

Tutorial

Practical GRIDportal with mpiBLAST

Martin Matusiak

30th November 2005



Contents

1	Logging into GRIDportal	5
2	Creating a BLAST job	9
3	Submitting a BLAST job	14
4	Monitoring a job	17
5	Getting job results	20
6	Viewing job results	23
7	Further reading	28
7.1	Getting help	28
7.2	References	28

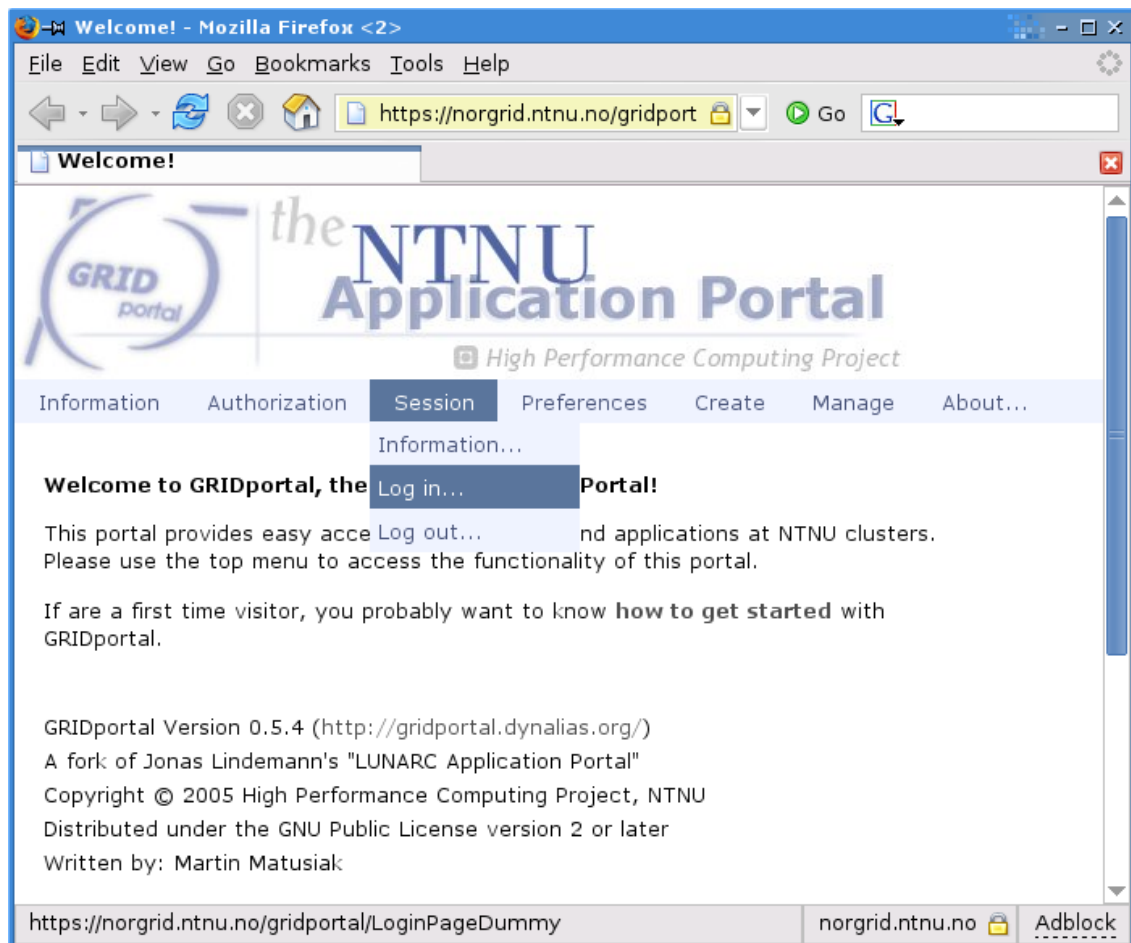
Preface

Firstly, a note on GRID access is in order. This tutorial does not discuss obtaining access to GRIDportal, it assumes access has been granted beforehand. This implies that either you already have an account with GRIDportal, or you will be using a temporary account granted for the purpose of this tutorial.

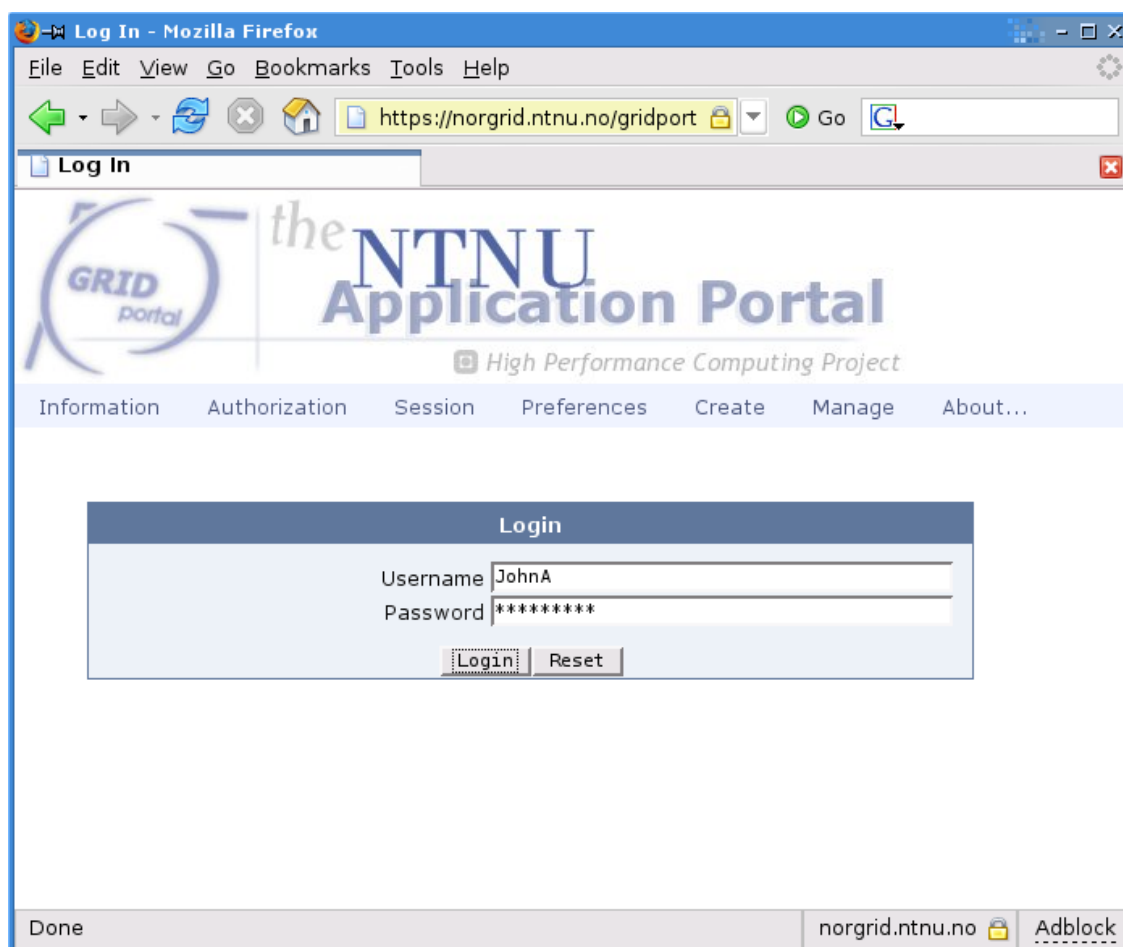
If you wish to apply for an account on GRIDportal, please consult the *Getting Started* guide at <https://norgrid.ntnu.no/gridportal/DocGettingStarted>.

1 Logging into GRIDportal

Open the GRIDportal website at <https://norgrid.ntnu.no/gridportal/>. From the menu, select *Session > Log in...*

