

GigaDB and Galaxy: revolutionizing data dissemination, organization and analysis

Peter Li

GigaScience

peter@gigasciencejournal.com



(GIGA)ⁿ SCIENCE

Journal and database for
large-scale data



in conjunction with



Editor-in-Chief: Laurie Goodman

Editor: Scott Edmunds

Commissioning Editor: Nicole Nogoy

Lead Curator: Tam Sneddon

Data Platform: Peter Li



www.gigasciencejournal.com



GigaScience aims to revolutionize data dissemination, organization, understanding, and use. An online open-access open-data journal, we publish 'big-data' studies from the entire spectrum of life and biomedical sciences. To achieve our goals, the journal has a novel publication format: one that links standard manuscript publication with an extensive database that hosts all associated data and provides data analysis tools and cloud-computing resources.

Our scope covers not just 'omic' type data and the fields of high-throughput biology currently serviced by large public repositories, but also the growing range of more difficult-to-access data, such as imaging, neuroscience, ecology, cohort data, systems biology and other new types of large-scale sharable data.

[Editorial Board](#) | [Editorial Team](#) | [Instructions for authors](#) | [FAQ](#)

Publication fees

There are currently no article processing charges (APCs) for articles published in *GigaScience* due to generous support from the [BGI](#).

[Submit a manuscript](#)[Register](#)[Sign up for article alerts](#)[Contact us](#)[Follow us on Twitter](#)[Support](#)

Articles

[Editor's picks](#)[Latest](#)[Most viewed](#)[Most forwarded](#)

The little piggy genome

An important model organism for medical research, the genome of the Wuzhishan inbred miniature pig is presented along with analysis of its unusual genetic features

GigaScience 2012, **1**:16



The Puerto Rican parrot genome

Community funded with an innovative approach to assembly, the critically endangered Puerto Rican Parrot (*Amazona vittata*) genome is presented here for the first time

GigaScience 2012, **1**:14

GigaScience is a new integrated [database](#) and journal co-published in collaboration between [BGI](#) [Shenzhen](#) and [BioMed Central](#), to meet the needs of a new generation of biological and biomedical research as it enters the era of "big-data." BGI (formerly known as Beijing Genomics Institute) was founded in 1999 and has since become the largest genomic organization in the world and has a proven track record of innovative, high profile research.

To achieve its goals, *GigaScience* has developed a novel publishing format that integrates manuscript publication with a database that will provide [DOI](#) assignment to every dataset. Supporting the open-data movement, we require that all supporting data and source code be publicly available in a suitable public repository and/or under a public domain [CC0 license](#) in the [BGI](#).



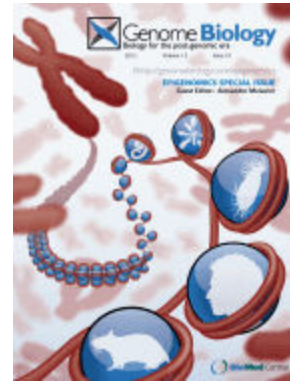
ADVERTISEMENT



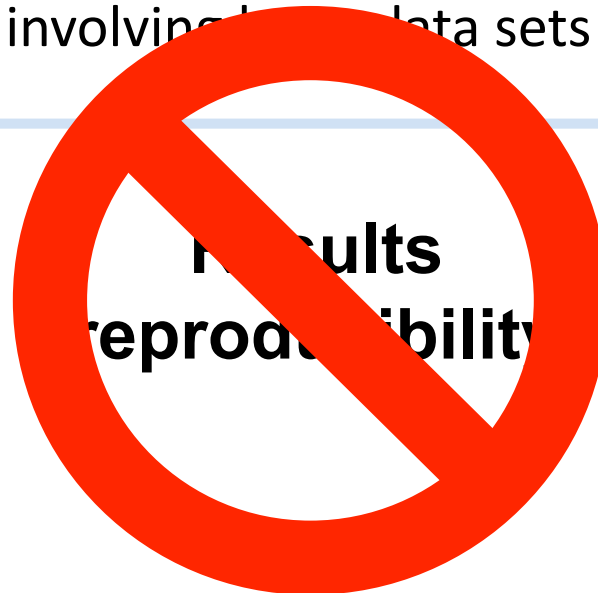
Recent articles include

Skene and Henikoff on nuclear reprogramming

Why another *omics journal?

