



BRIEF REPORTS

Open Access

# GTC: A web server for integrating systems biology data with web tools and desktop applications

Dan Tenenbaum<sup>1</sup>, J Christopher Bare<sup>1</sup>, Nitin S Baliga<sup>1,2\*</sup>

## Abstract

Gaggle Tool Creator (GTC) is a web application which provides access to public annotation, interaction, orthology, and genomic data for hundreds of organisms, and enables instant analysis of the data using many popular web-based and desktop applications.

## Background

There are hundreds of public databases for systems biology data, and an equal number of applications for working with that data. However, it is often difficult to work with data of interest in the desired applications. Databases may not offer programmatic access, or require special scripting. Software tools may imprison the data in such a way that it can only be analyzed by a particular application. Data available for one organism may not be available for another. Individuals may have to download their own copies of databases in order to work with them in nonstandard ways, forcing them into the role of curator. Software tools may not allow users to work with their own data. Applications may only accept data in arcane formats, requiring special conversion.

Here we describe Gaggle Tool Creator (GTC), a web application which addresses these problems by providing public data for hundreds of organisms, and making it instantly accessible to many popular, unrelated web resources and desktop applications. This in turn allows sophisticated analyses and novel discoveries to be achieved with just a few mouse clicks (Figure 1).

## Implementation

GTC is composed of a number of MySQL databases, regularly updated by scheduled scripts which download systems biology data from public sources. These sources, and the method used to download data from each, are as follows: NCBI (web services), STRING[1] (flat file

download), BioNetBuilder[2] (flat file download), KEGG [3] (web services), and the UCSC genome browser[4] (flat file download). We currently have data for 500 organisms, with the eventual goal of having data for all sequenced organisms. The core of GTC is a Java web application which makes available links to several applications suited to the analysis of particular types of data. These applications are launched using Java Web Start, a technology which seamlessly pushes software updates to the user's computer. All of these applications implement the Gaggle framework [5] for sharing data between applications and web sites (Figure 2).

## Results and Discussion

GTC's user interface allows the user to choose up to ten organisms to work with.

GTC displays several links for each organism, and the following list describes in greater detail what happens when the user clicks these links.

### Annotation

A desktop application displaying annotation data from NCBI.

### Synonyms

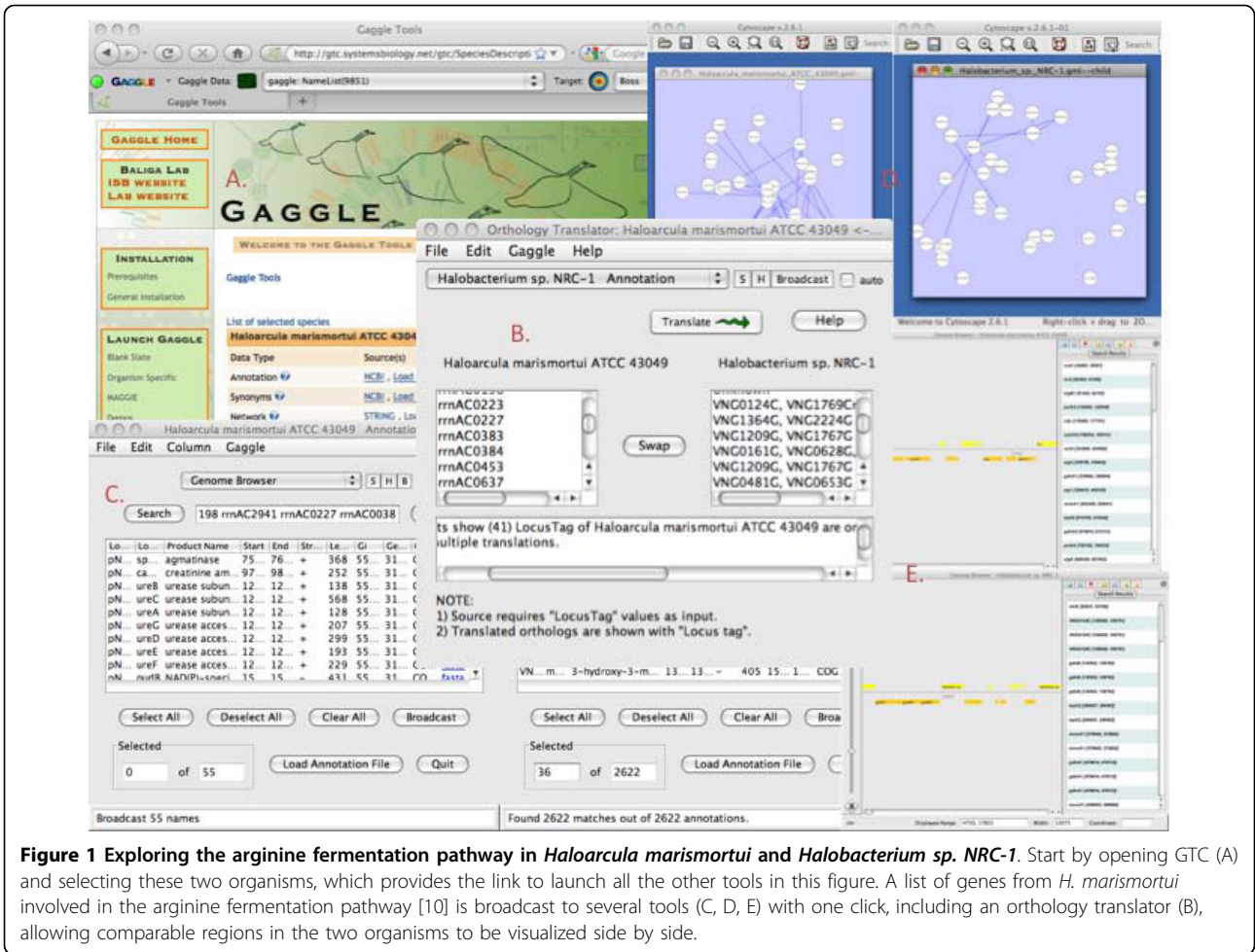
An application that can translate between various naming schemes for genes (Locus Tag, Gene Name, Product Name). Data come from either NCBI or BioNetBuilder.

