

Software

Open Access

Apollo2Go: a web service adapter for the Apollo genome viewer to enable distributed genome annotation

Kathrin Klee, Rebecca Ernst, Manuel Spannagl and Klaus FX Mayer*

Address: MIPS/Institute for Bioinformatics, GSF Research Center for Environment and Health, Ingolstädter Landstrasse 1; 85764 Neuherberg, Germany

Email: Kathrin Klee - Kathrin_Klee@web.de; Rebecca Ernst - Rebecca.Ernst@gmx.de; Manuel Spannagl - manuel.spannagl@gsf.de; Klaus FX Mayer* - Kmayer@gsf.de

* Corresponding author

Published: 30 August 2007

Received: 25 April 2007

BMC Bioinformatics 2007, 8:320 doi:10.1186/1471-2105-8-320

Accepted: 30 August 2007

This article is available from: <http://www.biomedcentral.com/1471-2105/8/320>

© 2007 Klee 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.

Abstract

Background: Apollo, a genome annotation viewer and editor, has become a widely used genome annotation and visualization tool for distributed genome annotation projects. When using Apollo for annotation, database updates are carried out by uploading intermediate annotation files into the respective database. This non-direct database upload is laborious and evokes problems of data synchronicity.

Results: To overcome these limitations we extended the Apollo data adapter with a generic, configurable web service client that is able to retrieve annotation data in a GAME-XML-formatted string and pass it on to Apollo's internal input routine.

Conclusion: This Apollo web service adapter, Apollo2Go, simplifies the data exchange in distributed projects and aims to render the annotation process more comfortable.

The Apollo2Go software is freely available from ftp://ftpmips.gsf.de/plants/apollo_webservice.

Background

The number of finished bacterial genomes and higher eukaryotic genomes is dramatically increasing. Consequently there is a pressing need for accurate annotation of the respective genomes and, even more important, for the maintenance of genome annotation. With the increasing complexity, and the large number of the genomes on one hand and the only limited availability of appropriate and sufficient resources for annotation and curation on the other there is a strong demand for solutions and tools to assist in community-centric, decentralized and distributed genome annotation for maintenance of high quality genome annotation. In addition, genome sequencing projects are often set up as distributed, collaborative projects with contributions from multiple research

groups. A necessity for such approaches is the implementation of common data exchange formats and an annotation system that enables remote users to edit and curate database entries. Such a system should be user friendly, flexible and easy to implement.

Several approaches have been proposed to address these problems. A popular and widely applied approach is to employ experts for an initial genome annotation within genome annotation jamborees. Examples comprise the annotation of the rice genome, the *Drosophila* genome and the Human Genome Project Consortium's Analysis Group (HGPCAG) [1-3]. These approaches aim to overcome the problem of collecting distributed data by bringing together scientists, harvest their joint expertise and

integrate the annotations into the respective database system.

A different approach is to distribute not only the annotation efforts but also the data management by using a distributed annotation system (DAS) browser that integrates and displays annotations from various sources [4]. DAS annotations are available for numerous genomes and provide a powerful means to integrate different and decentralized annotations.

However, to our knowledge currently there is no DAS client software available that provides functionality to display, integrate, annotate and re-write data in a standardized way.

Finally, a third approach is frequently used to integrate contributions from several groups. In this approach, a web accessible annotation interface is made available along with visualization tools and complementary analysis data. With the genome browser visualization interface along with complementary analysis data contributors are enabled to view and annotate database entries and finally to submit them back into the reference database. In our work we are following this approach by adopting and advance the Apollo genome annotation and visualization tool.

For annotation, Apollo is established as a widely used genome annotation and visualization tool. Beside local usage, Apollo has been used in distributed genome annotation projects [5]. After completion of the annotation process a database update is carried out by uploading the annotation files. Albeit robust the non-direct database upload via Apollo's intermediate GAME-XML formatted files is a tedious and laborious task and evokes problems of data synchronicity. To overcome these limitations we developed a web service adapter (Apollo2Go) for the Apollo genome annotation tool. Web services recently became widely used tools and provide important advantages, such as platform interoperability, only limited firewall problems, open standards and, importantly, wide acceptance in the scientific community.

The Apollo2Go adapter is a client program that enables the connection to specified web services that serve and accept the annotation data. The adapter combines the advantages of a well established and widely used annotation software with the convenience of directly accessing and writing data back into the database. Thereby the intermediate processing of GAME-XML files is omitted.

Implementation

To implement Apollo2Go's web service functionality we added several adapter classes to the Apollo code. We cre-

ated two GUI interfaces to communicate web service input and output parameters as well as two adapter classes, each concerning the respective web service client for reading and writing data from and to an individual data source. The web service adapter classes that contain the clients were connected to the existing GameXML adapter that allows data to be easily converted into an Apollo-compatible form (see also Figure 1). The web service adapter GUIs can be addressed via the data adapter chooser from Apollo's input and output/save forms.