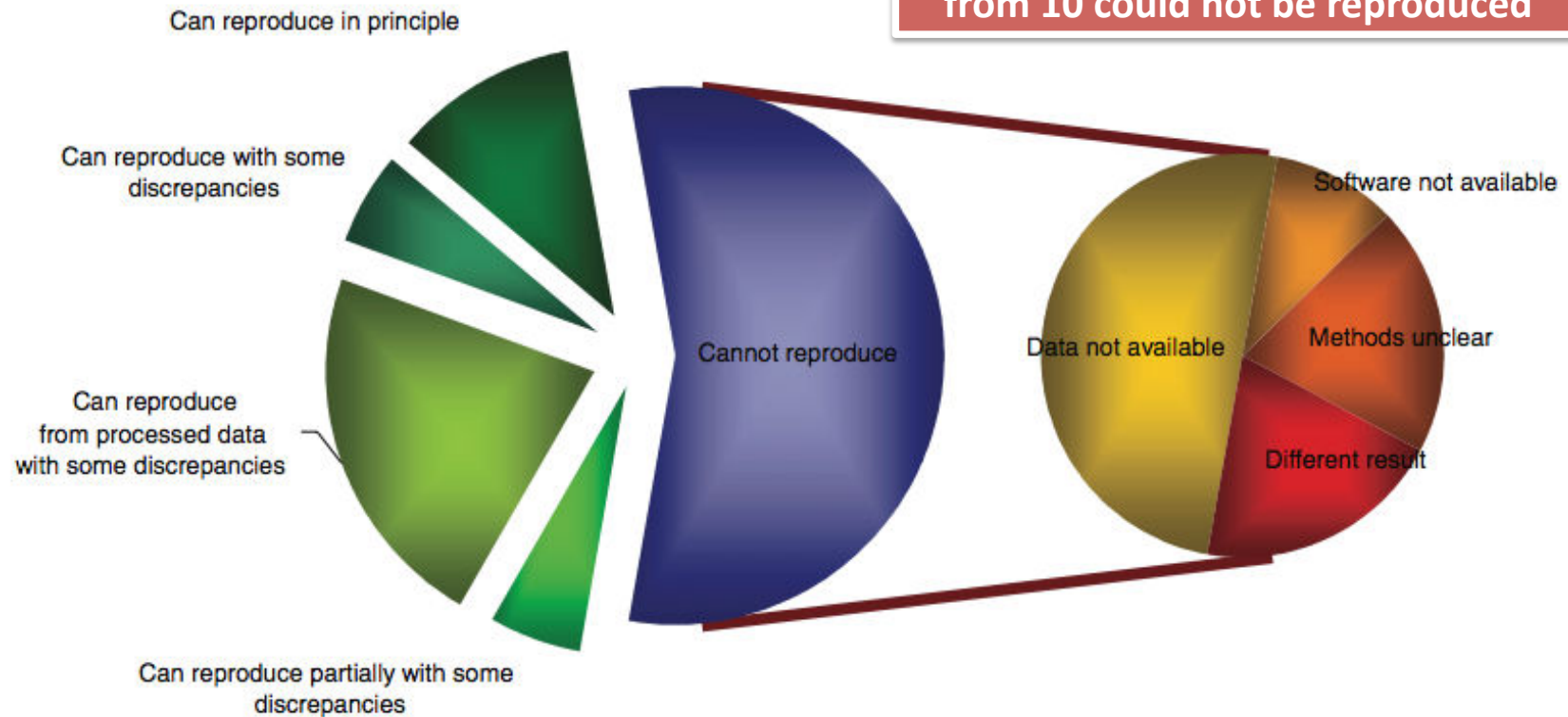


Already many journals publishing research
involving large data sets



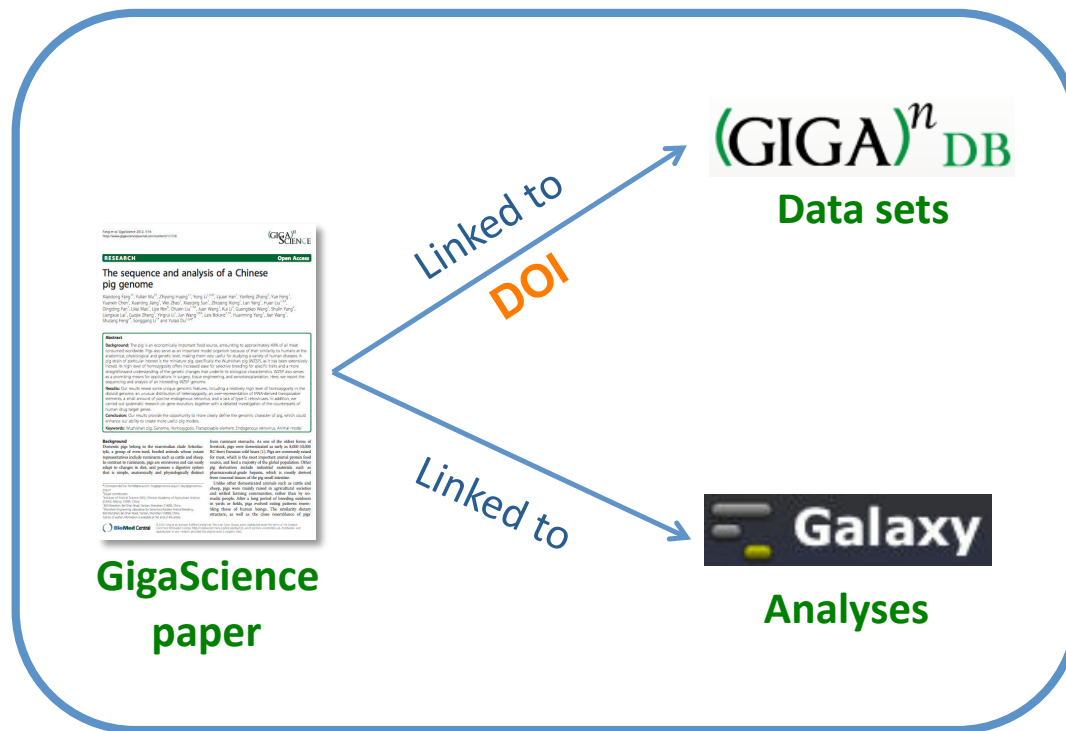
Unrepeatability of scientific results

Out of 18 microarray papers, results from 10 could not be reproduced



Ioannidis et al., 2009. Repeatability of published microarray gene expression analyses. *Nature Genetics* 41: 149-155.

How are we supporting data reproducibility?



Community tools for
data reproduction and reuse

DATA NOTE

Open Access

A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education

Taras K Oleksyk^{1*}, Jean-Francois Pombert², Daniel Siu³, Anyimilehidi Mazo-Vargas¹, Brian Ramos¹, Wilfried Guiblet¹, Yashira Afanador¹, Christina T Ruiz-Rodriguez^{1,4}, Michael L Nickerson⁴, David M Logue¹, Michael Dean⁴, Luis Figueroa⁵, Ricardo Valentin⁶ and Juan-Carlos Martinez-Cruzado¹

Additional file 15: Table S9. Bioinformatics tools and outputs for scaffold and gene annotation.

Additional file 16: Table S10. An example of annotation output produced by a student in the Genome annotation class using *A. vittata* genome.

Competing interests

Oleksyk TK, Pombert JF, Mazo A, Ramos B, Guiblet W, Afanador Y, Ruiz-Rodriguez CT, Nickerson ML, Logue D, Dean M, Figueroa L, Valentin R, and Martinez-Cruzado JC do not have competing interests. Siu D is employed by Axseq Technologies; the company which carried out the DNA Sequencing.

Authors' contributions

TKO, LF, RV, MD, MLN, DL and JCMC came up with the idea, and designed the experiments. TKO, WG, YA, CTRR and JCMC organized public support and raised the funds. TKO, AMV, BR, YA, CTRR and RV collected, extracted and quantified DNA. DS performed sequencing and assembly by SOAPdenovo. JFP performed assembly by Ray. TKO and WG designed the data browser webpage. TKO, JFP, MLN, DL, MD and JCMC wrote the paper. All authors read and approved the final manuscript.

6. Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, et al: **The genome of a songbird.** *Nature* 2010, **464**(7289):757–762.
7. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: An information aesthetic for comparative genomics.** *Genome Res* 2009, **19**(9):1639–45.
8. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, et al: **Hybrid error correction and de novo assembly of single-molecule sequencing reads.** *Nat Biotechnol* 2012, **30**(7):693–700.
9. Oleksyk TK, Guiblet W, Pombert JF, Valentin R, Martinez-Cruzado JC: **Genomic data of the Puerto Rican Parrot (*Amazona vittata*) from a locally funded project.** *GigaScience* 2012. <http://dx.doi.org/10.5524/100039>.
10. O'Brien SJ: **Genome empowerment for the Puerto Rican parrot – *Amazona vittata*.** *GigaScience* 2012, 1:13.

doi:10.1186/2047-217X-1-14

Cite this article as: Oleksyk et al: A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education. *GigaScience* 2012 1:14.

Linking of papers and data by citation of DOIs

← Data set DOI

← Paper DOI

Puerto Rican parrot » GigaDB

gigadb.org/puerto-rican-parrot/ Reader

gigasci xe java api citeulike blogs news misc work Import to Mendeley

www.gigasiencejournal.com/content/pdf/2047-217X-1-14.pdf

Puerto Rican parrot » GigaDB

(GIGA)ⁿ DB Beta

Revolutionizing data dissemination, organization, and use

Home About Contact Terms of use

search

Puerto Rican parrot

These data represent the first assembly of a genome sequence for a critically endangered parrot (*Amazona vittata*) endemic to the United States, and also the first genome of a species from the diverse and ecologically important genus *Amazona* native to South America and the Caribbean. One sample has been selected from the non-reproductive female at Rio Abajo Breeding Facility in Puerto Rico (IACUC#201109.1), and sequenced on Illumina HiSeq platform with both fragment and paired-end sequencing approaches, resulting in a total of 42,479,499,706 bases. We predicted a total coverage depth of 26.89X of the parrot's genome: 17.08X coverage for the short fragment reads, and 9.8X coverage for the mate pairs. The sequencing was initiated with the construction of two genome libraries: a short fragment library (~300 bp inserts) for sequencing the majority of the genome, and a long fragment library (~2.5 Kb inserts) to generate scaffolds to be used to order and assemble contigs derived from the short fragment library. The Illumina paired-end and mate-pairs reads were assembled together with Ray (<http://denovoassembler.sourceforge.net>), with the k-mer defined iteratively. In total, given that the genome size is predicted to be 1.58Gb, with the total scaffold length of 1,184,594,388 bp, the overall coverage of the genome is around 76%, a value that might be slightly overestimated given that some of the scaffolds may be overlapping but could not be assembled. Filtering followed by assembly resulted in 259,423 contigs (N50=6,983 bp, longest = 75,003 bp), which was further combined into 148,255 scaffolds (N50 = 19,470, longest = 206,462 bp). The database contains all of the contigs, scaffolds, corresponding assembly parameters, and the annotations for the known repeats and coding sequences. The assembled scaffolds allow basic genomic annotation and comparative analyses with other available avian whole-genome sequences.

