

Supplementary Material

MOWServ: a web client for integration of bioinformatics resources

Sergio Ramírez^{1,#}, Antonio Muñoz^{1,#}, Johan Karlsson¹, Maximiliano García¹, M. Gonzalo Claros² and Oswaldo Trelles¹

[1] Computer Architecture department, ETSI Informática

[2] Molecular Biology and Biochemistry, Faculty of Science

Campus de Teatinos s/n, University of Malaga, 29071 Málaga, Spain

These authors contributed equally

Homology Search and Phylogenetic Study

Given a query amino acid sequence, find similar sequences with a common evolutionary history.

Procedure:

STEP 1: Retrieve the query sequence (by ID) from databases (getAminoAcidSequenceUMA service):

STEP 2: Perform a homology search (Blast service)

STEP 3: Select the more similar sequences from the Blast report (getBestHitsFromBlast & getAminoAcidSequenceCollection)

STEP 4: Perform a Multiple Sequence Alignment (Clustalw service)

STEP 5: Built-up a Phylogenetic Tree (CreateTreeFromClustalw service)

Start the exercise by contact the MOWServ client at www.inab.org/MOWServ or use any of the available mirrors (e.g. <http://www.chirimoyo.ac.uma.es/MOWServ>)

The screenshot displays the MOWServ web application interface. At the top, there is a navigation menu with tabs for MISSION, PROJECTS, ORGANIZATION, RMB, RESOURCES, TRAINING, PROGRAM, NEWS, CAREERS, and CONTACT. Below the navigation is a search bar and a login section for an anonymous user. The main content area is titled 'Products' and lists several services: IWWE&M (Interactive Web Workflow Enactor & Manager), JORCA (Easily integrating), Magallanes, and MOWServ Workflow List. A left sidebar contains a tree view for Data Types, Services, and Namespaces.

MOWServ's main screen. The upper panel allows authenticated sessions to be established, providing user files system space. On the left, the browsing frame for service, datatype and namespaces identification (including text based searches, etc). The main panel is used for requesting service parameter, to display results, etc.

The exercise has been tested using the MOWServ client at the University of Malaga (UMA), and using the services provided by UMA authority at <http://www.chirimoyo.ac.uma.es/MOWServ> (in most cases, services have identified with the suffix UMA to facilitate the exercise reproduction)

STEP 1: Amino Acid Sequence retrieval.

The first step will retrieve the protein sequence from its identifier.

Browse the service tree (left panel): Services -> Service -> Bioinformatics -> Database -> Retrieving -> Getting sequences -> Getting AminoAcids

An alternative way is to use the search box (left panel) to filter services by keyword, or the Discovering engine (Magallanes) in the right hand side tab.

(Observe that several services with similar or even with the same name can be available, corresponding to services supplied by different authorities).

Click in the **getAminoAcidSequenceUMA** service to open the parameter interface. In this interface, there are two main groups of parameters (see Figure):

INPUT PARAMETERS: ID specification (use SMN_HUMAN as example).

OUTPUT NAME for the file that will contain the retrieved sequence. The system provides a by default name that can be modified by the user (default name; service name + Date + GM time). In this exercise we rename the output file to: "SMN_HUMAN Object"

Once the submit button is clicked, the progress of the service can be monitored in "User tasks" tab using reload in case of the service require more than few seconds, and when the process had finished, the result can be displayed in different formats also from this tab:

- XML: View the object in XML format like a BioMOBY object (click on the below icons to view how your output objects are shown).

STEP 3: Select the most similar sequences from Blast report

This step is aimed to produce a file with a set (Collection) of the most similar sequences from the Blast report.

This is solved in two sub-steps: (a) select the IDs of the most similar sequences using a threshold based on the Blast E-value; and (b) use the IDs to retrieve the sequences and store in a collection file.

For the selection of the most similar sequences we will execute the service [getBestHitsFromBlast](#) (Services -> Service -> ObjectHandling -> Parsing) using 0,02 and 5 as Threshold value and number of best hits to be retrieved (default parameters).

