Bi⚙techvana

Biotechvana Bioinformatics

The Join Alignments Server version 1.0

Futami, R. [1] Lloréns, C.[1,2] Vicente-Ripolles, M. [1,3] and Moya, A.[2,4]

1 - Biotechvana, Valencia, Spain
2 - Instituto Cavanilles de Biodiversitat i Biología Evolutiva Universitat de València, Spain
3 - Departament de Sistemes Informátics i Computació Universitat Politécnica de València, Spain
4 - CIBER de Epidemiología y Salud Pública (CIBERESP), Spain

Corresponding author: ricardo.futami@biotechvana.com

In this work we introduce the Join Alignments Server a free accessible web-based application that allows users to automatically join several files containing multiple alignments in Fasta format into a single file. Join Alignment Server PHP-script can be downloaded in open source for installation and source code adaptation.

Keywords: PHP | Multiple alignments | Fasta

OVERVIEW

DNA and protein alignments are an important substrate for inferring phylogenetic relationships of genes and proteins. A frequent aim in multiple alignment methodology is to identify the set of motifs common to all sequences with the same known function; this is the most conserved part (core) of a DNA and protein domain. In cases such as mobile genetic elements (1) it is usual to find Open Reading Frames (ORFs) coding for polyproteins that contain several protein domains. Each protein domain usually has a phylogenetic signal that may be, with obvious variations due to the different rates of evolution, more or less similar to that preserved by the remaining harbored by these proteins. In these cases and depending of the goal, researchers may be interested in evaluating the phylogenetic signal of one or more protein domains. For instance, in the Gypsy database Project (2) we have recently constructed a collection of hidden markov model (HMM) profiles (3) taking into consideration the clusters reported by the phylogenetic analysis of Ty3/Gypsy and Retroviridae LTR retroelements inferred based on all protein products encoded by the gag-pro-pol region. To achieve this analysis we obtained a number of multiple alignments (one for each protein domain), and concatenated all alignments in a single one to obtain a general perspective of the clustering. To facilitate this labor, which usually is manual we designed an automated tool, the Join Alignments Server (JAS) to automatically join several alignments.
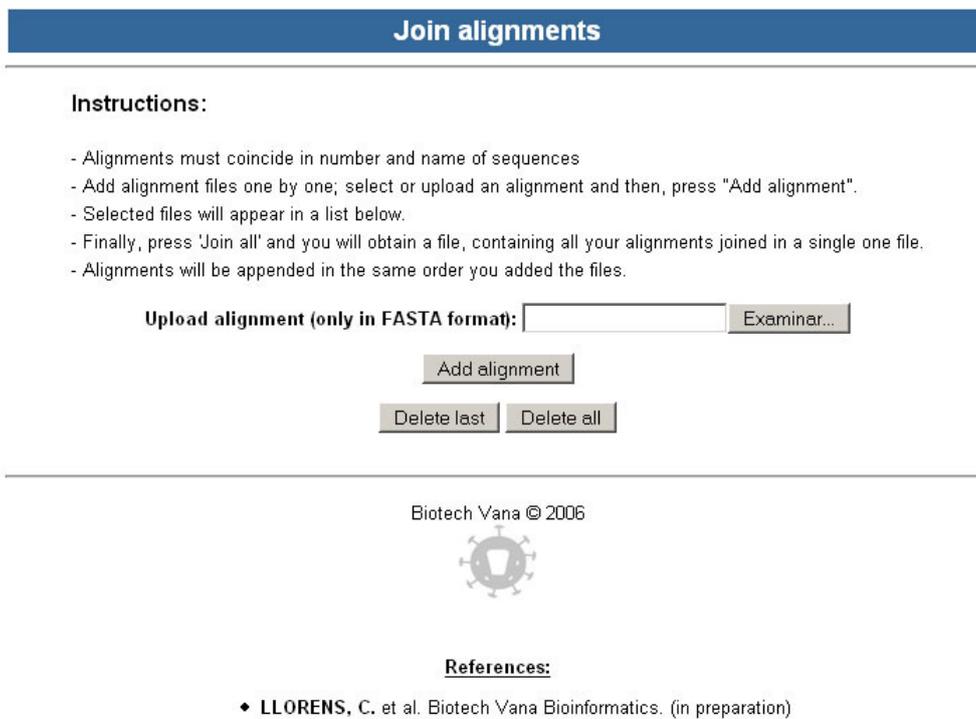

Figure 1. Screen shot of JAS web interface.

