

Genes4all and the Drupal Bioinformatic Server Framework Manual

Alexie Papanicolaou *Tuesday, April 20, 12:58 AM*

Prelude

The Drupal Bioinformatic Server Framework (dbsf) utilized the Drupal Content Management System (CMS) via genes4all, a GMOD compatible Drupal API. This manual aims to explain a bit more of what this sentence means and provide assistance in making the most out of it.

The manual makes a distinction between 4 types of users:

- End-users are those who are using the software on the web but do not have any administrative rights nor any interest in further developing the software;
- System Administrators are those responsible for the initial installation of Drupal, this module and installing any other software. They are also responsible for setting up/customizing the Drupal website. In Drupalesque, this is usually User 1, i.e. the first user who set up the installation;
- DBSF administrator users are those who have access/privileges to alter some of the settings of dbsf, e.g. to deploy new BLAST databases. Any number of users can have these privileges but be warned that there can be security implications;
- Developers are primarily those who code in PHP and can use/want to learn using the Drupal API to extend genes4all or code new modules. Individuals who want to contribute in another way, however, (perl scripting, CSS theming, beta-testing), are more than welcome;

For developers, please check out the **doxygen** documentation found elsewhere in the project directory

Contents

- [BioSoftware Bench End-user Manual](#)
- [Drupal installation notes](#)
- [BioSoftware Bench Administrator Manual](#)

BioSoftware Bench End-user Manual

This manual was written for release 0.95; last update on *Thursday, April 22, 4:19 PM*

This section is for end-users who wish to know how to use the web-pages. We have tried to make it as self-explanatory as possible but limited resources make it difficult. Once a brave person comes forward, we will add some on-page help buttons. Until then, if you have any questions, please refer to this manual.

Software server documentation

This is the server which is responsible for providing you with BLAST, annot8r, [InterProScan](#) or SSAHA2 searches. In this manual we will focus on BLAST as the other software are similar (Figure 1).

Currently only BLAST is available so will restrict ourselves to that. You can load some sample data by clicking on the sample data button (Figure 2) or you can download one of the FASTA files and uses to test the upload capabilities of the server (Figure 3). You will notice that you can submit multiple searches simultaneously. Each algorithm has its own query sequence and we find that this is useful to wet-lab biologists looking for homologies using both the full-length [CoDing](#) Sequence (CDS) and the protein. You can also select more than one reference database and/or combine it with a reference database you upload. Just select a FASTA file (Figure 4) and the server will do the formatting for you (caution: for computational reasons, the server will not verify you're uploading the right type of sequence). Furthermore, you can select any additional parameters you may wish to use (Figure 5).

Once you click Run Search you will be redirected to a screen while your output is being processed (Figure 6). After uploading and pre-processing any files, the URL of this waiting screen is stable and you can bookmark it: it will remain valid for at least 7 days. While different algorithms become available, the page will auto-refresh. During this time, your searches are being run via the Condor job management queue on a 96-node PC Farm. As the PC-Farm has many uses, there may be a short wait depending on the time of the day.

From the output you can immediately get a good glimpse of the quality of your search via the Bio::Graphics graph on the top (Figure 7). Clicking on it will redirect you to the complete HTML BLAST output of all your queries. The hit table underneath (Figure 8) allows you to get the BLAST output for a specific query and you have a range of output formats (XML for processing with software; HTML and TEXT for you to view and save).

By selecting one or more checkboxes you can select which HIT sequences you'd like to download. They will appear as FASTA text output on the next screen and you can choose to download a text file with them (Figure 9).

Tips: Unlike most servers, clicking the BACK button is functional and permitted. You can use it to navigate back to your results. You can also use it to navigate back to your search page. Sometimes, you have to skip the 'automatic upload page' (which will give you a non-important error - Figure 10 - you can ignore this error, just go back one more page). That way you can perform the same search against different database or change the parameters.

Screenshots

Playing with 0s and 1s

[Demonstrations](#)

[Tutorials](#)

[Contact](#)

[Login](#)

[Home](#)

Select dbsf software to use

NCBI's [BLASTALL](#)

BLAST from NCBI implementation from the C toolkit (BLASTALL)

Sanger's [SSAHA2](#)

SSAHA2 rapidly aligns reads of most sequencing platforms (ABI-Sanger, Roche 454, Illumina-Solexa) to reference databases

Playing with 0s and 1s

[Demonstrations](#)

[Tutorials](#)

[Contact](#)

[Login](#)

[Home](#)

dbsf BLAST Server

Demonstration

Click to [load some sample data](#) or you can download sample [protein FASTA](#) and [nucleotide FASTA](#) files from the Arabidopsis thaliana database (TAIR), suitable for uploading as queries or databases.

