
译自《WebLab 帮助文件》



WebLab 中文使用文档

姓名：靳进朴

学号：10811061

注：本文中的插图均来自《WebLab 帮助文件》

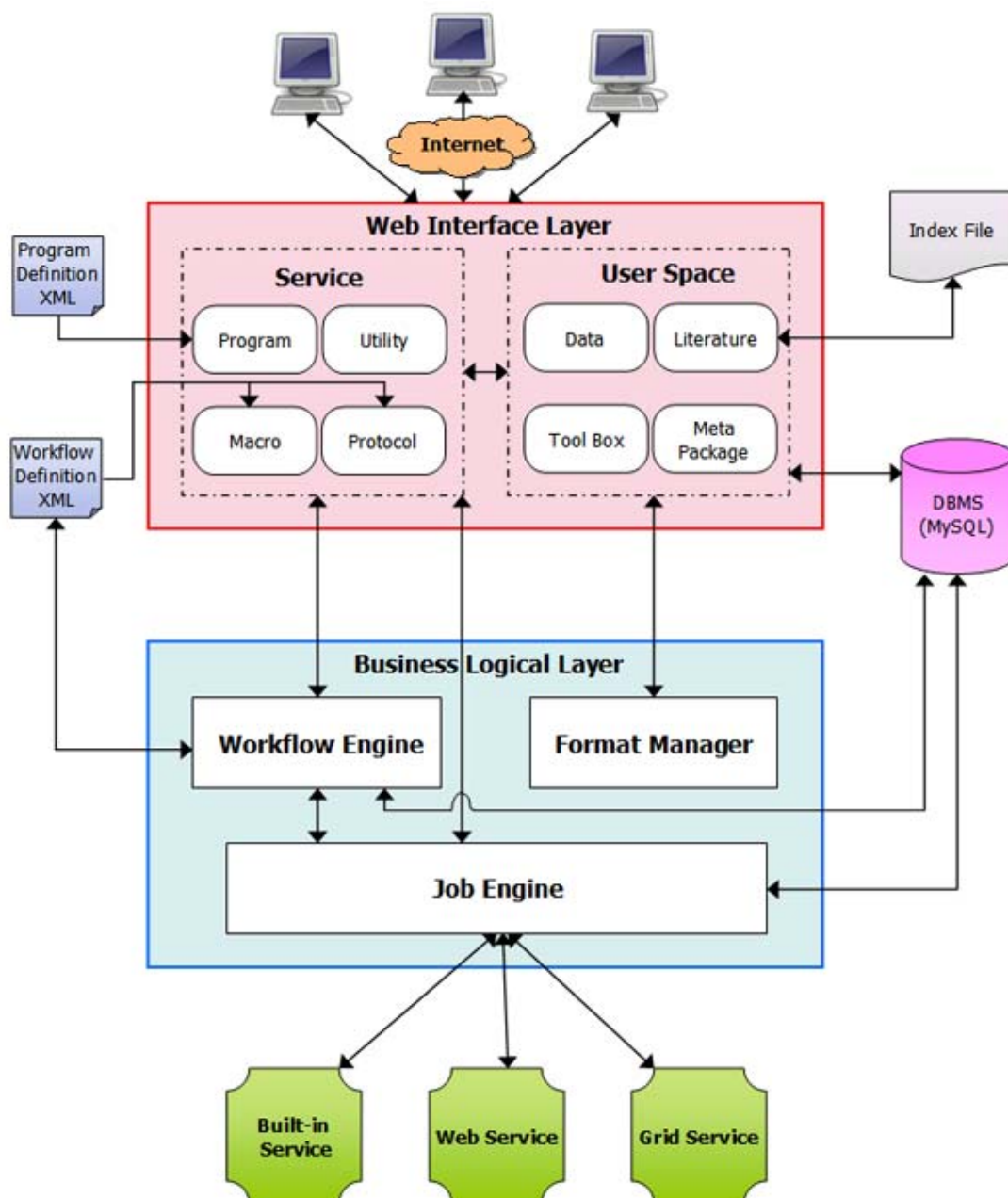
目录

1. WebLab 简介.....	3
2. 用户注册.....	4
3. 用户空间及数据管理.....	6
4. 如何在 WebLab 上工作.....	15
5. 和同事一起工作——分组策略和分享机制.....	32

1. WebLab 简介

伴随着生物数据的激增，出现了大量用于分析这些生物数据的工具和网站。然而安装和维护这么多生物信息软件是非常困难的，不但非常耗时，而且成本也很高。为了提高生物信息领域的研究效率，开发一个整合了各种分析工具，具有友好用户界面的多功能平台是非常有意义的。

开发 WebLab 的目的就是为了给生物医学研究人员提高一个“一站式”的生物信息分析平台。在 WebLab 上，用户不但可以使用上面的丰富的软件进行生物信息分析，而且可以利用 WebLab 上强大的数据管理功能，分组策略和数据共享机制来大大的提高工作效率。



如上图所示，WebLab 的架构是一个使用代理模式的层次性的拓扑结构。服务器的入口是一个具有友好 web 用户界面，它作为一个代理来接受用户的请求。WebLab 主要由用户界面显示层和内部逻辑处理层组成。在接受到入口服务器的请求后，相应的后台服务器就会运行要求的程序。在分析完成后，便将结果送回入口服务器并被存入入口服务器的数据库中。

一般来说，用户可以把 WebLab 当作一个松散联系的分布系统。

2. 用户注册

为什么每个用户都需要一个帐号？

首先说明一点，注册帐号是免费的。注册了账号以后，你就可以将序列和分析结果存入“my data”。这对于有大量数据要分析的用户来说是非常有用的。

其次，你还可以分别在“my literature”，“my toolbox”和“my metapackage”中保存和管理文献，喜欢的工具，meta-packages 等以便于进一步分析。这意味着只要你连上互联网就可以在地球的任何地方得到你的数据。

如果你的任务需要运行很长时间，不能立即得到结果，你还能监视已提交任务的状态。而且还易于控制已提交的任务。

你可以根据任务来运行和设计流程（protocol 和 macros）。

可通过分享机制和分组策略来合作研究。

因此，我们强烈推荐用户注册一个账号。你只需要填一个简单的表格点一下注册按钮便可拥有一个属于自己的账号。

注：填写一个正确的 email 地址是非常重要的。你需要通过 email 来激活你的账号，并且当你的长任务完成后，服务器会将运行结果发到你邮箱里。

The screenshot shows the WebLab interface. At the top, there's a banner with the text "Center of Biological your lab on the web" and "WebLab". On the right side of the banner, there are links for "Feedback", "Contact", and "Help".

On the left side, there's a navigation menu with the following items:

- Service
 - Program
 - Protocol
 - Macro
 - Utility
- User Space
 - My Data
 - My Literature
 - My MetaPackage
 - My Toolbox
 - History
- Account
 - Login in
 - Registration

The main content area has the heading "Why do I need an account" and two bullet points:

- For freely registered account, you can save sequence and analysis result in "user data space". It's especially useful if you have many sequences to analyze.
- You could monitor the status of your submitted work if it needs long time to compute and can't return result immediately.

Below this is the "ACCOUNT REGISTRATION" form with the following fields:

- * Account: [input field] 3-20 characters
- * Email: [input field] valid email address
- * Password: [input field] 5-20 characters
- * Confirm Password: [input field] 5-20 characters
- * Name: [input field]
- * Organization: [input field]

At the bottom right of the form, there is a "Submit" button. Below the form, it says "Fields with * are required to fill in."

登陆后，用户可以通过点击“My Account”来看自己的账号信息。在这，服务器允许用户更改账号的基本信息。



The screenshot shows the WebLab user profile page. The header includes the CBI logo and the text "Center of Bioinformatics your lab on the web WebLab". On the right, there are links for Feedback, Contact, and Help. The left sidebar contains navigation menus for Service, User Space, and Account. The main content area displays account information in a table:

Account:	wangjun
Email:	wangjun_kathleen@126.com
Name:	wangjun
Organization:	ecnu
Send long time job result to email box:	Yes
Register Date:	2006-10-11
Last Login Time:	2008-03-05 22:54:49

An "Edit My Account" button is located at the bottom right of the table.

在编辑页，用户可自由的更改信息。



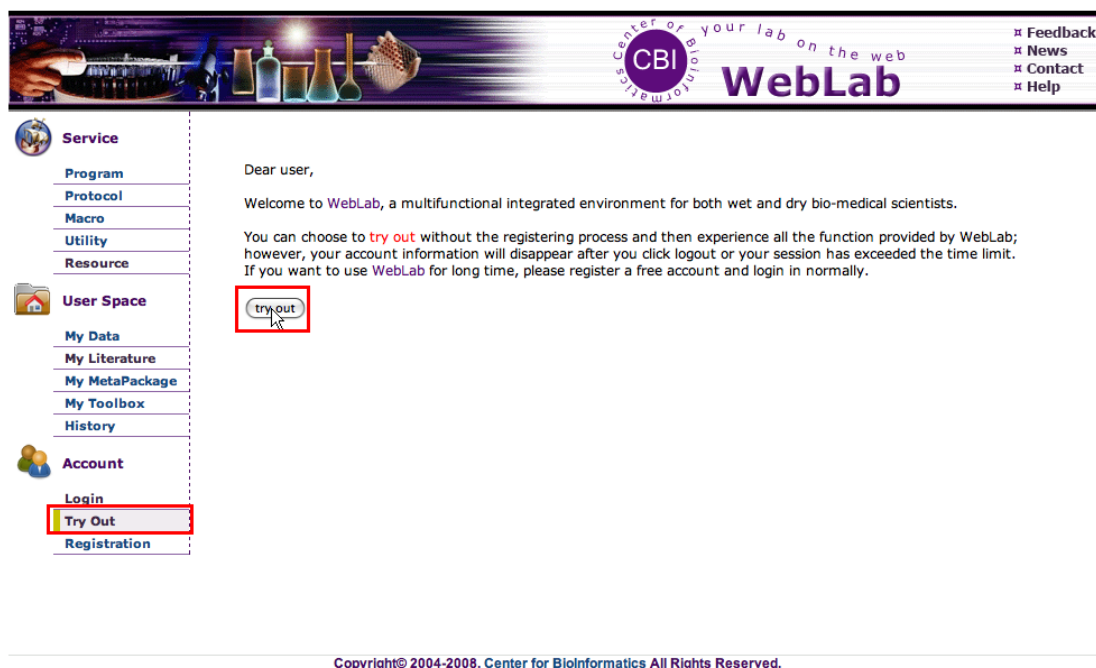
The screenshot shows the WebLab user profile page with the registration form. The header and sidebar are the same as in the previous screenshot. The main content area displays the registration form with the following fields:

Email:	xxx@xxx.xxx
Password:	*****
Confirm password:	*****
Name:	xxx
Organization:	xxx
Send long time job result to email box:	<input type="button" value="Yes"/>

A "submit" button is located at the bottom right of the form.

对于暂时不想注册的用户，WebLab 也提供了“try out”功能。

在没有注册的情况下，用户可通过“try out”体验 WebLab 提供的所有功能。不过，当你登出后或对话超出时间限制，你的账号信息将会消失。如果你经常使用 WebLab，最好能注册一个账号并在使用前登陆。



3. 用户空间及数据管理

1). My Data

“My Data”就像用户的网络硬盘。没有注册的用户只能从本地磁盘上传文件或粘入数据，而每一个注册的用户都有一个用户空间来存储序列进行后续的分析。

便于选择保存的文件进行不同的分析。

不受空间限制的继续做工作。例如，你可以在任何地方用不同的计算机分析这些数据，因为 WebLab 中保存你需要的数据。

分享数据给指定的用户。

基本操作：我们为用户提供了一些基本操作来管理自己的文件，包括目录操作，上传文件，修改，删除，分享和添加评论和标签。

目录操作：用户可以创建层次性的目录来保存数据。这对于管理大量数据和分析结果是非常有利的。此外，用户还可以删除，重命名和移动这些目录。

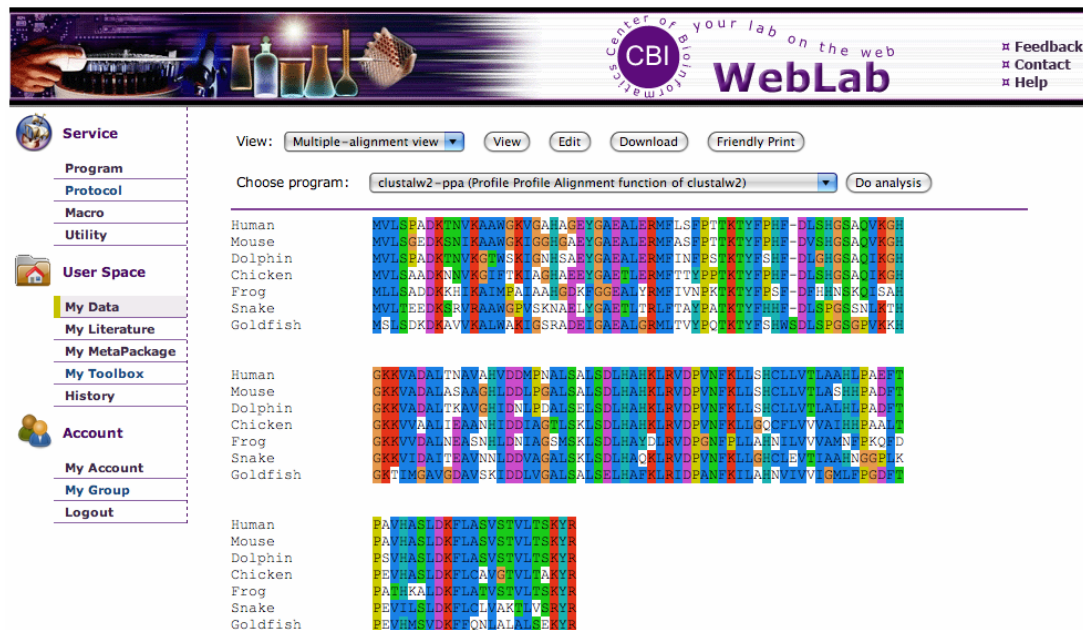
文件上传：有几种上传文件的方法。上传后，用户可以重命名，移动或删除这些文件，也能改变文件所对应的格式。

