BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 14 2011, pages 2001–2002
doi:10.1093/bioinformatics/btr304

Sequence analysis

Advance Access publication May 18, 2011

Java bioinformatics analysis web services for multiple sequence alignment—JABAWS:MSA

Peter V. Troshin, James B. Procter and Geoffrey J. Barton*
Biological Chemistry and Drug Discovery, College of Life Sciences, University of Dundee, Dundee DD1 5EH, UK
Associate Editor: Alex Bateman

ABSTRACT

Summary: JABAWS is a web services framework that simplifies the deployment of web services for bioinformatics. JABAWS:MSA provides services for five multiple sequence alignment (MSA) methods (Probcons, T-coffee, Muscle, Mafft and ClustalW), and is the system employed by the Jalview multiple sequence analysis workbench since version 2.6. A fully functional, easy to set up server is provided as a Virtual Appliance (VA), which can be run on most operating systems that support a virtualization environment such as VMware or Oracle VirtualBox. JABAWS is also distributed as a Web Application aRchive (WAR) and can be configured to run on a single computer and/or a cluster managed by Grid Engine, LSF or other queuing systems that support DRMAA. JABAWS:MSA provides clients full access to each application's parameters, allows administrators to specify named parameter preset combinations and execution limits for each application through simple configuration files. The JABAWS command-line client allows integration of JABAWS services into conventional scripts.

Availability and Implementation: JABAWS is made freely available under the Apache 2 license and can be obtained from: http://www.compbio.dundee.ac.uk/jabaws.

Contact: g.j.barton@dundee.ac.uk

Received on March 31, 2011; revised on May 5, 2011; accepted on May 11, 2011

1 INTRODUCTION

The explosion in biological data volume, diversity and complexity, has been paralleled by the development of computational methods to aid data analysis. Locally installed copies of programs give full access to their capabilities, and the ability to run long CPU or memory intensive jobs. However, configuring software to install easily on a range of operating systems is challenging. Instead, many developers make their techniques available through a custom web page, or increasingly as a web service, which provides a flexible, programmable and scriptable (Hull *et al.*, 2006) middle ground between installing software locally and accessing via a web browser. There is now a rich ecosystem of public web services (Bhagat *et al.*, 2010), but they do not always meet the needs of users. Their interfaces may have limited functionality and even the biggest service providers, such as EBI and NCBI, limit the computational resources (memory/CPU) available to an individual so that total demand can be met. Web services at the University of Dundee were created in 2005 as part of the Jalview project (Waterhouse *et al.*, 2009) to provide its users access to ClustalW (Larkin *et al.*, 2007), Mafft (Katoh and Toh, 2008) and Muscle (Edgar, 2004) multiple alignment programs and the JPred secondary structure predictor (Cole *et al.*, 2008). Although convenient, these suffer from the same public service limitations, and many of the >20 000 users of Jalview requested the ability to run alignments locally on their workstation or to deploy the same services on their institution's cluster. Although locally installable web service systems such as Opal 2 (Krishnan *et al.*, 2009) and SoapLab 2 (Senger *et al.*, 2008) are available, they require significant investment in time or expertise to install. In contrast, the JABAWS:MSA system introduced here, is an easy to install web services solution for multiple alignment that can run on a desktop or laptop computer, local server or computer cluster. JABAWS:MSA's rich interface lets the user track execution progress, retrieve results as a data structure instead of file(s) and provides SOAP services conforming to WS-I Basic Profile 1.2/2.0. While developed primarily for the Jalview Desktop, these can be scripted against, and a command-line client for this purpose is distributed with the system.

2 THE SERVER

The JABAWS:MSA server and command-line client are both implemented in Java, but the server exploits native programs to provide ClustalW (Larkin *et al.*, 2007), Muscle (Edgar, 2004), Probcons (Do *et al.*, 2005), T-coffee (Notredame *et al.*, 2000) and Mafft (Katoh and Toh, 2008) alignment services. These programs can be difficult to compile or incompatible with some operating systems, so a range of JABAWS:MSA distributions have been created to suit most environments. The basic JABAWS:MSA distribution is a pre-configured Java Web Application aRchive (WAR) conforming to the Servlet 2.4 specification, optionally bundled with the source and prebuilt binaries of its dependent bioinformatics programs. This option is most appropriate for experienced users wishing to set up multiple alignment services for their group or institution, or those working on Linux or Mac operating systems. For most other users, a 'Virtual Appliance' (VA) (http://en.wikipedia.org/wiki/Virtual_appliance) is supplied that is built on TurnKey Linux and contains a fully operational JABAWS:MSA server. This can be run on a freely available virtualization product such as VMware player (http://www.vmware.com/products/player) or Oracle VirtualBox (http://www.virtualbox.org).