Available algorithms

- [blastn](#)
- [blastp](#)
- [blastx](#)
- [tblastn](#)
- [tblastx](#)

[Run search](#)

[Reset data](#)

[\[+\] Feedback](#)

>gi|224576212|gb|ACN56799.1| phytochrome A [Arabidopsis thaliana]
DFEPVKPYEVPMTAAGALQSYKLAAKAITRLQSLPSGSMERLCDTMVQEVFELTYDRVMAYKFHEDDHGEVVSEVTKPG
LEPYLGLHYPATDIPQAARFLFMKNKVRMIVDCNAKHARVLQDEKLSFDLTWCGSTLRAPHSCHLQYMANMDSIASLVMA
VVVNEEDGEGDAPDATTQPQKRKRLWGLVVCHNTTPRFVPFPLRYACEFLAQVFVFIHVNKEVELDNQMVEKNI XRTQTLL
CDMLMRDAPLGI VSQSPNIMDLVKCDGAALLYKDKIWKLGTTPSEFHLQEI ASWLCEYHMDLTGLSTDLSLHDAGFPRLS
LGDSVCGMAAVRISSKDMIFWFRSHTAGEVRWGGAKHDPDDRDDARRMHPTSSFKAFLLEVVKTRSLPWKDYEMDAIHSLQ
LILRNAFKDSETTDVNTKVIYSKPNDLKIDG IQELEAVTSEMVR LIETATVPILAVDSDGLVNGWNTKIAELTGLSVDEA
IGKHFLTLEDSSVEIVKRMLENALEGTEEQNVQFEIKTHLSRADAGPISLVVNACASRDLENVVGVCFVAHDLTGQKT
VMDKFTRIEGDYKAI IQ

>gi|57791644|gb|AAW56590.1| phytochrome B [Arabidopsis thaliana]
MVSGVGGSGGGRGKKEEPSSTHTPNRRGGEQAQSSGKSLRPRSNTESMSKAIQQYTVDARLHAVFEQSGESGKSF DYS
QSLKTTTYSVPEQQITAYLSRIQRGGYIQPF GCMIAVDESSFR IIGYSENAREMLGLMPQSVPTLEKPEILAMGTDVR
SLFTSSSSILLERAFVAREITLLNPWVIHSKNTGKPFYAILHRIDVGVIDLEPARTEDPALSIAGAVQSQKLAVRAISQ
LQALPGGDIKLLCDTVVESVRDLTYDRVMVYKFHEDEHGEVVAESKRDDLEPYI GLHYPATDIPQASRFLFKQNRVRMI
VDCNATPVLVVQDDRLTQSMCLVGSTLRAPHGCHSQYMANMGSIASLAMAVIINGNEDDGSNVASGRSSMRLWGLVVCHH
TSSRCIPFPLRYACEFLMQAFGLQLNMELQLALQMSEKRVLRQTLLCDMLLRDSPAGIVTQSPSIMDLVKCDGAFLYH
GKYYPLGVAPSEVQIKDVVEWLLANHADSTGLSTD SLGDAGYPGAAALGDAVCGMAVAYITKRDFLWFRSHTAKEIKWG
GAKHHPEDKDDGQRMHPRSSFQAFLEVVKSRSPWETAEMDAIHSLQLILRDSFKES EAMNSKVVDGVVQPCRDMAGEQ
GIDELGAVAREMVR LIETATVP IFAVDAGGCINGWNAKIAELTGLSVEEAMGKSLVSDLIYKENEATVNKLLSRALRGDE
EKNVEVKLKTFSPELQGKAVFVVVNACSSKDYLNNIVGVCVVGQDVT SQKIVMDKFINI QGDYKAI VHSPNPLIPPIFAA
DENTCCLEWNNAMEKLTGWSRSEVI GKMI VGEVFGSCCMLKGPDALTKFMI VLHNAIGGQDTDKFPFPFFDRNGKFVQAL
LTANKRVSLGKVI GAFCLQIPSELQQALAVQRRQDTECF TKAKELAYICQVIKNPLSGMRFANSLLEATDLNEDQKQ
LLET SVSCEKQISRIIGMDLESIEDGSFVLKREEFFLGSVINAI VSQAMFLLRDRGLQLIRDIPEEIKSIEVFGDQIRI
QQLLAEFLLSIIRYAPSQEWVEIHLSQVSKQ MADGFAAIRTEFRMACPGEGLPPELVRDMFHSSRWTSPEGLGLSVCRKI
LKL MNGEVQYIRESERSYFLIILELPVPRKRPLSTATGSGDMMLMMPY

▼ [blastn](#)

Enter query sequence in simple text or FASTA format:

or upload query sequence in FASTA format:

/home/alexie/Desktop/test2.fsa

Browse...

Please give a text file, not a MS-Word or other document

Dataset(s):

Genomes

Red flour beetle T. castaneum genome
Drosophila melanogaster genome
Wasp Nasonia vitripennis genome
Silkworm Bombyx mori genome

multiples allowed (use option/control keys; likewise to deselect).

or upload your own database:

/home/alexie/Desktop/test.fsa

Browse...