几种方法：在文本框中粘入序列；从本地磁盘上传；从 WebLab 内嵌的 resource 模块来获得数据。

注：WebLab 包含一个数据格式表，该表中定义了生物医学领域中使用的各种数据格式。请在上传文件时指定一个恰当的数据格式，因为很多对数据的操作依赖于数据的格式。用户可以在“My data”中修改数据的格式。

数据操作：对于非二进制文件，用户可通过点击文件名浏览文件内容。

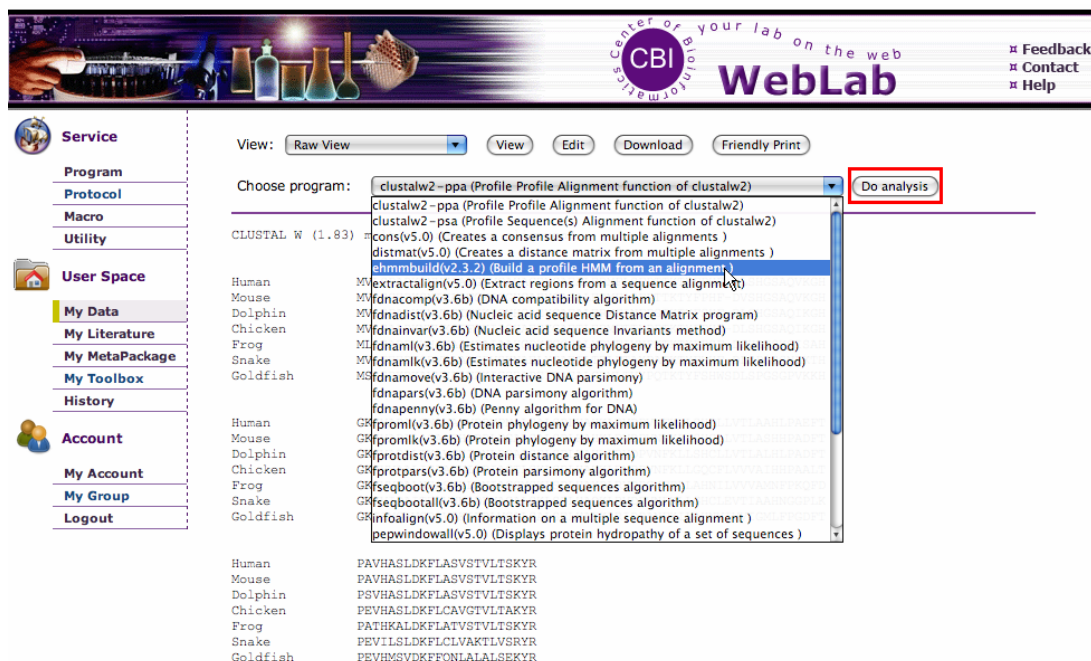
对一些特殊的数据格式，WebLab 提供了一个更直观的浏览方式，这是 WebLab 的呈递机制。目前安装了一些常用数据类型的呈送，包括 ClustalW，双重序列比对的输出，KOBAS 结果，Prints 和 Prosite 数据库的检索结果。用户可以通过选择合适的观看方式来更好的理解这些结果。



通过点击相应的按钮，用户可以直接的编辑或打印数据文件。

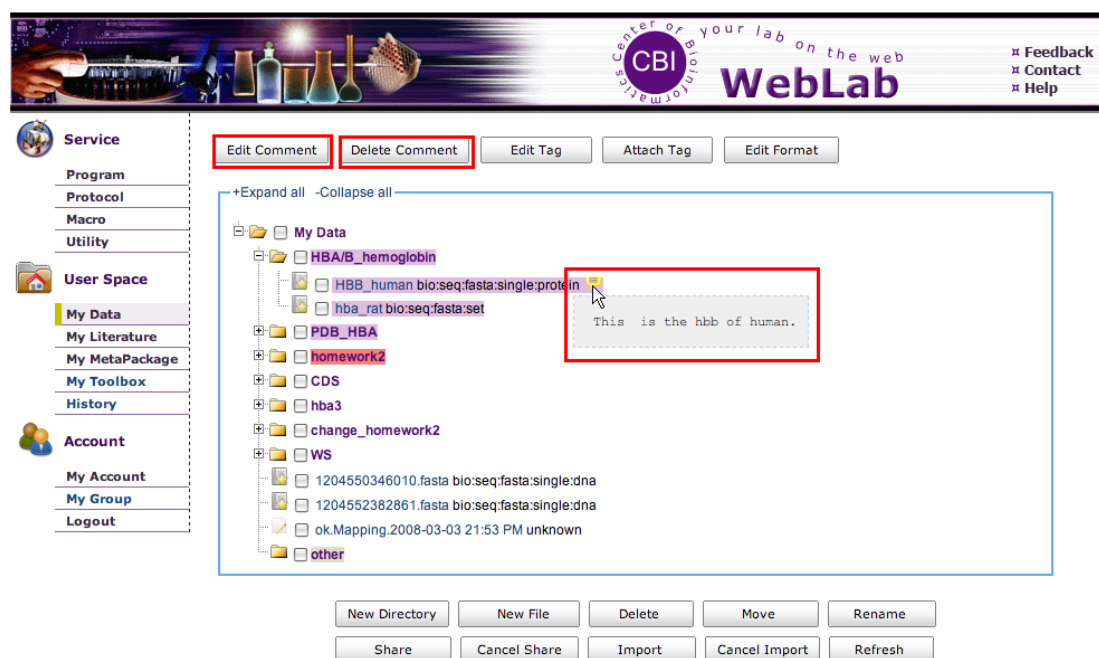
通过点击“do analysis”按钮，用户可以从程序列表中选择需要的程序做进一步的分析。

注：根据指定的数据类型，系统可以将处理该种数据类型的程序列出。因此，不同的数据类型有不同的程序列表。

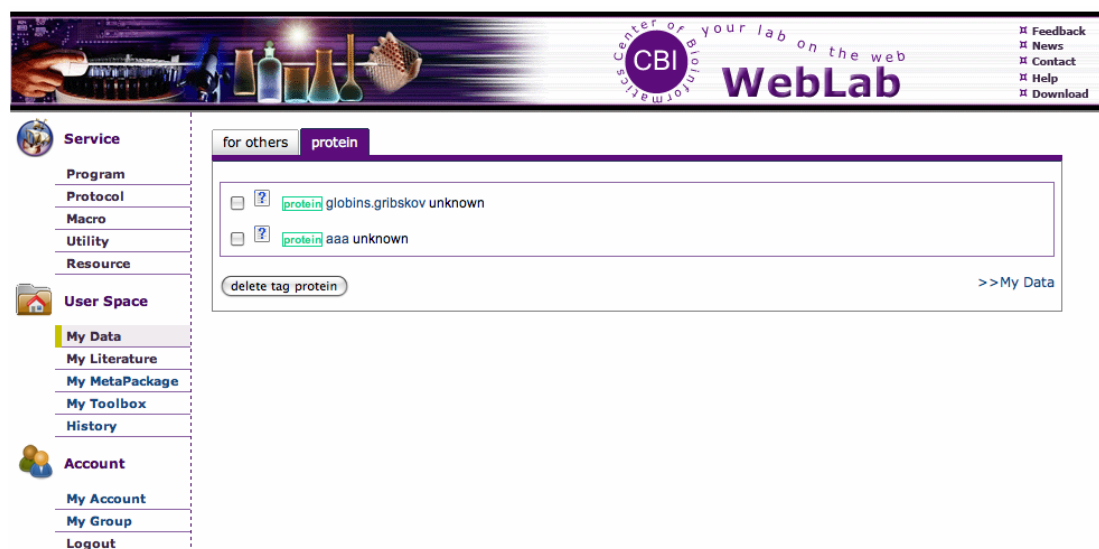


注释和标签:

注释: 注释是用户对文件的描述。它可以帮助用户在短时间内理解文件内容以提高工作效率。用户可通过添加，编辑或删除注释使文件更容易理解。然后，当把鼠标放到黄色图标上时，你将会看到关于该数据的简短描述。



标签: 除了用经典的树状结构来组织数据文件外，还可通过自定义的标签来浏览“my data”中的数据。



通过点击“edit tag”按钮，用户可创建或修改标签。然后通过“attach tag”按钮为文件添加自定义的标签。

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

分享: 在"my data", "my literature" and "my metapackage" 中的所有文件都可以分享给你的同事。

2). My Literature

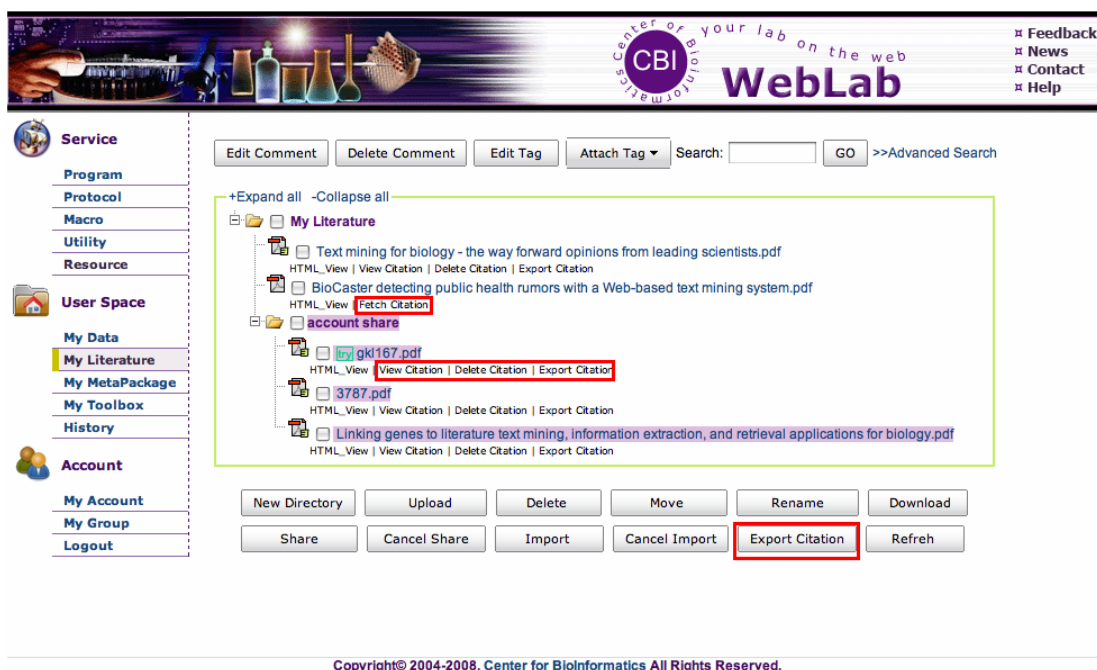
WebLab 提供了一个“my literature”让用户来管理自己的文献。“my literature”除了具有“my data”具有的功能外，还对全文检索，通过 EndNode 和 PubMed 等参考文献管理工具进行互操作等提供全面的支持。

文献上传: 用户可以上传文献到指定的目录。目前，我们只接受 txt, pdf, doc 格式文献。另外，我们支持批量上传。用户最多可一次上传 10 篇英文文献，每个不超过 10M。

引用信息: 除了文献本身，WebLab 还允许用户添加每篇文献的引用信息。

引用信息可提供给用户关于该文献的充分的细节。用户可通过引用信息来检索文献。更为重要的是，所有的引用信息可被导出成 Endnote 导入格式，bibtex 等著名的引用格式。

目前，WebLab 在“my literature”中提供获取，删除和导出引用信息等功能。



The screenshot shows the WebLab interface. At the top, there is a banner with the CBI logo and the text "your lab on the web" and "WebLab". On the right side of the banner, there are links for Feedback, News, Contact, and Help. Below the banner, there is a navigation menu on the left with categories like Service, User Space, and Account. The main content area shows a file list under "My Literature". One file, "BioCaster detecting public health rumors with a Web-based text mining system.pdf", has a "Fetch Citation" button highlighted in red. Below the file list, there are several action buttons, including "Export Citation", which is also highlighted in red. At the bottom of the page, there is a copyright notice: "Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved."