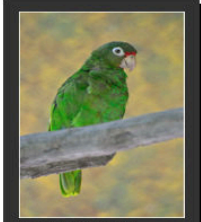
download

readme
[readme.txt](#)

Genome assembly
[Assembly-2011](#)

Additional data
[DATABASE.DATA.Repeats_to_scaffolds.txt](#)
[DATABASE.ANNOTATION.repeats_to_scaffolds.xls](#)

History
September 11, 2012: Data released.
September 14, 2012: Additional data added.



Puerto Rican
(cc) BY-NC

Citation

In accordance with our [terms of use](#), please cite this dataset as:
Oleksyk, TK; Guiblet, W; Pombert, JF; Valentin, R; Martinez-Cruzado, JC (2012): Genomic data of the Puerto Rican Parrot (*Amazona vittata*) from a locally funded project. GigaScience.
<http://dx.doi.org/10.5524/100039>

Related manuscript available at:
[doi:10.1186/2047-217X-1-14](https://doi.org/10.1186/2047-217X-1-14)

Accession codes associated with this data:
NCBI BioProject [PRJNA171587](#)
EBI Project [PRJEB225](#)

<http://gigadb.org>

[Search](#) [Advanced Search](#)

GigaDB is a new database integrated with the GigaScience journal to meet the needs of a new generation of biological and biomedical research as it enters the era of “big-data”... ([see more](#))

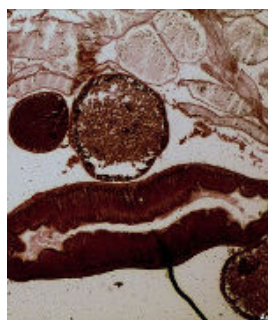
Genomic



[10.5524/100008](#)

Genomic data from the polar bear (*Ursus maritimus*).

2011-07-06



[10.5524/100017](#)

Genomic data from the roundworm *Ascaris suum*.

2011-11-12



[10.5524/100003](#)

Genomic data from the crab-eating macaque/cynomolgus monkey (*Macaca fascicularis*).

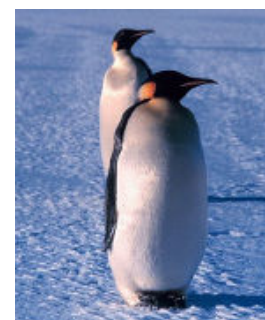
2011-07-06



[10.5524/100001](#)

Genomic data from *Escherichia coli* O104:H4 isolate TY-2482

2011-06-03



[10.5524/100005](#)

Genomic data from the Emperor penguin (*Aptenodytes forsteri*).

2011-07-06

Data released on 2011-11-12

Genomic sequence from an Aboriginal Australian.

Here we present Aboriginal Australian (*Homo sapiens*) genomic sequence obtained from a 100-year-old lock of hair donated by an Aboriginal man from southern Western Australia in the early 20th century. The nuclear genome was sequenced to an overall depth of 6.4-fold. It had a high degree of fragmentation, but no genotype level evidence of sample contamination by DNA from sample handlers of European descent was detected. Over 2 million SNPs were identified after genome mapping and genotyping, approximately 6 percent of which were not previously reported. The mitochondrial genome (mtDNA) was sequenced to an average depth of 338x and found to belong to a new subclade of haplogroup O.

In accordance with our [terms of use](#), please cite this dataset as:

Related manuscripts:

[10.1126/science.1211177](#) (PubMed: [21940856](#))




Accessions (data not in GigaDB):

SRA: [SRP006388](#)

Samples:

Sample ID	Taxonomic ID	Common name	Genbank name	Scientific name	Sample attributes			
-----------	--------------	-------------	--------------	-----------------	-------------------	--	--	--

Files ([FTP site](#)):

DOI	File Name	Sample ID	File Type	File Format	Size	Release Date	Download
10.5524/100010	100726_1198_FC204RWABXX_L3_HUMgspRBBSEW.fq.gz	FILE_CODE	Genome sequence	FASTQ 	5.05 GB	2011-11-12	Download
10.5524/100010	100726_1198_FC204RWABXX_L4_HUMgspRBCSEW.fq.gz	FILE_CODE	Genome sequence	FASTQ 	5.75 GB	2011-11-12	Download
10.5524/100010	100726_1198_FC204RWABXX_L5_HUMgspRBCSEW.fq.gz	FILE_CODE	Genome sequence	FASTQ 	5.42 GB	2011-11-12	Download



Faster download speeds

Aspera data transfer



BGI Datasets Get DOI[®]s

Invertebrate

Ant

- Florida carpenter ant
- Jerdon's jumping ant
- Leaf-cutter ant

Roundworm

Schistosoma

Silkworm

Parasitic nematode

Pacific oyster

Human

Asian individual (YH)

- DNA Methylome
- Genome Assembly
- **Transcriptome**

Cancer (14TB)

Single cell bladder cancer

HBV infected exomes

Ancient DNA

- Saqqaq Eskimo
- Aboriginal Australian



Vertebrates

Darwin's Finch

Giant panda Macaque

-Chinese rhesus

-Crab-eating

Mini-Pig

Naked mole rat

Parrot, Puerto Rican

Penguin

- **Emperor penguin**

- **Adelie penguin**

Pigeon, domestic

Polar bear

Sheep

Tibetan antelope

Released pre-publication

Paper published in GigaScience

Microbe

E. Coli O104:H4 TY-2482

T2D gut metagenome

Cell-Lines

Chinese Hamster Ovary

Mouse methylomes

PLANTS

Chinese cabbage

Cucumber

Foxtail millet

Pigeonpea

Potato

Sorghum

39 data sets





Currently: 39 public datasets
10 citations in references



Humans

Ancient D

- Aborigine

- Saqqaq Eskimo

Asian individual (YH)



Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome

Zhiyu Peng, Yanbing Cheng, Bertrand Chin-Ming Tan, Lin Kang, Zhijian Tian, Yuankun Zhu, Wenwei Zhang, Yu Liang, Xueda Hu, Xuemei Tan, Jing Guo, Zirui Dong, Yan Liang, Li Bao & Jun Wang