3 JABAWS:MSA CONFIGURATION OPTIONS

A native JABAWS:MSA installation can be tightly integrated with the local computing infrastructure. It allows fine-grained control

*To whom correspondence should be addressed.

over the calculations to be run on the server through three user-editable configuration files: 'Parameters', 'Limits' and 'Presets', for each alignment method. The Parameters file defines command-line parameters that are supported by the method, its valid ranges and which parameters are mandatory, while the Presets file allows combinations of parameter settings to be given a name and brief description to explain their use. For example, Mafft has six alternative presets, enabling a range of accuracy or speed-oriented options. The Limits file specifies the maximum length and number of sequences allowed for a method, and separate limits may be set for calculations to be run on the same server as Tomcat, or *via* a cluster queuing system. Further configuration files centralize the definition of application-specific directories, and any command-line flags needed to enable JABAWS to interact correctly with a queuing system. Together, these permit JABAWS to operate efficiently in nearly any compute environment, by executing jobs on the local machine, on a cluster through a queuing system or on whichever resource is most appropriate according to a job's size and the availability of local CPUs.

4 THE CLIENTS

Web services provided by a JABAWS:MSA installation can be accessed from any programming language, providing libraries are available for the consumption of SOAP web services. The service interface allows a client to discover the execution limits, command line parameters and named presets for a program, and once a calculation is submitted, check on its progress by retrieving any text normally output to the console by the program. Importantly, data structures and formats for all services are standardized, so the same client can submit data to, and handle the results from any of the alignment programs provided by any JABAWS:MSA server. A lightweight, ready to use JABAWS:MSA command-line client is provided which allows access to JABAWS alignment services. The client is Java based, and is also a useful guide for those who wish to develop or adapt their own clients or test new JABAWS:MSA installations. For non-programmers, a graphical client is available within Jalview 2.6. This client connects to a JABAWS:MSA server to discover which services are available, their parameters and presets, and then populates the user interface with these options.

5 CONCLUSION

A collection of portable web services, initially for multiple sequence alignment, has been developed. It is designed to interact with other software or services and be deployed easily on a variety of computing infrastructures, either directly or through the use of virtualization. JABAWS:MSA provides an in-house web services solution that is limited only by locally available computing resources. Furthermore, its client provides fault tolerance by allowing access to multiple server instances, while the service interface gives a high level of control over individual tasks. Comprehensive metadata provided by each service enables their integration with Jalview, to allow novices and experts alike to utilize their local compute resources efficiently. Finally, although JABAWS:MSA is focused on multiple alignment, the JABAWS framework is flexible, and will soon be extended to include analytical methods such as secondary structure and disorder prediction.

ACKNOWLEDGEMENTS

We thank Dr David Martin for his helpful advice and comments and Dr Tom Walsh for computer systems support.

Funding: European Network of Excellence ENFIN (contract LSHG-CT-2005-518254); Wellcome Trust Strategic Award No. 083481; UK Biotechnology and Biological Sciences Research Council Grants (BBS/B/14434, BB/G022682/1); Scottish Universities Life Sciences Alliance (SULSA).

Conflict of Interest: none declared.

REFERENCES

Bhagat,J. *et al.* (2010) BioCatalogue: a universal catalogue of web services for the life sciences. *Nucleic Acids Res.*, **38**, W689–W694.

Cole,C. *et al.* (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.

Do,C.B. *et al*. (2005) ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.*, **15**, 330–340.

Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

Hull,D. *et al*. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.

Katoh,K. and Toh,H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Brief. Bioinformat.*, **9**, 286–298.

Krishnan,S. *et al*. (2009) Design and evaluation of Opal2: a toolkit for scientific software as a service. In *Services - I, 2009 World Conference on Services*, IEEE Congress on Services, pp. 709–716.

Larkin,M. *et al*. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Notredame,C. *et al*. (2000) T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.*, **302**, 205–217.

Senger,M. *et al*. (2008) Soaplab2: more reliable Sesame door to bioinformatics programs. In *The 9th Annual Bioinformatics Open Source Conference.* SOAP Lab 2 project URL http://soaplab.sourceforge.net/soaplab2.

Waterhouse,A.M. *et al*. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.