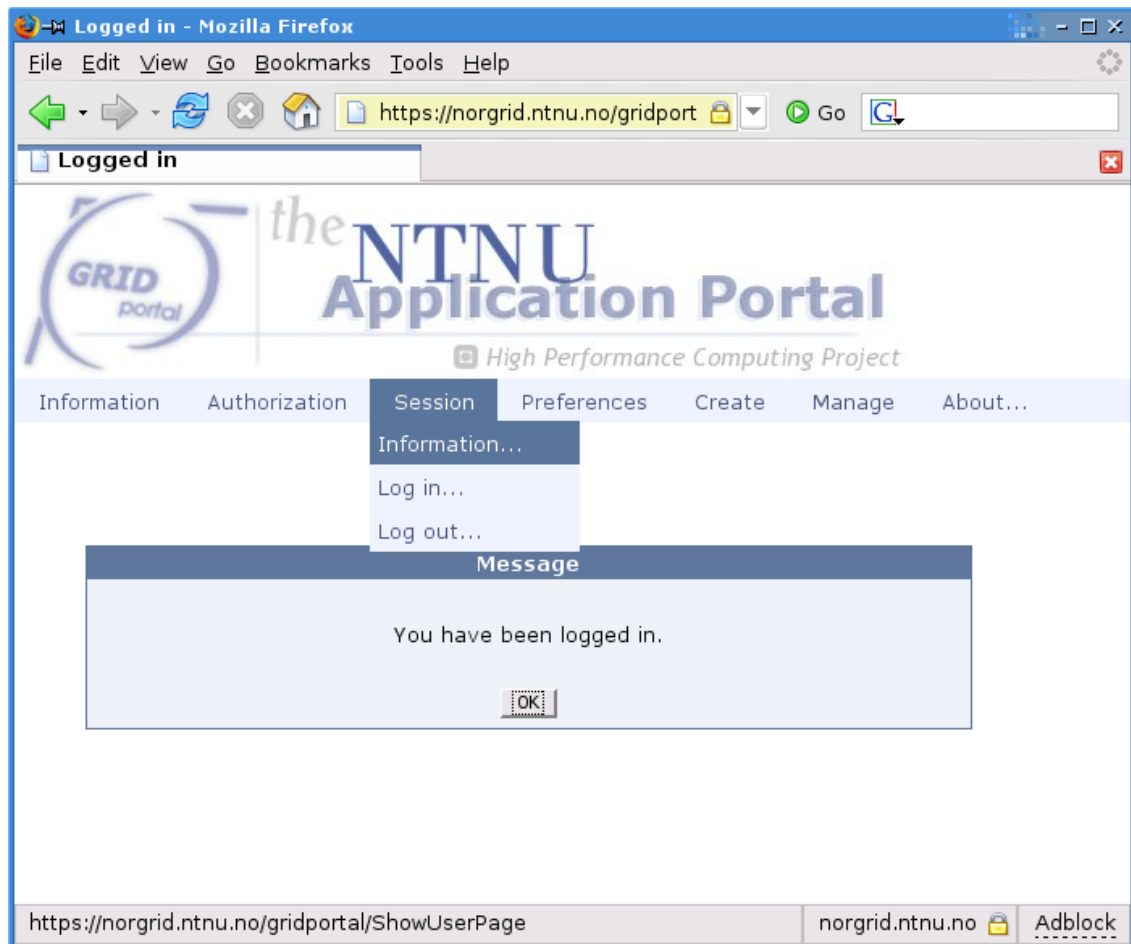


Now, enter the username and password for your GRIDportal account and click *Login*.



CHAPTER 1. LOGGING INTO GRIDPORTAL

You are now logged into GRIDportal. Choose *Session > Information...* from the menu to view some information about your session.



This page displays information about your session. The *DN* (Distinguished Name) field shows your unique identifier on the grid.

A note on sessions is in order. GRIDportal deals with two different types of sessions. One is your session on GRIDportal, which is valid for 60 minutes. This means that after 60 minutes of inactivity, the session expires and you will have to login anew. As long as you keep loading pages, the session will remain active. The status of this session is not shown on this page.

The other type of session is your GRID session, granted through a grid proxy. GRIDportal handles this for you, but the proxy is valid for a limited period of time, specifically 12 hours. *Proxy remaining* shows how much longer your GRIDportal session is valid in HH:MM:SS (hours, minutes, seconds). Although you can maintain your GRIDportal session indefinitely (as mentioned in the previous paragraph), your grid session will expire without fail. You will still be able to access GRIDportal, but not submit jobs or monitor them.


NOTE: If you receive an error saying "proxy expired", simply re-login (log out, then log in) to GRIDportal and a new proxy will be created in your name for another 12 hours. Your jobs will never be lost over an expired proxy.

ShowUserPage - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

https://norgrid.ntnu.no/gridport

ShowUserPage

 the NTNU Application Portal

High Performance Computing Project

Information Authorization Session Preferences Create Manage About...

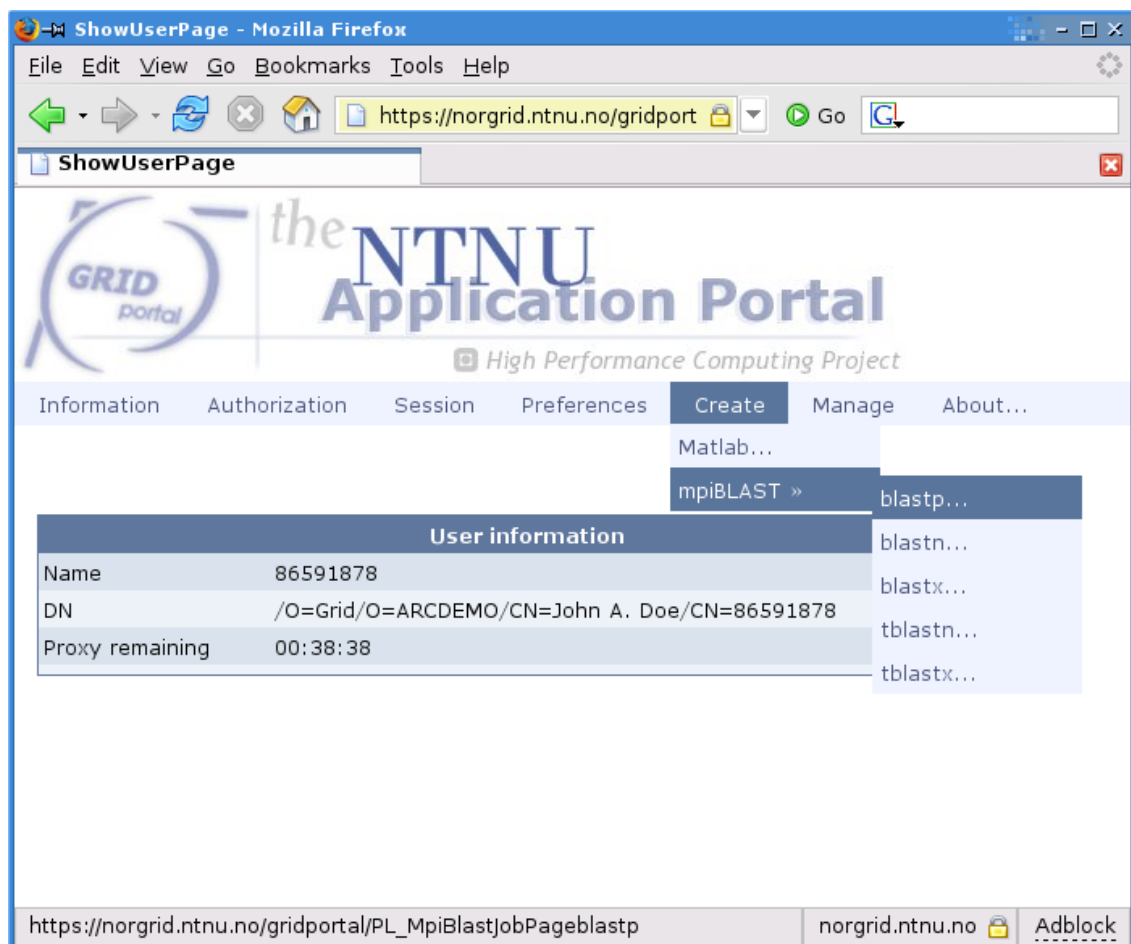
User information	
Name	86591878
DN	/O=Grid/O=ARCDemo/CN=John A. Doe/CN=86591878
Proxy remaining	00:38:38

Done norgrid.ntnu.no Adblock

2 Creating a BLAST job

To get started with mpiBLAST, choose *Create > mpiBLAST > blastp...* from the menu. mpiBLAST offers all BLAST variants except for megablast. This section discusses the use of blastp, but all BLAST types are very similar in use, so the material is easily transferrable to blastn, blastx etc.

As such, there is no difference between BLAST and mpiBLAST jobs, except that mpiBLAST was developed to run on several machines at once. The results from mpiBLAST jobs match those of BLAST.



Now fill in this form to set the options for your job. The fields marked with an asterisk (*) are mandatory. **NOTE: you may either enter query sequences into the field *Enter query sequences* or provide an input file with sequences using the field *Sequences input file*.**

Download the file https://norgrid.ntnu.no/gridportal/testjob_blastp.txt and select it as *Sequences input file*.

The choice of databases to search in is a reflection of which databases are installed. If you seek a certain database which isn't available, please request it.

The defaults on this page are made to match the defaults of NCBI's BLAST.

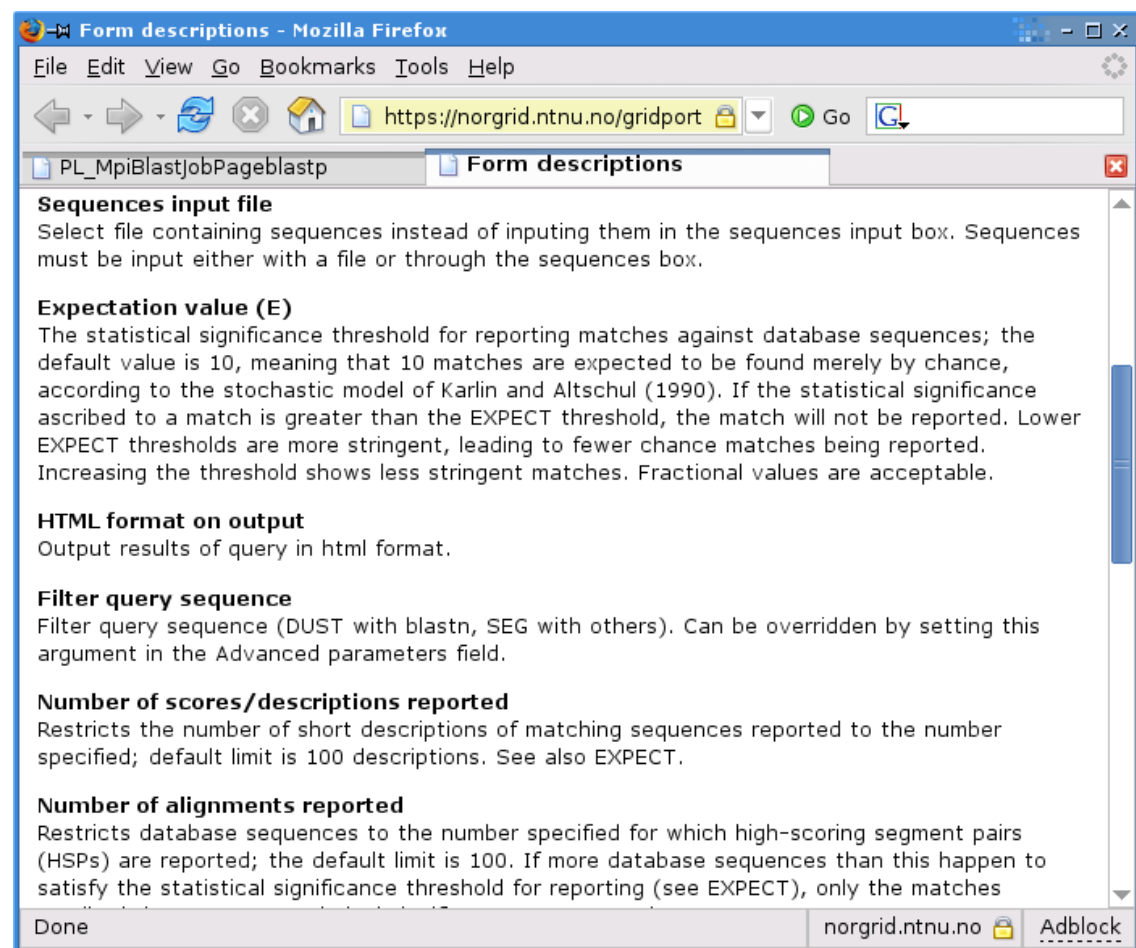
If you are unsure of certain options, click on the label of the field to bring up an explanation (next screenshot).