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

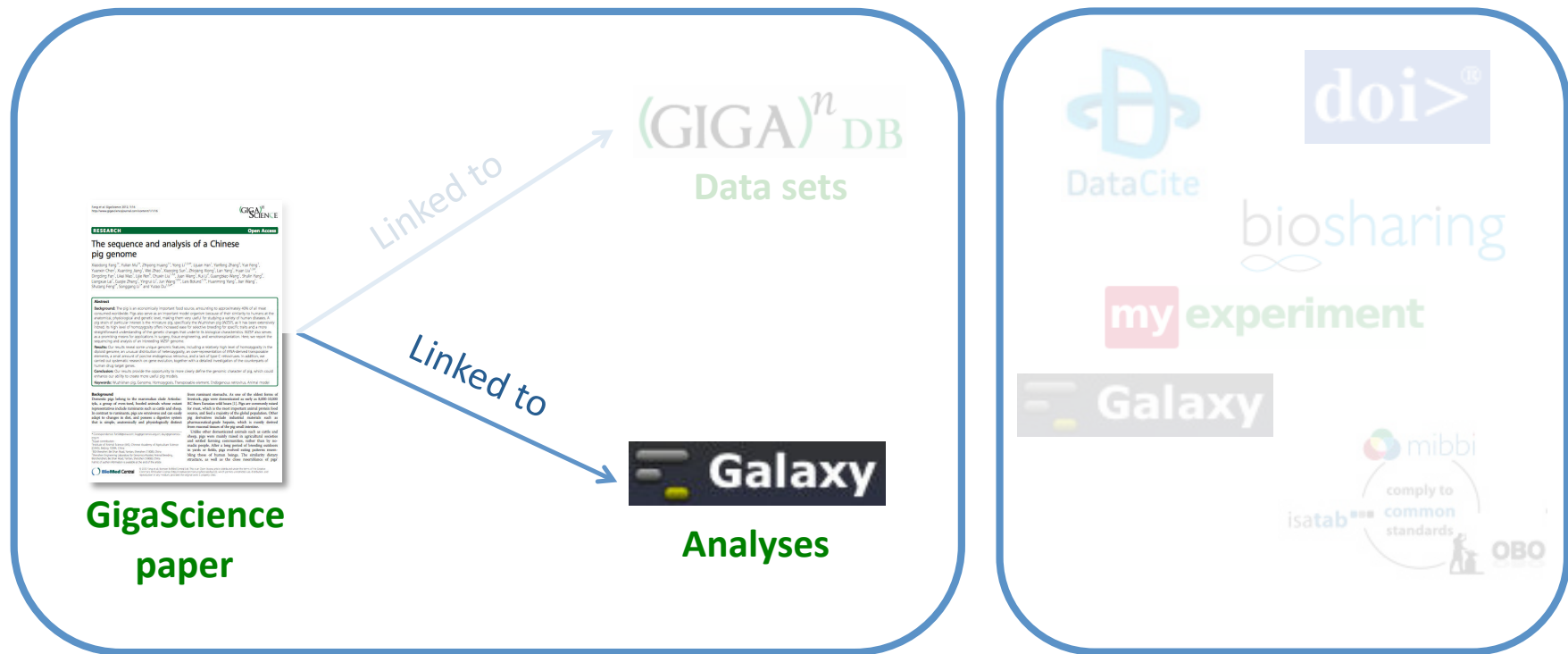
Nature Biotechnology 30, 253–260 (2012) | doi:10.1038/nbt.2122

Received 07 December 2011 | Accepted 17 January 2012 | Published online 12 February 2012

40. Tian, Z. *et al.* Transcriptome from a lymphoblastoid cell line taken from the YH Han Chinese individual. *Giga Sci.* <http://dx.doi.org/10.5524/100013>. (2011).

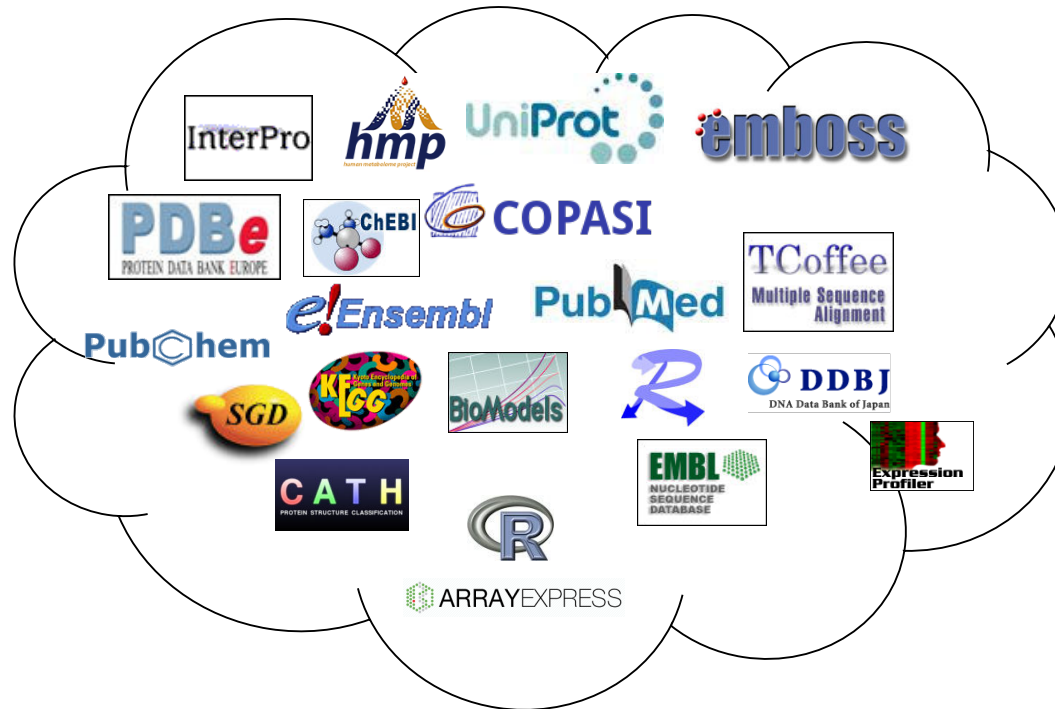


What about the analyses?

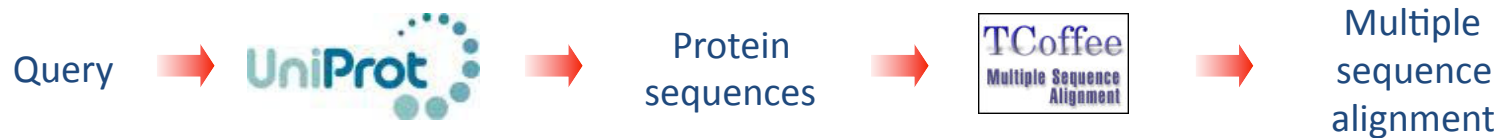


How will we make analyses available for downloading and execution?

Bioinformatics data analyses as workflows



Example workflow: Investigate the evolutionary relationships between proteins




Implement GigaScience workflows in a community-accepted format

Galaxy

Data intensive biology *for everyone.*


Galaxy is an open, web-based platform for data intensive biomedical research. Whether on the [free public server](#) or [your own instance](#), you can perform, reproduce, and share complete analyses.

Use Galaxy




[Use the free public server](#)

Get Galaxy




Install [locally](#) or [in the cloud](#)

Learn Galaxy



[Screencasts](#), [Galaxy 101](#), ...