对于没有引用信息的文献，用户可提供 PubMed id 或文献名通过“fetch citation”功能从 PubMed 中获取引用信息。WebLab 将会把获取的引用信息保存成中间格式。



The screenshot shows the WebLab interface with the "Fetch Citation" form. The form has two sections: "through title" and "through pubmed ID". The "through title" section has a text input field with the text "BioCaster detecting public health rumors with a Web-based text mining system" and a "submit" button. The "through pubmed ID" section has a text input field with the text "Pubmed ID (like 11748933):" and a "submit" button. At the bottom of the page, there is a copyright notice: "Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved."

用户可通过“view citation”来浏览引用信息。

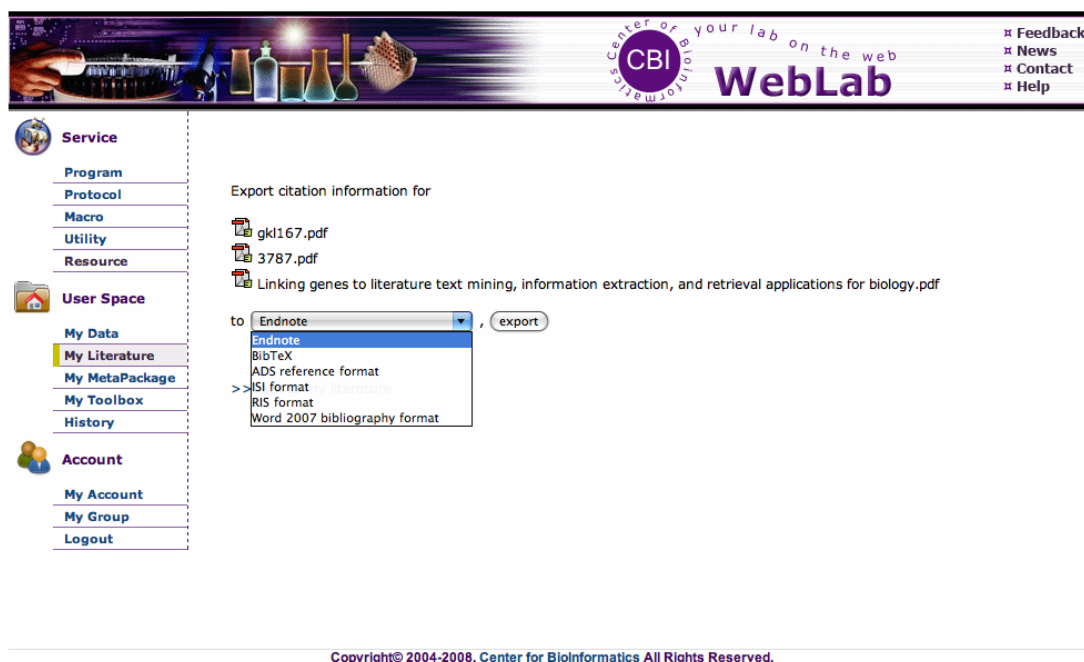
注：WebLab 只显示标题，杂志名，出版日期，卷，期，页，摘要，PubMed ID 等信息而不是所以的引用信息。



The screenshot shows the WebLab interface with a navigation menu on the left and a main content area. The main content area displays citation information for a PDF file named '3787.pdf'. The citation is: 'Cryobiology.2001-Jun;42(4):244-255.' Below the citation, the title of the article is 'Is cryopreservation a homogeneous process? Ultrastructure and motility of untreated, prefreezing, and postthawed spermatozoa of Diplodus puntazzo (Cetti)'. The authors listed are R Taddei, F Barbato, L Abelli, S Canese, F Moretti, J Rana, M Fausto, and M Mazzini. The abstract text follows, describing the study's methodology and findings. The PMID is 11748933.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

用户可以通过单个（每篇文献下面的按钮）和批量（整个文献空间下面的按钮）方式来导出文献。用户可选择导出的格式。目前 WebLab 支持 Endnote, BibTeX, ADS 文献格式, ISI 格式, RIS 格式和 word2007 参考书目格式。



The screenshot shows the WebLab interface with a navigation menu on the left and a main content area. The main content area displays the 'Export citation information for' dialog. The dialog lists three PDF files: 'gk1167.pdf', '3787.pdf', and 'Linking genes to literature text mining, information extraction, and retrieval applications for biology.pdf'. Below the list, there is a dropdown menu for selecting the export format, with 'Endnote' selected. An 'export' button is visible next to the dropdown menu.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

索引和检索: 在文献上传的过程中, WebLab 将会提取文献内容建立全文索引并建立标题, 作者, 杂志, 出版日期等信息的索引。

目前, WebLab 使用 Lucene 作为文献库的索引和恢复。

注: 有时, WebLab 不能提取全文信息和引用信息, 此时则不能建立索引 (WebLab 会尽力给出出错的原因)。例如, WebLab 不能对加密的 pdf 文件建立索引。WebLab 使用不同的图标来标记文献的索引状态, 粉红色表示该文献的全文或引用信息没建立索引。



对于已成功建立索引的文献，用户通过简单检索和高级检索来搜索相关信息。简单检索只能按关键词对文献全文搜索。

The screenshot shows the WebLab interface. At the top right, there are navigation links: Feedback, News, Contact, and Help. The main header features the CBI logo and the text 'your lab on the web' and 'WebLab'. On the left, there is a sidebar with categories: Service, Program, Protocol, Macro, Utility, Resource, User Space, My Data, My Literature (highlighted), My MetaPackage, My Toolbox, History, Account, My Account, My Group, and Logout. The main content area has a search bar with a 'GO' button and a link to 'Advanced Search'. Below the search bar, there is a list of search results under the folder 'My Literature'. The results include titles like 'Text mining for biology - the way forward opinions from leading scientists.pdf' and 'BioCaster detecting public health rumors with a Web-based text mining system.pdf'. At the bottom of the results area, there are several action buttons: New Directory, Upload, Delete, Move, Rename, Download, Share, Cancel Share, Import, Cancel Import, Export Citation, and Refresh.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

对于高级检索，用户可对全文和引用信息进行复杂的检索。

The screenshot shows the advanced search interface. It features a 'Full Text Search' section with a 'Full Text:' input field containing 'text mining' and a dropdown menu set to 'Any Words'. Below this is a 'Refine Search' section with fields for 'Title:', 'Authors:', and 'Journal:', each with a dropdown menu set to 'Any Words'. The 'Publish Date:' section includes a 'Specify a date range' dropdown, 'from' and 'To' labels, and dropdown menus for months and years (e.g., April 1988 to January 2008). At the bottom, there is a 'Under the directory' dropdown set to 'account share' and a 'search' button.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

搜索的结果如下图所示，检索的关键词被高亮显示。每篇文献的得分是指匹配得分。得分高的文献显示在前面。

Center of Bioinformatics
your lab on the web
WebLab

- Feedback
- News
- Contact
- Help

Service

- Program
- Protocol
- Macro
- Utility
- Resource

User Space

- My Data
- My Literature**
- My MetaPackage
- My Toolbox
- History

Account

- My Account
- My Group
- Logout

Under the directory [Literature Home/account share](#), using search expression `+body:(text mining)+publishdate:[1988-04 TO 2008-01]`, you get the following search results:

total count: 4 >>Advanced Search PageSize: 10

- [Linking genes to literature text mining, information extraction, and retrieval applications for biology.pdf](#) score:0.65855014

*Fulltext: ...Open Access 2008Krallinger et al.Volume 9, Suppl 2, Article S8 Review Linking genes to literature: **text mining**, information extraction, and retrieval applications for biology Martin Krallinger 1, Alfonso Valencia 1 and Lynette Hirschman 2 Addresses: 1 Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), C/Melchor F. Almagro, 3, E-28029 Madrid, Spain...

*Publish Date: **2006-07-01**
- [BioRAT extracting biological information from full-length papers.pdf](#) score:0.64416313

*Fulltext: In this paper, we present BioRAT, a new information extraction (IE) tool, specifically designed to perform biomedical IE, and which is able to locate and analyse both abstracts and full-length papers. BioRAT is a Biological Research Assistant for **Text mining**, and incorporates a document search ability with domain-specific IE. Results: We show first, that BioRAT performs as well as existing systems...

*Publish Date: **2004-11-22**
- [3787.pdf](#) score:0.56566197

*Fulltext: et al. (2004) The complete genome sequence of *Propionibacterium acnes*, a commensal of human skin. *Science*, 305, 671?673. Castillo-Davis,C.I. and Hart,D.L. (2003) GeneMerge?post-genomic analysis, data **mining**, and hypothesis testing. *Bioinformatics*, 19, 891?892. Cheng,J. et al. (2004a) NetAffx gene ontology **mining** tool: a visual approach for microarray data analysis. *Bioinformatics*, 20, 1462?1463...

*Publish Date: **2001-06**
- [gkl167.pdf](#) score:0.5652194

*Fulltext:, Diaz-Uriarte,R. and Dopazo,J. (2004) FatGO: a web tool for finding significant associations of Gene Ontology terms with groups of genes. *Bioinformatics*, 20, 578?580. 5. Masseroli,M., Martucci,D. and Pinciroli,F. (2004) GFINDER: Genome Function INtegrated Discoverer through dynamic annotation, statistical analysis, and **mining**. *Nucleic Acids Res.*, 32, W293?W300. 6. Beissbarth,T. and Speed,T.P....

*Publish Date: **2006-07-01**

1

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

HTML 预览: 对于存储在“[My literature](#)”中的每篇文献，WebLab 提供 HTML 预览功能便于浏览文献的内容。

Center of Bioinformatics
your lab on the web
WebLab

- Feedback
- News
- Contact
- Help

Service

- Program
- Protocol
- Macro
- Utility
- Resource

User Space

- My Data
- My Literature**
- My MetaPackage
- My Toolbox
- History

Account

- My Account
- My Group
- Logout

Edit Comment Delete Comment Edit Tag Attach Tag Search: GO >>Advanced Search

+Expand all -Collapse all

- My Literature
 - [Text mining for biology - the way forward opinions from leading scientists.pdf](#)
HTML_View | View Citation | Delete Citation | Export Citation
 - [BioCaster detecting public health rumors with a Web-based text mining system.pdf](#)
HTML_View | Fetch Citation
 - account share
 - [gkl167.pdf](#)
HTML_View | View Citation | Delete Citation | Export Citation
 - [3787.pdf](#)
HTML_View | View Citation | Delete Citation | Export Citation
 - [Linking genes to literature text mining, information extraction, and retrieval applications for biology.pdf](#)
HTML_View | View Citation | Delete Citation | Export Citation

New Directory Upload Delete Move Rename Download
Share Cancel Share Import Cancel Import Export Citation Refresh

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

Preview for automatic extraction of gene protein biological functions from biomedical text.pdf

BIOINFORMATICS ORIGINAL PAPER Vol. 21 no. 7 2005, pages 1227
doi:10.1093/bioinformatics/btl

Data and text mining

Automatic extraction of gene/protein biological functions from biomedical text

Asako Koike^{1,2}, Yoshiki Niwa² and Toshihisa Takagi¹

¹Department of Computational Biology, Graduate School of Frontier Science, The University of Tokyo, Kiban-3A1(CB01) 5-1-5, Kashiwanoha Kashiwa, Chiba 277-8561, Japan and ²Central Research Laboratory, Hitachi Ltd., 1-280 Higashi-koigakubo, Kokubunji City, Tokyo 185-8601, Japan

Received on April 16, 2004; revised on September 21, 2004; accepted on October 5, 2004
Advance Access publication October 27, 2004

ABSTRACT (Fukuda et al., 1998; Collier et al., 2000; Tanabe and W
Motivation: With the rapid advancement of biomedical science and
2002), protein-protein interactions (Friedman et al., 2001; Koike
2002).