The screenshot shows a web-based configuration form for the service 'getBestHitsFromBlast'. The form is titled '[www.bioinfo.uma.es] getBestHitsFromBlast: Get the best hits from a Blast report'. It is divided into 'INPUT PARAMETERS' and 'OUTPUT NAME' sections. Under 'INPUT PARAMETERS', there are four fields: 'Threshold_type' (String, Secondary) with a dropdown menu set to 'Expected_value'; 'Threshold_value' (Float, Secondary) with a text input field containing '0.02'; 'Hits' (Integer, Secondary) with a dropdown menu set to '5'; and 'Blast_report' (BLAST-Text) with a dropdown menu set to 'Blast SMN_HUMAN'. There are 'Create' and 'Upload' buttons to the right of these fields. Under 'OUTPUT NAME', there is a 'NAME' field with a text input containing 'getBestHitsFromBlast-2010-1-7-01:18:43' and a 'TYPE' dropdown set to 'Object (collection)'. Below this is a 'Submit' button and a 'Let the scheduler choose for you' dropdown menu. At the bottom, there are 'Home' and 'Service Tutorial' links.

When we have the ID list we will need to retrieve the corresponding sequences from databases. In this way we will use [getAminoAcidSequenceCollection](#) service (Services -> Service -> Database -> Retrieving -> GettingSequences -> GettingAminoacids) in order to retrieve the aminoacid sequences from the previous IDs/databases (object collection) to produce an amino acid sequence collection. When the service has finished we can look at the created objects in the html viewer. The output file can be renamed as "AA Collection SMN_HUMAN" to distinguish from the rest of objects.

Type: AminoAcidSequence **Namespace:** GenBank **Id:** Q16637

Type	articleName	Content
Integer	Length	294
String	SequenceString	MAMSSGGSGG GVEPQEDSUL FRRGTGQSDD SDIWDOTALI KAYDKAVAS F KHALKNGD IC ETSQKPKTTP KRKPAKQKKS QKQNTAASLQ QWVQDKCSA IWSDDGCIYP ATIASIDFKR ETCVVYTY GWRREQLSD LLSFICEVAN NIEQNAQENE MESQUSTDES EMRSFPGMS DNIKPKAPW NSILPPFPFH PGPFLGPKP GLKTRGPPFP PFPFPHLLS CULFPFSPG PIIPFPFIC PDSLDDADAL GSHLSWIMS GYHTGFMGF RQWKEGRCS KSLM

Type: AminoAcidSequence **Namespace:** GenBank **Id:** P97801

Type	articleName	Content
Integer	Length	288
String	SequenceString	MAMSGGAGS EQEDTULFR GTGQSDSD I WDDTALIKAY DKAVASFGHA LKNGD ICETP DKPKGTARRK PAKKMSQKK NATTPLKQVK VGDKCSAOWS EDGCIYPAT I TSIDFKRET C VVVYTYGWR EDQLSLLS PFCVANSYE QMTQEMESVU STDDSEHSSR SLRSKANSK KAAPWMSIL PFPFPHGSL GPKGLKRW GPPFPPLFP PFLPCWMPF PFGFPPIIFP PFPISPDCLD DTDALGSM I SWMSGYHTG YMGFRQKX EKCSHTM

Type: AminoAcidSequence **Namespace:** GenBank **Id:** O02771

Type	articleName	Content
Integer	Length	287
String	SequenceString	MGGGGGLPEP EDSULFRRGT GQSDSD IVD D TALIKAYDK AVASFKHALK NGDISEASDK FRSTPKRKA KRNKQKRW T TALKRQWVQ DECSAVWSED GC IYPAT IAS IDFKRET CVV VITGWRRE QWVQDKCSA CFWANRQTD TQEMESQ I STDESSESR SPGRPKIK SKAAPMSIL PFPFPHGSL LQPKGPKT GPPFPFPFP HILSCMLPFP PFGFPPIIFP PPICPDCLD ADALGSM I SWMSGYHTG YMGFRQKX EKCSHTM

STEP 4: Multiple Alignment.

Now we will run the ClustalW program with the most similar found sequences to our query sequence to build up a multiple alignment with them.