Display of annotation data from remote resources

To use Apollo's web service capability, i.e. to be able to display annotation data of selected regions or complete contigs or chromosomes from remote resources, the data provider has to make its data available in Apollo compatible GAME-XML format via a web service. Beside a particular database that hosts annotation data also specified directories that contain the required GAME-XML files can serve as data sources for the web service. Once a web service is set-up, annotations can be accessed using the Apollo2Go web service from any remote computer.

In addition to the Apollo2Go web service-enabled Apollo, the remote user only has to fill in a configuration file which contains the web service dependent variables.

Writing annotation data to remote resources

After remote display and data edits in Apollo, modified annotations can be inserted into the reference database. Typically, annotation data are stored in a dedicated database that, within the framework of a distributed annotation project, will often be embodied by a non-local resource.

Since there is a variety of different database schemas and concepts, direct insertion of remote, Apollo-curated annotation is impossible. Therefore, each database provider has to make a web service available that is capable to insert annotation data in GAME-XML format into the respective database. All logic that is associated with database insertion like consistency checks, user authorization, as well as insertion itself takes place within the web service.

Thus, similar to the concept of displaying remote data, the configurable Apollo2Go web service client has been integrated into Apollo's 'save' functionality. The authorized user solely has to complete a configuration file to write data back into the specified database. Due to the uncoupling of database schema and display no restrictions with respect to the database schema are apparent for the enhanced Apollo version. An overview on the web service

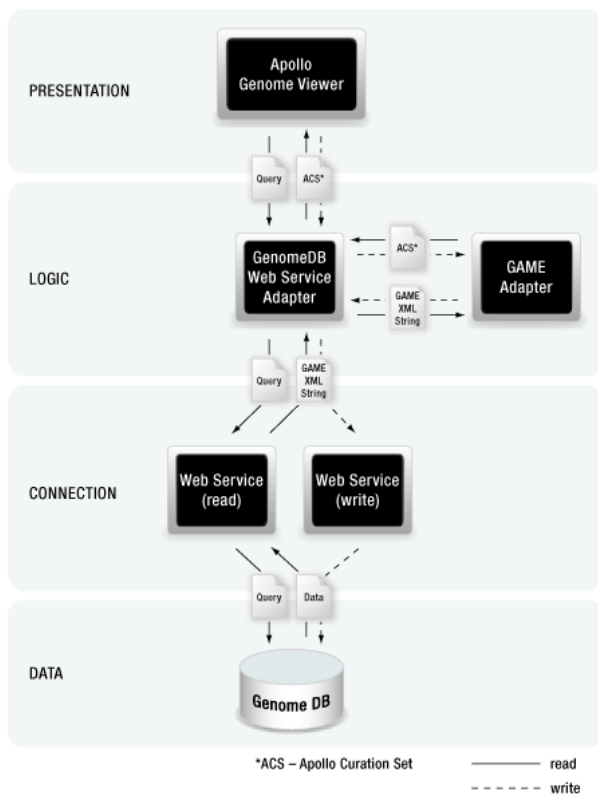


Figure 1
Apollo web service adapter overview. To load data into Apollo the Apollo GUI calls the web service adapter GUI, which calls the web service adapter. The web service adapter invokes the web service client that communicates with the web service. The web service is deployed on an in house server that directly accesses the database. The database returns the information of interest to the web service and the web service returns it to the web service adapter. The web service returns the information as a GameXML formatted String and the GAME-XML adapter transforms the information into Apollo readable format. Subsequent the information can be loaded from the web service adapter into the Apollo main window. **To write data** back into the database or in a flat file the web service adapter receives the information as CurationSet from Apollo. This CurationSet will be send to the GAME-XML adapter to retrieve a GameXML formatted string, which can be transferred to the web service. The web service directs the information into the database or in a dedicated flat file.

enabled Apollo version and the interplay with the components involved is depicted in Figure 1.

Results

Use case: Implementation and Application of Apollo web services for MIPS' PlantsDB

We developed a web service that receives annotation data from the plant databases developed and hosted by our group (MIPS PlantsDB) [6,7]. PlantsDB hosts multiple databases for different genomes and the web service receives data for the given genome database, the respective contig name and the selected sequence range. The Apollo2Go built-in client invokes this web service with arguments derived from the adapter's graphical user interface. The client retrieves the result as a GAME-XML-formatted string and passes it to the Apollo input methods. For data not provided by a database system, the web service also enables to receive data provided as flat files.

After retrieval, the data can be subjected to curation and annotation using Apollo's annotation and visualization interface. In our implementation before write-back into the database all data are tested for redundancy, versioned and inserted into the database. 'Old' data will not be overwritten but are set to a non-visible mode and only more recent and up-to-date annotations are displayed via the public interface. However this functionality has to be provided by the individual web service implementation.