The screenshot shows a web browser window titled "PL_MpiBlastJobPageblastp - Mozilla Firefox". The address bar shows the URL "https://norgrid.ntnu.no/gridport". The page content is a form titled "Create a blastp job". The form contains the following fields and options:

- * CPU time (min): 30
- * Job name: blastpJob
- Email notification: (empty text box)
- * Enter query sequence(s): (empty text box)
- Sequences input file: /home/alex/martin_testset.txt (with a "Browse..." button)
- Location on query sequence: [10,400] (text box)
- Database: refseq_protein (dropdown menu)
- Filter query sequence: Yes (dropdown menu)
- Expectation value (E): 10.0 (dropdown menu)
- Word Size: 3 (dropdown menu)
- Matrix: BLOSUM62 (dropdown menu)
- HTML format on output: Yes (dropdown menu)
- Number of alignments reported: 250 (dropdown menu)
- Number of scores/descriptions reported: 500 (dropdown menu)

The browser's status bar at the bottom shows "Done", "norgrid.ntnu.no", and "Adblock".

This page of descriptions explains every input field on the job creation page.



If you wish to set any parameters which are not available on this form, you may enter them in the *Advanced parameters* field. You may also use this field to override any parameters set above.

Finally, when you are satisfied with all the options, click *Create*.

The screenshot shows a web browser window titled "PL_MpiBlastJobPageblastp - Mozilla Firefox". The address bar shows the URL "https://norgrid.ntnu.no/gridport". The page content is a form for creating a BLAST job. The form includes the following fields and options:

- Database:
- Filter query sequence:
- Expectation value (E):
- Word Size:
- Matrix:
- HTML format on output:
- Number of alignments reported:
- Number of scores/descriptions reported:
- Restrict search of database to list of GI's:
- Advanced parameters:

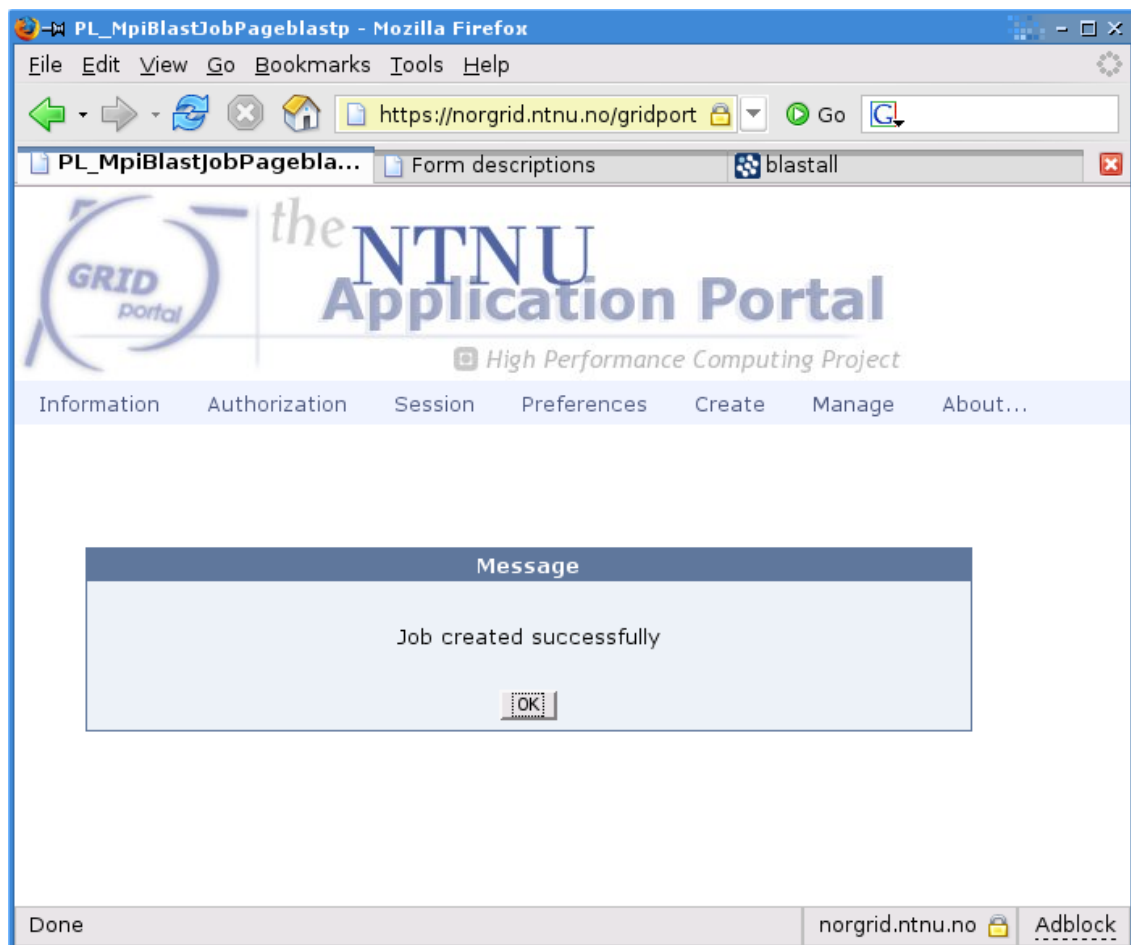
Below the form, there is a note: "Fields marked with an asterisk (*) must be filled in. Note that query sequences must be supplied either in the input box or in an input file. Defaults displayed on this page reflect the defaults of NCBI BLAST."

At the bottom of the form, there are two buttons: "Create" and "Reset".

The browser's status bar at the bottom shows "Done", the address "norgrid.ntnu.no", and an "Adblock" extension icon.

CHAPTER 2. CREATING A BLAST JOB

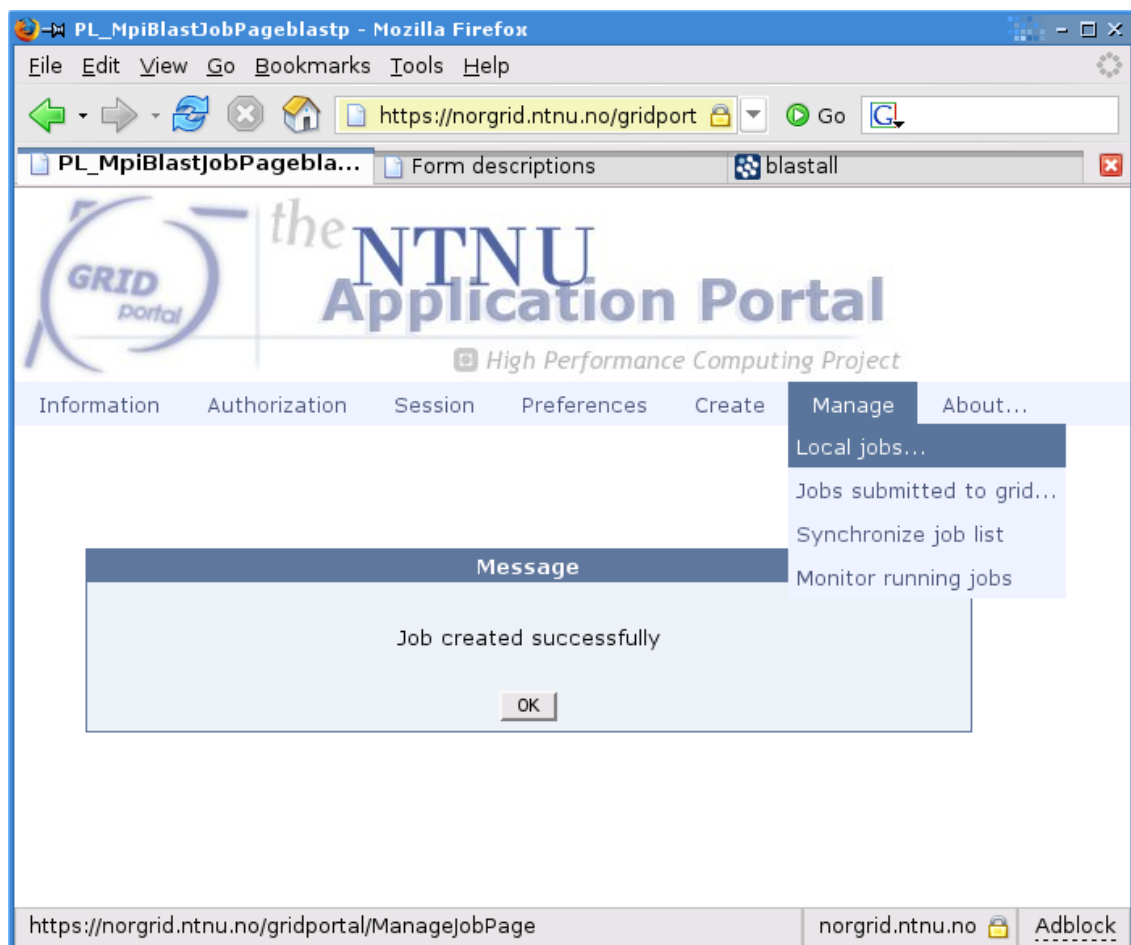
Your job is now created. Once a job exists, it can be changed and submitted any number of times.