### Network

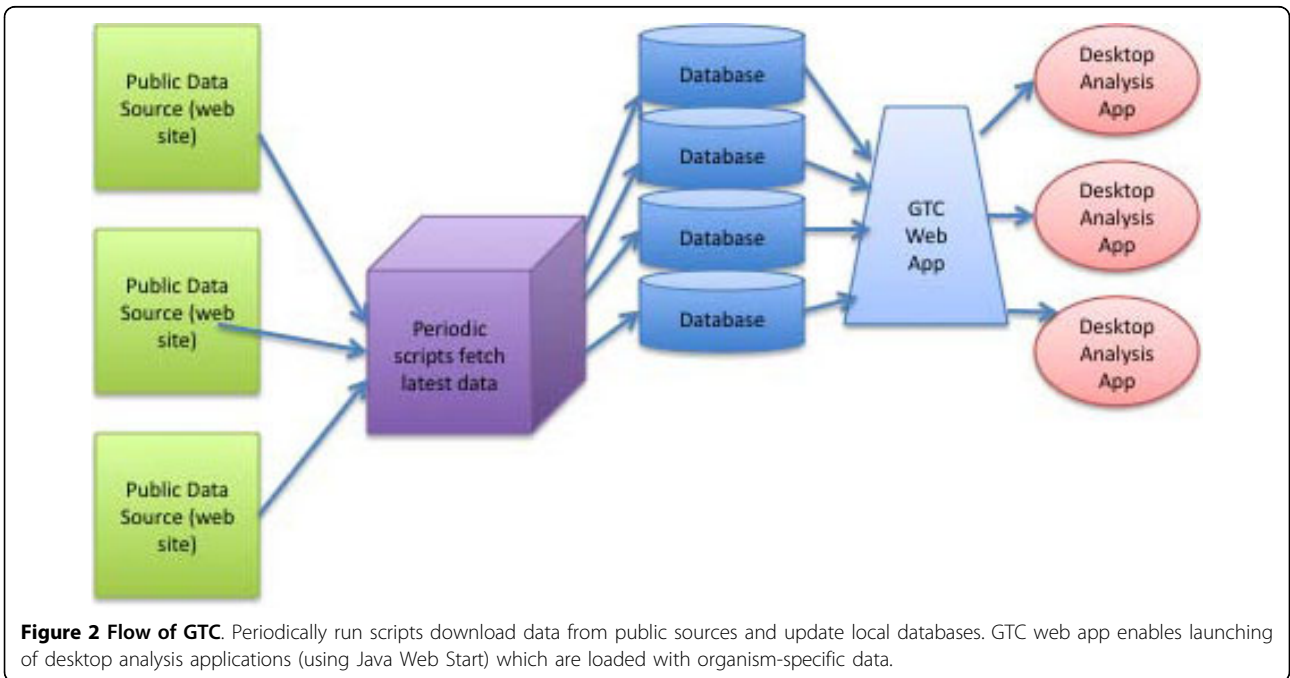
The Cytoscape network viewer [6], with data from either STRING or BioNetBuilder.

