Topic 2. Free BLAST to your hands with customized databases and well result viewer.

主題二：高效率平行BLAST之個人化序列搜尋平台、客製化比對資料庫及搜尋結果管理系統

2010.10.07.

by Linda Lu

MyBLAST

Myblast is the web-application of the customized biological sequence alignment tool.

Motivation of MyBlast

- In most public BLAST service, users can only search a pre-existing public database and find their interesting targets from massive blast results

- Here we construct a **web–accessible service**, **myBLAST**, **for running blast on customized databases** **and** **managing these databases** **as well as the blast results**

- **With the versatile BLAST result parser, users can browse the output filtered by options on strand-match and top n ranked hits**

- **A blast service can be shared for small group / personal use**

From Similarity Search and Phylogentic Analysis PDF file page14 ,2009,Dec.

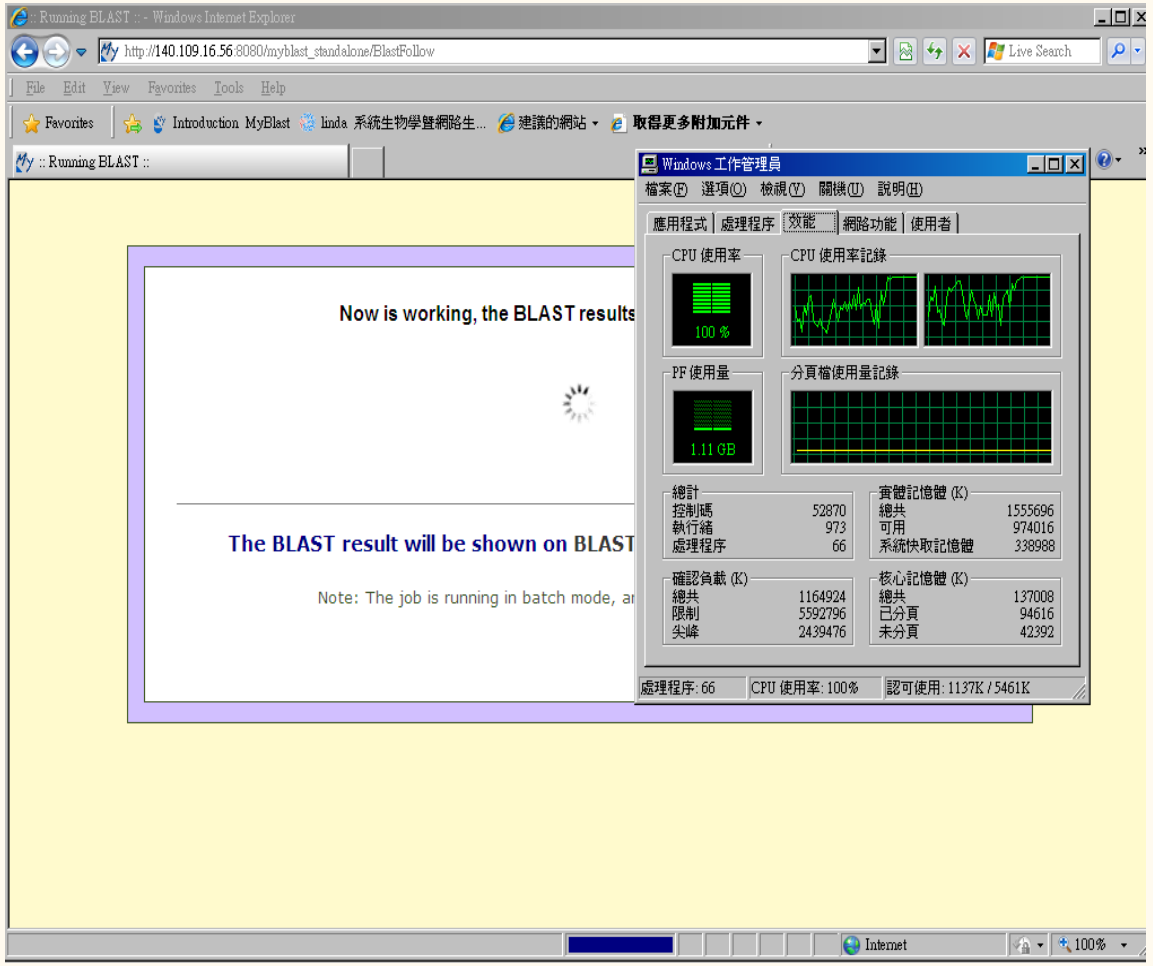
	myBLAST Web	myBLAST (standalone)	NCBI Web BLAST	NCBI BLAST (standalone)
Graphic Interface	✓	✓	✓	✗
Personalize Database	✓	✓	✗	✓
Parallel Computing	✓	✓	✓	✓
Batch Query	✓	✓	✓	✓
Database manager	✓	✓	✗	✗
Result Parser and manager	✓	✓	✗	✗
User manager	✓	✗	✓	✗

From Similarity Search and Phylogentic Analysis PDF file page15 ,2009,Dec.