Please give a text file, not a MS-Word or other document

— ▶ [Advance Search Parameters](#)

— ▶ [blastp](#)

— ▶ [blastx](#)

— ▶ [tblastn](#)

[+] [Feedback](#)

Scoring matrix:

BLOSUM62 ▾

Filtering:☒ Low complexity regions☐ Lower-case letters**E-value cutoff:**

1e-3 ▾

Advanced BLAST options:☐ ungapped☒ megablast**query genetic code:**

Standard ▾

DB genetic code:

Standard ▾

Frame shift penalty:**Number of results:**

100 ▾

[\[+\] Feedback](#)

Playing with 0s and 1s

[Demonstrations](#)

[Tutorials](#)

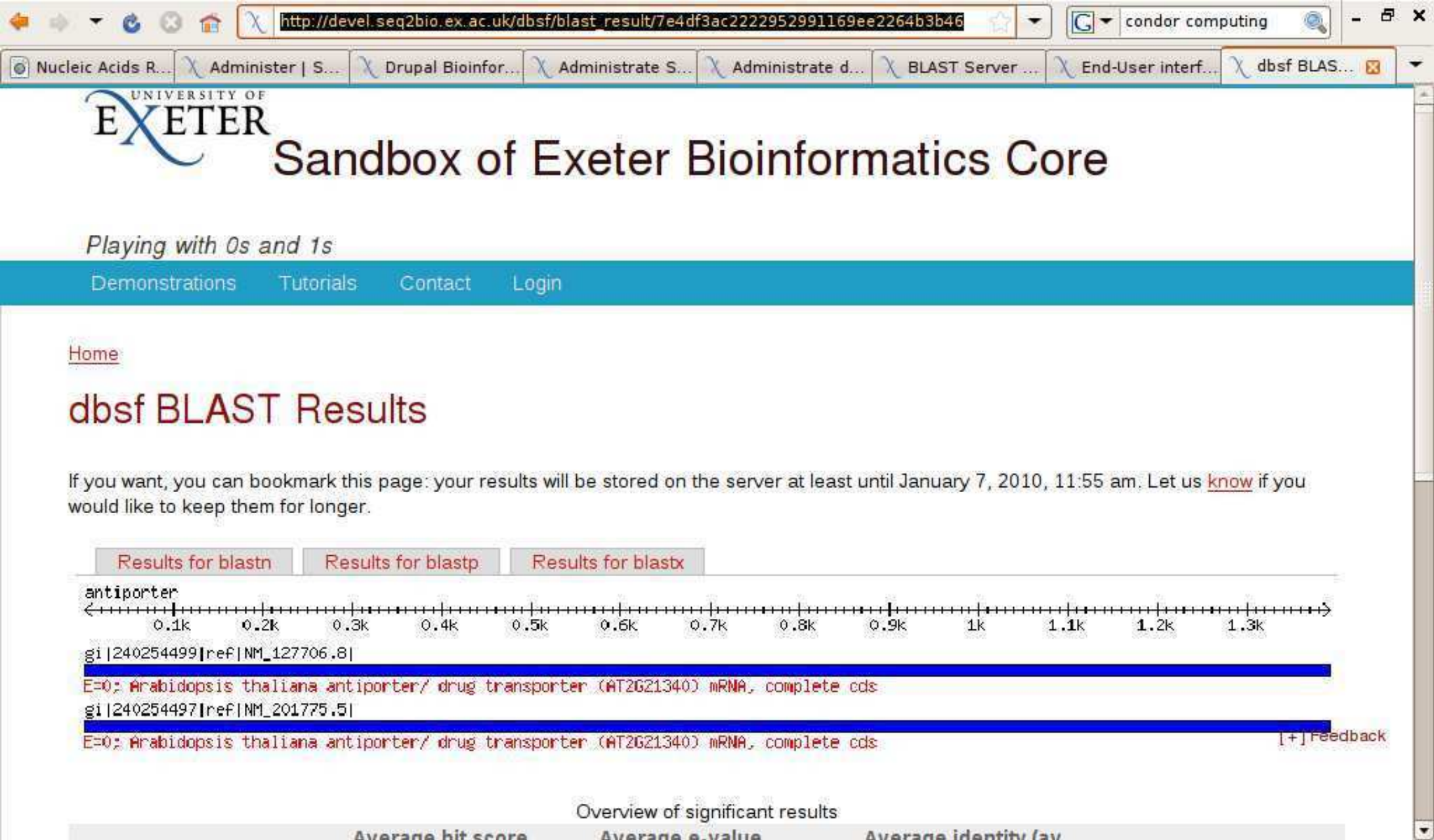
[Contact](#)

[Login](#)

[Home](#)

Preparing data needed for BLAST jobs...











gi|240254499|ref|NM_127706.8|

E=0; Arabidopsis thaliana antiporter/ drug transporter (AT2G21340) mRNA, complete cds

gi|240254497|ref|NM_201775.5|

E=0; Arabidopsis thaliana antiporter/ drug transporter (AT2G21340) mRNA, complete cds

Overview of significant results

<input type="checkbox"/> Query ID	Hit ID	Average bit score (top)	Average e-value (lowest)	Average identity (av. similarity)	Links
<input type="checkbox"/> antiporter	NM_127706	2759.93 (2759.93)	0.00e+0 (0.00e+0)	100.00 % (100.00 %)	  
<input type="checkbox"/> antiporter	NM_201775	2696.49 (2759.93)	0.00e+0 (0.00e+0)	99.35 % (99.35 %)	   Report as text

Download hits of selected results as

If the reference database has been indexed then you can download the hits. Warning, the server will refuse to serve you with very large hits (such as a whole chromosome of a genome).