3). My Meta Package

“Meta Package” 是用户自己的数据库。用户可在上面建立自己的数据库。目前我们提高两种类型的程序来生成数据库: `reverse_ePCR_db_builder` 和 `formatdb`。然后用户可以 blast 他们自己的数据库。这使得实验变得更加的灵活。

注: 绿色箭头链向可产生这种数据库的程序。另外, 每一个 meta-package 文件不仅仅是一个文献, 而是一个压缩包。用户可下载并解压缩来得到这些文件。

My MetaPackage

- blastdb
 - hba7 weblab.metadata:blastdb:protein
 - hbb weblab.metadata:blastdb:protein
- reElectronicPCRdb
 - rat weblab.metadata:reElectronicPCRdb

New Directory Delete Move Rename

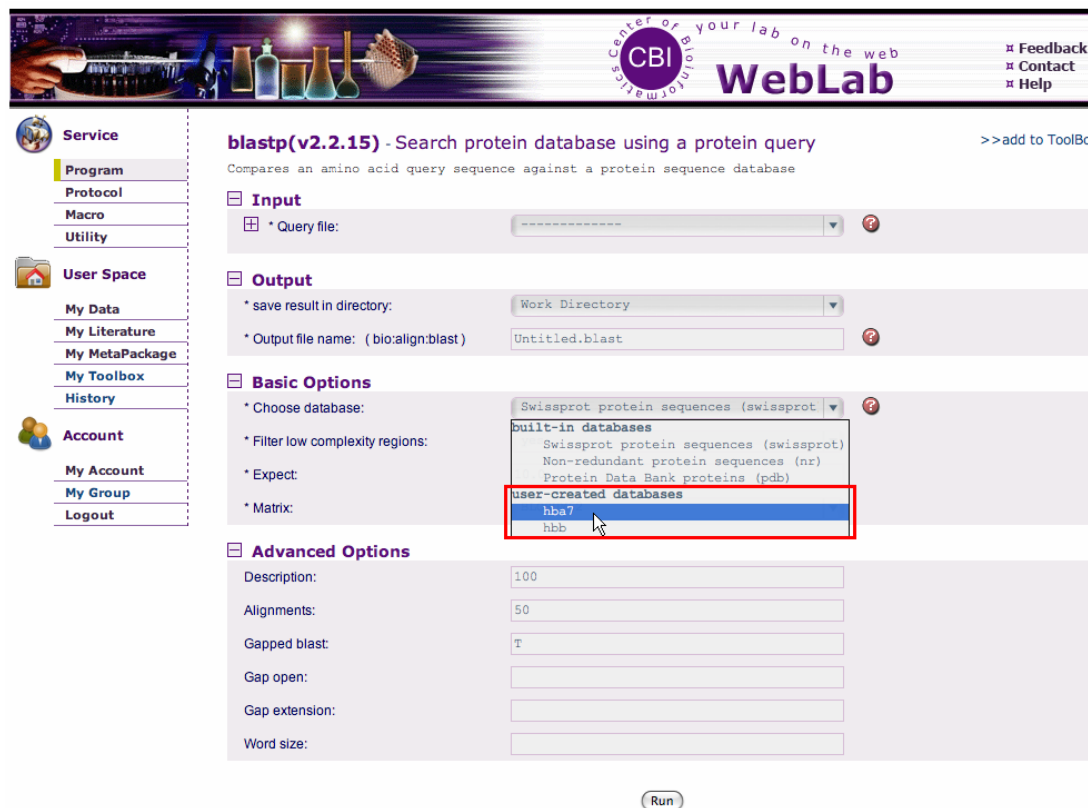
Share Cancel Share Import Cancel Import Refresh

生成 Meta Package: 在 Meta Package 中有两个目录: `blastdb` 和 `reElectronicPCRdb`。程序 `Formatdb` 用来对蛋白和核酸进行格式化得到 `blast` 数据库。`reverse_ePCR_db_builder` 用来生成要求的数据库文件以通过 STS 来检索序列数据库。在这里我们以 `reverse_ePCR_db_builder` 为例。

注: 如果程序可生成 Meta-package, 则输出路径将自动设为 “Meta package”。



使用 Meta Package: 当用户运行相关的程序时可选用自己的数据库。比如假如我生成了一个蛋白数据库，那么我在运行 `blastp` 程序时可以从用户创建的数据库中选择该数据库。通过该功能，用户可在他们的实验中选择更加合适的数据库。



4. 如何在 WebLab 上工作

1). Utility——一些有用的工具

“Utility”是一些用于生物信息学分析的小部件。这些“Utility”主要由 javascript, Java 插件和其它基于浏览器的客户端程序组成。它们一般运行的时间不是太长不过非常有用。截至目前,在 WebLab 中共有 66 “utilities”,其中大部分都集中在序列编辑,数据可视化等方面。

注:在运行 WebLab 上的“utility”之前,请设置浏览器允许运行 Java 和 JavaScript,并要安装 Java 运行环境。

2). Program——构造元件

WebLab 中的程序包含三种不同类型的服务:

- 本地程序:这类程序在 WebLab 的计算节点上运行,其中大部分程序来自 EMBOSS, NCBI BLAST 等最通用的生物信息分析软件包。

- Web 服务程序:现在 Web 服务在生物信息领域变得越来越热门,因此 WebLab 也为一些著名的生物信息的网站的 web 服务提供了客户端程序。通过我们提供的 web 界面你可以向远程 web 服务器提交任务。

- 基于网格的服务程序:WebLab 也提供一些基于网格的程序来处理像 blast 这样的计算和输入输出密集型服务。

到目前为止,WebLab 上有 264 个程序。用户可通过它们的软件包,功能及字母顺序进行浏览。

程序的运行: WebLab 为不同的程序服务提供了统一的符合感官的界面便于用户的使用。

在程序提交页面,具有相似功能的参数聚成一组。在提交任务以前,用户需要注意下面几个方面:

- 对于输入文件,未注册用户可通过上传文件或直接贴入数据,而注册用户还可以通过“my data”来输入数据。

- 注册用户需要指定输出的路径。

- 带*号的参数为必须填的项目。

- 如果你不清楚那个参数的意义,可以通过点击右面的“?”获得帮助。

- 如果你想知道程序的具体细节,请点击程序名查看程序文档。

在为每个参数输入一个恰当的值后,你就可以点击“Run”来提交任务了。

clustalw2-msa - Multiple Sequence Alignment function of clustalw2 >> add to ToolBox

Clustal W is a general purpose multiple alignment program for DNA or proteins. clustalw2-msa is used to do the multiple sequence alignment task.

01 Input

* input sequences: [dropdown] **04** ?

OR upload file from local disk: [Choose File] no file selected

OR paste into window: [text area]

file for old guide tree: [dropdown] ?

Comparison matrix file (optional) for pairwise alignment: [dropdown] ?

Comparison matrix file (optional) for multiple alignment: [dropdown] ?

02 Output

* save result in directory: [Work Directory]

03 * sequence alignment file name: (bio:align:emboss.msf) [Untitled.msf] ?

Dendrogram (tree file) from clustalw output file: (bio:report:dendrogram) [Untitled.dendrogram] ?

General settings

- fast pairwise alignments
- slow pairwise alignments
- multiple alignments

Run

当你提交任务时，前端服务器会查找合适的计算节点然后向其发送请求。在接到请求后，后台服务器便会运行要求的程序。当分析完成后，结果被送回前端服务器。

一个任务可以有五个状态。通过这些工作状态你可以控制在 WebLab 上运行的任务。

当程序顺利跑完后，你可以在“my data”中看到结果文件。当然你可以在“result”的下拉菜单中选择不同的结果浏览。

3). 工作流（Protocol 和 macro）

- 什么是工作流

生物信息分析是要涉及大量生物信息软件的复杂工作。传统的工具平台缺乏不同软件间的联系和互用，用户不得不在不同的软件间传递数据并一个个的运行。而 WebLab 提供的工作流机制可自动的根据一系列规则将相关的软件组合在一起。

- WebLab 上的两种工作流

根据交互能力的不同，WebLab 将工作流分为两种类型。

a). Protocol: 在运行过程中需要用户的介入，它给了用户改变程序参数和执行顺序的机会。

b). Macro: 它可自动的执行，大大简化了分析工作流程。在输入命名参数后，无需用户介入便可执行玩整个工作流程。

- WebLab 上的工作流模型

workflows 模型在服务器的文件系统中是以 XML 格式表示的，在内存中是以 DAG（有向无环图）。在 DAG 中，矩形节点表示一个程序或 macro——这个表示 WebLab 支持定义递归 workflow，你可以使用 macro 作为其它 workflow 的节点。菱形节点表示算子，边表示数据流。算子控制着 workflow 的执行顺序。目前在模型中有两种类型的算子。

a). 选择算子：这类算子将更加输入条件判断下一步要执行的 analysis。

注：目前在 WebLab 中有一种类型的选择算子——序列类型选择算子，通过该算子判断序列是蛋白序列还是核酸序列，然后选择不同的 analysis 步骤。

选择算子由选择分支与选择终止组成。分支算子将 workflow 分成不同的分支，当然只有其中的一个分支被选择执行。选择终止算子将不同的分支合为一支。

b). 平行算子：该算子可以使得几个程序同时分析同一输入。这对 analysis 具有相同功能的不同软件的结果非常有用。

平行算子由平行分支和平行终止组成。平行分支让输入数据流向几个相应的程序。平行终止将几个数据流合在一起。

- 预定义 workflow

WebLab 已定义好了几个 workflow，包括 DNA 分析，蛋白分析，系统发育分析和代谢通路识别等以方便用户的使用。

workflow 的运行：

- Protocol 的运行

下面的例子将示范如何在 WebLab 上运行 workflow。例中的 workflow 仅包含两个程序。第一个为 emma，一个多序列比对的程序。第二个为 cons，它将从 emma 生成的多序列比对中提取保守区域。原则上，你可以点击任何程序来运行，但一般来说，你应该点击起始程序（没有父程序的程序），在该例中通过 emma 来启动整个 workflow。

注：在每个 workflow 中有且仅有一个起始程序。

a). 工作流程序的提交界面和单个程序的界面类似。唯一的不同就是你可以在页面上半部看到 workflow 的 DAG。在填写了合适的参数后，点运行来提交工作流程序。

demonstration_protocol

```

graph TD
    input((input)) --> emma[1-emma(v5.0)]
    emma --> cons[2-cons(v5.0)]
    cons --> output((output))
  
```

emma(v5.0) - Multiple alignment program - interface to ClustalW program >> add to ToolBox

Input

- * sequence: |-human_hba.fasta
- Dendrogram (tree file) from clustalw file (optional):
- Comparison matrix file (optional):
- Comparison matrix file (optional):

Output

Basic Options

Advanced Options

Run

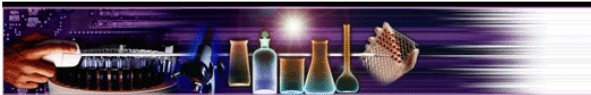
b). 你将会看到下面的界面，点击超链接可以看到工作流的运行状态。


You can view the status of protocol from

<http://weblab0.cbi.pku.edu.cn/protocol.status.do?protocolID=3801>

or from "History" menu.

c). 你会看到表示 emma 的那个矩形填充成粉红色，这表示这个程序以成功运行完。如果程序正在运行，在矩形为橄榄绿，如果出错，则为红色。你可以在下面的表格中查看提交程序的详细运行信息。





your lab on the web

WebLab

- Feedback
- Contact
- Help

Service

- Program
- Protocol
- Macro
- Utility

User Space

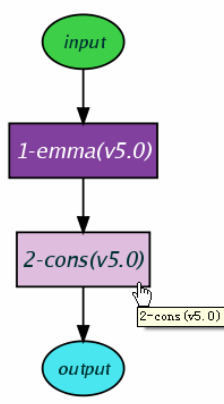
- My Data
- My Literature
- My MetaPackage
- My Toolbox
- History

Account

- My Account
- My Group
- Logout

demonstration_protocol

this mini protocol include only two activites to demonstrate how to run protocol



```

graph TD
    input((input)) --> step1[1-emma(v5.0)]
    step1 --> step2[2-cons(v5.0)]
    step2 --> output((output))
            
```

ID	PROGRAM	STATUS	INPUT FILE	OUTPUT FILE	START TIME	RUN TIME	ERROR MESSAGE
56320	emma(v5.0) (Multiple alignment program - interface to ClustalW program)	Finished	hba15.fasta	Untitled.dendrogram.2008-03-08 20:53 PM Untitled.msf.2008-03-08 20:53 PM	2008-03-08 20:53 PM	1.0s	

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

根据上面的步骤你可以选择下一个要执行的程序直到 Protocol 结束。要说明的是在 Protocol 执行的过程中，你可以调整每个程序的参数。

• Macro 的运行

运行 macro 的过程与运行 Protocol 的过程完全不同。唯一的相似点就是在运行 Protocol 中的 a), b), c) 步骤仍然是必要的。不过，有一些不同的地方需要注意。

- 在 macro 中，只有一个提交页面。因此在 macro 运行前需要填入所有必须的参数。
- 一旦 macro 被提交，内嵌的工作流引擎将根据相应的 DAG 执行 macro 中定义的所有程序。
- 在你首次进入查看 macro 运行状态的页面时，在大多数情况下，你的 macro 还没结束。一般都会有几个程序正在运行。在所有的程序执行完或出错结束前，你需要不断的载入该页面查看 macro 的状态。

Macro 最显著的特征是你只需提交一个无需干涉便可运行完整个工作流。

用户自定义工作流

虽然 WebLab 提供了一些预定义的 Protocol 和 Macro，但这对于实际的分析工作来说是远远不够的。因此，WebLab 提供了让注册用户自定义 Protocol 和 Macro 的机制。

注：WebLab 将引导你定义一个正确能运行的工作流。但你要保证你的工作流是有效的，这是 WebLab 所不能担保的。

• 定义工作流

注册用户即可定义 Protocol 又可定义 Macro。定义的过程是有所不同的。下面是一个构建 Macro 的例子。因为构建 Macro 比构建 Protocol 相对复杂些，所以当看完这个例子后你应该就会构建 Protocol 了。

