun Wang, the executive director of BGI-Shenzhen, will give a keynote speech for the ceremony.

Organizing Centres

Centre for Genomic Sciences

Li Ka Shing Faculty of Medicine

The University of Hong Kong

The Centre for Genomic Sciences (formerly known as Genome Research Centre) was established to provide leadership in genome research in Hong Kong and the region by developing expertise and infrastructure for studies in genomics, proteomics, and bioinformatics. It also aims to facilitate the translation of knowledge into applications for the understanding of disease mechanisms and for the development of diagnostic and therapeutic measures.

The Centre has participated in the International HapMap Project, in collaboration with other major genome centres around the world. During the outbreak of SARS in 2003, the Centre supported the Departments of Microbiology and Zoology in the identification and subsequent sequencing of the agent of the disease. The Centre also undertook the whole genome sequencing project of *Laribacter hongkongenesis*. Centre scientists are developing methodologies for studying the role of genetic variations in disease, and have identified genetic loci that contribute to Mendelian and complex disorders.

The Centre contributes to multidisciplinary research through experimental design, data generation, data analysis and bioinformatics support. The scope of our collaborative projects covers genetic linkage and association studies, transcriptome analysis of various tissues and diseases, novel mutation detection by exome and other targeted sequencing, metagenomics, and *de novo* assembly of new genomes.

The Centre also offers professional Core Services with advanced high-throughput technology platforms such as the Affymetrix GeneChip, the Sequenom MassArray, the Bio-Plex suspension array and the Pyrosequencer. Other fundamental tools such as DNA sequencer, real-time PCR, bioanalyzer, fluorescence image scanner and oligonucleotide synthesis are also provided. Next-generation sequencing platforms using Illumina GAIIx and Roche 454 GS FLX, as well as the Raindance (for target enrichment) are also available. The Core Service platforms have served over 600 researchers since its establishment with a current monthly average of 360 jobs.

The Centre is very active in outreach to promote genomics in the forms of public forums, mass media interviews, conferences, workshops, technology seminars, and education and training to students, postdoctoral fellows, scientists and clinicians.

Organizing Centres

HKU-BGI Bioinformatics Algorithms and Core Technology Research Laboratory

Faculty of Engineering

The University of Hong Kong

Since 2008, the Computer Science Department of HKU and BGI, the world's largest genomic institute, based in China, have been working closely on the algorithmic, analytics and engineering aspects of computing technologies for the enhancement of the throughput and quality of the analysis of sequencing data. The research outputs have led to state-of-the-art software such as SOAP2 (2008) and SOAP3(2011) for short read alignment; SOAPsplice (2010) and SOAPfusion (2011) for RNA-Seq analysis. Some of these softwares are now the core analysis tools in BGI's sequencing workflow.

For deeper collaboration towards research advancement, the Computer Science Department of HKU and BGI commit to establish a joint research laboratory, namely HKU-BGI Bioinformatics Algorithms and Core Technology Research Laboratory (HKU-BGI BioAlgo Lab or BAL in short) to enhance their cooperation in research and education. BGI will invest HK$10 Million to setup one laboratory located at BGI Shenzhen while HKU will finance and setup another one in the HKU Campus.

Mission of the lab:

- To be a leader in the research and development of algorithmic and analytic techniques and computing technologies for high-throughput analysis of sequencing data.
- To pioneer the standardization of the core analysis tools for sequencing output.
- To support and facilitate other HKU research units on computational biology & genomics with advanced analytics techniques and software.
- To provide opportunities of postgraduate studies in bioinformatics computing technologies with joint supervision from HKU and BGI.

Also, our web page is www.bal.hku.hk

Program

Inauguration Ceremony	
10:00 – 10:15am	**Opening Speeches by Guests of Honors** ***Professor Norman C Tien*** *Dean* *Faculty of Engineering, HKU* ***Professor Pak Chung Sham*** *Director* *Centre for Genomic Sciences* *Chair Professor of Psychiatric Genomics* *Head of Department of Psychiatry* *Faculty of Medicine, HKU* ***Professor Frederick C C Leung*** *Professor* *School of Biological Sciences* *Faculty of Science, HKU*
10:15 – 11:00am	**Keynote Speech** **Life is a Game of Evolution** ***Professor Jun Wang*** *Executive Director* *BGI-Shenzhen*