FASTA

[+] Feedback

next

Playing with 0s and 1s[Demonstrations](#)[Tutorials](#)[Contact](#)[Login](#)[Home](#)

dbsf Retrieve sequences

You can download a [FASTA file](#) or copy-paste it from below.

```
>gi|240254499|ref|NM_127706.8| Arabidopsis thaliana antiporter/ drug transporter (AT2G21340) mRNA, complete cds
CTATCAGTTAATCAACGCAACGCCGAGAAGCTTTTT CAGGGGTTTGGTGC GTTGAATTTGTGGATCAAGAAACAATGCAA
ATTCAATGCAAAACCCTAACTTTTACAGTTTCTTCGATTCCTTGTAACCCAAAGCTACCATTCCCATTCTCCCTCACTTT
ACGGTCATGGAATCCTTCATTCCCGAGTTTCAGGAGCTCCGCTGTTCCGGACCAAAGTCATCACTGAAGCTGAACAGGT
TTTTAAGAAACTGTGCAAGTACGAATCAGGAGCTTGTTGTTGATGGAGAAACCGGAAATGGGTCGATTTCCGAGCTCAA
GGAGATGCAGCAAATGGTTCGATTTCCGCCGTGGAAGTGGAAGCAGAAGTAGAAGAAGTGAAGGTAGATGATTTGGCGAC
TCAGAGCATTTGGGGACAGATGAAAGAGATCGTCATGTTTACCGGACCTGCCGCGGGATTGTGGCTATGTGGGCCGTTGA
TGAGTCTCATTGATACGGCGGTGATTGGTCAAGGAAGCTCACTCGAACTCGCTGCTTTAGGTCCTGCTACCGTCATCTGT
GATTATTTGTGTTATACGTTTCATGTTCTCTCAGTTGCGACTTCAAATCTTGTTGCTACCTCTCTTGCTCGGCAGGATAA
AGATGAAGTACAACATCAGATATCGATCTTGCTTTTCATTGGGTTGGCTTGTTGGAGTCACGATGATGGTGTGACAAGAC
TGTTTGGTTCTCGGGCACTAACTGCTTTTACAGGGGTAAAGAATGCCGACATTGTTCCAGCAGCTAATAAATATGTTAG
ATTCGTGGTTTAGCATGGCCAGCTGTTCTCATTGGATGGGTTGCTCAAAGTGCAAGTCTTGGTATGAAAGACTCATGGGG
ACCTCTTAAGGCTTTGGCGGTTGCTAGTGAATAAACGGTGTTGGTGATGTGGTCTTATGCACCTTTCTAGGATATGGTA
TAGCAGGTGCAGCTTGGGCAACTATGGTGTACAAAGTTGTTGCTGCTTATATGATGATGGACGCATTGAACAAGAAAGGA
TACAGCGCATTCTCATTCTGTGTTCTTCTCCAAGTGAACTTTTGACGATTTTTGGAATCGCTGCCCTGTCTTTATAAC
TATGATGTCAAAGGTTTTGTTCTATACGCTCCTTGTTGTTGCTACTTTGCTACATCAATGGGTACAAATATCATAGCTGCTCATC
AGGTTATGCTTCAGATATATACCATGAGTACGGTTTGGGGGAGCCTCTCTCTCAAAGTGCACAGTCCTTTATGCCTGAG
CTTTTATTCCGAATCAATCGTAATTTGCCTAAAGCTAGGGTGCTTCTGAAGTCACTAGTTATCATCGGAGCTACGCTAGG
AATAGTAGTCGGAACCATGGCACAGCAGTCCATGGCTGTTCCCTGGCATCTTACACGTGACAAGGTTGTACATCCG
AGATGCACAAGGTCATAATACCGTATTTTCTTGCTTTATCCATCACTCCAAGTACTCACAGTCTTGAAGGCACCTTACTG
```

[\[+\] Feedback](#)



Sandbox of Exeter Bioinformatics Core

Playing with 0s and 1s

[Demonstrations](#)

[Tutorials](#)

[Contact](#)

[Login](#)

[Home](#)

Preparing data needed for BLAST jobs...

Your BLAST submission encountered an error.

Please continue to [the error page](#).

An HTTP error 403 occurred. /batch?id=14&op=do

Drupal installation notes

Preliminary steps

- Disable SELINUX (security enabled Linux which is default for Fedora). Personally, I never used it as we code in Ubuntu but we do know that it requires a specific level of expertise in order to successfully install pretty much anything. If you are not comfortable with SELINUX, I recommend you switch it off.
- In order to install Drupal, you have to unpack the tar archive into your Apache web server (or equivalent) Document Root or set it as your document root.
- You can name the main directory however you wish (e.g. Drupal-6.xx is the default, where xx is the minor version number). Let's assume you call it drupal. We also assume that your apache directory is called [/var/www](#) but this varies with each operating system
- You will also benefit if your Apache has mod_rewrite installed/enabled. This will allow the conversion of the ?q=page URLs to /page, which is more user friendly.
- Find out the name of your apache user (e.g. apache or www-data)