Get Involved



[Mailing lists](#), [Tool Shed](#), [wiki](#)

[Search all resources](#)

The [Galaxy Team](#) is a part of [BX](#) at [Penn State](#), and the [Biology](#) and [Mathematics and Computer Science](#) departments at [Emory University](#). The Galaxy Project is supported in part by [NHGRI](#), [NSF](#), [The Huck Institutes of the Life Sciences](#), [The Institute for CyberScience](#) at [Penn State](#), and [Emory University](#).

Open source

Over 20,000 main
Galaxy server users

Over 500 papers
citing Galaxy use

Over 55 Galaxy
servers deployed

<http://galaxyproject.org>

Galaxy

[Analyze Data](#)
[Workflow](#)
[Shared Data](#)
[Visualization](#)
[Cloud](#)
[Help](#)
[User](#)

Using 0%

The cluster on which many NGS tools run will be down for maintenance from 4 PM, Monday, Nov. 19 until 9 AM the following day (EST5EDT, UTC-0400). Jobs running on that cluster at that time

Tools

EMBOSS

NGS TOOLBOX BETA

NGS: QC and manipulation

NGS: Mapping

- Map with BWA for Illumina
- Map with BWA for SOLID

ILLUMINA

- Map with Bowtie for Illumina

ROCHE-454

- Lastz map short reads against reference sequence
- Megablast compare short reads against htgs, nt, and wgs databases

Parse blast XML output

- AB-SOLID

Map with Bowtie for SOLID

NGS: SAM Tools

NGS: GATK Tools (beta)

NGS: Variant Detection

NGS: Indel Analysis

NGS: Peak Calling

NGS: RNA Analysis

NGS: Picard (beta)

BEDTools

snpEff

RGENTICS

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

SNP/WGA: Statistical Models

Map with BWA for Illumina (version 1.2.3)

Will you select a reference genome from your history or use a built-in index?

Use a built-in index

Select a reference genome:

Arabidopsis lyrata: Araly1

Is this library mate-paired?:

Single-end

FASTQ file:

FASTQ with either Sanger-scaled quality values (fastqsanger) or Illumina-scaled quality values (fastqillumina)

BWA settings to use:

Commonly Used

For most mapping needs use Commonly Used settings. If you want full control use Full Parameter List

Suppress the header in the output SAM file:

☐

BWA produces SAM with several lines of header information

Execute

What it does

BWA is a fast light-weighted tool that aligns relatively short sequences (queries) to a sequence database (large), such as the human reference genome. It is developed by Heng Li at the Sanger Insitute. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754-60.

Know what you are doing

There is no such thing (yet) as an automated gearshift in short read mapping. It is all like stick-shift driving in San Francisco. In other words = running this tool with default parameters will probably not give you meaningful results. A way to deal with this is to **understand** the parameters by carefully reading the [documentation](#) and experimenting. Fortunately, Galaxy makes experimenting easy.

History

0 bytes

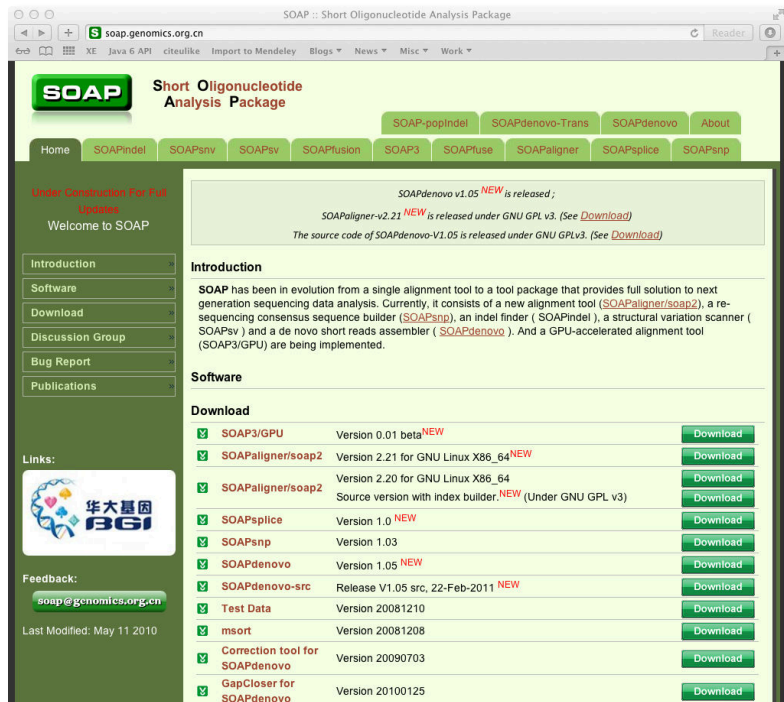
Your history is empty. Click 'Get Data' on the left pane to start

Tool list

Tool parameterisation

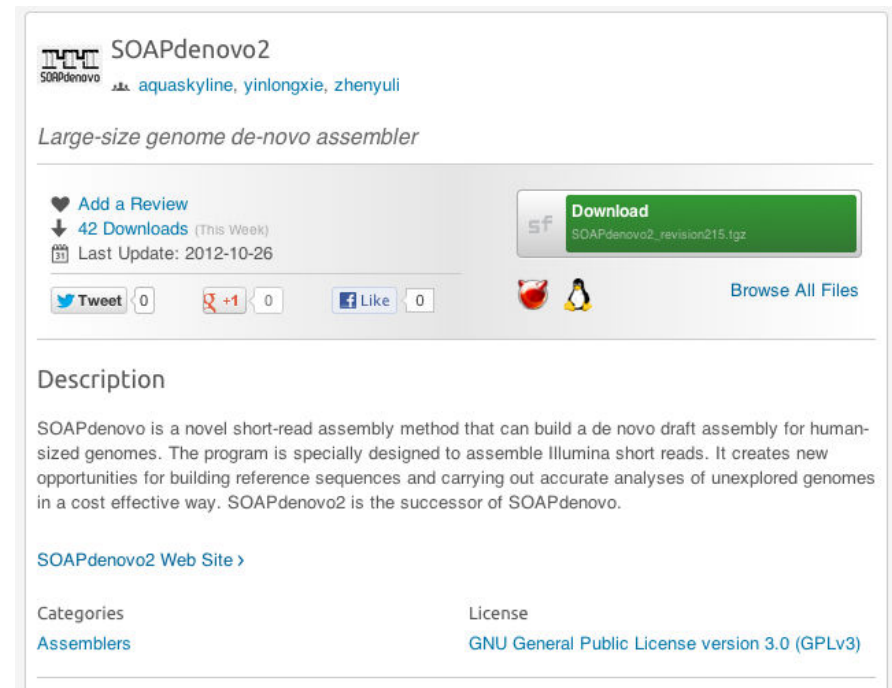
Results panel

Pilot project - Integrate BGI SOAP package into Galaxy



The screenshot shows the SOAPgenomics.org.cn website. The header includes the SOAP logo and the title "Short Oligonucleotide Analysis Package". A navigation bar lists various tools: SOAPpindel, SOAPsnv, SOAPsv, SOAPfusion, SOAP3, SOAPfuse, SOAPaligner, SOAPsplice, and SOAPsnp. The main content area features a "Download" section with a table of software releases. The table includes columns for the tool name, version, and a "Download" button. The tools listed are SOAP3/GPU, SOAPaligner/soap2, SOAPaligner/soap2 (Source version with index builder), SOAPsplice, SOAPsnp, SOAPdenovo, SOAPdenovo-src, Test Data, msort, Correction tool for SOAPdenovo, and GapCloser for SOAPdenovo. The website also has a sidebar with links to Introduction, Software, Download, Discussion Group, Bug Report, and Publications.