Program

Workshop	
Session 1	
Chairperson:	***Dr H F Ting*** *Associate Professor* *HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory* *Department of Computer Science* *Faculty of Engineering, HKU*
11:30am – 12:00pm	**Keynote Speech** **Meta-genomics Assembly and Binning** ***Professor Francis Y L Chin*** *Associate Dean of Engineering* *Chair of Computer Science* *Taikoo Professor of Engineering* *Faculty of Engineering, HKU*
12:00 – 12:25pm	**FaSD: A Fast and Accurate SNP Detection Algorithm for Next-generation-sequencing Data** ***Dr Junwen Wang*** *Assistant Professor* *Centre for Genomic Sciences* *Department of Biochemistry* *Faculty of Medicine, HKU*
12:30 – 2:00pm	Lunch
Session 2	
Chairperson:	***Dr Stacey Cherny*** *Assistant Professor* *Centre for Genomic Sciences* *Department of Psychiatry* *Faculty of Medicine, HKU*
2:00 – 2:25pm	**Personal Genomes are Personalized** ***Mr Yingrui Li*** *Deputy Director* *BGI Research Institute*

Program

2:25 – 2:50pm	**Efficient SNP-sensitive Alignment and Database-assisted SNP Calling for Low Coverage Samples** ***Dr H F Ting*** *Associate Professor* *HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory* *Department of Computer Science* *Faculty of Engineering, HKU*
2:50 – 3:15pm	**A Comprehensive Bioinformatics Framework for Disease Gene Identification Using Exome Sequencing Data and its Application** ***Dr Miaoxin Li*** *Postdoctoral Fellow* *Centre for Genomic Sciences* *Department of Psychiatry* *Faculty of Medicine, HKU*
3:15 – 3:30pm	Tea Break
Session 3	
Chairperson:	***Dr S M Yiu*** *Associate Professor* *HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory* *Department of Computer Science* *Faculty of Engineering, HKU*
3:30 – 3:55pm	**A Metagenome-wide Association Study of Gut Microbiota in Type 2 Diabetes** ***Dr Junjie Qin*** *BGI-Shenzhen*
3:55 – 4:20pm	**Systems Biology of High Energy, Fast-growing Plant** ***Dr Wallace B L Lim*** *Associate Professor* *School of Biological Sciences* *Faculty of Science, HKU*
4:20 – 4:45pm	**SOAP3-dp: A GPU-based Dynamic Programming Tool for Short Read Alignment** ***Dr Thomas K F Wong*** *Postdoctoral Fellow* *HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory* *Department of Computer Science* *Faculty of Engineering, HKU*


bioinformatics
Talks

Keynote Speeches

Life is a Game of Evolution

Jun Wang
BGI-Shenzhen

About the Speaker

WANG Jun (WJ), who was born on June 4 1976, graduated Ph.D. from the Peking University in 2002, the same year as he received the national excellent Ph.D. thesis for highest academic standing from the Ministry of Education, China.

The Bioinformatics Department of Beijing Genomics Institute (BGI) was founded by his efforts in 1999. It has been moved to Shenzhen, and has been referred as BGI Shenzhen since 2007. It is now widely recognized as one of China's premier research facilities, committed to excellence in genome sciences. WJ has been leading the scientific direction and daily operation of BGI genomics and its informatics part since 2002. In 2003, WJ was appointed as associate director and professor at BGI. During 2004-06 he was for 24 months invited guest professor at the University of Aarhus, Denmark. WJ was then appointed as Ole Rømer professor of University of Southern Denmark from September 2006 to July 2009, the professorship has been transferred to University of Copenhagen since October 2009. In 2005, he was appointed professor of genomics (personal chair) at the life science college, Peking University and since 2007 he has been chairing the same position at the University of Aarhus.