3 Submitting a BLAST job

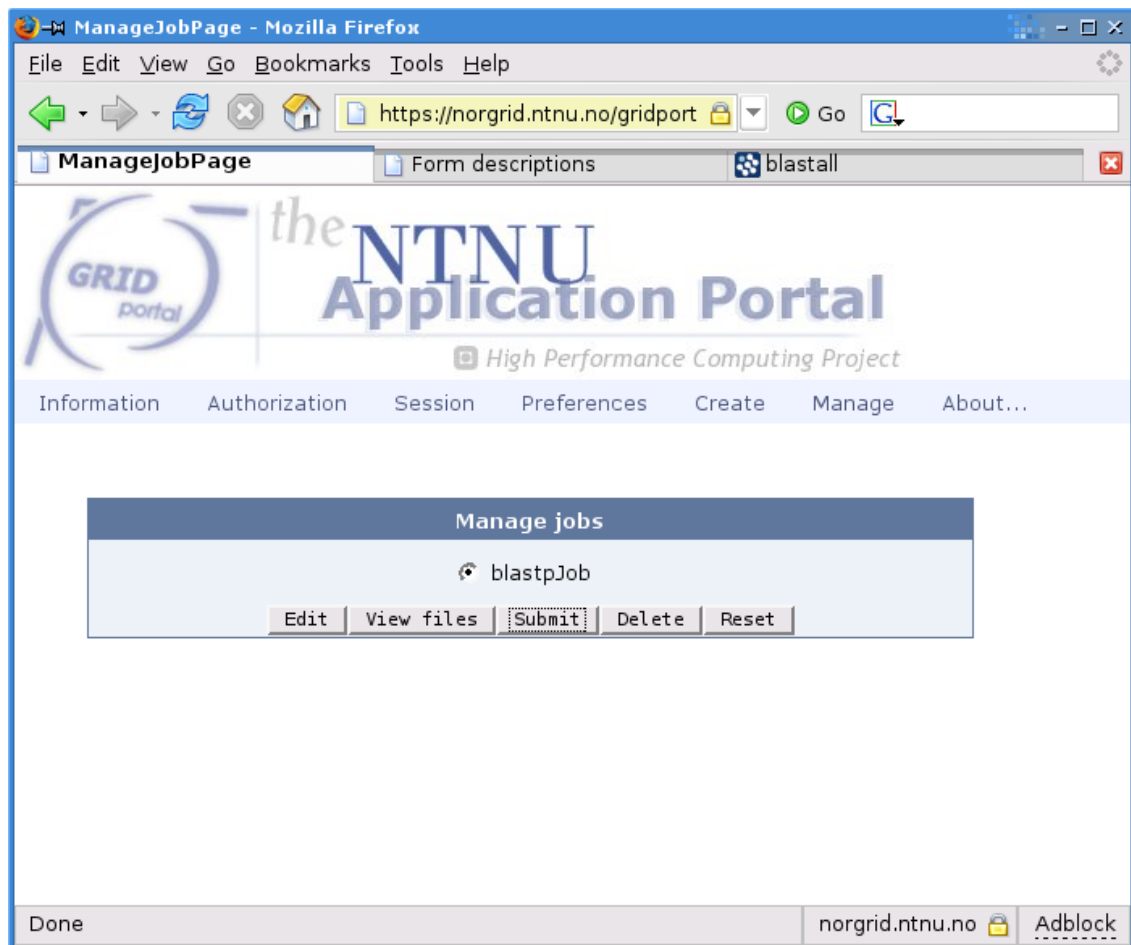
Submitting (or running) a job is sending the job from GRIDportal to the grid, where it will be run.

From the menu, select *Manage > Local jobs....*

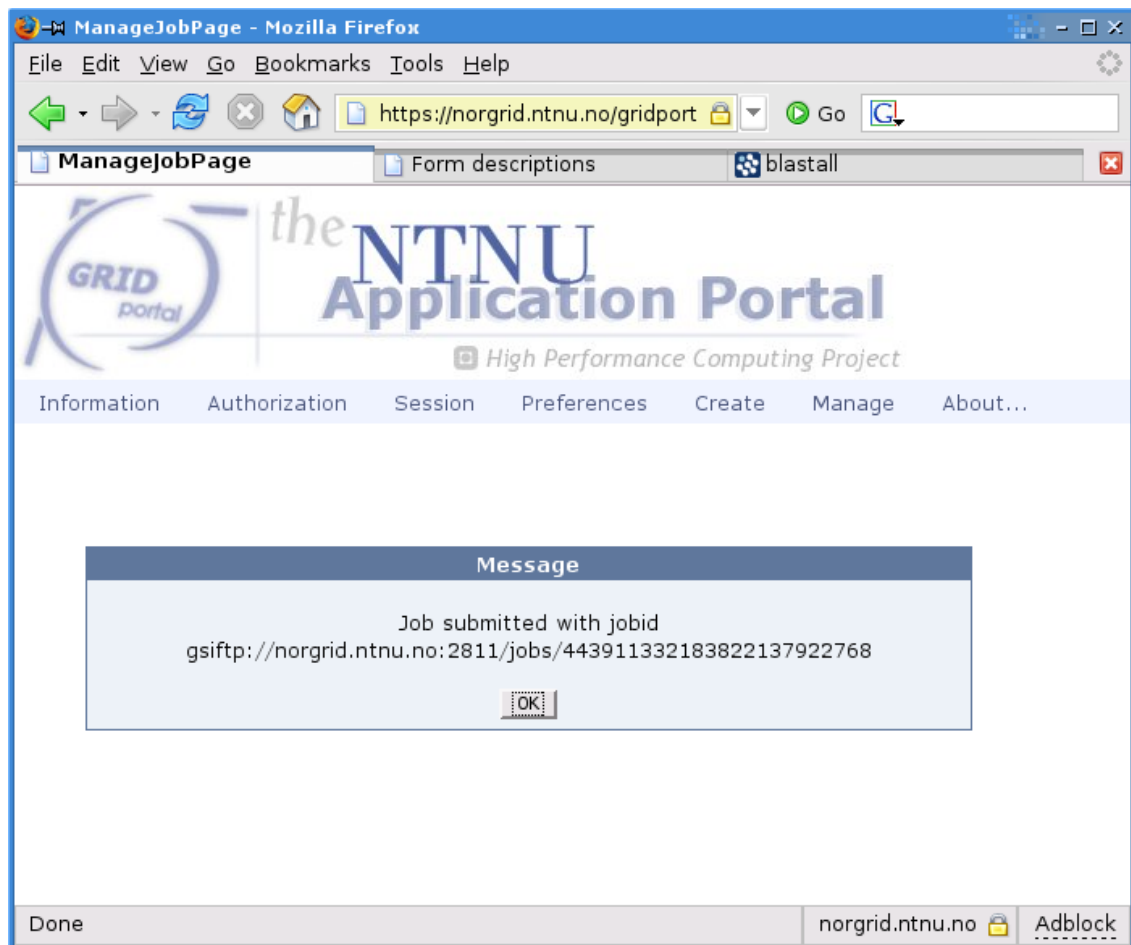


CHAPTER 3. SUBMITTING A BLAST JOB

This page shows all the jobs you have created. Select the job to submit and click *Submit*.