首先，点击“create your Macros”按钮来开始构建工作流。



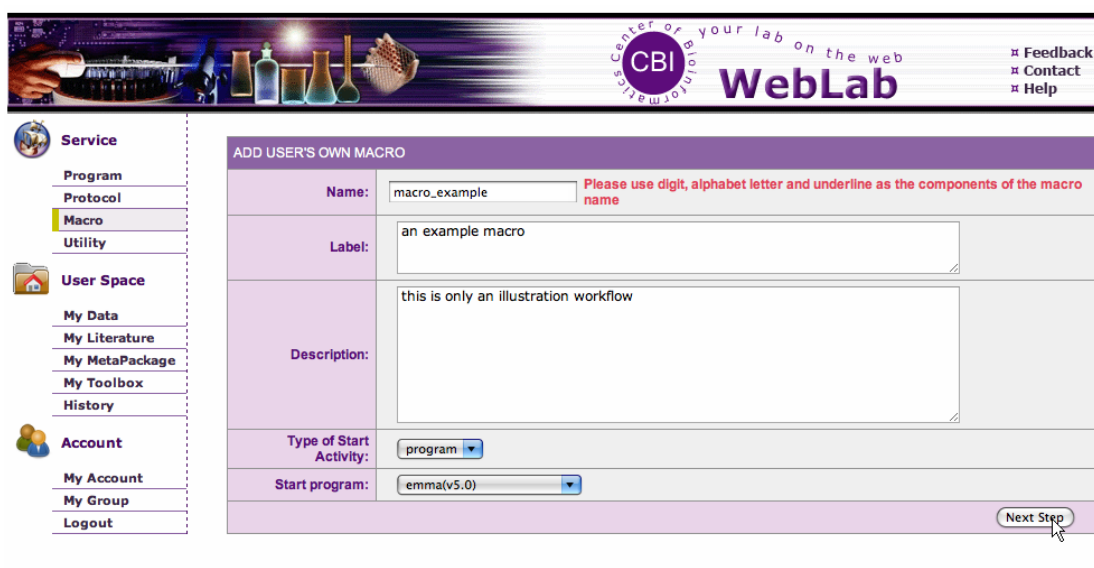
The screenshot shows the WebLab interface with a sidebar on the left containing navigation options like Service, User Space, and Account. The main content area displays a table of pre-defined macros. A red box highlights the 'Create Your Macros' button in the top right corner.

MACRO NAME	DESCRIPTION
ANDJoinerTest (a demo workflow to test ANDJoiner)	this workflow is used to test the workflow which can have ANDJoiner operator
demo (just a demo workflow)	this workflow is used to test the workflow which can have more than one input
profunction (Protein function analysis based on sequence similarity)	This macro is focused on prediction of protein functions based on sequence similarity, where you have a gene sequence or several one in hand. You may begin with this protocol from nucleotide sequence(s), either cDNAs/ESTs or genomic DNA, as long as there will be protein coding region(s) existed. You will follow family identification, sensitive sites finding, and distant homologue search processes. Each step will be performed by well-known software or programs. The flow chart will draw a sketch for you to understand it.
MSAnalysis (Multiple sequences analysis workflow)	This workflow is designed to do some simple analysis for multiple sequences. At first, the input sequences will be aligned using emma. Then, weblogo, cons and ealstat will give you some statistical or intuitive information for the aligned sequences. Besides, ehmmbuild will build a HMM model and at last two incidental programs will run to generate sequences based on this model and calibrate the model.
phylogenetic (Construct phylogenetic tree by Neighbor-Joining method)	This workflow is focused on constructing phylogenetic tree by Neighbor-Joining method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, distance of each sequence is calculated from result of multiple sequence alignment. Then, the phylogenetic tree is built using Neighbor-Joining method. Finally, the strict consensus tree is calculated.

(1). 构建 Macro/Protocol 的第一步是要填写一些关于要构建的工作流的基本信息，包括名字，标签（简短的描述），详细的描述。

(2). 选择开始程序，工作流（系统预定义的或自己已定义的）或算子。在该例中，我们选择程序 emma。

注：Macro/protocol 的名字由数字，字母和下划线来构成。



The screenshot shows the 'ADD USER'S OWN MACRO' form in the WebLab interface. The form fields are filled with example data: Name: macro_example, Label: an example macro, Description: this is only an illustration workflow, Type of Start Activity: program, Start program: emma(v5.0). A 'Next Step' button is visible at the bottom right.

(3). 点击“next step”，设置所选程序的参数。你可以看到你正在构建的 Macro/Protocol 当前的 DAG。

在 Macro 执行前，通过复选框选择你想调整的参数，也可对参数默认值进行修改。

注：对于没有默认值的必须参数，WebLab 将替你选中。因为假如没有这些参数，Macro 不能正确的运行。

通过点击“save parameters and continue”接着构建工作流，点击“finish activity selection”完成创建。现在我们选中“save parameters and continue”。

Please select the parameters which you want to justify before submitting the macro, and then provide the default value. For the mandatory parameters having no default value, we have checked the boxes for you and recommend that you hold them; otherwise your macro may not work normally during the execution.

emma(v5.0) - Multiple alignment program - interface to ClustaW program

Basic Options

- * output format: Clustal
- * Do alignment using an old dendrogram: N
- * Protein pairwise alignment matrix: blosum
- * DNA pairwise alignment matrix: iub
- * Protein multiple alignment matrix: blosum
- * Nucleotide multiple alignment matrix: iub
- * Slow pairwise alignment: gap opening penalty: 10.0
- * Slow pairwise alignment: gap extension penalty: 0.1
- * Fast pairwise alignment: similarity scores: K-Tuple size: 1
- * Fast pairwise alignment: similarity scores: gap penalty: 3
- * Fast pairwise alignment: similarity scores: number of diagonals to be considered: 5
- * Fast pairwise alignment: similarity scores: diagonal window size: 5
- * Fast pairwise alignment: similarity scores: suppresses percentage score: N
- * No residue specific gaps: N
- * List of hydrophilic residues: GPSNDQEKRR
- * No hydrophilic gaps: N

Advanced Options

save parameters and continue finish activity selection

1-emma(v5.0)

(4). 现在你进入了扩展工作流的页面。在此你也能看到你当前构建工作流的 DAG。在页面的上面，选择工作流中已有的一个源程序和一个目标程序。目标程序可为新程序或一个结束算子来形成一个新的指向路径。在页面的下面，可选择去除工作流中已有的程序节点。

注：你无法删除有后子节点的程序节点。

Center of your lab on the web
CBI WebLab
Feedback
Contact
Help

MACRO: AN EXAMPLE MACRO

From: Activity selected: 1-emma(v5.0)

Target: Activity which will accept emma(v5.0) output as input: cons(v5.0)

Operator which is already in the macro: -----

Add it

1-emma(v5.0)

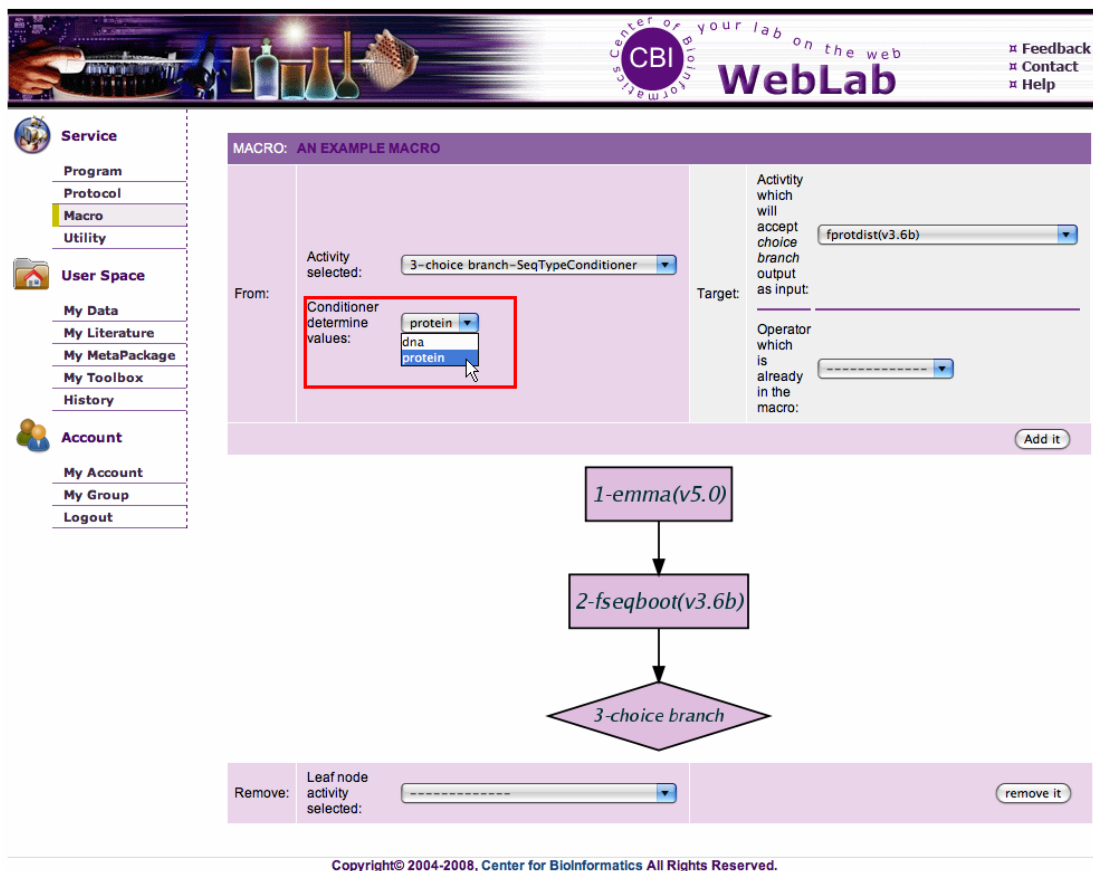
Remove: Leaf node activity selected: ----- remove it

多注意一下这一步。

注：WebLab 将对子节点能否接受父节点的输出做基本的判断。如果子节点的输入与父节点的输出矛盾，下拉菜单的该程序被标为红色。不过，如果你认为这个流程是合理的，请在选择这个标为矛盾的程序时不要犹豫。



注：当选择分支作为父节点时，你需要确定一个判断值以运行不同的程序。对 SeqTypeConditioner 来说，判断值是“protein”和“dna”。



然后点击“add it”，进入新选择程序的参数设置页面。重复上面的两个步骤知道完成工作流的构建。

对 Protocol 来说，点击“finish Protocol”完成 Protocol 的构建。

对 Macro，在点击“finish activity selection”后，还有重要的事要做。

Please select the parameters which you want to justify before submitting the macro, and then provide the default value. For the mandatory parameters having no default value, we have checked the boxes for you and recommend that you hold them; otherwise your macro may not work normally during the execution.

fconsense(v3.6b) - Majority-rule and strict consensus tree

Output section

Additional section

save parameters and continue

finish activity selection

```

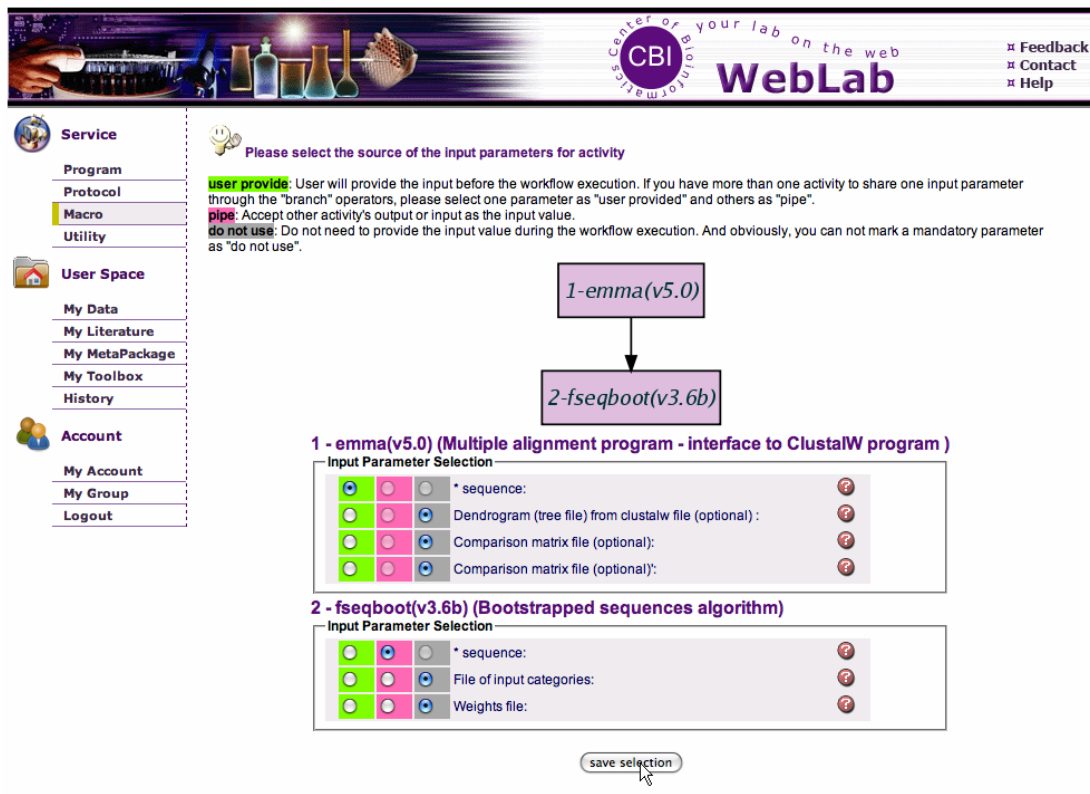
graph TD
    A[1-emma(v5.0)] --> B[2-fseqboot(v3.6b)]
    B --> C{3-choice branch}
    C --> D[5-fdnadist(v3.6b)]
    C --> E[4-fprotdist(v3.6b)]
    D --> F{6-choice end}
    E --> F
    F --> G[7-fneighbor(v3.6b)]
    G --> H[8-fconsense(v3.6b)]
  