Apache configuration

For example, to set the Drupal directory inside the document root ([/var/www/drupal](#)), your apache configuration file will look similar to this:

```
<VirtualHost *:80>
    ServerAdmin webmaster@localhost
    ServerName localhost
    DocumentRoot /var/www/
    Alias /gbrowse "/var/www/gbrowse/"
    <Directory /var/www/>
        Options Indexes FollowSymLinks MultiViews
        AllowOverride None
        Order allow,deny
        allow from all
    </Directory>
    <Directory /var/www/drupal>
        Options Indexes FollowSymLinks MultiViews
        AllowOverride None
        Order allow,deny
        allow from all
        RewriteEngine on
        RewriteBase /drupal
        RewriteCond %{REQUEST_FILENAME} !-f
        RewriteCond %{REQUEST_FILENAME} !-d
        RewriteRule ^(.*)$ index.php?q=$1 [L,QSA]
    </Directory>
    <Directory "/var/www/gbrowse">
        AllowOverride None
        Order allow,deny
        allow from all
    </Directory>

    ScriptAlias /cgi-bin/ /usr/lib/cgi-bin/
    <Directory "/usr/lib/cgi-bin">
        AllowOverride None
        Options +ExecCGI -MultiViews +SymLinksIfOwnerMatch
        Order allow,deny
        Allow from all
    </Directory>
```

Where my cgi-bin is really in [/usr/lib/cgi-bin/](#), my Document root is [/var/www](#) and I have gbrowse installed in [/var/www/gbrowse](#) (you can have it anywhere, not the Alias /gbrowse line).

If the drupal directory is your [DocumentRoot](#) then it will look like this:

```
DocumentRoot /var/www/drupal
```

```
<Directory /var/www/drupal>
Options Indexes FollowSymLinks MultiViews
AllowOverride None
Order allow,deny
allow from all
RewriteEngine on
RewriteBase /
RewriteCond %{REQUEST_FILENAME} !-f
RewriteCond %{REQUEST_FILENAME} !-d
RewriteRule ^(.*)$ index.php?q=$1 [L,QSA]
</Directory>
```

Note the differences in [RewriteBase](#)

Restart the apache server (apache2ctl or apachectl restart)

Drupal configuration

You will need the following

- A postgres server
- an unprivileged postgres user, created via the createuser <name>. They should have no privileges. For this tutorial, we will call her www-db
- a postgres database, say you call it drupal. It is owned by www-db (createdb drupal -O www-db)
- go to the drupal site (localhost/drupal or localhost/) and follow the instruction on how to install the database. Ensure you've selected the postgres engine. You will need to cp a file (in [/var/www/drupal/site/default/](#)) and edit its permissions so that they are writable by the apache user. You will also need to create the directory sites/default/files and make it writable by the apache user
- continue with the online instructions. Once the database is installed, you must remove the permissions of the settings.php file so that it only writable by the root user but readable by all. The files/ directory must remain writable by the apache user. Ensure this directory exists and is NOT writable by the apache user [/var/www/drupal/site/all/modules/](#)
- You've installed Drupal! Now install the modules by unpacking them into the [/var/www/drupal/site/all/modules/](#) directory so that genes4all.module is [/var/www/drupal/site/all/modules/genes4all/genes4all.module](#) etc. Read the README files to get any dependencies (sorry, but you have to fetch them manually, the Drupal CVS server does not allow us to upload third party utilities, even if their licenses allow it...)
- Go to <http://localhost/drupal/admin/build/modules/> and you will see a list of modules. Select the relevant checkboxes (you will also need any dependencies modules) and save. That's it, you've installed the modules. Now you need to configure them as per the README files.

BioSoftware Bench Administrator Manual

This manual was written for release 0.95; Screenshots were prepared for version 0.5. Last update on *Thursday, April 22, 4:19 PM*

Setting up a software

The end-user interface is aimed to be simple and intuitive because your users don't have time to learn new software all the time. The real power of the framework, however, can be seen by the administrators' interface. If you have been given a sandbox password then you can try it yourself, otherwise, you can see the screenshots attached.

The main administrator menu is accessible from your Drupal administration menu (Figure 1). It allows you (Figure 2) to setup/activate which software you'd like to use; which datasets and to link those datasets with each software independently. For example, you cannot link a protein database with a nucleotide-based tools such as SSAHA2. Also some tools do not require linking, (e.g. annot8r, [InterProScan](#)).

Setting up the software is easy once they have been installed on your server (sadly, **DBSF** cannot do that for you!). Click on the software settings and check the software that is installed (Figure 3) and then add the necessary path information for each software (Figure 4). You can then start adding the databases/datasets.

First of all, you have to setup two necessary variables: at least one directory path where the databases reside and at least one group/category (used in the end-user interface to group common databases; Figure 5). We have implemented a security feature where an additional permission token is needed to add directories, if you would like other users (except user 1) to be able to do this, then you have to give them permission via Drupal (Administrators -> User Management -> Permissions). We recommend that you don't give this privilege to anyone you don't trust explicitly as they may add directories they should not. Further, note that the apache web server user (e.g. www-data on Ubuntu) needs to have read permissions on the directory and the files. In the case of SSAHA2, we save space by compressing the .head file. The server uncompresses it when it is needed and keeps it for 7 days after it was last accessed. Therefore, you will need to give write privileges to the ssaha2 dataset directory (we keep them separately from our other datasets).