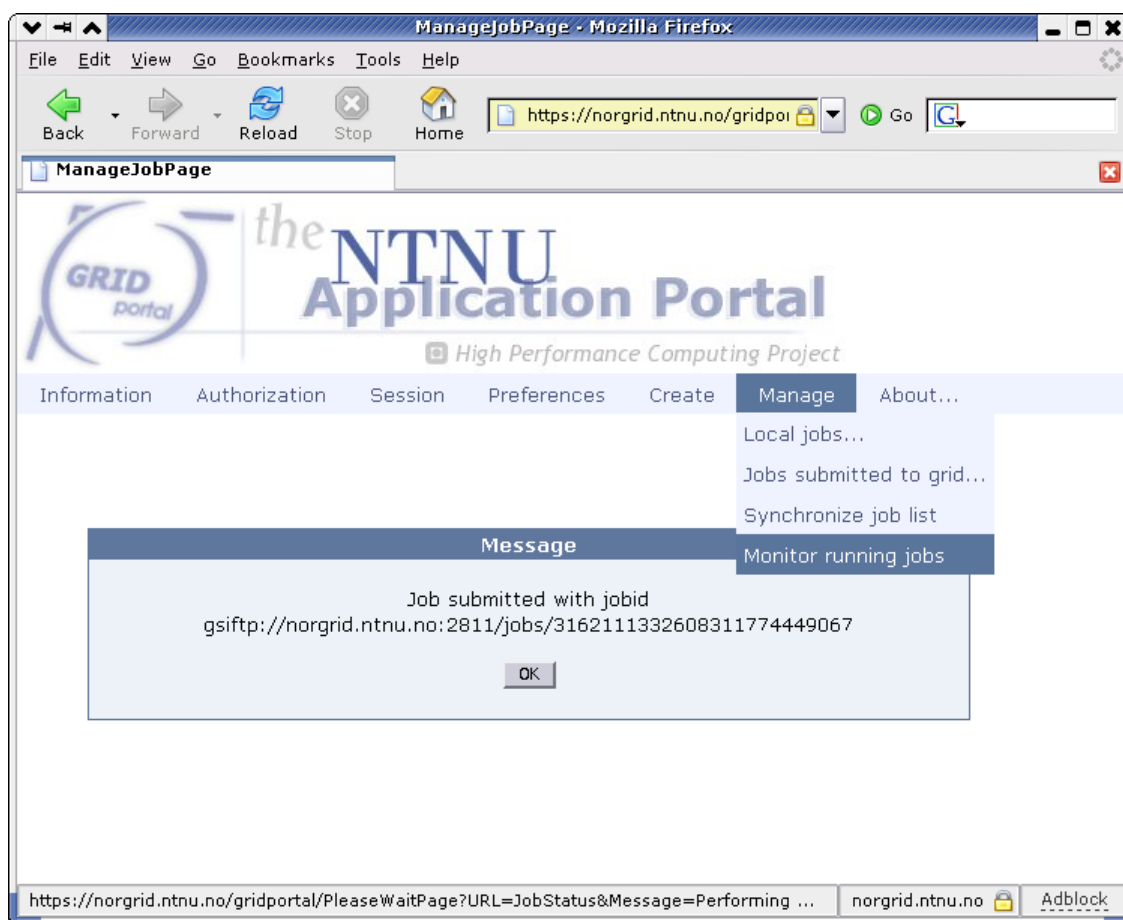


Your job is now submitted.

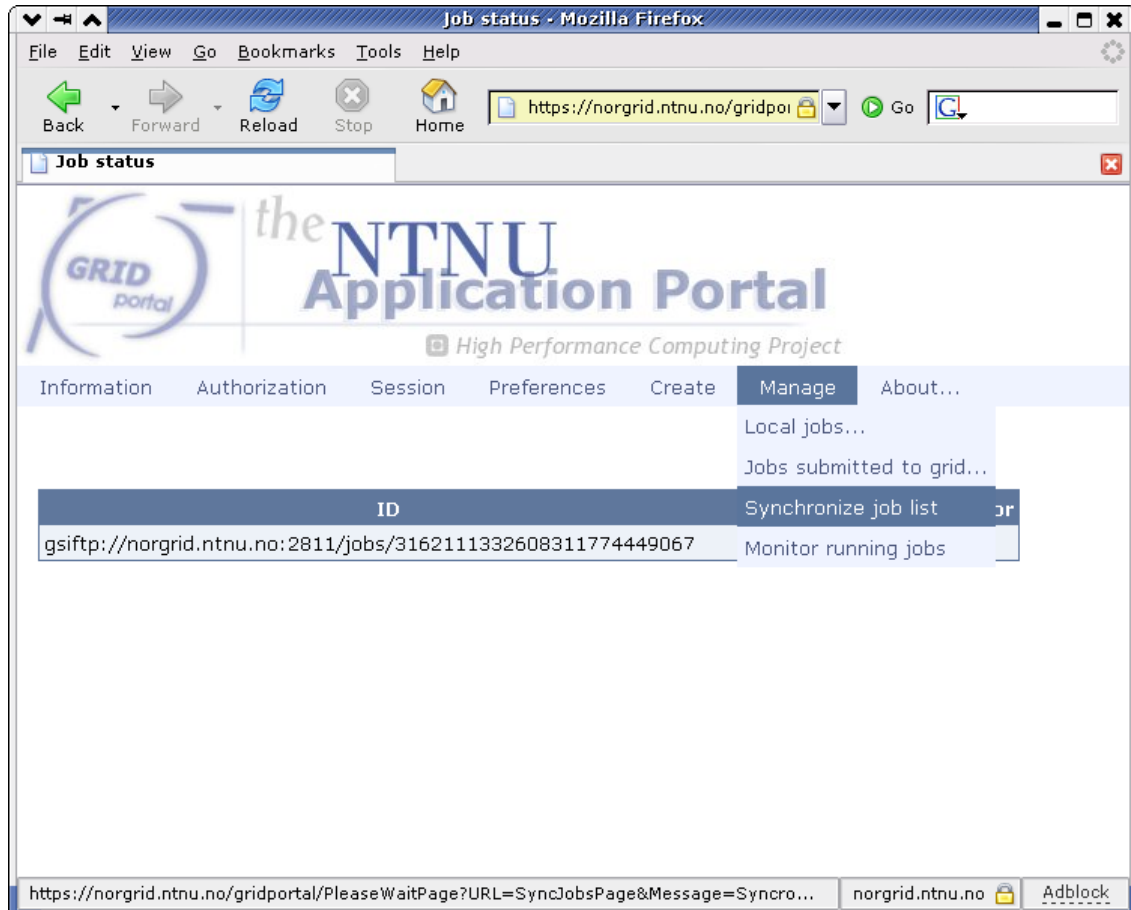


4 Monitoring a job

Once a job has been submitted, you will want to keep an eye on it to know what is happening to it. From the menu, choose *Manage > Monitor running jobs*.

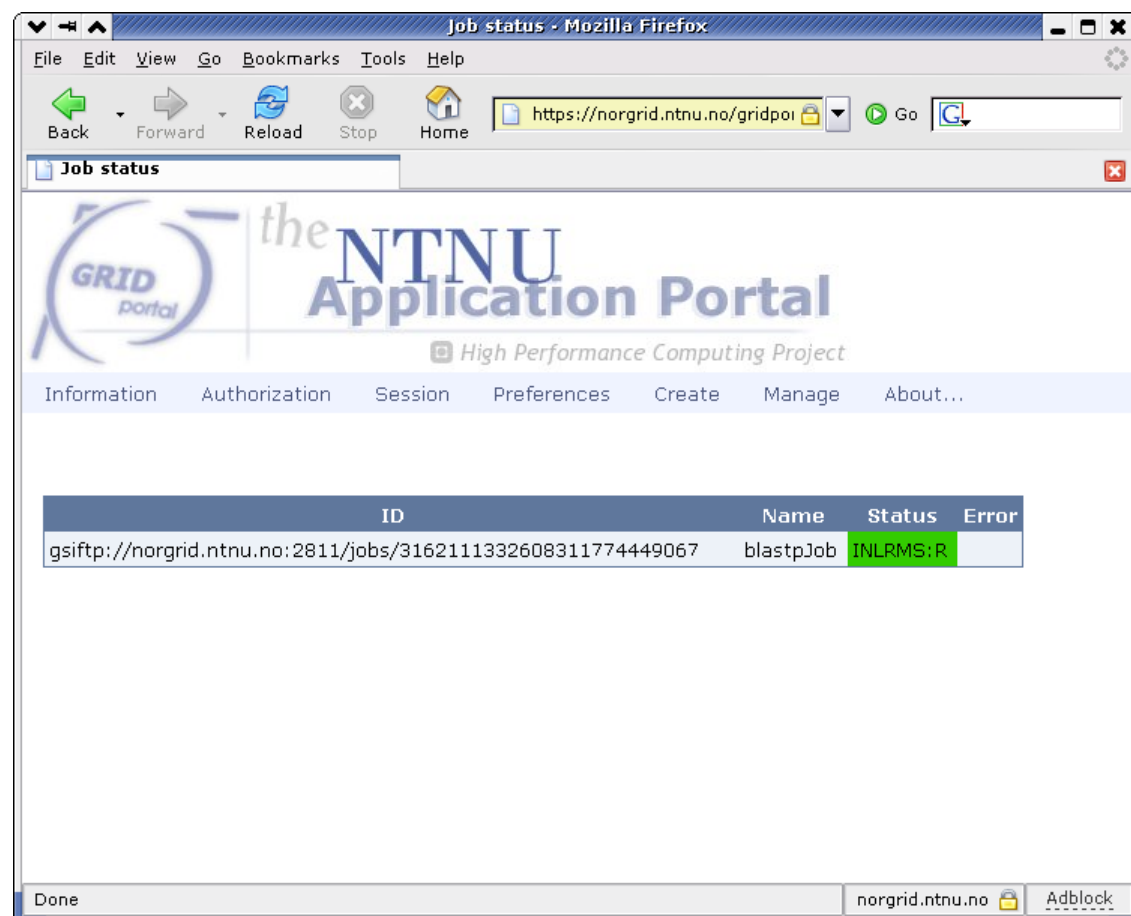


If your job does not appear immediately, choose *Manage > Synchronize job list* from the menu. If it is still missing, wait a few minutes and try again. There is a slight delay from the time the job is sent to the grid to the time the grid confirms that it has been submitted, which explains why jobs sometimes seem to be missing at first.