MyBlast is **Multi-Thread Processing**.

-More efficiency.

-Faster then before.



MyBLAST standalone installation

- **Lab of Systems Biology & Network Biology website**

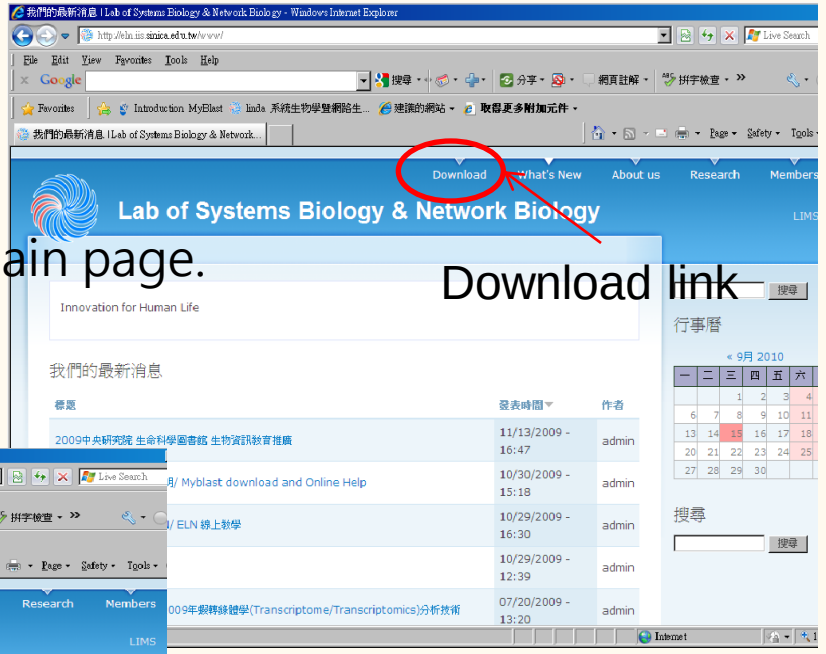
  http://eln.iis.sinica.edu.tw

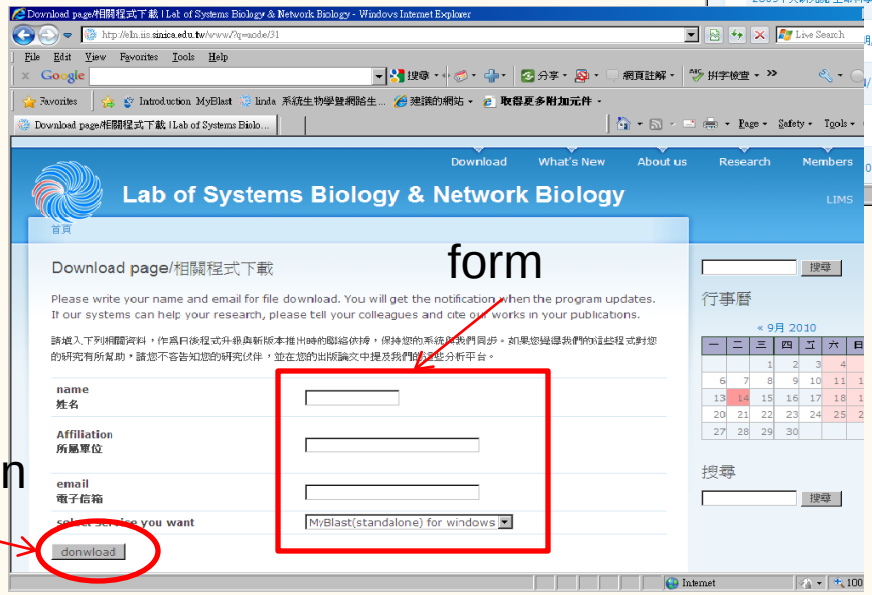
- **Download link.**

1.Click the Download link on the top of the main page.

2.Fill in the form, then submit it.

The download process starts immediately.



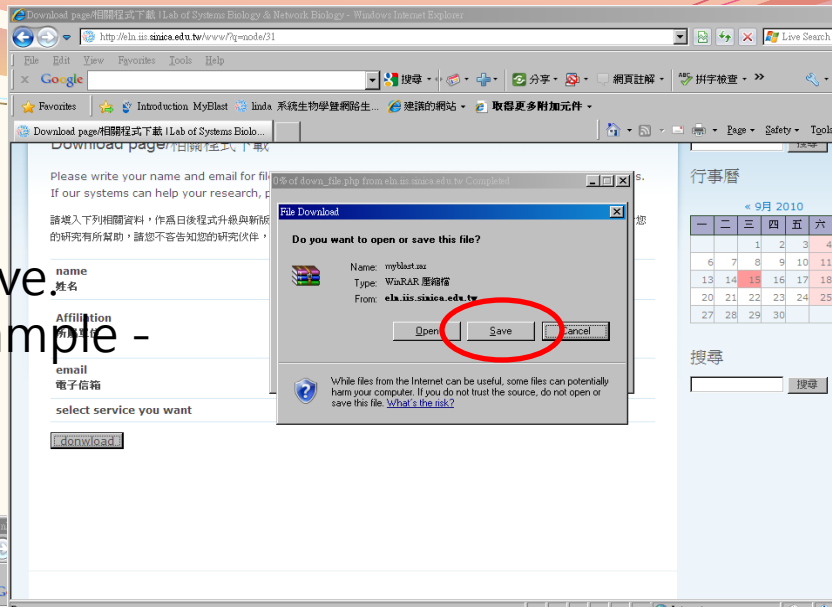


- **Download process**

1.The browser pop up the question. Click Save.