Once you give at least one directory and one category name, you can start adding datasets and naming them with a friendly name that the users will see (Figure 6). A table underneath the 'Available datasets' tab helps you by showing you a list of the databases that have been found in the directory paths you have registered (Figure 7). Since the screenshot was taken, we're showing also the date the file was last modified and there is a tab to allow you to edit the variables you have stored for the datasets.

Not every dataset is fit for public consumption: some datasets are meant to be private! Even within your research group, some users may want to have restricted access. So you can use the Drupal role management to set database permissions for each dataset independently (Figure 8). By default all newly added databases are visible only to 'authenticated users' i.e. to any user with a login (who has actually logged in). Anonymous users are visitors to your site who have not logged in. So your public databases should have permissions for them. Moreover, you can create new roles and assign a subset of your users (Figure 9) but this has to be done by a Drupal administrator who has the necessary permissions.

In any case, when you have decided which datasets you want and for which users, you have to link them to your software. You can either link them via the dataset administration page (Figure 10) or go to the software specific settings (Figure 11). The BLAST server does not have any special settings so this page currently only allows you to link datasets to BLAST. The API framework, however, could allow you to place specific options for other software plugins.

That is it! You can add and modify your settings at any time and I hope you can see that it is extremely quick. If you have any queries at all, please contact us or click the feedback button.

Screenshots

Sandbox of Exeter Bioinformatics Core

nd 1s

[Tutorials](#) [Contact](#) [Logout](#)

[Home](#)

Administer

[By task](#)

[By module](#)

Welcome to the administration section. Here you may control how your site functions.

[Hide descriptions](#)

Content management

Manage your site's content.

[Books](#)

Manage your site's book outlines.

[Comments](#)

List and edit site comments and the comment moderation queue.

[Content](#)

View, edit, and delete your site's content.

[Content types](#)

Manage posts by content type, including

[/full-text.](#)

[Taxonomy](#)

Manage tagging, categorization, and classification of your content.

Site building

Control how your site looks and feels.

[Blocks](#)

Configure what block content appears in your site's sidebars and other regions.

[Menus](#)

Control your site's navigation menu, primary links and secondary links, as well as rename and reorganize menu items. [\[+\] Feedback](#)

[Modules](#)

Enable or disable add-on modules for your

Views are customized lists of content on your system; they are highly configurable and give you control over how lists of content are presented.

User management

Manage your site's users, groups and access to site features.

[CAPTCHA](#)

Administer how and where CAPTCHAs are used.

Site configuration

Adjust basic site configuration options.

[Administration theme](#)

Settings for how your administrative pages should look.

[Clean URLs](#)

Enable or disable clean URLs for your site.

[Date and time](#)

Settings for how Drupal displays date and time, as well as the system's default timezone.

[Date Popup Configuration](#)

Allows the user to configure the Date Popup settings.

[Drupal Bioinformatic Server framework settings](#)

Configure the dbsf module

[Error reporting](#)

Control how Drupal deals with errors including 403/404 errors as well as PHP error reporting.

- [My account](#)
- ⊕ [Create content](#)
- [Recent posts](#)
- ⊖ [Administer](#)
 - ⊕ [Content management](#)
 - ⊕ [Site building](#)
 - ⊖ [Site configuration](#)
 - [Administration theme](#)
 - [Clean URLs](#)
 - [Date Popup Configuration](#)
 - [Date and time](#)
 - [Drupal Bioinformatic Server framework settings](#)
 - [Error reporting](#)
 - [File](#)

Drupal Bioinformatic Server framework settings

Setup the software available

Specify which software to activate for your search server

Setup available reference datasets/databases

Load databases and datasets, specify which software and which users will have access to them

Setup the BLAST server

Specify the necessary variables and databases to use with NCBI's BLAST

Setup the SSAHA2 server

Specify the necessary variables and databases to use with Sangers's SSAHA2

Administrate Software for the dbsf

Software blastall is active.
Software condor is active.
Software formatdb is active.
Software fastacmd is active.
Software ssaha2 is active.
Software ssaha2Build is active.

Available software

BLASTALL

SSAHA2

Availabe software

Please specify which software will be active.

☒ blastall

BLAST from NCBI implementation from the C toolkit (BLASTALL)

☒ condor

If installed, searches will be launched via Condor

☒ formatdb

Format reference sequence as a BLAST database - NCBI C toolkit. Required for blastall to work properly.

☒ fastacmd

Retrieve data from formatted BLAST databases - NCBI C toolkit. Required for blastall to work properly.

☒ ssaha2

SSAHA2 rapidly aligns reads of most sequencing platforms (ABI-Sanger, Roche 454, Illumina-Solexa) to reference databases

☒ ssaha2Build

Format reference sequence as a SSAHA2 database. Needed for creating your own SSAHA2 databases

[+] [Feedback](#)

[Home](#) » [Administer](#)

Administrate Software for the dbsf

[Available software](#)[BLASTALL](#)[SSAHA2](#)

BLAST core settings

Please set some important settings for the blastall program.

BLASTALL executable path: *

Please provide the **full path** to the **blastall** executable on the server, e.g. /usr/bin/blastall .