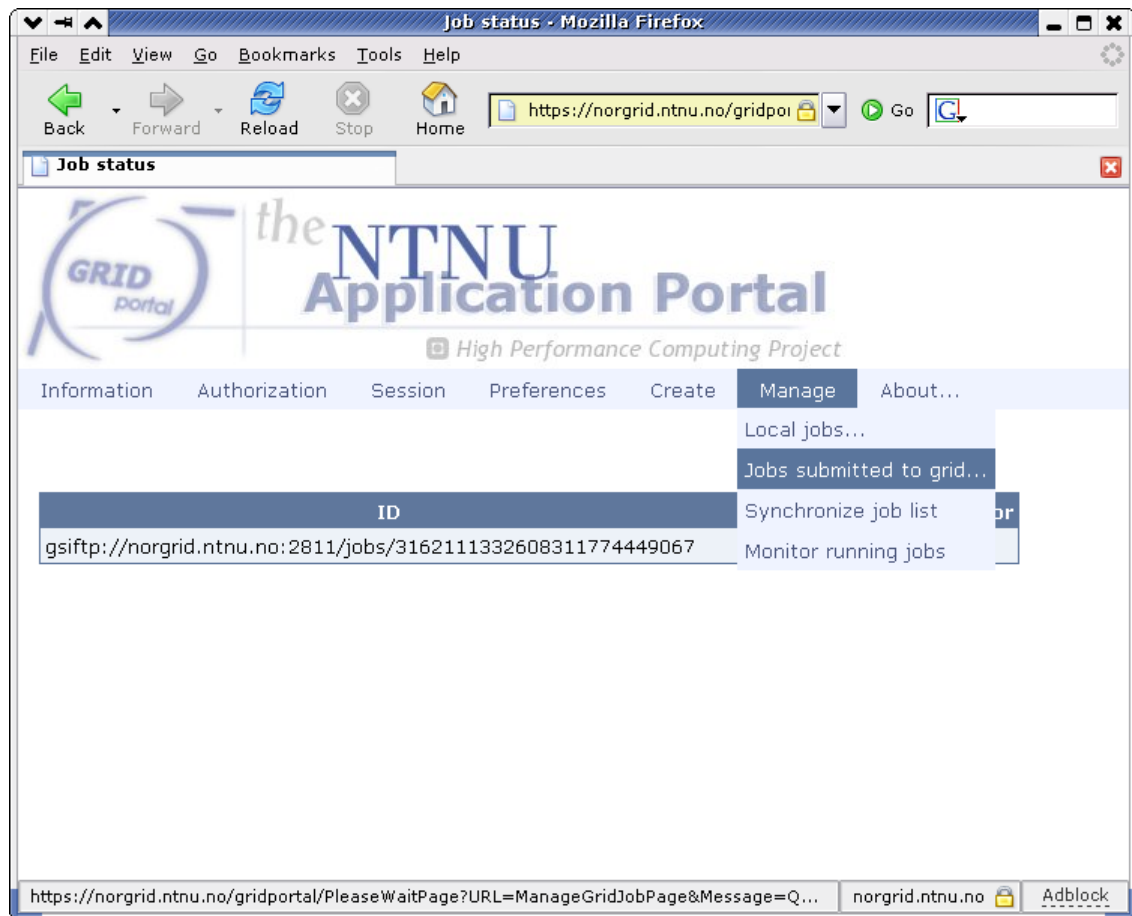
Here the job is shown running. The different job states are explained in the table below.

Job state	Explanation
ACCEPTED	job submitted but not yet processed
PREPARING	input files are being retrieved
SUBMITTING	interaction with grid ongoing
INLRMS: Q	job is queued by grid
INLRMS: R	job is running
FINISHING	output files are being transferred
FINISHED	job is finished
CANCELING	job is being canceled
DELETED	job is removed due to expiration time



5 Getting job results

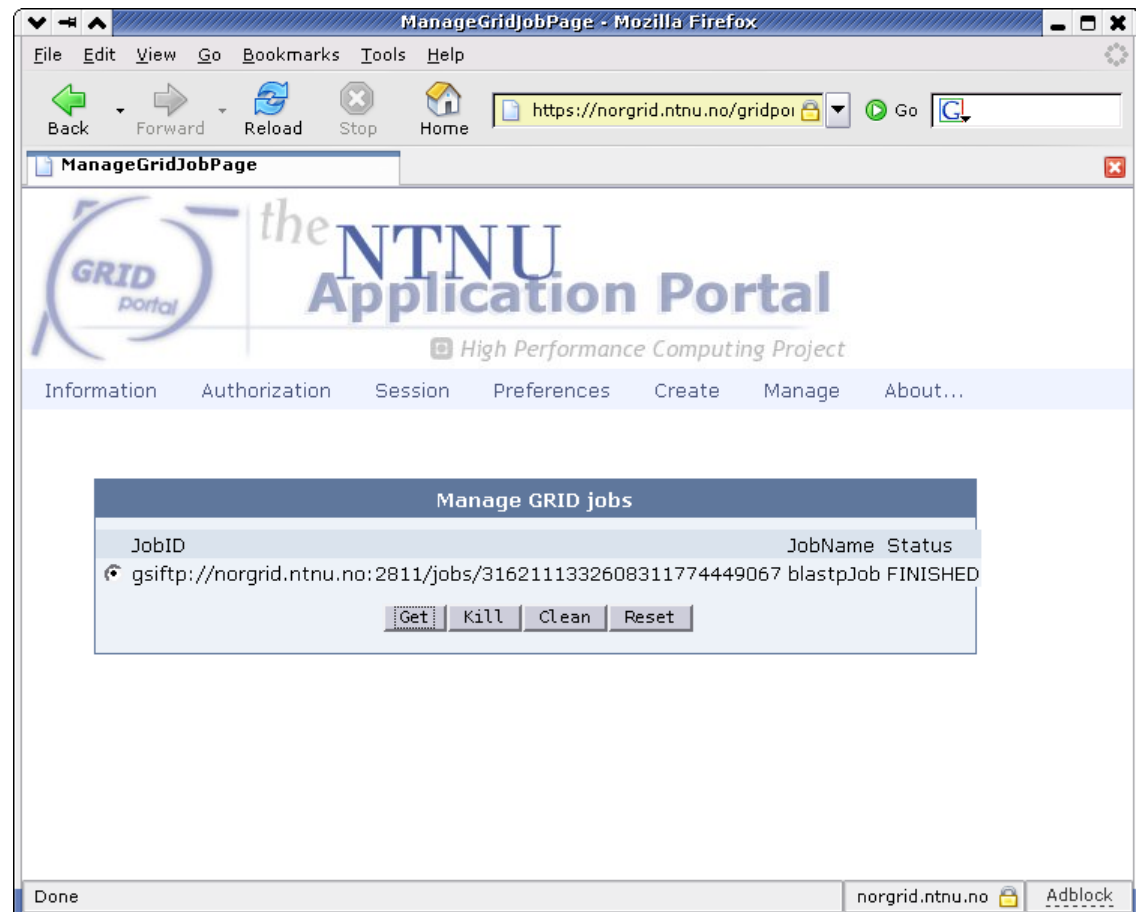
Once a job has finished, choose *Manage > Jobs submitted to grid...* from the menu.



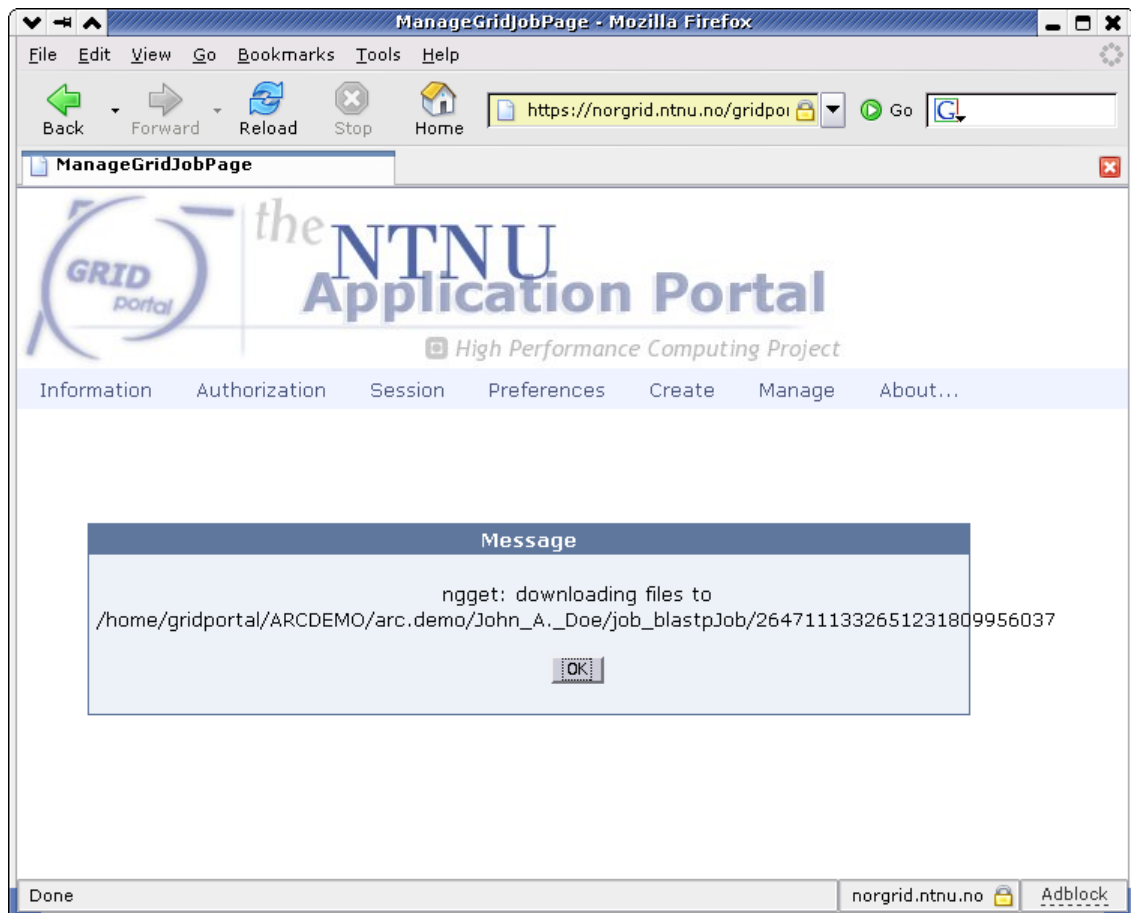
Select the job you wish to get and click *Get*. *Get* removes the job from the grid and stores results from the job on GRIDportal.