```

(5). 为了每一个 Macro 程序在运行中都能得到恰当的输入，你需要为每个程序选择输入参数的类型。这里有三种输入参数的类型。

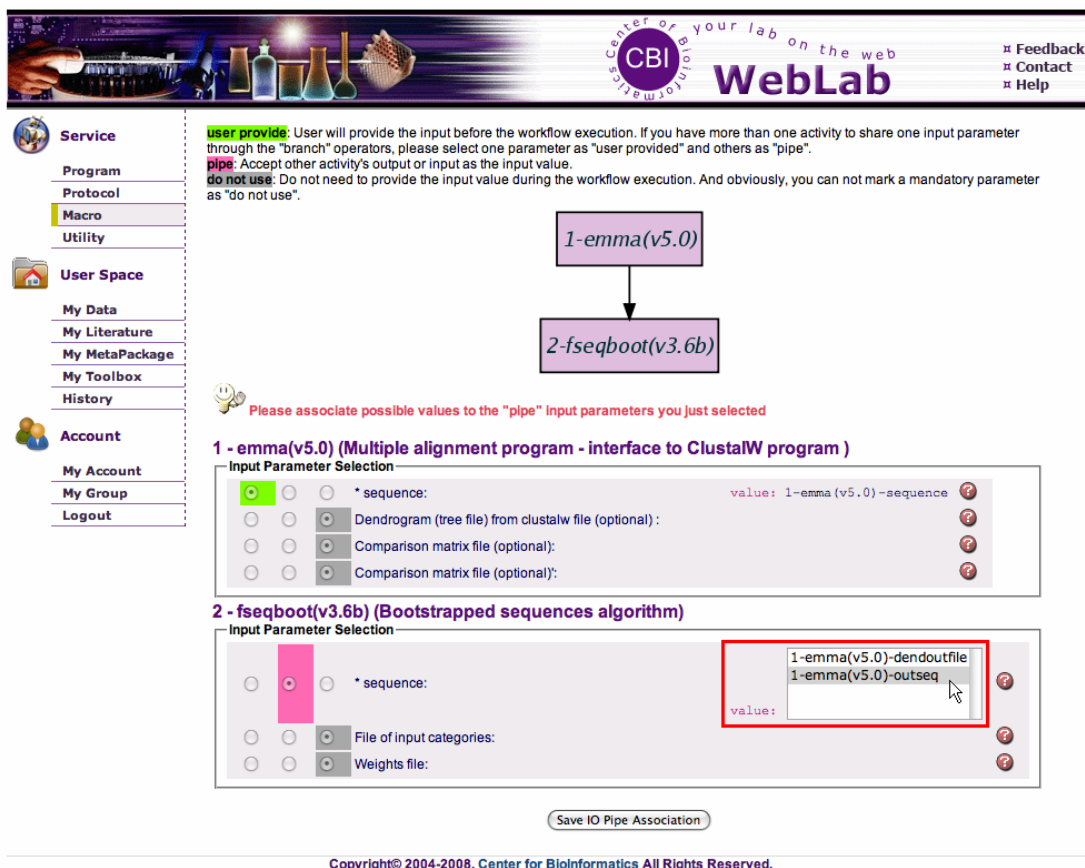
- **user provided:** 在提交 Macro 时，用户需提供该类型参数的值。也就是说这些参数是 Macro 的全局输入参数。
- **pipe:** 这种类型的参数或者接受其它程序的输入或者使用“user provided”参数。在下部中 WebLab 就会引导你为 pipe 类型的参数指派数据源。
- **do not use:** 在 Macro 执行中，这些参数是可以不用的。一般来说，这些参数不是必须参数。

注：WebLab 给每一个输入参数都指定了一个默认类型。

指定完参数类型后，点“save selection”按钮。



(6). 现在进入 Macro 构建的最后一步。在本步中，为 pipe 类型的参数指定数据源。
 注：WebLab 为每个 pipe 参数都指定了一个默认的数据源。
 现在，点击“Save IO Pipe Association”完成 Macro 的构建。



- 删除和编辑 workflow

你可以通过点击“remove”来删除相应的用户自定义的 Protocol 或 Macro。而且你还可以通过“edit”按钮来编辑相应的工作流。



Center of your lab on the web
CBI WebLab

- Feedback
- Contact
- Help

Service

- Program
- Protocol
- Macro
- Utility

User Space

- My Data
- My Literature
- My MetaPackage
- My Toolbox
- History

Account

- My Account
- My Group
- Logout

Pre-defined Macros Create Your Macros

MACRO NAME	DESCRIPTION
ANDJoinerTest (a demo workflow to test ANDJoiner)	this workflow is used to test the workflow which can have ANDJoiner operator
demo (just a demo workflow)	this workflow is used to test the workflow which can have more than one input
profuction (Protein function analysis based on sequence similarity)	This macro is focused on prediction of protein functions based on sequence similarity, where you have a gene sequence or several one in hand. You may begin with this protocol from nucleotide sequence(s), either cDNAs/ESTs or genomic DNA, as long as there will be protein coding region(s) existed. You will follow family identification, sensitive sites finding, and distant homologue search processes. Each step will be performed by well-known software or programs. The flow chart will draw a sketch for you to understand it.
MSAnalysis (Multiple sequences analysis workflow)	This workflow is designed to do some simple analysis for multiple sequences. At first, the input sequences will be aligned using emma. Then, weblogo, cons and ealstat will give you some statistical or intuitive information for the aligned sequences. Besides, ehmmbuild will build a HMM model and at last two incidental programs will run to generate sequences based on this model and calibrate the model.
phylogenetic (Construct phylogenetic tree by Neighbor-Joining method)	This workflow is focused on constructing phylogenetic tree by Neighbor-Joining method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, distance of each sequence is calculated from result of multiple sequence alignment. Then, the phylogenetic tree is built using Neighbor-Joining method. Finally, the strict consensus tree is calculated.
EmbedMacro (a demo workflow which is used to test embedded macro)	this workflow is used to test the workflow which has other macros as component activity
phylogenetic_max_parsimony (Construct phylogenetic tree by maximum parsimony method)	This workflow is focused on constructing phylogenetic tree by maximum-parsimony method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, the phylogenetic tree is built using maximum parsimony method. Finally, the strict consensus tree is calculated.

User Customized Macros Distribute Macros Cancel Distribute Macros Import Macros

MACRO NAME	DESCRIPTION	ACTION
<input type="checkbox"/> example_macro (an example macro)	this macro is an illustration workflow	Remove Edit

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

编辑过程是 workflow 创建过程的一个延伸。你可以在此通过“remove it”删除特定的节点和加入新的节点。如果你修改的是 Macro 工作流，在最后还要设置参数的类型及 pipe 参数的数据源。



Center of your lab on the web
CBI WebLab

- Feedback
- Contact
- Help

Service

- Program
- Protocol
- Macro
- Utility

User Space

- My Data
- My Literature
- My MetaPackage
- My Toolbox
- History

Account

- My Account
- My Group
- Logout

MACRO: AN EXAMPLE MACRO

From: Activity selected: 2-fseqboot(v3.6b)

Activity which will accept fseqboot(v3.6b) output as input: [dropdown]

Target: [dropdown]

Operator which is already in the macro: [dropdown]

Add it

```

graph TD
    A[1-emma(v5.0)] --> B[2-fseqboot(v3.6b)]
  
```

Remove: Leaf node activity selected: [dropdown]

remove it

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

4). 任务控制——你是你的任务的主人！

任务状态

一个任务可以是如下五个状态之一：

- submitted: 当用户输入参数发送请求运行程序，一个任务被创建被设为“submitted”状态。
- running: 当通过参数验证并存入数据库，请求被推进入任务调度，任务的状态改为“running”。
- in queue: 如果拥有请求程序的所有节点都在高负荷运行，任务状态设为“in queue”。
- finished: 任务顺利完成，结果存入“my data”，状态改为“finished”。
- failed: 如果由于某些原因造成任务失败或非正常退出。状态改为“failed”并记录和显示详细的警告信息。

任务历史

注册用户可以在“History”中监视提交任务的最新状态。

The screenshot shows the WebLab interface. The 'History' tab is selected in the left navigation menu. The main content area displays a table of task history with the following data:

WORKFLOW	JOB ID	SERVICE	STATUS	INPUT FILE	OUTPUT FILE	START TIME	RUN TIME	ERROR MESSAGE
	56406	infoseq(v5.0)	Finished	1205061914268.fasta	Untitled.infoseq.2008-03-09 19:25 PM	2008-03-09 19:25 PM	1.0s	
3802 demonstration_protocol	56320	emma(v5.0)	Finished	hba15.fasta	Untitled.dendrogram.2008-03-08 20:53 PM Untitled.msf.2008-03-08 20:53 PM	2008-03-08 20:53 PM	1.0s	
3801 demonstration_protocol	56319	emma(v5.0)	Fail	human_hba.fasta	Untitled.dendrogram.2008-03-08 20:45 PM Untitled.msf.2008-03-08 20:45 PM	2008-03-08 20:45 PM		?
	56235	cons(v5.0)	Finished	abiview.msf	Untitled.fasta	2008-03-08 15:32 PM	1.0s	
	55987	clustalw2- msa	Finished	hba15.fasta	hba.msf hba.dendrogram	2008-03-06 20:42 PM	1.0s	

total count: 1824 PageSize: 5 Order: Start Time Type: Desc

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 >>

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

任务控制

对于一些运行了很长时间的任务，你可以选择杀掉该进程。

注：你最多可同时运行 30 个任务。



total count: 1825 PageSize: 5 Order: Start Time Type: Desc

WORKFLOW	JOB ID	SERVICE	STATUS	INPUT FILE	OUTPUT FILE	START TIME	RUN TIME	ERROR MESSAGE
	56474	blastp(v2.2.15)	Running	human_hba.fasta	Untitled.blast.2008-03-10 14:59 PM	2008-03-10 14:59 PM		
	56406	infoseq(v5.0)	Finished	1205061914268.fasta	Untitled.infoseq.2008-03-09 19:25 PM	2008-03-09 19:25 PM	1.0s	
3802 demonstration_protocol	56320	emma(v5.0)	Finished	hba15.fasta	Untitled.dendrogram.2008-03-08 20:53 PM Untitled.msf.2008-03-08 20:53 PM	2008-03-08 20:53 PM	1.0s	
3801 demonstration_protocol	56319	emma(v5.0)	Fail	human_hba.fasta	Untitled.dendrogram.2008-03-08 20:45 PM Untitled.msf.2008-03-08 20:45 PM	2008-03-08 20:45 PM		?
	56235	cons(v5.0)	Finished	abiview.msf	Untitled.fasta	2008-03-08 15:32 PM	1.0s	

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 > >>

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

5). 我的工具箱

工具箱中存放你最爱的服务或程序。“my toolbox”是一个类似“my data”的用户空间。注册用户可用它来方便的收集和管理 WebLab 中的程序，Protocols, macros 和 utilities。



+Expand all -Collapse all

- My Toolbox
 - analysis
 - tool2
 - isPCR(v33x2) weblab:service:program
 - profuction weblab:service:macro
 - phylogenetic weblab:service:protocol
 - jalview weblab:service:utility

New Directory New Tool Delete Move Rename

使用工具箱

WebLab 上所有服务的提交页面上都有一个“add to Toolbox”按钮。如果你任务这个服务有用，点击此按钮便可将其加入工具箱。



6). 资源——获取源数据

WebLab 提供 resource 模块供用户方便的提取生物数据库中大量的数据。而且，还可将这些数据保存入“my data”中供进一步分析。



提取数据

目前，WebLab 提供提取 BioMart, plantfdb, h1pp 和 SRS 中数据的用户界面。

• BioMart

BioMart 是 OiCR 和 EBI 共同开发的一个数据检索管理系统。

通过 BioMart 获得数据，你需要首先选择要检索的数据库。

BioMart is a query-oriented data management system developed jointly by the Ontario Institute for Cancer Research (OICR) and the European Bioinformatics Institute (EBI). Please select the dataset and start your query.