It should be emphasized that usability of the Apollo2Go web service functionality is not restricted to local or in-house use but has successfully been used for non-local annotation of PlantsDB database entries.

Conclusion

The Apollo2Go web service adapter is a powerful extension to the Apollo genome editor. It is ideally suited for manual genome annotation and curation in distributed environments. By combining the widely applied Apollo genome editor with web service functionality, the data exchange is more efficient and the annotation process is more comfortable. The implementation has been made generic. Therefore the method is not restricted to a specific genome data provider or database schema but can be used for different settings and environments. In addition, extended functionality that allows for complex features like security checks before insertion of updated data into the database or data repository may be added to the web service.

Within the annotation process curators are enabled to connect to a remote database, collect the sequences of interest and to display and edit within Apollo. After finishing of the annotation process, the modified data can be directly inserted into the remote database. The necessary read and write access to the database can be secured via password controlled access. Any center hosting genome and annotation data can enable access to the respective

data by hosting web services to either serve the data that allows other groups to display it in their Apollo installation or to accept annotated data from approved curators.

Availability and requirements

Project name: Apollo2Go

Project home page: ftp://ftpmips.gsf.de/plants/apollo_webservice

Operating systems: Platform independent (Apollo is now set up to run with JDK1.5 and it runs on Windows, Unix, Linux and Mac platforms).

Programming language: Java; Other requirements: Java 1.5

License: The Apollo JAVA code is Open Source; Apollo is distributed under the terms of the Artistic License <http://www.opensource.org/licenses/artistic-license.php> and the Apollo source code is available at SourceForge http://sourceforge.net/project/showfiles.php?group_id=27707 as part of the GMOD project <http://sourceforge.net/projects/gmod/>. We plan to include the Apollo web service adapter package as soon as possible at this place.

Any restrictions to use by non-academics: see terms of the Artistic License <http://www.opensource.org/licenses/artistic-license.php>.

The Apollo2Go software package is available via anonymous FTP from ftp://ftpmips.gsf.de/plants/apollo_webservice.

The zipped TAR file contains Apollo2Go, the latest executable Apollo genome annotation curation tool (Version 1.6.5, last updated June 29, 2006) extended with the build-in web service adapter (including the example to connect to PlantsDB at MIPS). The Apollo2Go software does not need any additional local installation of Apollo. Memory requirements are unchanged to Apollo's memory requirements and largely depend on the genomic region selected for annotation. At least 164 Mb of RAM, preferably more, are recommended.

Authors' contributions

KK carried out the programming work, designed the web service and integrated the web service into Apollo. RE conceived the work and participated in the web service design and draft manuscript. MS contributed the database connectivity of Apollo2Go and contributed to the draft manuscript. KFXM conceived and supervised the work and wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank the Apollo consortium for supporting the development of the web service adapter. This work has been in part funded by grants from the European Commission in frame of the PLANET and GLIP projects (QLRI-CT-2001-00006 and FOOD-2004-506223)

References

- Ohyanagi H, Tanaka T, Sakai H, Shigemoto Y, Yamaguchi K, Habara T, Fujii Y, Antonio BA, Nagamura Y, Imanishi T, Ikeo K, Itoh T, Gojobori T, Sasaki T: **The Rice Annotation Project Database (RAP-DB): hub for *Oryza sativa* ssp. japonica genome information.** *Nucleic Acids Res* 2006, **34(Database issue):D741-4.**
- Pennisi E: **Ideas fly at gene-finding jamboree.** *Science* 2000, **287(5461):2182-2184.**
- Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrum J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissole SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramsay J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglu S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409(6822):860-921.**
- Dowell RD, Jokerst RM, Day A, Eddy SR, Stein L: **The distributed annotation system.** *BMC Bioinformatics* 2001, **2:7.**
- Lewis SE, Searle SM, Harris N, Gibson M, Lyer V, Richter J, Wiel C, Bayraktaroglu L, Birney E, Crosby MA, Kaminker JS, Matthews BB, Prochnik SE, Smithy CD, Tupy JL, Rubin GM, Misra S, Mungall CJ, Clamp ME: **Apollo: a sequence annotation editor.** *Genome Biol* 2002, **3(12):RESEARCH0082.**
- Schoof H, Spannagl M, Yang L, Ernst R, Gundlach H, Haase D, Haberer G, Mayer KF: **Munich information center for protein sequences plant genome resources: a framework for integrative and comparative analyses I(W).** *Plant Physiol* 2005, **138(3):1301-1309.**

- Spannagl M, Noubibou O, Haase D, Yang L, Gundlach H, Hindemitt T, Klee K, Haberer G, Schoof H, Mayer KF: **MIPSPplantsDB--plant database resource for integrative and comparative plant genome research.** *Nucleic Acids Res* 2007, **35(Database issue):D834-40.**

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