2.Select the location to place the file.(for example - desktop)

3.Wait until the process done.

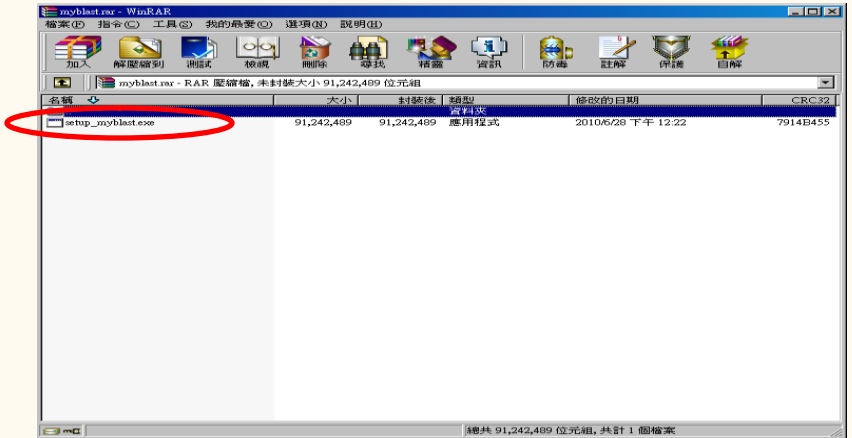


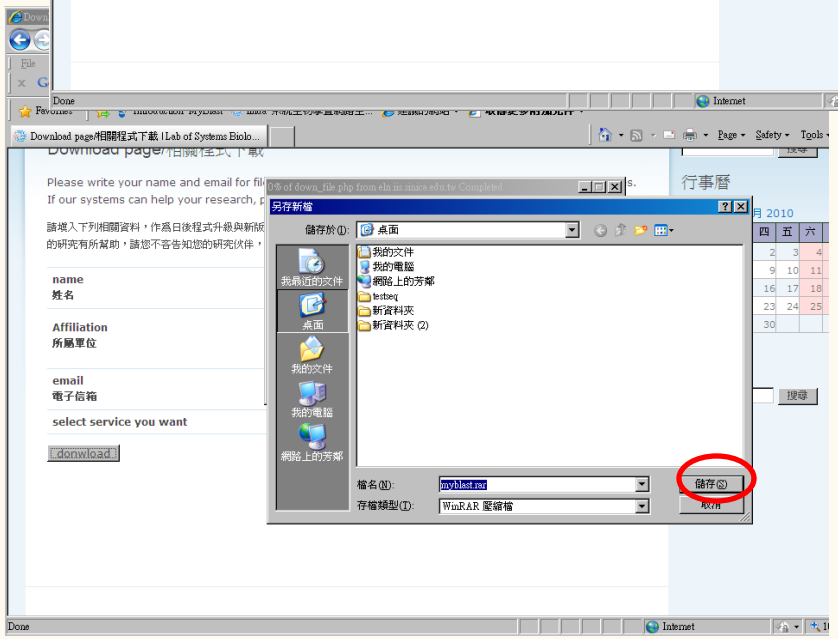
- **Execute setup_myblast.exe**

1.The setup_myblast.exe is inside the myblast.rar

Double click the myblast.rar

2.Execute this setup_myblast.exe