Dataset

Gene(Ensembl & VEGA)

- aaegypti_gene_ensembl
- agambiae_gene_ensembl
- btaurus_gene_ensembl
- celegans_gene_ensembl
- cfamiliaris_gene_ensembl
- cintestinalis_gene_ensembl
- cporcellus_gene_ensembl
- csavignyi_gene_ensembl
- dmelanogaster_gene_ensembl
- dnovemcinctus_gene_ensembl
- drerio_gene_ensembl
- ecaballus_gene_ensembl
- europaeus_gene_ensembl
- etelfairi_gene_ensembl
- fcatus_gene_ensembl
- gaculeatus_gene_ensembl
- ggallus_gene_ensembl
- hsapiens_gene_ensembl
- lafricana_gene_ensembl
- mdomestica_gene_ensembl

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

然后 BioMart 为你收集相关的属性和过滤。

Dataset

celegans_gene_ensembl

Attribute

Features

GENE:

- Ensembl Gene ID
- Chromosome Name
- Transcript Start (bp)
- External Transcript ID
- Ensembl Peptide length
- Ensembl Transcript ID
- Gene Start (bp)
- Transcript End (bp)
- External Transcript DB
- Transcript count
- Ensembl Peptide ID
- Gene End (bp)
- Gene name
- Ensembl CDS length
- % GC content
- Description
- Strand
- Gene DB
- Ensembl cDNA length
- Biotype

EXTERNAL:

PROTEIN:

GENOMIC REGION:

Structures

Homologs

Sequences

Filter

chromosome_name: V

id_list_filters: -----

only excluded

在选择一些属性和过滤后，BioMart 会给出结果。此外，你可以设置文件名，指定数据格式，选择存储数据的目录。你也可以选择保存检索条件。然后点击“save to my data”，结果就会存入你的数据空间。

Save Data

Your File Name: save the query condition

Specify File Format:

Select the directory you want to save:

Attribute: ensembl_gene_id chromosome_name percentage_gc_content
Filter: chromosome_name = V

Results:

B0348.5	V	51
B0348.6	V	44
B0348.4	V	37
B0348.2	V	33
B0348.1	V	39
Y38C9B.2	V	36
Y38C9B.1	V	33
Y38C9B.3	V	42
F56E10.2	V	38
F56E10.3	V	31
F56E10.4	V	40
F56E10.1	V	42
Y38C9A.1	V	35
Y38C9A.2	V	36
W03F9.1	V	41
W03F9.2	V	40
W03F9.10V	V	39
W03F9.3	V	45

• PlantTFDB

PlantTFDB 是 CBI 开发的一个综合的植物转录因子数据库。检索 PlantTFDB 的过程与检索 BioMart 类似。不同的是在这我们还提供 blast 功能。

Choose your query condition

Field

Gene_Model
 Family
 Domain_ID
 Domain_Description
 Expression
 Protein_Length
 Protein_Sequence
 CDS_Sequence

Filter

Gene_Model:

Family:

Domain_ID:

Domain_Description:

Expression:

Protein_Length:

All_Text:

Do Blast

Program:

Sequence:

Expect_Threshold:

Matrix:

- SRS

序列检索系统 SRS 是世界上第一个整合，分析和显示生物数据的生物信息学工具。在此，我们提供两种检索方式。

- SRS ID/AC 检索

用户需要选择数据库，数据格式，填入感兴趣的序列的编号。每行填一个序列编号。

- SRS 关键字检索

选择数据库和数据格式，输入关键字进行检索。

检索结果如下图所示。用户可选择感兴趣的序列并保存。

ENTRY NAME	PRIMARY ACCESSION	DESCRIPTION	SEQUENCE LENGTH
<input type="checkbox"/> UNIPROT:OLF1_APILI	Q26419	Olfactory receptor-like protein Hba1 (Fragment).	165
<input type="checkbox"/> UNIPROT:Q7Z6G4_HUMAN	Q7Z6G4	HBA2 (Fragment).	31
<input type="checkbox"/> UNIPROT:Q148I4_BOVIN	Q148I4	Similar to brain acyl-CoA hydrolase isoform hBACHa.	338
<input type="checkbox"/> UNIPROT:P95636_RHOPA	P95636	4-hydroxybenzoyl-CoA reductase HbaC subunit.	774
<input type="checkbox"/> UNIPROT:P95635_RHOPA	P95635	4-hydroxybenzoyl-CoA reductase HbaB subunit (4-hydroxybenzoyl-CoA reductase, first of three subunits) (EC 1.3.99.20).	163
<input type="checkbox"/> UNIPROT:P95637_RHOPA	P95637	4-hydroxybenzoyl-CoA reductase HbaD subunit (4-hydroxybenzoyl-CoA reductase, third of three subunits) (EC 1.3.99.20).	327
<input type="checkbox"/> UNIPROT:Q9RFR5_RHOPA	Q9RFR5	Transcriptional activator HbaR (Hydroxybenzoate anaerobic degradation regulatory protein HbaR, Crp/Fnr family).	241
<input checked="" type="checkbox"/> UNIPROT:A7M7S6_MOUSE	A7M7S6	Hemoglobin X, alpha-like embryonic chain in Hba complex (Fragment).	156
<input checked="" type="checkbox"/> UNIPROT:Q78PA4_MOUSE	Q78PA4	Zeta globin (Hemoglobin X, alpha-like embryonic chain in Hba complex).	142
<input checked="" type="checkbox"/> UNIPROT:A8KBN4_DANRE	A8KBN4	Hbae1 protein.	143

5. 和同事一起工作——分组策略及分享机制

1). 概述

生物学者通常需要和合作者分享他们的分析数据，感兴趣的文献和工作经验。为了便于组内的合作研究，WebLab 提供了分享和分组机制。组内成员之间几乎可以方便的分享任何 WebLab 上的东西。

2). 分组——你不是一个人在工作

WebLab 的分组机制

组是一群用户的集合。每一个注册用户可以点击“**My Group**”查看组的状态。你可以检查你所创建组（上面的粉红框内）和参加组（下面的绿框内）的详细信息。

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

创建组

每一个注册用户都有权创建并管理自己的组。点击“**Create Your Group**”进入创建组页面。

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

按照以下步骤创建自己的组：

- 填写组的基本信息：包括组名及组的简短描述。
- 添加组员：输入组员的账号（每次一个），加入组中。在创建组的过程中，你可以动态的修改组员列表。注意，一个组内最多不得超过 15 个成员。

注意：如果你想分享该组的数据和知识请记着把自己加入该组中。

- 点击“save group”完成组的创建。

CREATE NEW GROUP

Name:

Description:

Please input the user name:

Member list: (maximum 15)

liuxq

luojc

zhaosq

kongl

gaoge

wangj

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

组操作

- 作为所创建组的管理员，你可以点击“delete”删除相应的组，或点击“edit”修改组的基本信息和添加删除组员。
- 如果你发现参加的组非常无聊，可以点击“quit”退出该组。

注：组操作会改变 WebLab 系统中分享的数据和知识信息。

User Created Group >>Create Your Group

NAME	DESCRIPTION	MEMBERSHIP	ACTION
test	test group	kongl xiaoqiao wangjun qingwa gaoge alygunia	delete edit
weblab_group	the group is used to share literature in weblab development group	luojc kongl wangj zhaosq liuxq gaoge	delete edit

User Attended Group

NAME	DESCRIPTION	CREATOR	MEMBERSHIP	ACTION
test	test group	qingwa	kongl xiaoqiao wangjun qingwa gaoge alygunia	quit

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

3). 分享

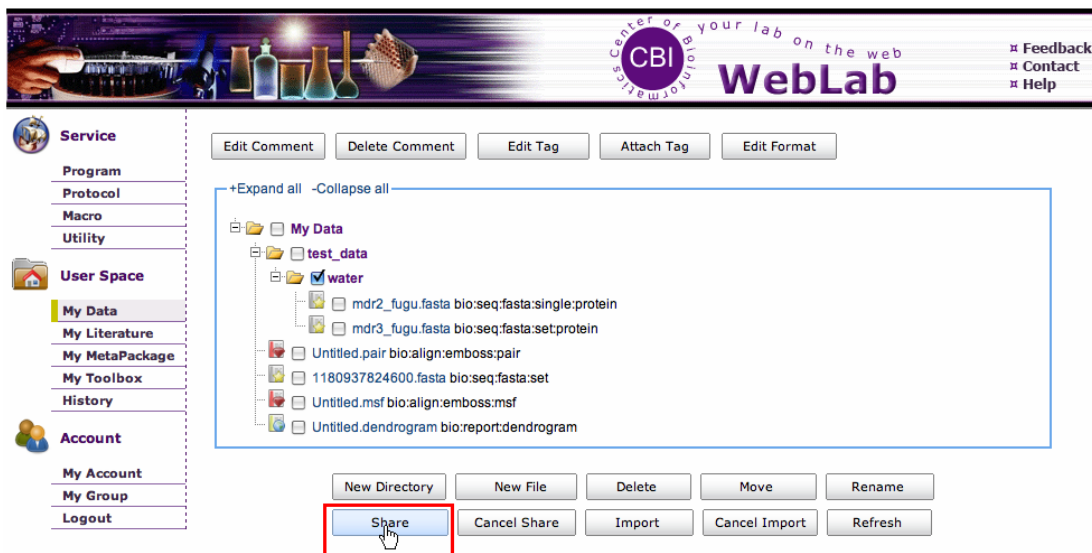
分享数据，文献和 MetaPackage

数据，文献和 MetaPackage 的分享模型是链接分享。在数据库中只存在一分拷贝和他的分享链接信息。分享内容的所有改动都能被提供者和分享者同时看到。

- 如何分享数据，文献和 MetaPackage

选择目录 My Data, my literature 或 my MetaPackage, 然后点 “share” 分享。

注：用户既不能分享根目录，也不能导入根目录。



Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

选择目录及权限（只读，读写）分享给其它组或用户。

- 只读权限只允许授权用户读取数据而不能修改文件或向分享目录写入文件。
- 读写权限允许用户读，修改文件并能向分享目录写入文件。



根据分享类度的不同，分享的目录被标记成不同的颜色。

分享给单个用户。

分享给整个组。

既分享给单个用户又分享给一个组。

如果想取消分享，点击 “cancel share” 按钮。

The screenshot shows the WebLab interface with a file tree under 'My Data'. The tree includes folders like 'test data', 'ii', 'rice', and 'Hemoglobin alpha chain'. Below the tree are buttons for file management: 'New Directory', 'New File', 'Delete', 'Move', 'Rename', 'Share', 'Cancel Share', 'Import', 'Cancel Import', and 'Refresh'. The 'Cancel Share' button is highlighted with a red box and a mouse cursor.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

你可以获得你所分享数据的所有信息。点击“cancel share”停止某个分享。

The screenshot shows the 'Account Based Sharing' and 'Group Based Sharing' sections. The 'Account Based Sharing' table lists shared directories and their permissions. The 'Group Based Sharing' table lists shared directories and their permissions. A mouse cursor is pointing at the 'cancel share' link in the 'Group Based Sharing' table.

SHARED DIR	ACCOUNT	PRIVILEGE	ACTION
cirdna	xiaoqiao	Read only	cancel share
blast	xiaoqiao	Read only	cancel share

SHARED DIR	GROUP	PRIVILEGE	ACTION
ePCR	test	Read and Write	cancel share

>> Back to My Data

- 如何导入数据，文献和 MetaPackage

当知道某人分享给你或你所属的组东西，可通过点击“import”按钮来查看与你有关的分享信息。

The screenshot shows the WebLab interface with a sidebar on the left containing navigation options like Service, User Space, and Account. The main area displays a file tree under 'My Data' with folders like 'test data', 'rice', and 'Hemoglobin alpha chain'. Below the tree is a toolbar with buttons for 'New Directory', 'New File', 'Delete', 'Move', 'Rename', 'Share', 'Cancel Share', 'Import', 'Cancel Import', and 'Refresh'. The 'Import' button is highlighted with a red box.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

你可以查看与你有关的所有分享的详细信息，然后选择导入你感兴趣的数据。

The screenshot shows the 'Account Based Import' and 'Group Based Import' sections. The 'Account Based Import' table has columns: AVAILABLE DIR, FROM ACCOUNT, PRIVILEGE, and ACTION. The 'Group Based Import' table has columns: AVAILABLE DIR, AS MEMBER OF GROUP, FROM ACCOUNT, PRIVILEGE, and ACTION. Both tables show importable resources with an 'Import' button in the ACTION column.