The used service in this step is [runClustalwFastUMA](#), located on the tree branch: Services -> Service -> Bioinformatics -> Alignment -> Multiple_Sequence_Comparison. This service has several input parameters that we will leave in its default values (it can be changed in order to modify the final alignment). The service has also an input where we select the previous sequence collection.

Now we can choose an output name (e.g. Clustalw SMN_HUMAN) and submit the task.

The screenshot shows the web interface for the service `runClustalwFastUMA`. It features a navigation bar with tabs for 'Main', 'User Objects', 'User Tasks', 'Workflows', and 'Discovering'. The main content area is titled '[chirimoyo.ac.uma.es] runClustalwFastUMA: Run a clustal from a collection of generic sequences'. Below this is a section for 'INPUT PARAMETERS' with a table of settings:

NAME	TYPE	VALUE
topdiag	Integer (Secondary)	5
gapdist	Integer (Secondary)	4
gapopen	Float (Secondary)	
pair gap	Integer (Secondary)	3
endgaps	String (Secondary)	off
matrix	String (Secondary)	BLOSUM
window	Integer (Secondary)	5
transitions_weight	Float (Secondary)	0.5
ktup	Integer (Secondary)	1
gapext	Float (Secondary)	
sequences	GenericSequence	⌘ Collection SMN_HUMAN

Below the input parameters is an 'OUTPUT NAME' section with a text input field containing 'Clustalw SMN_HUMAN'. There are 'Create' and 'Upload' buttons to the right of the input parameters, and 'Submit' and 'Reset' buttons below the output name field. At the bottom of the interface are links for 'Home' and 'Service Tutorial'.

When the service has finished we can look at the created object in the html viewer. This displays the multiple alignments formatted by the Mview program.

```
CLUSTAL 2.0.9 multiple sequence alignment

GenBank-002771      ----MGGGGGLPEPEDSVLFRRTGQSDSDIWDDTALIKAYDEAVASFRKHALKNGDIS
GenBank-018870      ----MGGGGGGFPEPEDSVLFRRTGESDSDSVWDDTALIKAYDEAVASFRKHALKNGDIS
GenBank-Q16637      MAMS3GGGGGGVPEQEDSVLFRRTGQSDSDIWDDTALIKAYDEAVASFRKHALKNGDIC
GenBank-F97801      --MAMSGGGAGSEQEDTVLFRRTGQSDSDIWDDTALIKAYDEAVASFRKHALKNGDIC
GenBank-035876      --MAMSGGGAGSEQEDTVLFRRTGQSDSDIWDDTALIKAYDEAVASFRKHALKNGDMC
      ..** . * ..*****:*****:*****:*****:*****:*****:
      ..** . * ..*****:*****:*****:*****:*****:*****:

GenBank-002771      EASDKFRSTPFRKPAKKNKSRKKNATTALKQKVKVGRKCSAVWSEDCIYPATIASIDFRR
GenBank-018870      EASEKFRGTFRKSAK-NKSRKNTTSPSKQKVKVGRKCSAVWSEDCIYPATIASIDFRR
GenBank-Q16637      ETSGRKTIPTFRKPAKKNKSRKKNATLAAASLQKVKVGRKCSAVWSEDCIYPATIASIDFRR
GenBank-F97801      ETPDRFGTARRKPAKKNKSRKKNATTLKQKVKVGRKCSAVWSEDCIYPATIASIDFRR
GenBank-035876      ETSDKFRGTARRKPAKKNKSRKKNATLFLKQKVKVGRKCSAVWSEDCIYPATIASIDFRR
      *..** . * ..*****:*****:*****:*****:*****:*****:

GenBank-002771      ETCVVVYTYGNREEQNVSDLLSPACEVANNVEQDTQENENESQISTDESENSSRSFGNK
GenBank-018870      ETCVVVYTYGNREEQNLSDLLSPTSEVAN-IEQNAQENENESQISTDESENSSRSFLNK
GenBank-Q16637      ETCVVVYTYGNREEQNLSDLLSPTCEVANNTIEQNAQENENESQVSTDESENSSRSFGNK
GenBank-F97801      ETCVVVYTYGNREEQNLSDLLSPTCEVANSTEQNTQENE--SQVSTDDSEHSRSLRSK
GenBank-035876      ETCVVVYTYGNREEQNLSDLLSPTCEVANTEQNTQENE--SQVSTDDSEHSRSLRSK
      *****:*****:*****:*****:*****:*****:*****:

GenBank-002771      FNNIKSKAAPWNSFLPPFPMSSGLGPGKFGVKFSGPPPPPPPP--HFLSKWLPFFPSG
GenBank-018870      FNNIRRAAPWNSFLPPFPMRSGSLGPGKSLMTSGPPPPPPPP--HFLSKWLPFFPSG
GenBank-Q16637      SNNIKRKAAPWNSFLPPFPMRSGRLGPGKFLRNGPPPPPPPPPPHLLSKWLPFFPSG
GenBank-F97801      AHSK-SKAAFWNSFLPPFPMRSGSLGPGKFLRNGPPPPPPPPPPFLCWMPPFPSG
GenBank-035876      AHSK-SKAAFWNSFLPPFPMRSGSLGPGKFLRNGPPPPPPPPPPFLCWMPPFPSG
      ..** . * ..*****:*****:*****:*****:*****:*****:

GenBank-002771      FPIIPFPFPCPSLDDADALGSMLISWYMSGYHTGYMGFKQKQEGRCRSHFN-
GenBank-018870      FPMIPFPFPCPSLDDADALGSMLISWYMSGYHTGYMGFKQKQEGRCRSHFN-
GenBank-Q16637      FPIIPFPFPCPSLDDADALGSMLISWYMSGYHTGYMGFRQKQEGRCRSHLN-
GenBank-F97801      FPIIPFPFPCPSLDDADALGSMLISWYMSGYHTGYMGFRQKQEGRCRSHLN-
GenBank-035876      FPIIPFPFPCPSLDDADALGSMLISWYMSGYHTGYMGFRQKQEGRCRSHLN-
      ..** . * ..*****:*****:*****:*****:*****:*****:

```

STEP 5: Phylogenetic Tree

Finally we will build a phylogenetic tree in phylip format which can be visualized with a specific viewer.

To construct the phylogenetic tree from the clustalw output, we will run the [runCreateTreeFromClustalw](#) service (Services -> Service -> Bioinformatics -> Distances -> Phylogenetics -> Phylogenetics_Tree_Computing). This service has an only input parameter:

Clustalw_output: We will select the before clustalw_report from the object list in the interface (Clustalw SMN_HUMAN).

Again we have to include an output name and submit the task.

The screenshot shows a web interface for the service [\[www.bioinfo.uma.es\] runCreateTreeFromClustalw: Produces a phylogenetic tree with dnd format from a ClustalW multiple alignment](#). The interface is divided into two main sections: INPUT PARAMETERS and OUTPUT NAME.

INPUT PARAMETERS:

NAME	TYPE	VALUE
Clustalw_output	Clustalw_Text	Clustalw SMN_HUMAN

Buttons: Create, Upload

OUTPUT NAME:

TYPE	NAME
Hewick_Text	tree from SMN_HUMAN

Buttons: Submit, Reset

URL: http://chirimoyo.ac.uma.es/cgi-bin/services/MobyEd_dispatcher.cgi

Home Service Tutorial

Now, when the service finishes, we can see the created object, as a phylogenetic tree, with the html viewer. Here, the initial query sequence (SMN_HUMAN or Q16637; Survival motor neuron protein) is mostly linked with the other proteins from the same family (O02771, O18870, rat and mouse homologues, O35876, and P97801, and one fish protein Q16637.

In short, with this set of services we have obtained the SMN_HUMAN homologues from Swiss-Prot database and their phylogenetic relations, but also have the relations with other distant proteins sharing domains with our query sequence.

Additionally we could have executed other intermediated services, if they are considered useful for our analysis.

Contact person:

Antonio Muñoz-Mérida: <amunoz@uma.es>

Oswaldo Trelles <ortrelles@uma.es>