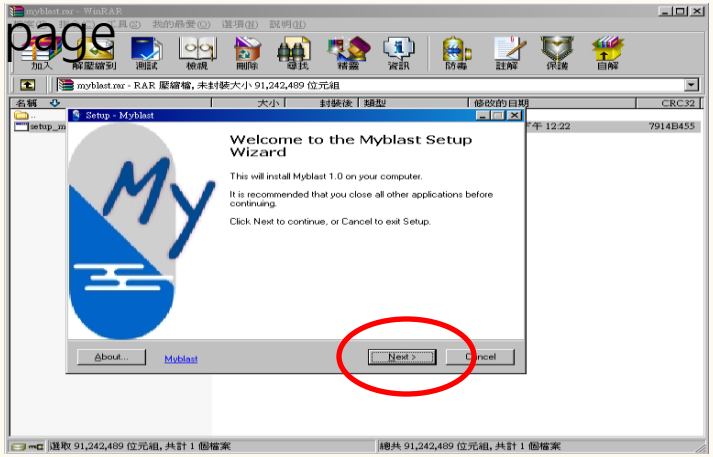




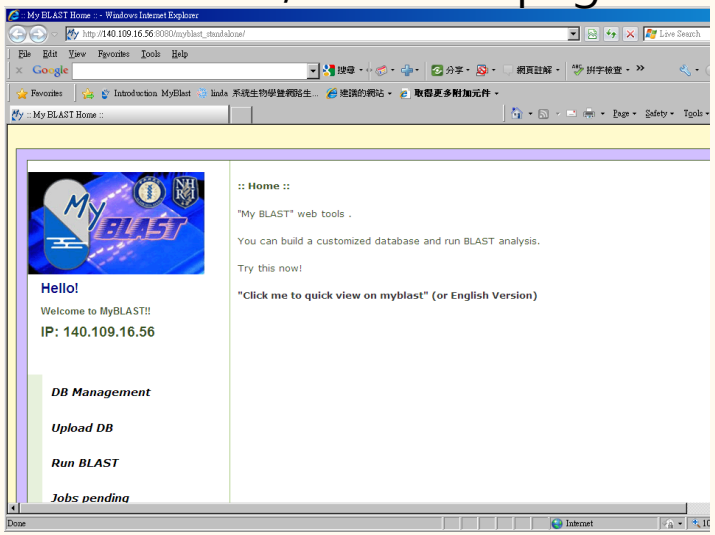
Myblast.rar shows on the desktop

- **Follow the setup wizard**

1.Start from the welcome page



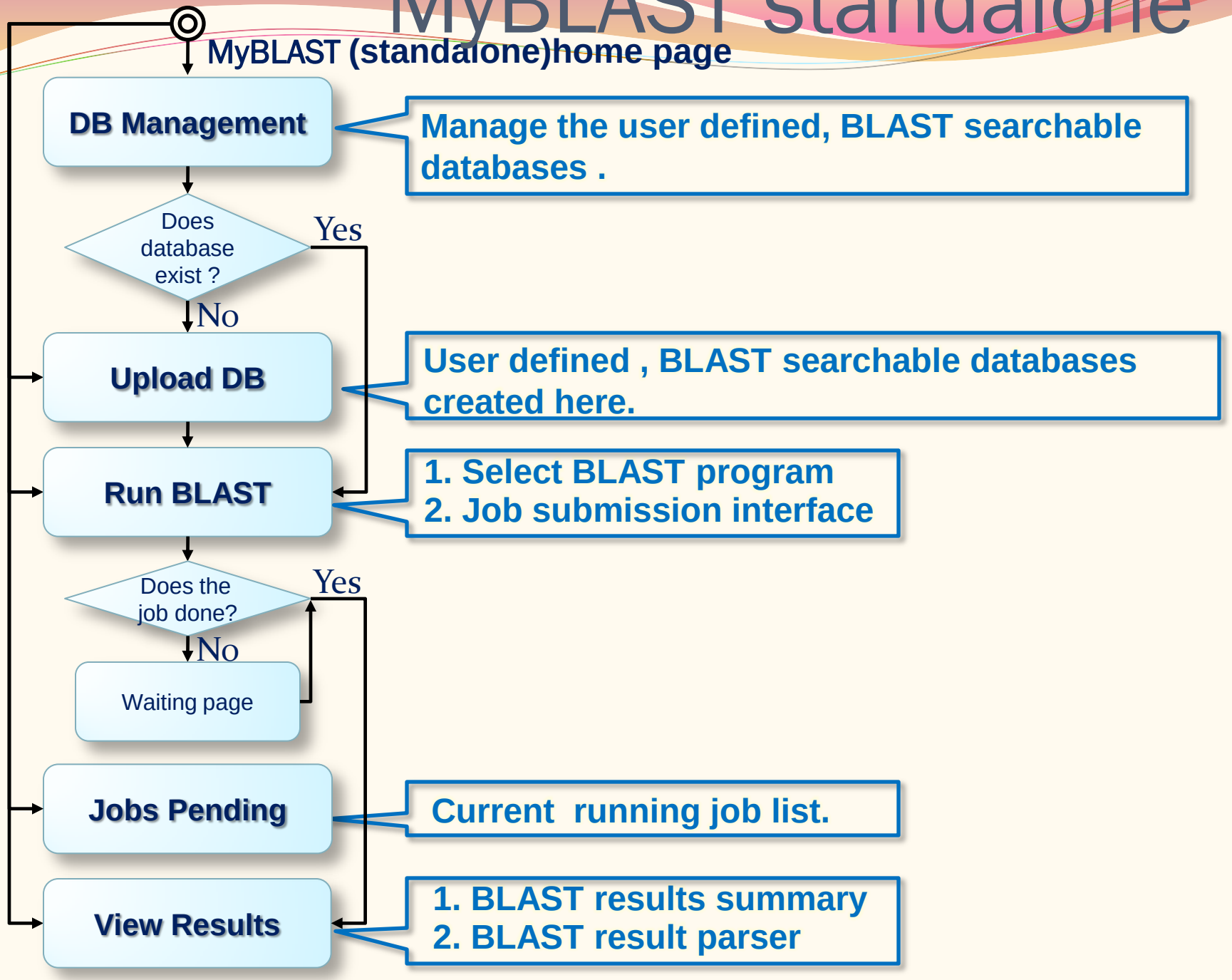
2.At the end, the main page and the control panel start automatically.



