GigaGalaxy: A GigaSolution for reproducible and sustainable genomic data publication and analysis

Scott C Edmunds[1,2,*], Peter Li[1,2], Huayan Gao[3,4], Ruibang Luo[2], Dennis Chan[1], Alex Wong[1], Zhang Yong[2], Tin-Lap Lee[3,4]

1 BGI-Hong Kong Ltd., 16 Dai Fu Street, Tai Po Industrial Estate, NT, Hong Kong SAR, China.
2 BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, China.
3 School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.
4 CUHK-BGI Innovation Institute of Trans-omics, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.

* Corresponding Author: scott@gigasciencejournal.com

Today's next generation sequencing (NGS) experiments generate substantially more data and are more broadly applicable to previous high-throughput genomic assays. Despite the plummeting costs of sequencing, downstream data processing and analysis create financial and bioinformatics challenges for many biomedical scientists. It is therefore important to make NGS data interpretation as accessible as data generation. GigaGalaxy (http://galaxy.cbiit.cuhk.edu.hk) represents a NGS data interpretation solution towards the big sequencing data challenge. We have ported the popular Short Oligonucleotide Analysis Package (http://soap.genomics.org.cn) as well as supporting tools such as Contiguator2 (http://contiguator.sourceforge.net) into the Galaxy framework, to provide seamless NGS mapping, de novo assembly, NGS data format conversion and sequence alignment visualization. Our vision is to create an open publication, review and analysis environment by integrating GigaGalaxy into the publication platform at *GigaScience* and its GigaDB database that links to more than 40 Tetrabytes of genomic data. We have begun this effort by re-implementing the data procedures described by Luo et al., (*GigaScience* 1: 18, 2012) as Galaxy workflows so that they can be shared in a manner that can be visualized and executed in GigaGalaxy. We hope to revolutionize the publication model with the aim of executable publications, where data analyses can be reproduced and reused.

GigaGalaxy: http://galaxy.cbiit.cuhk.edu.hk
GigaDB: http://gigadb.org/
Code: https://github.com/gigascience

Licenses:
Galaxy/GigaGalaxy = Academic Free License v 3.0
GigaDB = BSD
SOAPdenovo2 = GPLv3