For the last 8 years WJ has been leading the genomics institution of 1000+ people engaged in studies of genomics and its informatics. In 1999, WJ finished the analysis of a 1% region of human genome, which is sequenced by BGI. Then he devoted himself to genomics and its analysis, including genome assembly, annotation, expression, genome duplication, comparative genomics, molecular evolution, transcriptional regulation, genome variation analysis, database construction as well as related methodology development such as the sequence assembler and alignment tools. He also focuses on interpretation of the definition of "gene" by expression and conservation study. In 2003, Jun Wang was also been involved in the SARS genome analysis and the silkworm genome assembly and analysis in cooperation with Chinese Southeast Agricultural University. The Pig Genome Survey Project was also completed with his leading effort at BGI. He has led a group finishing the chicken genome variation map and the TreeFam in collaboration with the Sanger Institute. Recently, he and his group have finished the first Asian diploid genome, the 1000 genome project, large genome association study of Diabetes; metagenomics project of gut microbita. He is also leading genome variation projects of rice, silkworm, and pig genome project, as well as the molecular mechanisms of the domestication process. His research focuses on genomics and related bioinformatics analysis of common diseases and agricultural crops, with the goal of developing applications using this genomic information.

WJ has been mentor for 12 academics who have defended their Ph.D. degrees, and is supervising 15 Ph.D. students. He has three years of bioinformatics teaching at the graduate school of CAS and Peking University, and has been awarded the "CAS Excellent Course and Teaching" award 2004. He has authored 100+ peer-reviewed original papers – of which 32 are published in Nature (including Nature series), and Science. (12 as cover story); among those 32, WJ is the first/co-first author or corresponding/co-corresponding author for 19 (8 as cover) of them.

For the scientific achievements, WJ has been recognized with Award from His Royal Highness Prince Foundation, Young Elite Scientist from the Danish Research Council, Lundbeck Talent Price, Outstanding Science and Technology Achievement from the Chinese Academy of Sciences, Top 10 Scientific Achievements

In China, The first "TopSUN" Scientific Paper Award from Peking University, Tan Jiazhen Life Science Award from Fudan University, and Prize for Important Innovation and Contribution from Chinese Academy of Sciences.

Abstract

Nothing in biology makes sense except in the light of evolution. All the creatures on the planet can be drawn into a "Tree of Life" by their genomes and other omics data, while the comparison between these creatures shed light on how genomes are functioning in adaptation to different bio-niches.

These information would help us build the digital library of life, understand the relationship between traits and genes, and breed desired varieties under guidance of genetic information. Within a species, all the individuals can be organized into a "Tree of Individuals". Population and statistical genetics decipher how natural selection impact on genes. In the case of human, these impacts underlie medical relevant phenotypes and quantitative traits and thus help understand the genetics and molecular biology of Mendelian and complex disorders. Finally, within an organism, cells contain somatic mutations and epigenetic changes and are also under selection, driving the development of different tissues and organs.

Single-cell analysis would map the "tree of cells", which addresses essential questions such as embryo development and tumor progression. Here we demonstrate genomics studies in the three hierarchies of evolution and promote a new way of organized thinking about the future of omics.

Keynote Speeches

Meta-genomics Assembly and Binning

Francis Y L Chin

Department of Computer Science, Faculty of Engineering, HKU

About the Speaker

Professor Chin is the Associate Dean of Engineering, Taikoo Professor of Engineering and Chair of Computer Science at the University of Hong Kong. Professor Chin received the B.A.Sc. degree from the University of Toronto, Canada, in 1972, and the M.S., M.A. and Ph.D. degrees from Princeton University, in 1974, 1975, and 1976, respectively. Since 1975,

he has taught at the University of Maryland, Baltimore County, University of California, San Diego, University of Alberta, Chinese University of Hong Kong, and University of Texas at Dallas. He joined the University of Hong Kong (HKU) in 1985, where he is the Chair of the Department of Computer Science and was the founding Head of the department from its establishment until December 31, 1999. Between 1992-1996, he served as the Associated Dean of Graduate School. In 1996, Prof. Chin was elected to the grade of IEEE Fellow. Professor Chin is currently serving as Manager Editor of the International Journal of Foundations of Computer Science and is also on the editorial boards of several journals. He has served on many program committees, including RECOMB 2013, and as conference chairman, including APBC 2007 and TAMC 2013, of numerous international workshops and conferences. Professor Chin's research interests are bioinformatics, design and analysis of algorithms and on-line algorithms.