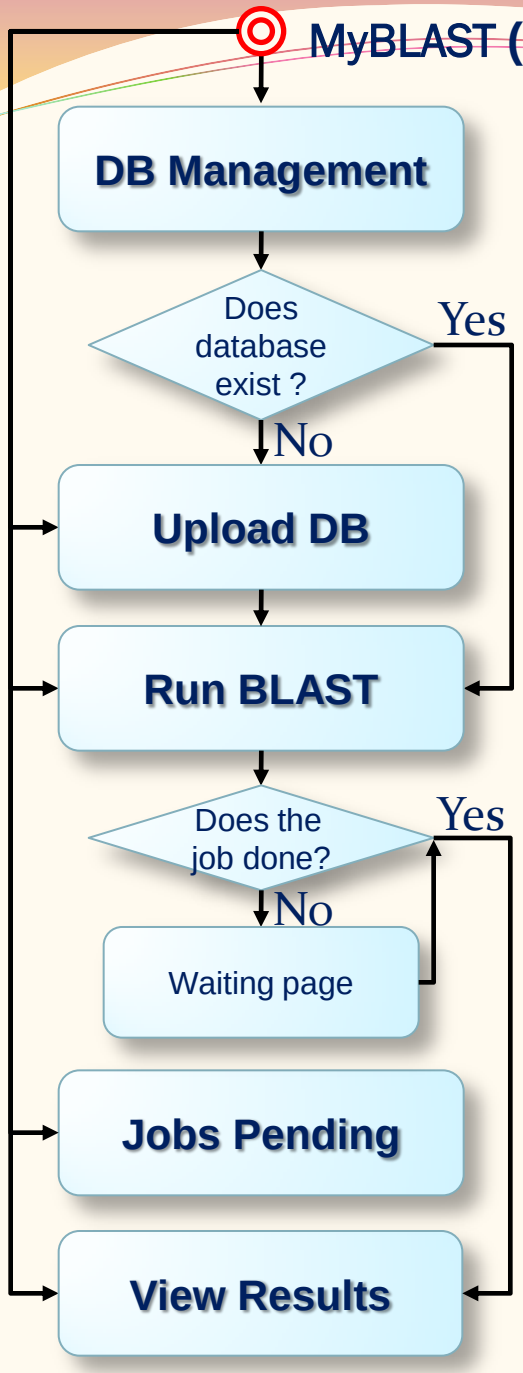
Information of web link

- Online website: (use MyBLAST online)
  http://mybioweb.nhri.org.tw/myblast
- Standalone version: (need to install MyBLAST in your computer)
  http://eln.iis.sinica.edu.tw
- MyBLAST video demo : (watch online)
  http://eln.iis.sinica.edu.tw

MyBLAST standalone



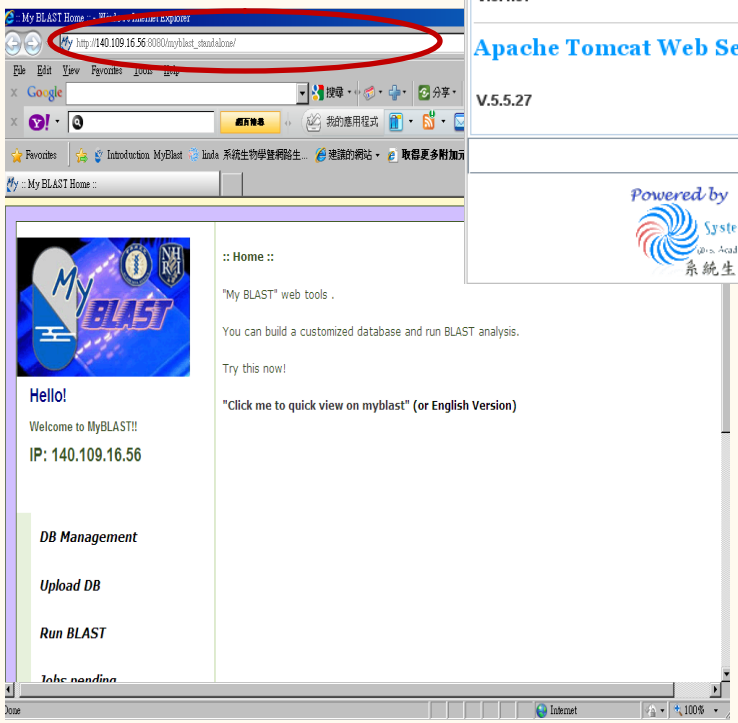
MyBLAST (standalone)home page



Start point : Go to MyBLAST home page.