\* Correspondence: [nbaliga@systemsbiology.org](mailto:nbaliga@systemsbiology.org)

<sup>1</sup>Institute for Systems Biology, 1441 N. 34th St., Seattle, WA USA



**Figure 1 Exploring the arginine fermentation pathway in *Haloarcula marismortui* and *Halobacterium sp. NRC-1*.** Start by opening GTC (A) and selecting these two organisms, which provides the link to launch all the other tools in this figure. A list of genes from *H. marismortui* involved in the arginine fermentation pathway [10] is broadcast to several tools (C, D, E) with one click, including an orthology translator (B), allowing comparable regions in the two organisms to be visualized side by side.



**Figure 2 Flow of GTC.** Periodically run scripts download data from public sources and update local databases. GTC web app enables launching of desktop analysis applications (using Java Web Start) which are loaded with organism-specific data.

### GenomeMap

A genome browser, preloaded with genomic data from the UCSC genome web service.

### Orthology Translator

Finds orthologs between any two organisms (invoked when the user chooses exactly two organisms to work with).

Each of these links is paired with another called "Load Your Own" which allows the user to provide their own data. Because Gaggle is a message-passing framework, the user is not limited to the applications listed above and in fact can use any Gaggle-enabled tool; and, using Gaggle's Firegoose extension [7] for the Firefox browser, web sites such as KEGG, STRING, EntrezGene, Entrez-Protein, and DAVID [8][9]. (GTC will work with any modern web browser but Firegoose enables two-way communication with Gaggle-enabled web sites.)

### Availability and Requirements

**Project name:** GTC (Gaggle Tool Creator)

**Project home page:** <http://gaggle.systemsbiology.net/gtc>

**Operating system:** Platform independent web site

**Programming languages:** Java, Javascript

**Other requirements:** Java 1.5 or higher and a web browser. While GTC can be viewed with any modern web browser, Firefox and the Firegoose extension are required for full interoperability with Gaggle-enabled applications and websites.

<http://getfirefox.com>

<http://gaggle.systemsbiology.net/docs/geese/firegoose/>

**License:** GNU LGPL

**Any restrictions to use by non-academics:** None

### Acknowledgements

We thank Optra Systems for engineering assistance, and Sarah Killcoyne for modifications to the Gaggle plugin for Cytoscape.

**Funding:** This work was supported by grants from NSF (DBI-0640950), DOE (DE-FG02-07ER64327), NIH (P50GM076547) and Battelle (PNNL subcontract 39527 to NSB).

### Author details

<sup>1</sup>Institute for Systems Biology, 1441 N. 34th St., Seattle, WA USA.

<sup>2</sup>Departments of Microbiology, and Molecular and Cellular Biology, University of Washington, Seattle, WA, USA.

### Authors' contributions

DT developed the specification and design of the software and drafted the manuscript. JCB contributed to the design of the software. NSB conceived of the project, participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

### Competing interests

The authors declare that they have no competing interests.

Received: 14 April 2010 Accepted: 13 July 2010 Published: 13 July 2010

### References

1. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C: **STRING 8—a global view on proteins and their functional interactions in 630 organisms.** *Nucleic Acids Res* 2009, **37**.
2. Avila-Campillo I, Drew K, Lin J, Reiss DJ, Bonneau R: **BioNetBuilder: automatic integration of biological networks.** *Bioinformatics* 2007, **23**.
3. Kanehisa M, Araki M, Goto S, *et al*: **KEGG for linking genomes to life and the environment.** *Nucleic Acids Res* 2008, **36** Database: D480.
4. Karolchik D, Hinrichs AS, Kent WJ: **The UCSC Genome Browser.** *Curr Protoc Bioinformatics* 2007, **Chapter 1**(Unit 1.4).
5. Shannon PT, Reiss DJ, Bonneau R, Baliga NS: **The Gaggle: An open-source software system for integrating bioinformatics software and data sources.** *BMC Bioinformatics* 2006, **7**:176.
6. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks.** *Genome Res* 2003, **13**(11):2498-504.
7. Bare JC, Shannon PT, Schmid AK, Baliga NS: **The Firegoose: two-way integration of diverse data from different bioinformatics web resources with desktop applications.** *BMC Bioinformatics* 2007, **8**:456.
8. Huang DW, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources.** *Nature Protoc* 2009, **4**(1):44-57.
9. Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA: **DAVID: Database for Annotation, Visualization, and Integrated Discovery.** *Genome Biol* 2003, **4**(5):P3.
10. Baliga NS, Bonneau R, Facciotti MT, Pan M, Glusman G, Deutsch EW, Shannon P, Chiu Y, Weng SW, Gan RR, Hung P, Date SV, Marcotte E, Hood L, Ng WW: **Genome sequence of *Haloarcula marismortui*: A halophilic archaeon from the Dead Sea.** *Genome Res* 2004, **14**:2221-2224.

doi:10.1186/1751-0473-5-7

**Cite this article as:** Tenenbaum *et al.*: GTC: A web server for integrating systems biology data with web tools and desktop applications. *Source Code for Biology and Medicine* 2010 **5**:7.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