Abstract

Meta-genomics is the study of the collective genomes of all microorganisms from an environmental sample (also known as environmental genomics, or community genomics), for example, the diversity of microbes in humans is found to be associated with some common diseases such as gastrointestinal disturbance and inflammatory bowel disease. High-throughput next-generation sequencing (NGS) techniques enable researchers to directly sequence the genomes of multiple species obtained from such an environmental sample for analysis. In this talk, we address two important problems in meta-genomic analysis based on NGS data: (1) assembly - to reconstruct the genomes of each species; (2) binning - to group DNA fragments (or reads) from similar species together. Both of these problem are very difficult especially when up to 99% of the species found in environmental samples are unknown. The bioinformatics team at HKU has developed a series of IDBA assembly and MetaCluster binning tools for meta-genomics study. The main ideas of these software tools will be presented in this talk.

Talks

FaSD: A Fast and Accurate SNP Detection Algorithm for Next-generation-sequencing Data

Junwen Wang

Centre for Genomic Sciences, Department of Biochemistry,
Faculty of Medicine, HKU

About the Speaker

Dr. Junwen John Wang is an Assistant Professor at Department of Biochemistry and Centre for Genome Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong. He obtained his MS in Computer Science from the University of Pennsylvania, and his PhD in Bioscience from University of Washington. He was a staff scientist at National Institute of Cancer and a postdoc fellow at Center for Bioinformatics, University of Pennsylvania, USA. Dr. Wang was a recipient of Claire & Egtvedt Fellow and NRSA Computational Genomics Fellow. His major research areas are in developing analysis pipelines for next generation sequencing, gene regulatory network inference and protein-DNA interactions. His research was supported by grants from the Research Grants Council of Hong Kong, the Food & Health Bureau of Hong Kong, and NSFC of China and RGC of Hong Kong joint scheme.

Abstract

Tools have been developed to call Single Nucleotide Polymorphism (SNP) from Next-generation-sequencing (NGS) data. However, most of them require high sequencing depth for satisfactory performance, which is expensive to obtain. Here, we propose a fast and accurate SNP-detection program, FaSD, which utilizes a binomial distribution based algorithm and a transition/transversion ratio based posterior probability to detect SNPs. We extensively assessed the program on normal and cancer NGS data from The Cancer Genome Atlas (TCGA) project, and trio's data from the 1000 Genome Project. We also compared several state-of-the-art programs for SNP calling quality and analysed the pros and cons of these programs. We found FaSD highly accurate in SNP detection, particularly when the sequence depth is low. The program is also very fast, finishing SNP calling within four hours for ten-fold human genome NGS data (total 30 gigabases) on a standard desktop computer.

Feng Xu[1,2,†], Weixin Wang[1,2,†], Panwen Wang[1,2], Pak Chung Sham[3,4], and Junwen Wang[1,2,3*]

[1] Department of Biochemistry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

[2] Shenzhen Institute of Research and Innovation, The University of Hong Kong, Shenzhen, China.

[3] Centre for Genomic Sciences, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

[4] Department of Psychiatry, LKS Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China.

[*] Correspondence should be addressed to Junwen Wang
(Tel: +852 2831 5075; Fax: +852 2855 1254; Email: junwen@hku.hk)

[†] Both authors contribute equally to this work

Talks

Personal Genomes are Personalized

Yingrui Li

BGI Research Institute

About the Speaker

Yingrui Li (Li Y), born on May 24,1986, was the chairman of the Student Union when he studied at Chengdu Sichuan NO.7 middle school in 2001-2004, then he was recommended to College of Life Sciences, Peking University, in the winter camp of national Biology Olympiad Competition training team in 2004. In 2006, he came to BGI as a trainee when he was still a sophomore. From then on, he involved himself in the research of algorithms and data structures, computer programming, mathematical statistics, modeling methods, data mining techniques and genomics analysis. The bioinformatics analytical platform based on the next-generation sequencing of BGI was constructed with his great help. In the past few years, he engaged or organized a number of national projects, including Yan Huang Project, 1000 Genomes Project, Yan Huang Whole Genome Methylation, Cancer Genome Project and so on. 36 papers were published in high-impact journal including *nature* and *science* until now, impact factors of 24 was more than 25, he first/co-first author for 11 of 36 papers.