1. Typing URL
http://localhost:8080/myblast_standalone/

2. Or by clicking the **MyBlast Home** button in the MyBlastControlPanel.



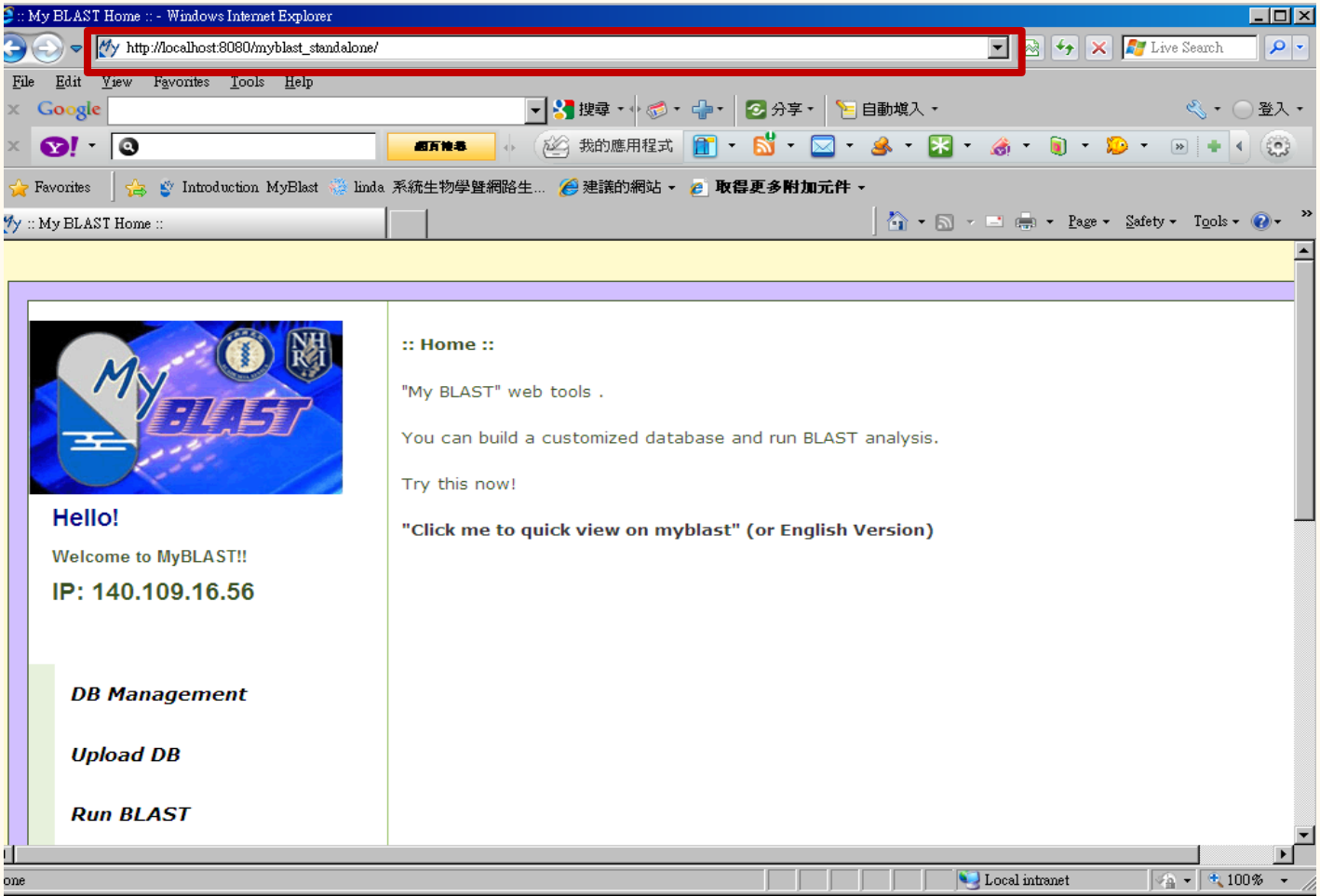
MyBLAST (standalone)home page

Start point : Go to MyBLAST home page.