Kill can only be performed a job which has not yet finished. *Kill* aborts a job in progress on the grid, job results are lost. The job itself is not deleted on GRIDportal, only on the grid, which means the job is still listed on the page *Manage > Local jobs...*

Clean can only be performed on jobs which have finished. *Clean* is the opposite of *Get*, it removes the job from the grid and discards the job results.

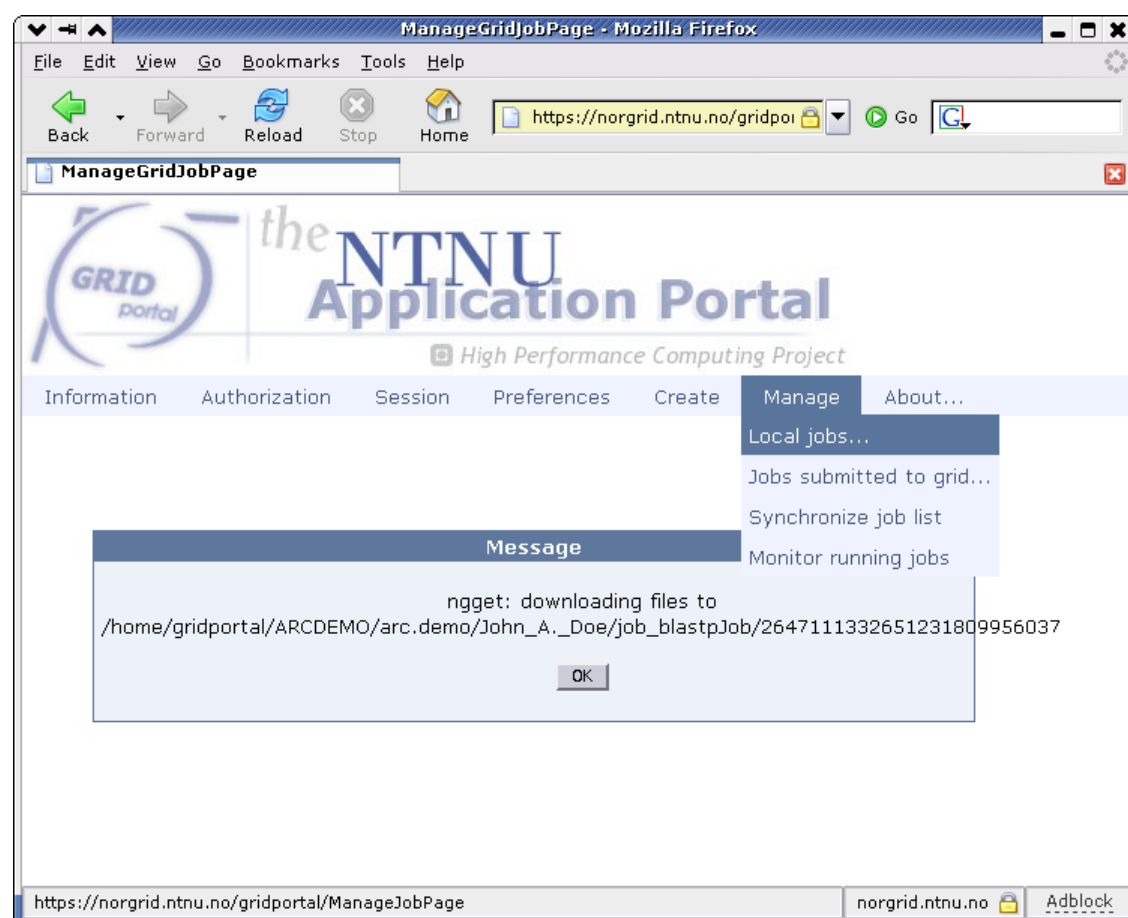


The job results are now transferred to GRIDportal.

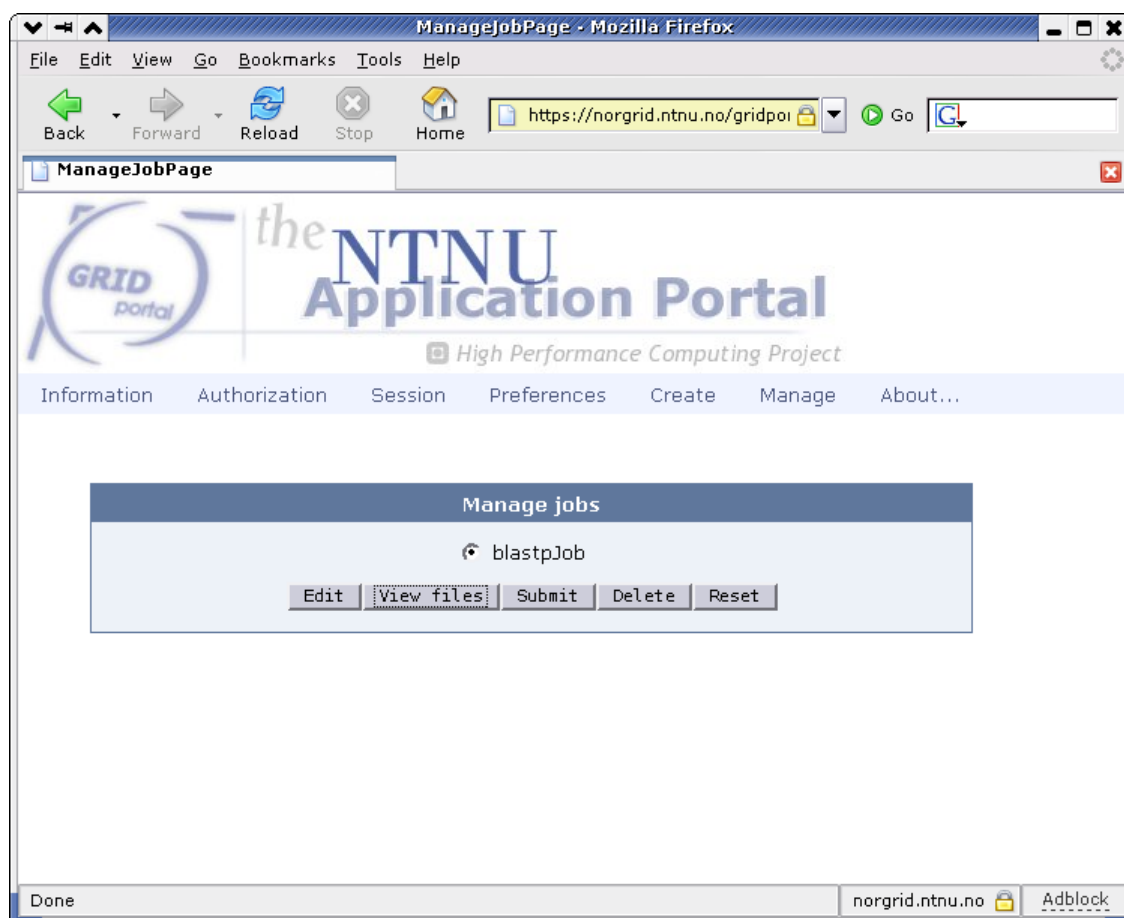


6 Viewing job results

Once job results have been transferred to GRIDportal, they remain stored on GRIDportal until you remove them. To view job results, choose *Manage > Local jobs...* from the menu.