At present, Li Y lead a team with hundreds of people engaging in DNA sequence assembly, protein sequence assembly, functional element annotation, comparative genomics, cancer genome, complex diseases, molecular evolution and breeding of animals and plants, epigenetics, metagenomics, stem cell application techniques, single cell operational. On the basis of high throughput sequencing data, he and his team make great effort to explore the new frontier of life science and develop the application of bio-technology combine the mathematical statistics, data algorithms and structures, computer software development, data mining and modeling as well as the biology idea.

Abstract

Advances in technology has reduced the cost of DNA sequencing down by orders of magnitude and brought the practice of personal genomes and personalized medicine into immediate future. Nevertheless, there is no consensus on the concept of personal genomes and what are the substantial scientific questions and applications. Here we first challenge the misleading re-sequencing approach on personal genomes by a full spectrum of genetic variations and their indications on functions. We then demonstrated that a complete picture by large-scale data acquirement and new analysis of personal genetic, epigenetic, transcriptomics, proteomics, metabolic, and environmental information would dramatically help understanding complex diseases and cancer.

Efficient SNP-sensitive Alignment and Database-assisted SNP Calling for Low Coverage Samples

H F Ting

HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory, Department of Computer Science, Faculty of Engineering, HKU

About the Speaker

Dr. Hingfung Ting is an associate professor at the Department of Computer Science, University of Hong Kong. He obtained his MPhil at the University of Hong Kong, and his PhD at Princeton University. His main research areas are design and analysis of algorithms and computational complexity. Recently, his research is focused on designing efficient algorithms and developing practical software tools for various short read alignment problems such as SNP-sensitive alignment and alignment of bisulfite-converted DNA.

Abstract

We have designed and implemented an efficient read alignment tool that is sensitive to a given set of SNPs. In particular, it returns alignments that permit mismatches at these SNPs. We then make use of it to develop a new SNP detection tool, which allows user to provide annotated SNPs classified in previous studies and use them to guide the detection process. By focusing on alignments covering these SNPs, our tool greatly accelerates the discovery of SNPs at prescribed loci. These annotated SNPs also help us distinguish sequencing error from authentic SNP alleles more easily. We have implemented our method and compared it with existing method on several applications. We found that our method have higher accuracy, especially for sample with low coverage. It is faster, and can be about two orders of magnitude faster for some applications.

Talks

A Comprehensive Bioinformatics Framework for Disease Gene Identification Using Exome Sequencing Data and its Application

Miaoxin Li

Centre for Genomic Sciences, Department of Psychiatry,

Faculty of Medicine, HKU

About the Speaker

Dr. Miaoxin Li is a Postdoctoral Fellow at Department of Psychiatry and Centre for Genome Sciences, Li Ka Shing Faculty of Medicine, the University of Hong Kong. He obtained his Bachelor of Science degree at Hunan Normal University, and his PhD at Biochemistry department of University of Hong Kong. Before he moved to Hong Kong, he worked at the Shanghai Center for Bioinformation Technology for 2 years. His main research interests and areas are the development of advanced statistical and bioinformatics methods for identifying genetic factors responsible for human diseases, and their application to specific human diseases, in which he has spent over ten years and published over 40 peer-reviewed papers. At HKU, he developed a series of novel integrative approaches and software tools (e.g., KGGSeq, KGG, and IGG) to combine the bioinformatics and statistical genetics information to more powerfully detect genetic risk factors of human diseases. Currently, he is working on the development of integrative knowledge-based approaches to isolate genetic loci or genes predisposing to complex diseases using next-generation sequencing data, including whole-exome sequencing, whole-genome sequencing and whole-transcriptome shotgun sequencing.

Abstract

Exome sequencing strategy is promising for finding novel mutations of human genetic disorders. However, isolating the casual mutation(s) from the large amount of exome sequencing variants in a small sample is still a big challenge. Here, we propose a multi-layer filtration and prioritization bioinformatics and statistical framework to identify the casual mutation(s) in exome sequencing studies. This comprehensive framework was able to efficiently narrow down the whole exome variants to tiny numbers of candidate variants in the proof-of-concept examples and has been successfully applied to a number of exome sequencing projects. The proposed framework has been implemented in a user-friendly software package, named KGGSeq (http://statgenpro.psychiatry.hku.hk/kggseq). KGGSeq has been downloaded by a lot of international genetic investigators since its release to the public.