Biotechvana Bioinformatics

JAS is very simple and consists in a web-based interface (Figure 1) that allows the user to select and upload a number of alignment files to be joined. Users can specify the order in which the files should be joined. Once all files have been uploaded a single result file is given. This file contains all alignments joined in the same order indicated by the user. JAS has two requisites: the sequence number and name of sequences must be identical in each file in order for successful joining, and alignments must be provided only in Fasta format. A short example of joining alignment files is shown in figure 2. The application verifies that each file contains the same set of sequences (if the files do not contain the same number of Fasta sequences labeled by the same names the algorithm ends). Once this step is verified, starting from the first file (A), the application creates a virtual structure containing a name for each sequence, and creates in the server's memory the corresponding sequence. The next file (B) is then loaded to create a new collection of structures, which are compared with file A structures one by one. It then appends sequences of file B where they correspond immediately downstream (or at C-terminal) to file A sequences in the server memory structure provided by file A. This process is repeated for each additional file to be joined. Finally, a result file containing all concatenated alignments is given in a plain file. The flow diagram of this process is shown in figure 3.

FILE_1
```
>SEQ_1 NAME_A
RAWPQWVPGPMGGRGDAVD
>SEQ_2 NAME_B
GPVKQRLGKRGALSQSLGL
>SEQ_3 NAME_C
RRKSRHHPRPRGRKDGSCL
```

FILE_2
```
>SEQ_1 NAME_A
PDKPSLRTPTEGTRVHAPN
>SEQ_2 NAME_B
EPLKMDRKTHSEWSQFQEK
>SEQ_3 NAME_C
STKISHPERMSRRHQLHQE
```

FILE_3
```
>SEQ_1 NAME_A
GSPATNISGDAKQGRACAA
>SEQ_2 NAME_B
HMRTPPITVDLIINNVANA
>SEQ_3 NAME_C
MQHGSPMGKRHAPSVTERQ
```

OUTPUT_FILE
```
>SEQ_1 NAME_A
RAWPQWVPGPMGGRGDAVD
PDKPSLRTPTEGTRVHAPN
GSPATNISGDAKQGRACAA
>SEQ_2 NAME_B
GPVKQRLGKRGALSQSLGL
EPLKMDRKTHSEWSQFQEK
HMRTPPITVDLIINNVANA
>SEQ_3 NAME_C
RRKSRHHPRPRGRKDGSCL
STKISHPERMSRRHQLHQE
MQHGSPMGKRHAPSVTERQ
```

Figure 2. Joining FASTA sequence files example

INSTALLATION

First, confirm that a Web server engine and a PHP application server are properly installed on your system. You can download and install a web server like Apache from URL 2 for Windows and Linux platforms; or an IIS (Internet Information Services) Web server for Windows platforms, which comes included in Windows server versions. Next, install the

PHP application server, which can be downloaded at URL 3; Installation instructions are provided on its corresponding Web sites. Once a Web server engine and a PHP application server are properly installed and working, unpack the package containing this script on your server's public folder, specified in its documentation. A folder named 'jas' will be created. To access to JAS, open a Web browser and type the following URL location: http://localhost/jas/main.php.

REQUIREMENTS

The online version of JAS requires only a web browser and an Internet connection. To install JAS as a private server, you will need a PHP application server and a web server. A web server is a computer program responsible for accepting HTTP requests from a user's web browser and returning contents such as HTML web pages, images, and other files. An application server is software that helps a web server to process web pages containing server-side scripts that cannot be processed by a regular web server. When a dynamic page is requested by a visitor's browser, the web server calls the application server for processing the scripts before sending the page to the browser. Alternatively, an external dynamic web hosting service can be used to host JAS; this allows the user to bypass installation and configuration of a Web server and a PHP application server on the local computer.

ACKNOWLEDGMENTS

Figure 3. Flow diagram of JAS server

Biotechvana Bioinformatics
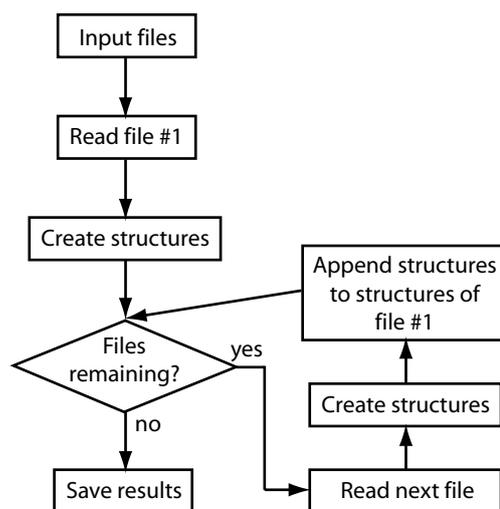
LITERATURE

1. Craig,N.L., Craigie,R., Gellert,M. and Lambowitz,A.M. (2002) Washington, DC.

2. Llorens,C., Futami,R., Bezemer,D. and Moya,A. (2008) Nucleic Acids Research (NAR) 36 (Database-Issue):38-46

3. Eddy,S.R. (1998) Bioinformatics, 14, 755-763.

URLS

1. **Open Source License:** http://biotechvana.com/loader.php?section=contents&page=terms_ocl
2. **PHP programming language:** http://php.net
3. **Apache web server:** http://www.apache.org
4. **SCSIE, Universitat de València:** http://scsie.uv.es

Biotechvana Bioinformatics

SPONSORS