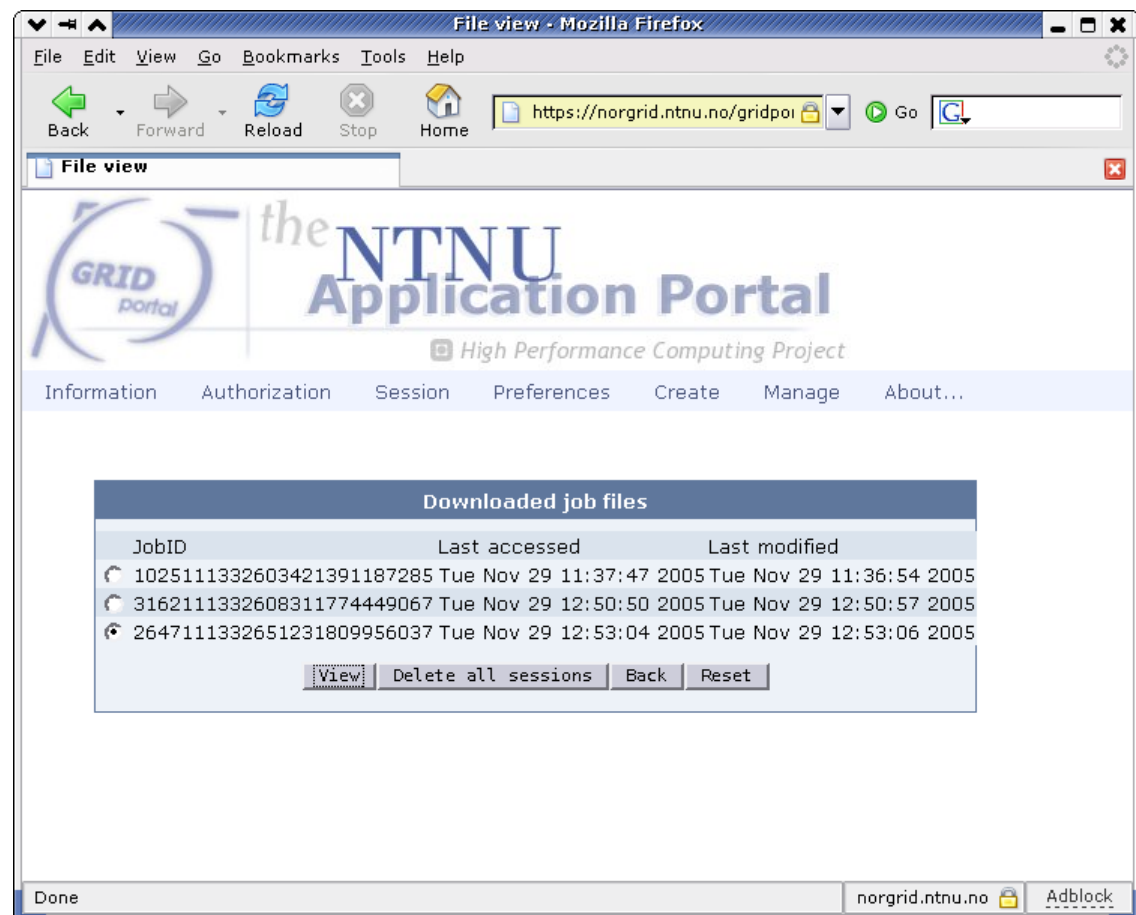
Choose the job whose results you want to examine and click *View files*.



CHAPTER 6. VIEWING JOB RESULTS

Here I have run this job three times. I choose the last job run and click *View* to look at the files resulting from the job.

Delete all sessions deletes all job results for this job, but not the job itself.

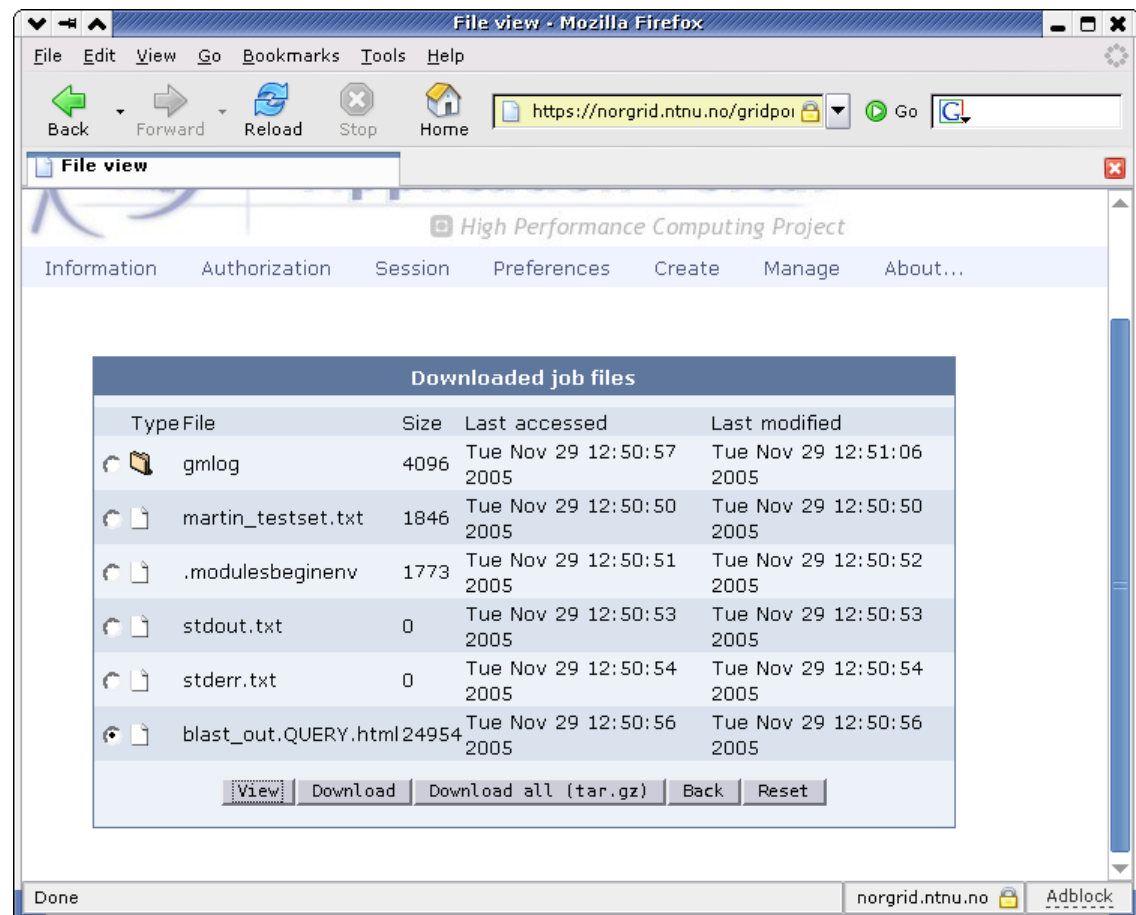


Now select the file to view and click *View*.

Download downloads the selected file to your computer.

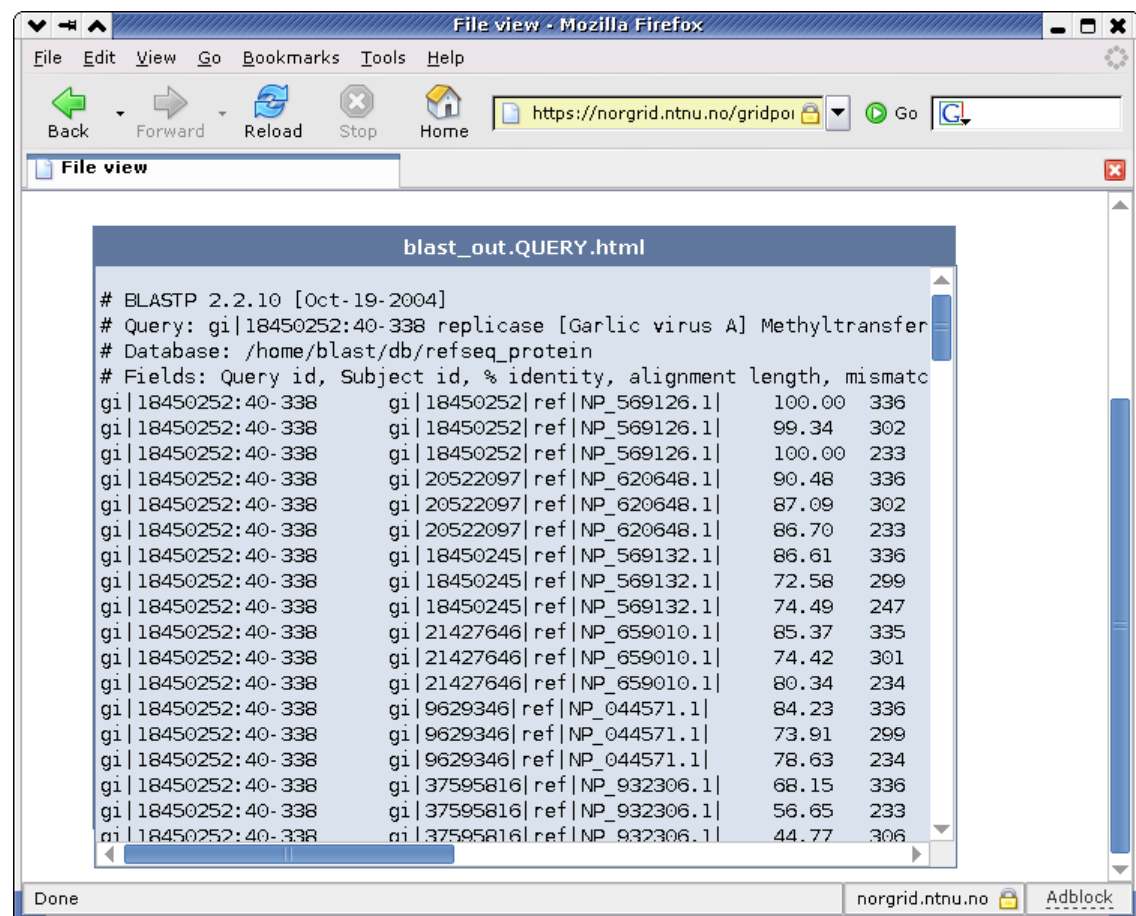
Download all (tar.gz) compresses all the files shown and lets you download the compressed archive.

gmlog is a directory containing debugging information about the job and can be safely ignored.



CHAPTER 6. VIEWING JOB RESULTS

This page shows a single file from the job results directory.



7 Further reading

7.1 Getting help

The material in this tutorial is discussed more thoroughly in the *User's Guide*, available at <https://norgrid.ntnu.no/gridportal/DocUserGuide> (or simply choose *Information > User's Guide* from the menu).

7.2 References

- GRIDportal: the NTNU Application Portal: <https://norgrid.ntnu.no/gridportal>
- GRIDportal project: <http://gridportal.dynalias.org/>
- mpiBLAST: <http://mpiblast.lanl.gov/>