A Metagenome-wide Association Study of Gut Microbiota in Type 2 Diabetes

Junjie Qin

BGI-Shenzhen

About the Speaker

Dr. Junjie Qin received his Ph.D. from the Beijing Institute of Genomics, Chinese Academy of Sciences in 2011. From 2006 to now, Dr. Qin was involved in or led many projects in BGI, such as Silkworm genome fine map, TreeFam, The first Asian genome (Yanhuang), Human Pan-genome and a lot of microbial genome projects, especially the 2011 Germany E. coli O104:H4 project. In 2008, he joined into the MetaHIT (Metagenomics of the Human Intestinal Tract), which is a large project financed by European Commission under the 7th FP program. Since that time, Dr. Qin has mainly focused on studying the human gut microbiome, especially the development of bioinformatics tools and biological analysis based on the omics data.

Abstract

Assessment and characterization of gut microbiota has become a major research area in human disease, including Type 2 Diabetes (T2D), the most prevalent endocrine disease worldwide. To carry out analysis on gut microbial content in T2D patients, we developed a protocol for a Metagenome-Wide Association Study (MGWAS) and undertook a two-stage MGWAS based on deep shotgun sequencing of the gut microbial DNA from 345 Chinese individuals. We identified and validated ~60,000 T2D-associated markers and showed that T2D patients were featured by a moderate-degree of gut microbiota dysbiosis. Our data provide insight into the characteristics of the gut metagenome related to T2D risk, a paradigm for future studies of the pathophysiological role of the gut metagenome in other relevant disorders, and the potential usefulness for a gut–microbiota-based approach for assessment of individuals at risk of such disorders.

Talks

Systems Biology of High Energy, Fast-growing Plant

Wallace B L Lim

School of Biological Sciences, Faculty of Science, HKU

About the Speaker

Dr. Lim obtained a Ph.D. degree in Biochemistry from the University of Oxford and is now an Associate Professor of the School of Biological Sciences at the University of Hong Kong. He has published almost 50 original articles in plant sciences and molecular biology. His current research interests include systems biology and carbon flow of plants, phosphorus cycle in the biosphere, and agrobiotechnology. He has filed several patents on plant genetic engineering, which help to improve the energy harvest process of plants. These inventions were shown to speed up plant growth and improve the seed yield by 50% in model plants. Dr. Lim is currently exploring the impact of his inventions in food crops and bioenergy. His invention can be viewed here: http://www.youtube.com/watch?v=8xS8iVqToK8

Abstract

Photosynthesis is the ultimate source of energy for most life forms on this planet. Energy harvest in plants starts from photosystems, and subsequently involves Calvin cycle, glycolysis, TCA and mitochondrial respiratory chain. These processes are regulated by three regulatory mechanisms: gene transcription, enzyme regulation by redox or allosteric regulation of enzymes. Most of the energy conversion processes are carried out in two key endosymbiotic organelles: chloroplasts and mitochondria. AtPAP2 is a purple acid phosphatase dual-targeted to both organelles. Over-expression (OE) lines of AtPAP2 grew faster and produced more seeds (http://www.youtube.com/watch?v=8xS8iVqToK8). The level of ATP and ADP in the OE lines are 100% and 35% higher than their levels in wild-type plants respectively, which could be the driving force of their fast-growth phenotypes. The energy-rich AtPAP2 OE line provides a tool for studying the regulation of energy system in plant. While the transcription of PSI, PSII and Lhca genes are unaltered, changes in Lhcb genes are sufficient for achieving a higher energy harvesting efficiency. The gene expression of most enzymes of the Calvin cycle was unaltered, indicating these enzymes are mainly regulated by light/redox status, but not at transcription level. Similarly, the gene expressions of almost all enzymes of glycolysis and TCA cycle were unaltered, indicating the activities of these enzymes are regulated by allosteric modulation through the products (citrate, ATP, etc). Nonetheless, our results indicates that transcriptional regulation do play a role in sucrose and starch metabolism, nitrogen, potassium, iron uptakes, amino acids metabolism and secondary metabolites. Integration of proteomics, metabolomics and sRNA profiling data provides further insights on the systems biology of this high energy and fast-growing plant.