FORMATDB executable path: *

Please provide the **full path** to the **formatdb** executable on the server, e.g. /usr/bin/formatdb .

FASTACMD executable path: *

Please provide the **full path** to the **fastacmd** executable on the server, e.g. /usr/bin/fastacmd .

Data directory: *

Please provide the **full path** to the **blastall data** directory on the server, e.g. /usr/share/ncbi/data/ . This is the directory where the BLOSUM62 etc matrixes reside; if you do not know how to find it, try the command "locate BLOSUM62"

[\[+ \] Feedback](#)

sandbox

- [My account](#)
- ⊕ [Create content](#)
- [Recent posts](#)
- ⊖ [Administer](#)
 - ⊕ [Content management](#)
 - ⊕ [Site building](#)
 - ⊕ [Site configuration](#)
 - [Administrate datasets/db for the dbsf](#)
 - [BLAST Server administration](#)
 - [SSAHA2 Server administration](#)
 - ⊕ [User management](#)
 - [Advanced](#)

[Home](#) » [Administer](#)

Administrate datasets/db for the dbsf

[Available datasets](#)[User permissions](#)[Dataset variables](#)[Link dataset w/ software](#)

Dataset/DB paths

Full path to directory containing your datasets:

Here you can register a directory on the server that contains one or more databases for BLAST or other software.

[Register directory](#)

File Groups

Group name:

In order to help users when viewing a large number of datasets, you can create groups for them.

[Create group](#)

- [+ Create content](#)
- [Recent posts](#)
- [-] [Administer](#)
 - + [Content management](#)
 - + [Site building](#)
 - + [Site configuration](#)
 - [Administrate datasets/db for the dbsf](#)
 - [BLAST Server administration](#)
 - [SSAHA2 Server administration](#)

Administrate datasets/db for the dbst

[Available datasets](#)[User permissions](#)[Dataset variables](#)[Link dataset w/ software](#)

Database settings

You can register new datasets to the system by providing the filename and the directory on the server. In order to remove datasets, just provide the **friendly name** you gave before and click **remove DB**.

Filename:

Please provide the basename of the database (e.g. as you would pass it to the -d option of BLASTALL). **Avoid any suffix created by a particular software (e.g. for BLAST, .pal.nal can be created).**

Friendly name: *

This will be shown to users instead of the filename

Directory:

Molecule type:

- ☒ protein
- ☐ nucleotide
- ☐ both

Select what molecule types are supported by this basename. You can select *both* if, for example, a nucleotide and protein BLAST database share the same basename.

Group:

[\[+\] Feedback](#)[Add DB](#)[Remove DB](#)

Add DB

Remove DB

Files in your registered directories which could act as databases

Directory	Filename	Formatted for	Suffixes found
/projects/dbsf/dbs	uniref100.fasta	blastall	pal
/projects/dbsf/dbs	human_proteins_nr	blastall	pal
/projects/dbsf/dbs	arabidopsis_protein	blastall	pal
/projects/dbsf/dbs	Dmel_ref_fbase_uniprot_100	blastall	pal
/projects/dbsf/dbs	silkpro_V2_ref_db_uniprot_100	blastall	pal
/projects/dbsf/dbs	Tribolium_ref_db_uniprot_100	blastall	pal
/projects/dbsf/dbs	uniref90.fasta	blastall	pal
/projects/dbsf/dbs	Agambiae_ref_vbase_uniprot_100	blastall	pal
/projects/dbsf/dbs	Amellifera_ref_db_uniprot_100	blastall	pal
/projects/dbsf/dbs	tribolium_genome_3.0	blastall	nal
/projects/dbsf/dbs	arabidopsis_genome	blastall	nal
/projects/dbsf/dbs	dmel_genome_5.16	blastall	nal
/projects/dbsf/dbs	nasonia_genome_1.0	blastall	nal
/projects/dbsf/dbs	sanger_univec.ssahadb	blastall	nal
/projects/dbsf/dbs	insecta_extra.repeats_nr95	blastall	nal
/projects/dbsf/dbs	silkworm_genome_v2.0	blastall	nal
/projects/dbsf/dbs	apis_genome_4.0	blastall	nal
/projects/dbsf/dbs	arabidopsis_mrna	blastall	nal
/projects/dbsf/dbs	hmelpomene_genome_0.001	blastall	nal
/home/alexie/personal_blast	IC77259AeApep	blastall	pal
/home/alexie/personal_blast	IC77259AeEread	blastall	nal
/home/alexie/personal_blast	IC77259AeEcon	blastall	nal
/home/alexie/personal_blast	IC77259A0Fread	blastall	nal

[+] Feedback

■ [Recent posts](#)

☐ [Administer](#)

⊕ [Content management](#)

⊕ [Site building](#)

⊕ [Site configuration](#)

■ [Administrate datasets/db for the dbsf](#)

■ [BLAST Server administration](#)

■ [SSAHA2 Server administration](#)

Dataset Agambiae_ref_vbase_uniprot_100 saved.

[Available datasets](#)

[User permissions](#)

[Dataset variables](#)

[Link dataset w/ software](#)

Data access settings

If you wish, you can protect datasets (e.g. databases for BLAST or other software) by asking that they are accessible only to users belonging to specific user roles. By default, all new datasets are accessible only to logged in users (authenticated). If you wish for all users to be able to see the database, add permissions for the anonymous group too; alternatively, you may remove permissions from the authenticated group and add them to a new private one you created via Drupal.