1. Typing URL

http://localhost:8080/myblast_standalone/

2. Or by clicking the **MyBlast Home** button in the MyBlastControlPanel.

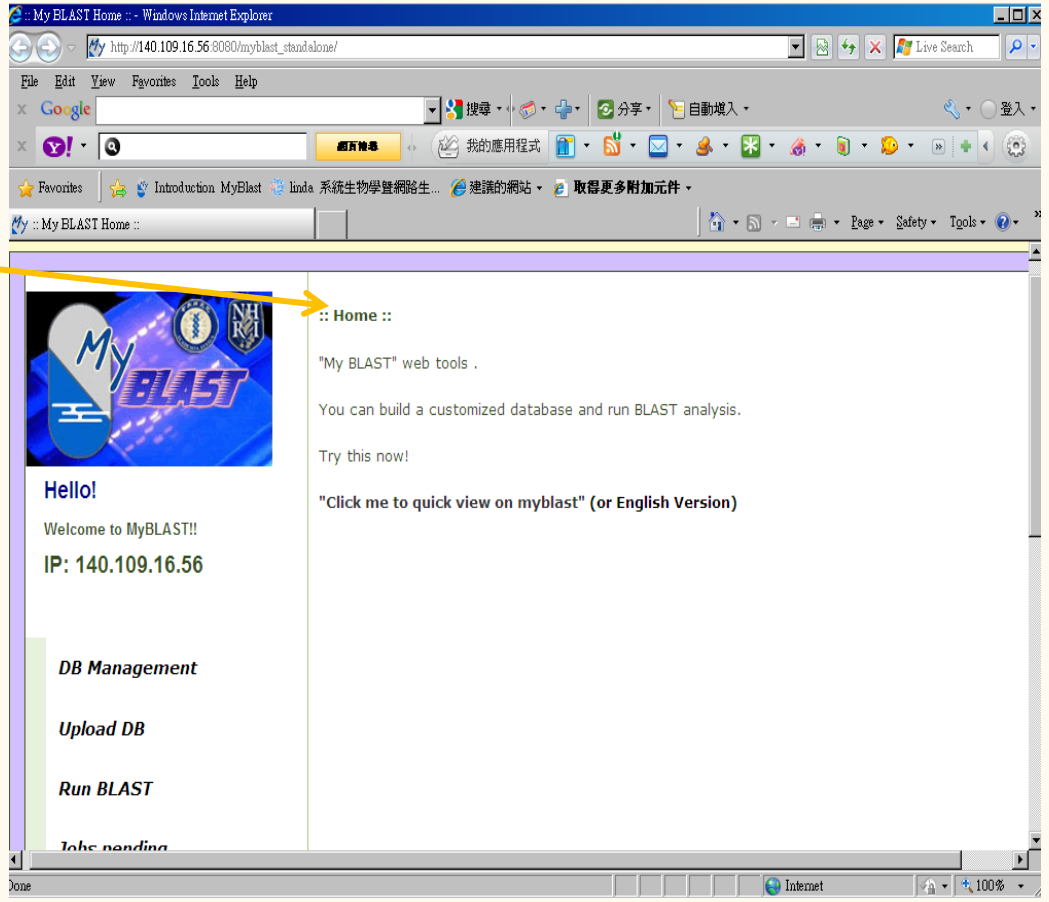


MyBLAST (standalone)home page

Start point : Go to MyBLAST home page.

1. Typing URL **http://localhost:8080/myblast_standalone/**

2. Or by clicking the **MyBlast Home** button in the MyBlastControlPanel. **(Don't need to type URL.)**



MyBLAST (standalone)home page

① MyBLAST home link
② DB Management link
③ Upload DB link
④ Run BLAST link
⑤ Jobs pending link
⑥ View Results link
⑦ User Guide link
⑧ Contact Us link
⑨ Flash Demo link
⑩ Websites link
⑪ Lab of System Biology & Network Biology website link

①

Hello!

Welcome to MyBLAST!!

IP: 140.109.16.56

DB Management ②

Upload DB ③

Run BLAST ④

Jobs pending ⑤

View Results ⑥

User Guide ⑦

Contact Us ⑧

:: Home ::

"My BLAST" web tools .

You can build a customized database and run BLAST analysis.

Try this now!

⑨

"Click me to quick view on myblast" (or English Version)