SOAP3-dp: A GPU-based Dynamic Programming Tool for Short Read Alignment

Thomas K F Wong

HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory, Department of Computer Science, Faculty of Engineering, HKU

About the Speaker

Dr. Thomas K.F. Wong received his Ph.D. degree in Computer Science from the University of Hong Kong in 2011. He is currently a postdoctoral fellow at the department of Computer Science, University of Hong Kong. His research interest is bioinformatics.

Abstract

Existing short read alignment tools either are not fast enough or cannot handle gaps well. By a skillful exploitation of whole genome indexing and dynamic programming on a GPU, we devised a GPU-based tool called SOAP3-dp that can find alignments involving mismatches and INDELs, and it achieves a drastic improvement in speed and better sensitivity over all existing tools. In our experiments, when compared to BWA and the newly released Bowtie2, the speedup gained by SOAP3-dp ranges from 12 to 23 times. SOAP3-dp can align more reads than both tools for both single reads and paired-end reads. SOAP3-dp is also faster than its predecessor SOAP3 (allowing mismatches only) by 1.6 times, showing that GPU-based dynamic programming coupled with indexing can be much more efficient in dealing with reads with gaps and more mismatches.

Notes

Notes

CENTRE FOR GENOMIC SCIENCES

LI KA SHING FACULTY OF MEDICINE
THE UNIVERSITY OF HONG KONG

One Centre...

Invitrogen **ABI 3730xL DNA Analyzer**

Sanger sequencing:

- Read length up to 1,000 bp
- Fragment sizing

Illumina **iScan System**

Genotyping, Gene expression and Methylation studies

NEW

Qiagen **EZ1 Advanced XL**

Fully automated extraction of DNA/RNA from various sample types

Affymetrix **GeneChip System**

- Gene expression profiling
- miRNA analysis
- Genome-wide genotyping
- Copy number analysis

IDT **Oligonucleotide Ordering Portal**

Quality oligos from Integrated DNA Technologies via online IDT-CGS Portal

ORDER IDT Oligos via CGS
www.IDTDNA.com/UHK-GRC/
OLIGO@GENOME.HKU.hk

Qiagen **PSQ 96MA**

Pyrosequencing:

- Quantitative methylation analysis
- Genotyping
- Allelic quantification

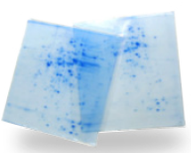
Invitrogen **ABI Prism 7900HT Sequence Detection System**

Realtime PCR for:

- Transcript quantitation
- Genotyping

2D Gel Electrophoresis System

Protein expression profiling



Differential Gel Electrophoresis (DiGE) System

Multiplex fluorescent-labeled 2D gel

Sequenom MassARRAY System

- Genotyping
- Somatic mutation profiling

AB/Sciex 4800 MALDI TOF/TOF™ Analyzer

Mass spectrometer analysis:

- Protein identification
- Biomolecules mass determination

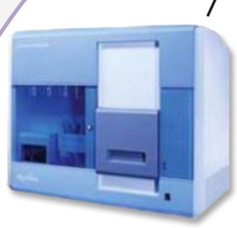
Agilent 2100 Bioanalyzer

DNA/RNA quality assessment

Bio-Rad Bio-Plex Suspension Array System

- Multiplex cytokine profiling
- Gene expression analysis

Solexa Genome Analyzer IIx



Next generation sequencing:

- Read length up to 150bp
- output 95GB/run

RainDance RDT 1000

Droplet PCR-based target enrichment:

- Size 1MB – 10MB
- Coverage >99%

Bioinformatics Support

Data analysis support with servers, PC workstations and software

Roche GS FLX System

Next generation sequencing:

- Modal read length 450bp
- Output 450MB/run

...Many Solutions!

HKU-BGI Bioinformatics Algorithms & Core Technology Research Laboratory

Faculty of Engineering
The University of Hong Kong

http://www.bal.hku.hk

Centre for Genomic Sciences

Li Ka Shing Faculty of Medicine
The University of Hong Kong

6th Floor
The Hong Kong Jockey Club Building
for Interdisciplinary Research
5 Sassoon Road
Pokfulam, Hong Kong

Tel: +852 2831 5500
Fax: +852 2818 5653
genome@hku.hk
http://genome.hku.hk