Account Based Import			
AVAILABLE DIR	FROM ACCOUNT	PRIVILEGE	ACTION
arabidopsis	bobbaby	Read only	Import

Group Based Import				
AVAILABLE DIR	AS MEMBER OF GROUP	FROM ACCOUNT	PRIVILEGE	ACTION
water	weblab_group	liuxq	Read only	Import

>>My Data

导入的目录根据分享粒度和权利的不同被标上不同的颜色。

导入的目录是分享给单个用户的只有读权限的目录。

导入的目录是分享给单个用户的有读写权限的目录。

导入的目录是分享给一个组的只有读权限的目录。

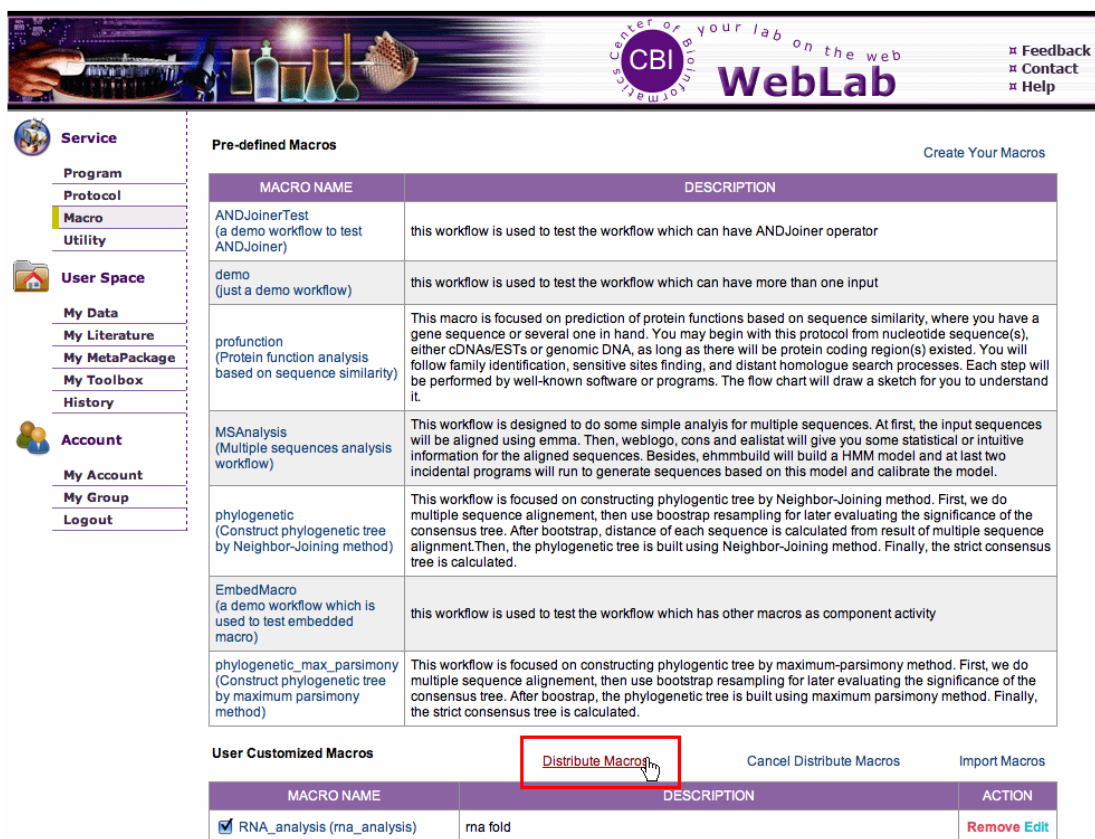
导入的目录是分享给一个组的有读写权限的目录。

注：对于 Metapackage，你既可以在“My MetaPackage”中看到导入的资源，又可以在相应程序的提交界面的上把其作为输入。

分享知识：分发工作流

工作流的分享机制与数据，文献和 MetaPackage 的分享机制不同。

- 工作流的分享是基于单个工作流（macro 或 Protocol）而不是一个目录。
- 由于 WebLab 提供递归的工作流设计，为了避免影响分享者的递归工作流，WebLab 用分发代替了分享。分发工作流意味着一旦用户导入工作流，则他得到的是这个工作流的一分拷贝而不是一个链接。你可以对此工作流做任何改动而不会改变分发者的工作流。
- 如何分发工作流
注册的用户可以分发任何自定义的或导入的工作流。



Pre-defined Macros Create Your Macros

MACRO NAME	DESCRIPTION
ANDJoinerTest (a demo workflow to test ANDJoiner)	this workflow is used to test the workflow which can have ANDJoiner operator
demo (just a demo workflow)	this workflow is used to test the workflow which can have more than one input
profuction (Protein function analysis based on sequence similarity)	This macro is focused on prediction of protein functions based on sequence similarity, where you have a gene sequence or several one in hand. You may begin with this protocol from nucleotide sequence(s), either cDNAs/ESTs or genomic DNA, as long as there will be protein coding region(s) existed. You will follow family identification, sensitive sites finding, and distant homologue search processes. Each step will be performed by well-known software or programs. The flow chart will draw a sketch for you to understand it.
MSAnalysis (Multiple sequences analysis workflow)	This workflow is designed to do some simple analysis for multiple sequences. At first, the input sequences will be aligned using emma. Then, weblogo, cons and ealstat will give you some statistical or intuitive information for the aligned sequences. Besides, ehmmbuild will build a HMM model and at last two incidental programs will run to generate sequences based on this model and calibrate the model.
phylogenetic (Construct phylogenetic tree by Neighbor-Joining method)	This workflow is focused on constructing phylogenetic tree by Neighbor-Joining method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, distance of each sequence is calculated from result of multiple sequence alignment. Then, the phylogenetic tree is built using Neighbor-Joining method. Finally, the strict consensus tree is calculated.
EmbedMacro (a demo workflow which is used to test embedded macro)	this workflow is used to test the workflow which has other macros as component activity
phylogenetic_max_parsimony (Construct phylogenetic tree by maximum parsimony method)	This workflow is focused on constructing phylogenetic tree by maximum-parsimony method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, the phylogenetic tree is built using maximum parsimony method. Finally, the strict consensus tree is calculated.

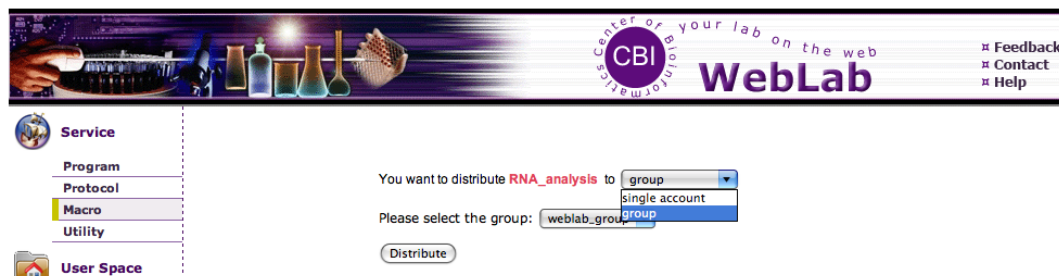
User Customized Macros Cancel Distribute Macros Import Macros

MACRO NAME	DESCRIPTION	ACTION
<input checked="" type="checkbox"/> RNA_analysis (rna_analysis)	rna fold	Remove Edit

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

选中的工作流可以分发给某个用户或你所参加的组。

注：在工作流的分发上没有权限选项。



You want to distribute **RNA_analysis** to

Please select the group:

根据分发粒度的不同，分发的 workflows 被标上不同的颜色。

- 分发给单个用户。
- 分发给整个组。
- 既分发给单个用户又分发给一个组。

如果你想停止分发，只要点击“cancel distribute macro”或“cancel distribute Protocol”按钮便可。

Pre-defined Macros Create Your Macros

MACRO NAME	DESCRIPTION
ANDJoinerTest (a demo workflow to test ANDJoiner)	this workflow is used to test the workflow which can have ANDJoiner operator
demo (just a demo workflow)	this workflow is used to test the workflow which can have more than one input
profuction (Protein function analysis based on sequence similarity)	This macro is focused on prediction of protein functions based on sequence similarity, where you have a gene sequence or several one in hand. You may begin with this protocol from nucleotide sequence(s), either cDNAs/ESTs or genomic DNA, as long as there will be protein coding region(s) existed. You will follow family identification, sensitive sites finding, and distant homologue search processes. Each step will be performed by well-known software or programs. The flow chart will draw a sketch for you to understand it.
MSAnalysis (Multiple sequences analysis workflow)	This workflow is designed to do some simple analysis for multiple sequences. At first, the input sequences will be aligned using emma. Then, weblogo, cons and ealstat will give you some statistical or intuitive information for the aligned sequences. Besides, ehmmbuild will build a HMM model and at last two incidental programs will run to generate sequences based on this model and calibrate the model.
phylogenetic (Construct phylogenetic tree by Neighbor-Joining method)	This workflow is focused on constructing phylogenetic tree by Neighbor-Joining method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, distance of each sequence is calculated from result of multiple sequence alignment. Then, the phylogenetic tree is built using Neighbor-Joining method. Finally, the strict consensus tree is calculated.
EmbedMacro (a demo workflow which is used to test embedded macro)	this workflow is used to test the workflow which has other macros as component activity
phylogenetic_max_parsimony (Construct phylogenetic tree by maximum parsimony method)	This workflow is focused on constructing phylogenetic tree by maximum-parsimony method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, the phylogenetic tree is built using maximum parsimony method. Finally, the strict consensus tree is calculated.

User Customized Macros Distribute Macros Cancel Distribute Macros Import Macros

MACRO NAME	DESCRIPTION	ACTION
<input type="checkbox"/> RNA_analysis (ma_analysis)	ma fold	Remove Edit

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

你可以看你所分发的所有工作流的信息。点“cancel distribute”按钮停止分发。

Group-Based Distributed Macro:

ACCEPT GROUP	MACRO NAME	ACTION
weblab_group	RNA_analysis	cancel distribute

>> macro home

• 如何导入工作流

当知道有人分发给你或你所属的组工作流后, 点击“Import Macros”/“Import Protocols”按钮来查看与你有关的分发信息。

The screenshot shows the WebLab interface with a navigation menu on the left and two macro tables. The top navigation bar includes 'Feedback', 'Contact', and 'Help' links. The left menu has categories like 'Service', 'User Space', and 'Account'. The main content area is divided into 'Pre-defined Macros' and 'User Customized Macros'.

Pre-defined Macros

MACRO NAME	DESCRIPTION
ANDJoinerTest (a demo workflow to test ANDJoiner)	this workflow is used to test the workflow which can have ANDJoiner operator
demo (just a demo workflow)	this workflow is used to test the workflow which can have more than one input
profuction (Protein function analysis based on sequence similarity)	This macro is focused on prediction of protein functions based on sequence similarity, where you have a gene sequence or several one in hand. You may begin with this protocol from nucleotide sequence(s), either cDNAs/ESTs or genomic DNA, as long as there will be protein coding region(s) existed. You will follow family identification, sensitive sites finding, and distant homologue search processes. Each step will be performed by well-known software or programs. The flow chart will draw a sketch for you to understand it.
MSAnalysis (Multiple sequences analysis workflow)	This workflow is designed to do some simple analysis for multiple sequences. At first, the input sequences will be aligned using emma. Then, weblogo, cons and ealstat will give you some statistical or intuitive information for the aligned sequences. Besides, ehmmbuild will build a HMM model and at last two incidental programs will run to generate sequences based on this model and calibrate the model.
phylogenetic (Construct phylogenetic tree by Neighbor-Joining method)	This workflow is focused on constructing phylogenetic tree by Neighbor-Joining method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, distance of each sequence is calculated from result of multiple sequence alignment. Then, the phylogenetic tree is built using Neighbor-Joining method. Finally, the strict consensus tree is calculated.
EmbedMacro (a demo workflow which is used to test embedded macro)	this workflow is used to test the workflow which has other macros as component activity
phylogenetic_max_parsimony (Construct phylogenetic tree by maximum parsimony method)	This workflow is focused on constructing phylogenetic tree by maximum-parsimony method. First, we do multiple sequence alignment, then use bootstrap resampling for later evaluating the significance of the consensus tree. After bootstrap, the phylogenetic tree is built using maximum parsimony method. Finally, the strict consensus tree is calculated.

User Customized Macros

MACRO NAME	DESCRIPTION	ACTION
<input type="checkbox"/> RNA_analysis (ma_analysis)	ma fold	Remove Edit

Buttons: Distribute Macros, Cancel Distribute Macros, **Import Macros** (highlighted), Remove, Edit.

Copyright© 2004-2008, Center for Bioinformatics All Rights Reserved.

你可以查看与你有关的所有分发的详细信息，然后选择你感兴趣的工作流导入。

The screenshot shows the 'Import Macros' section of the WebLab interface. It features two tables for macro import: 'Account Based Import' and 'Group Based Import'. The left navigation menu is visible, and the 'Import Macros' button from the previous screenshot is highlighted.

Account Based Import

AVAILABLE MACRO	FROM ACCOUNT	ACTION

Group Based Import

AVAILABLE MACRO	AS MEMBER OF GROUP	FROM ACCOUNT	ACTION
RNA_analysis	weblab_group	lluxq	Import (highlighted)

>> Macro Home

一旦你导入了某个工作流，它和你自己定义的工作流一样，你可以编辑，删除和分发。