Role:

anonymous user

These roles are defined by [Drupal's role management](#)

Dataset:

Arabidopsis thaliana genome
Arabidopsis thaliana proteins
Arabidopsis thaliana genes
Vector contamination
Insect repeats extended
African swallowtail Papilio dardanus proteins
African swallowtail Papilio dardanus reads in contigs
African swallowtail Papilio dardanus all reads (pre-processed)
African swallowtail Papilio dardanus EST contigs
Anopheles gambiae mosquito proteome

Which dataset(s) to protect.

Toggle permission setting

Databases currently registered

Linked with	Role permitted	File	Name shown to users	Grouping	Database type Feedback
blastall	authenticated user, anonymous user	/projects/dbsf/dbs/uniref100.fasta	UniRef 100	UniProt	protein

Roles

Roles allow you to fine tune the security and administration of Drupal. A role defines a group of users that have certain privileges as defined in [user permissions](#). Examples of roles include: anonymous user, authenticated user, moderator, administrator and so on. In this area you will define the *role names* of the various roles. To delete a role choose "edit".

By default, Drupal comes with two user roles:

- Anonymous user: this role is used for users that don't have a user account or that are not authenticated.
- Authenticated user: this role is automatically granted to all logged in users.

Name	Operations
anonymous user	locked edit permissions
authenticated user	locked edit permissions
P.dardanus	edit role edit permissions
site-admin	edit role edit permissions
A new role	<input type="button" value="Add role"/>

Users

Drupal allows users to register, login, log out, maintain user profiles, etc. Users of the site may not use their own names to post content until they have signed up for a user account.

[\[more help...\]](#)

Show only users where

☐ role

is

☐ permission

☐ status

Update options

<input type="checkbox"/>	Username	Status	Roles	Member for	Last access	Operations
<input checked="" type="checkbox"/>	Iva Fukova	active	• P.dardanus	17 hours 15 min	17 hours 15 min ago	edit
<input type="checkbox"/>	sandbox	active	• site-admin	22 hours 36 min	3 min 30 sec ago	edit

Administrate datasets/db for the dbsf

Permission granted.

[Available datasets](#)[User permissions](#)[Dataset variables](#)[Link dataset w/ software](#)

Link datasets with software

Here you can link datasets you have submitted earlier with activated software

Dataset: *

African swallowtail Papilio dardanus proteins
 African swallowtail Papilio dardanus reads in contigs
 African swallowtail Papilio dardanus all reads (pre-processed)
 African swallowtail Papilio dardanus EST contigs
Anopheles gambiae mosquito proteome

Which dataset/database to link.

Software:

blastall

These are the activated software

[Toggle link](#)

Databases currently registered

Linked with	Role permitted	File	Name shown to users	Grouping	Database type
blastall	authenticated user, anonymous user	/projects/dbsf/dbs/uniref100.fasta	UniRef 100	UniProt	protein [+] Feedback
blastall	authenticated user, anonymous user	/projects/dbsf/dbs/uniref90.fasta	Uniref 90	UniProt	protein

- [My account](#)
- ⊕ [Create content](#)
- [Recent posts](#)
- ⊖ [Administer](#)
 - ⊕ [Content management](#)
 - ⊕ [Site building](#)
 - ⊕ [Site configuration](#)
 - [Administrate datasets/db for the dbsf](#)
 - [BLAST Server administration](#)
 - [SSAHA2 Server administration](#)
 - ⊕ [User management](#)
 - [Advanced help](#)
 - [Help](#)
- [Log out](#)
- [New project](#)
- [My projects](#)
- [My tasks](#)
- [Tasks for all projects](#)

[sandbox](#)

- [My account](#)
- ⊕ [Create content](#)
- [Recent posts](#)
- ⊖ [Administer](#)
 - ⊕ [Content management](#)
 - ⊕ [Site building](#)
 - ⊕ [Site configuration](#)
 - [Administrate datasets/db for the dbsf](#)
 - [BLAST Server administration](#)
 - [SSAHA2 Server administration](#)
 - ⊕ [User management](#)
 - [Advanced help](#)
 - [Help](#)
- [Log out](#)
- [New project](#)
- [My projects](#)

[Home](#) » [Administer](#)

BLAST Server administration

Dataset Anopheles gambiae mosquito proteome linked with blastall.

[Link datasets to BLASTALL](#)

Link datasets with software

Here you can link datasets you have submitted earlier with activated software

Dataset: *

- ☐ UniRef 100
- ☐ Uniref 90
- ☐ Red flour beetle T. castaneum genome
- ☐ Drosophila melanogaster genome
- ☐ Wasp Nasonia vitripennis genome
- ☐ Silkworm Bombyx mori genome
- ☐ Human proteins
- ☐ Drosophila melanogaster proteins
- ☐ Red flour beetle T. castaneum proteins
- ☐ Silkworm Bombyx mori proteins
- ☐ Anopheles gambiae mosquito proteins
- ☐ Honey bee Apis mellifera proteins

[+] Feedback