Tool	Version	Download
SOAP3/GPU	Version 0.01 beta NEW	Download
SOAPaligner/soap2	Version 2.21 for GNU Linux X86_64 NEW	Download
SOAPaligner/soap2	Version 2.20 for GNU Linux X86_64	Download
SOAPaligner/soap2	Source version with index builder NEW (Under GNU GPL v3)	Download
SOAPsplice	Version 1.0 NEW	Download
SOAPsnp	Version 1.03	Download
SOAPdenovo	Version 1.05 NEW	Download
SOAPdenovo-src	Release V1.05 src, 22-Feb-2011 NEW	Download
Test Data	Version 20081210	Download
msort	Version 20081208	Download
Correction tool for SOAPdenovo	Version 20090703	Download
GapCloser for SOAPdenovo	Version 20100125	Download



The screenshot shows the SOAPdenovo2 GitHub repository page. The header includes the repository name "SOAPdenovo2" and the authors "aquaskyline, yinlongxie, zhenyuli". The description states "Large-size genome de-novo assembler". The page shows 42 Downloads (This Week) and a Last Update of 2012-10-26. There is a "Download" button for "SOAPdenovo2_revision215.tgz". The page also has a "Description" section, a "SOAPdenovo2 Web Site" link, and a "License" section indicating "GNU General Public License version 3.0 (GPLv3)".

SOAPdenovo2
aquaskyline, yinlongxie, zhenyuli

Large-size genome de-novo assembler

42 Downloads (This Week)
Last Update: 2012-10-26

[Download](#)
SOAPdenovo2_revision215.tgz

[Add a Review](#)
[42 Downloads](#) (This Week)
[Last Update: 2012-10-26](#)

[Tweet](#) 0 [+1](#) 0 [Like](#) 0

[Browse All Files](#)

Description

SOAPdenovo is a novel short-read assembly method that can build a de novo draft assembly for human-sized genomes. The program is specially designed to assemble Illumina short reads. It creates new opportunities for building reference sequences and carrying out accurate analyses of unexplored genomes in a cost effective way. SOAPdenovo2 is the successor of SOAPdenovo.

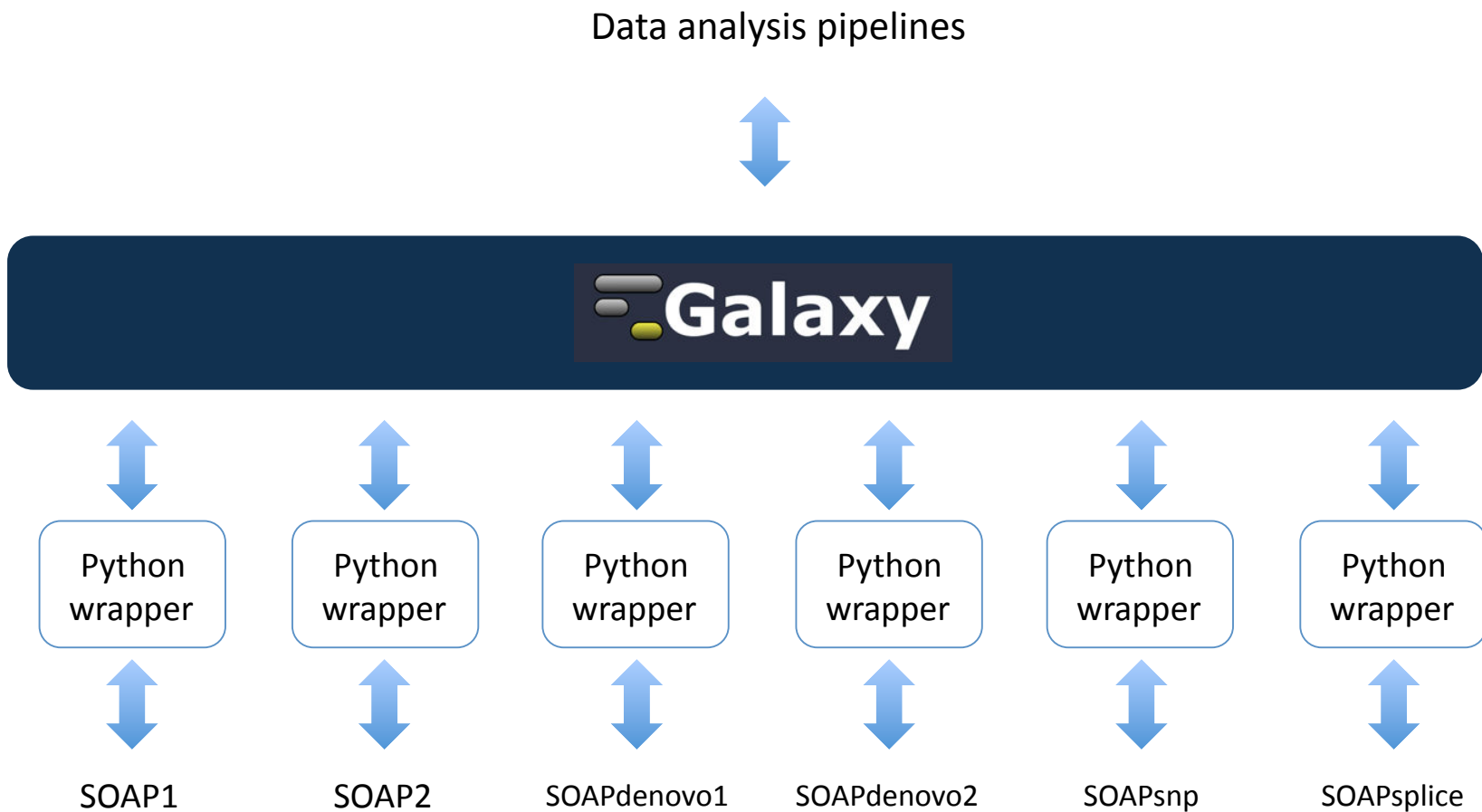
[SOAPdenovo2 Web Site](#)

Categories
[Assemblers](#)

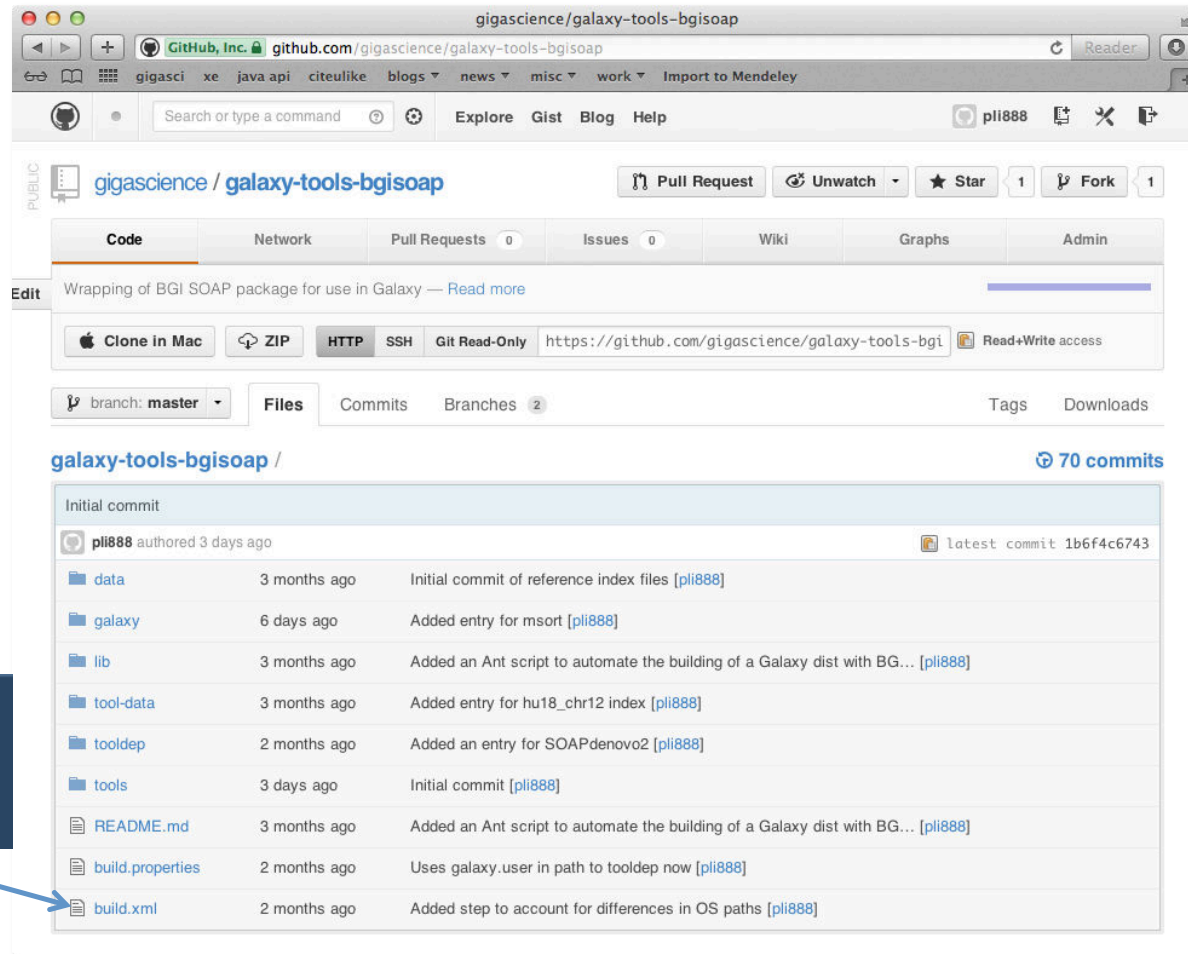
License
GNU General Public License version 3.0 (GPLv3)

Enable SOAP tools to be used from within Galaxy workflows

Integrate BGI SOAP package into Galaxy



GitHub open code repository



gigascience / galaxy-tools-bgisoap

Public

Wraping of BGI SOAP package for use in Galaxy — Read more

Clone in Mac ZIP HTTP SSH Git Read-Only <https://github.com/gigascience/galaxy-tools-bgi> Read+Write access

branch: master Files Commits Branches 2 Tags Downloads

galaxy-tools-bgisoap / 70 commits

File	Time	Description
Initial commit		
pli888 authored 3 days ago		latest commit 1b6f4c6743
data	3 months ago	Initial commit of reference index files [pli888]
galaxy	6 days ago	Added entry for msort [pli888]
lib	3 months ago	Added an Ant script to automate the building of a Galaxy dist with BG... [pli888]
tool-data	3 months ago	Added entry for hu18_chr12 index [pli888]
tooldep	2 months ago	Added an entry for SOAPdenovo2 [pli888]
tools	3 days ago	Initial commit [pli888]
README.md	3 months ago	Added an Ant script to automate the building of a Galaxy dist with BG... [pli888]
build.properties	2 months ago	Uses galaxy.user in path to tooldep now [pli888]
build.xml	2 months ago	Added step to account for differences in OS paths [pli888]



<https://github.com/gigascience>

Galaxy - Mozilla Firefox

Galaxy

localhost:8080

Galaxy

Analyze Data Workflow Shared Data Visualization Help User

Using 1.0 MB

Tools

BGI SOAP PACKAGE BETA

NGS: Mapping

- Map with SOAP 1
- Map with SOAP 2

NGS: Indel Analysis

NGS: De Novo Assembly

DE NOVO ASSEMBLY TOOLS

- SOAPdenovo - perform de novo genome assembly
- SOAPdenovo2 - perform de novo genome assembly
- SOAPdenovo-trans - perform de novo transcriptome assembly
- gapcloser - close gaps in SOAPdenovo scaffolds

SOAPDENOVO2 MODULES

- pregraph
- contig
- map
- scaff

CORRECTION PACKAGE

- kmerfreq - a kmer frequency counter
- corrector - corrects short reads
- mergepairlist - Extract pairs from a list of 2 files

NGS: Splice Detection

NGS: Somatic Mutation Detection

NGS: Supporting tools

Workflows

pregraph (version 0.1)

Select a config file from history or create a new one?:

Use one from history

Use one from history

Create new config file

84: scaffold statistics

SOAP settings to use:

Default

Default settings is suitable for most mapping needs. If you want full control, use Full parameter list

Execute

What it does

SOAPdenovo is a novel short-read assembly method that can build a de novo draft assembly for the human-sized genomes. The program is specially designed to assemble Illumina GA short reads. It creates new opportunities for building reference sequences and carrying out accurate analyses of unexplored genomes in a cost effective way.

The pregraph step in SOAPdenovo2 is responsible for loading raw read data into memory and constructing the Bruijn graph structure which is used to represent overlap between reads.

Output files from pregraph

- kmerFreq. Each row shows the number of Kmers with a frequency equals the row number. Note that those peaks of frequencies which are the integral multiple of 63 are due to the data structure.
- edge. Each record gives the information of an edge in the pre-graph: length, Kmers on both ends, average kmer coverage, whether it's reverse-complementarily identical and the sequence.
- markOnEdge and path. These two files are for using reads to solve small repeats.
- preArc. Connections between edges which are established by the read paths.
- vertex. Kmers at the ends of edges.
- preGraphBasic. Some basic information about the pre-graph: number of vertex, K value, number of edges, maximum read length etc.
- Configuration file.

Authors

Ruibang Luo and Zhenyu Li, 2012-7-10

History

Unnamed history 1.0 MB

84: scaffold statistics

83: bubbles in scaffolds

82: contig positions in scaffolds

81: scaffold sequences

55 lines
format: text, database: ?
Info: Status complete

```
>scaffold1 7.6
GTTTTATCGCTTCCATGACGCGAGAAGTTAACACTTT
AATTAAATCGAAGTGGACTGCTGGCGGAAAATGAGA
CGATTCTGTCAAAACTGACGCGTTGGATGAGGAGA
TG6TAGAGATTCTCTTTGACATTTTAAAGAGCC
CTGAACAATCCGTACGTTTCCAGACCGCTTTGGCCT
```

80: gap sequences between contigs

79: scaff

78: scaff gap

77: links

76: new contig index

75: read in gap

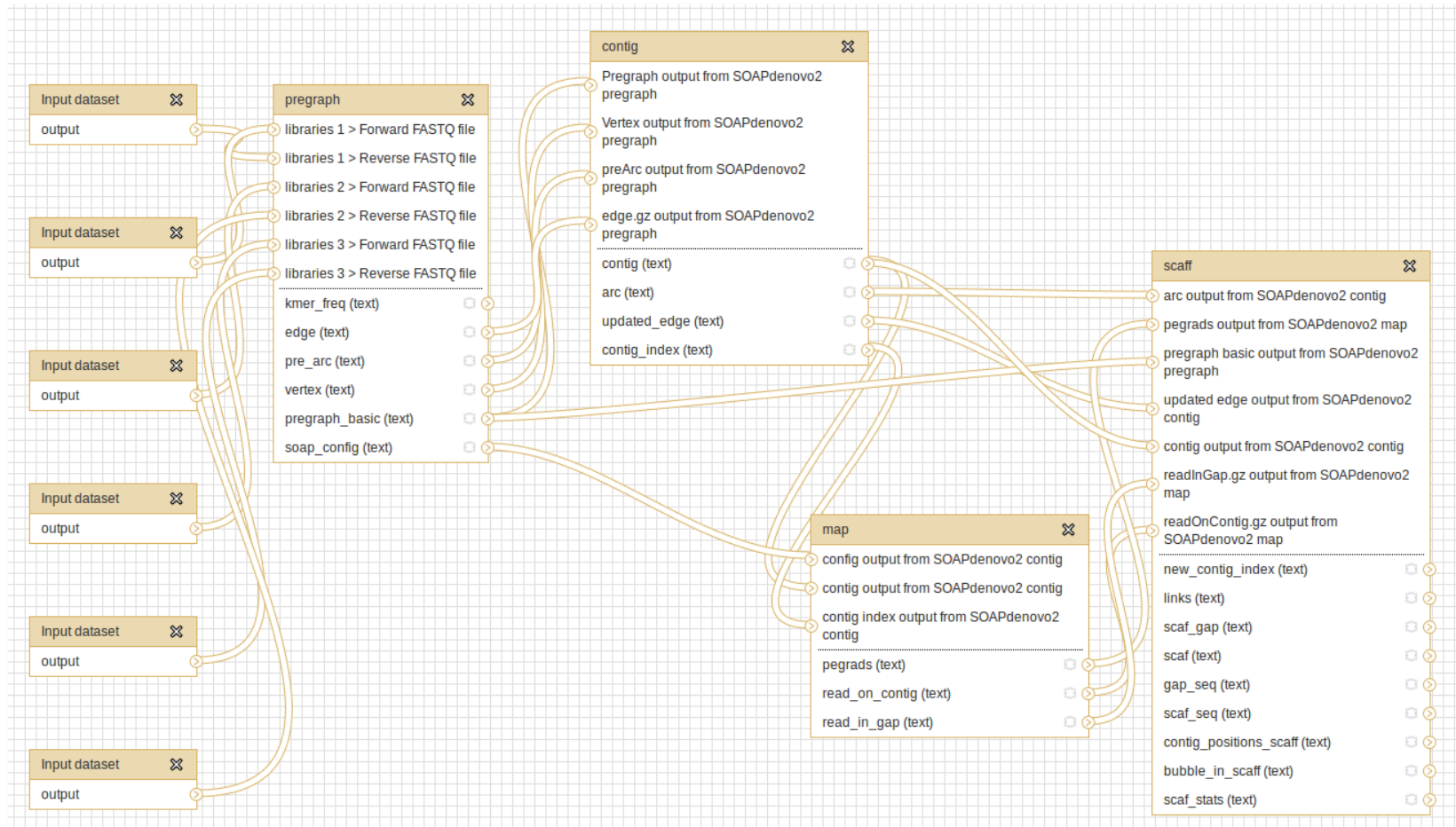
74: read on contig

Tool list

Tool parameterisation

Results panel

SOAPdenovo2 Galaxy workflow



[About](#) | [Mailing List](#) | [Publications](#)

[Logout](#) | [Give us Feedback](#) | [Invite](#)

[Home](#) | [Users](#) | [Groups](#) | **Workflows** | [Files](#) | [Packs](#) | [Services](#) | [Topics](#)

Home > Workflows

Search filter terms

Filter by type

☐ Tavema 2 1016
☐ Tavema 1 573
☐ RapidMiner 223
☐ Kepler 43
☐ Bioclipse Scri... 34
☐ LONI Pipeline 26
☐ GWorkflowDL 24
☐ BioExtract Ser... 17
☐ Tesla 11
☒ **Galaxy** 10

Filter by tag

☐ galaxy 4
☐ ngs 2
☐ cage 1
☐ counts 1
☐ data integration 1
☐ exons 1
☐ FASTA 1
☐ maf 1
☐ reverse 1
☐ snps 1

Filter by user

☐ David De Roure 5
☐ Peter Li 4
☐ Helen Hulme 1

Filter by licence

☐ by-sa 2

Sort by: Rank

Showing 10 results. Use the filters on the left and the search box below to refine the results.

Type: Galaxy ✕

Remove all filters

Galaxy

Original Uploader
David De Roure

Basic Illumina Reads Quality (Functional Genomics Workshop 2012) (v1)

Created: 16/07/12 @ 21:28:47 | Last updated: 16/07/12 @ 21:31:52

License: No license

From the RNA-Seq analysis tutorial during the Functional Genomics Workshop 2012
<https://caps.osu.edu/pfg-workshop> Workflow published by mejia-guerre on Galaxy Jun 22, 2012 imported to myExperiment Jul16, 2012 during demonstration of Galaxy-myExperiment integration

Rating: 0.0 / 5 (0 ratings) | Versions: 1 | Reviews: 0 | Comments: 0 | Citations: 0
Viewed: 18 times | Downloaded: 4 times
This Workflow has no tags!

Galaxy

Original Uploader
Peter Li

Perform QC on FastQ data (v1)

Created: 14/06/12 @ 07:44:09 | Last updated: 14/06/12 @ 07:46:11

License: No license

This Galaxy pipeline performs quality control analyses on FastQ data. A report is generated which provides various measures of the quality of the sequence reads.

Rating: 0.0 / 5 (0 ratings) | Versions: 1 | Reviews: 0 | Comments: 0 | Citations: 0
Viewed: 40 times | Downloaded: 7 times
Tags (2):
galaxy | ngs

New/Upload

Workflow

Peter Li

[My Profile](#) [edit]
[My Messages](#)
[My Memberships](#)
[My History](#)
[My News](#)

1 new friendship request

a mihailonita_me

My Stuff

28 Friends | 7 Groups | 70 Workflows | 3 Packs | 1 Files

Friends

Alan Williams
 Alex Nenadic
 Andy Tseng
 Antoon Goderis
 Bayu Jayawardhana
 Bhamecham
 Carole Coble

My Favourites

0 favourites

My Tags

45 tags

beanshell | bioconductor | chebi |
cheminformatics |
chemspider | data_integration |

<http://www.myexperiment.org>

Why publish in GigaScience?

Benefit

- Data hosted in GigaDB
- Allocation of DOIs to data
- Metadata in isa-tab format
- Galaxy tool integration
- Use of tools in Galaxy workflows

Added value

- No need to use own servers
- Citable data
- Aids reuse of data
- Supports reuse of tools
- Improves documentation
- Shows how tool can be used with other bioinf. software

Thanks to:

- Tin-Lap Lee and Huayan Gao - CUHK
- Tam, Jesse, Scott, Nicole & Laurie - GigaScience

peter@gigasciencejournal.com