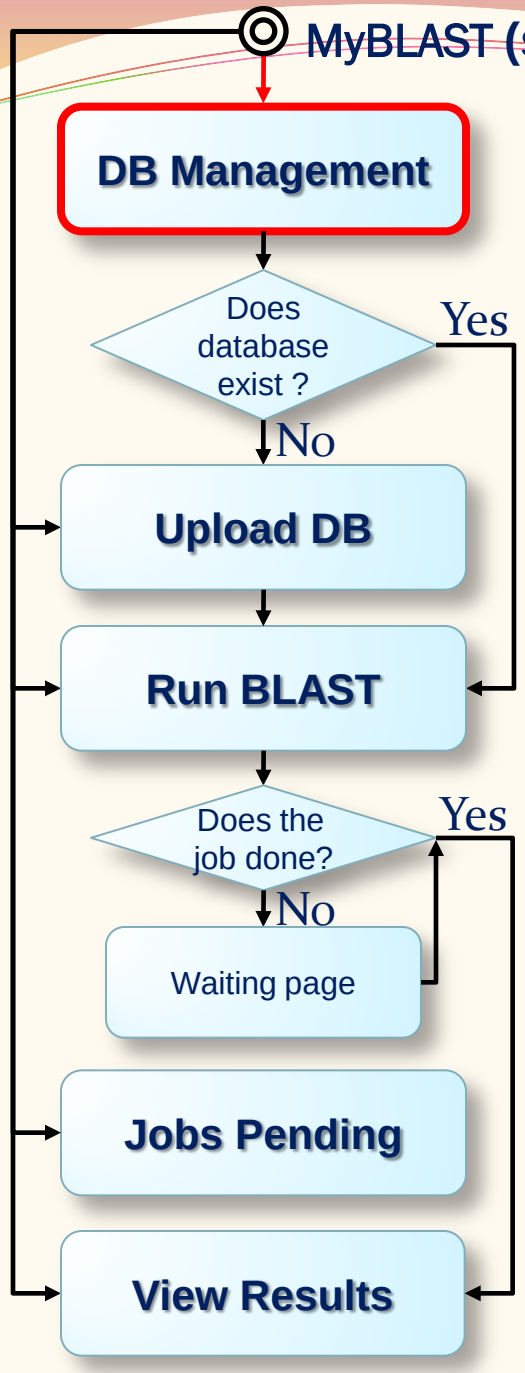
⑩

Lab of Systems Biology and Network Biology, Institute of Information Science, Academia Sinica, Taipei, TAIWAN &

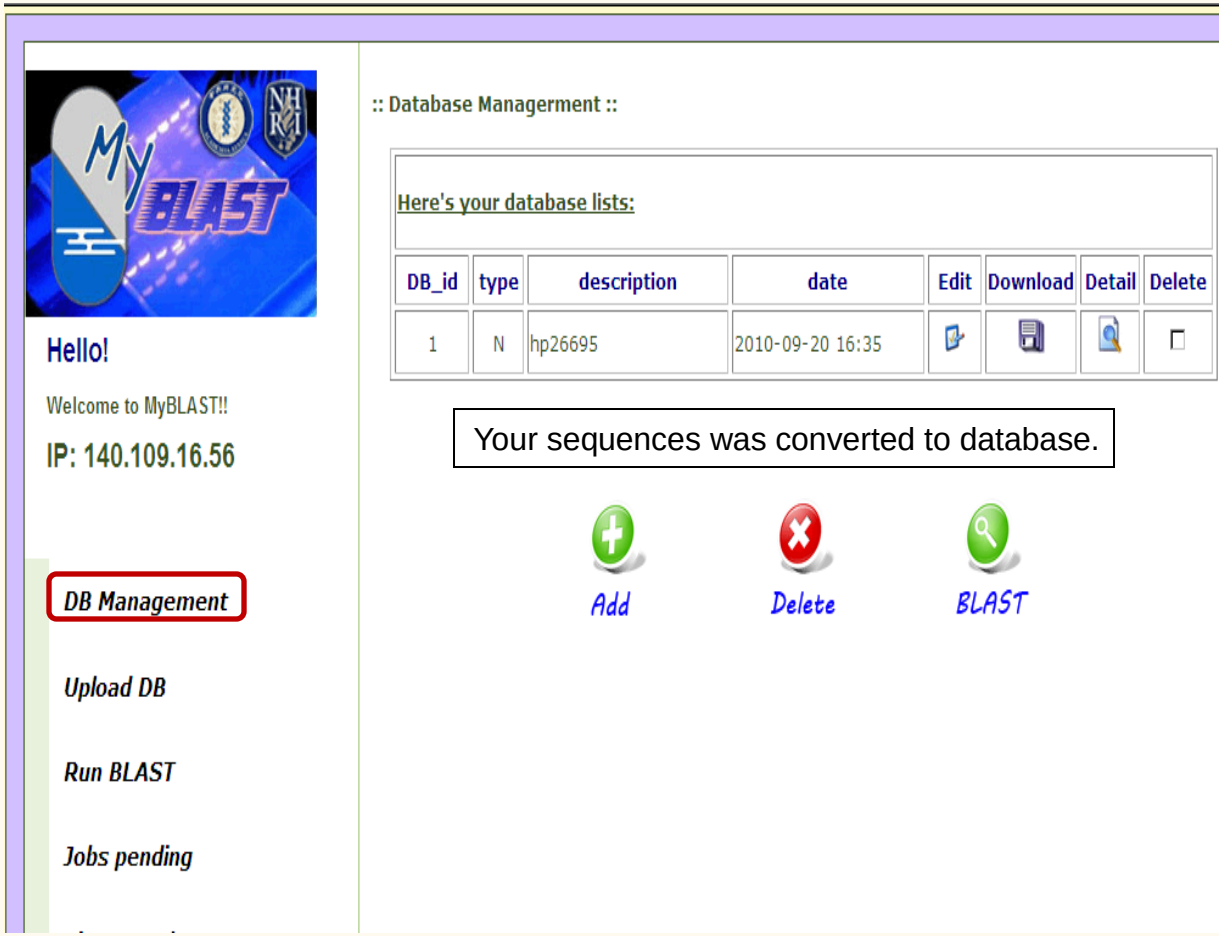
Division of Biostatistics and Bioinformatics, National Health Research Institutes, Miao-Li, TAIWAN © 2009 All rights

LAB OF System Biology & Network Biology
中央研究院資訊科學研究所 @iis, Academia Sinica, T

MyBLAST (standalone)home page



Manage the user defined, BLAST searchable databases .



DB Management

Manage the user defined, BLAST searchable databases .

① Edit icon: Modify the particular uploaded DB data.

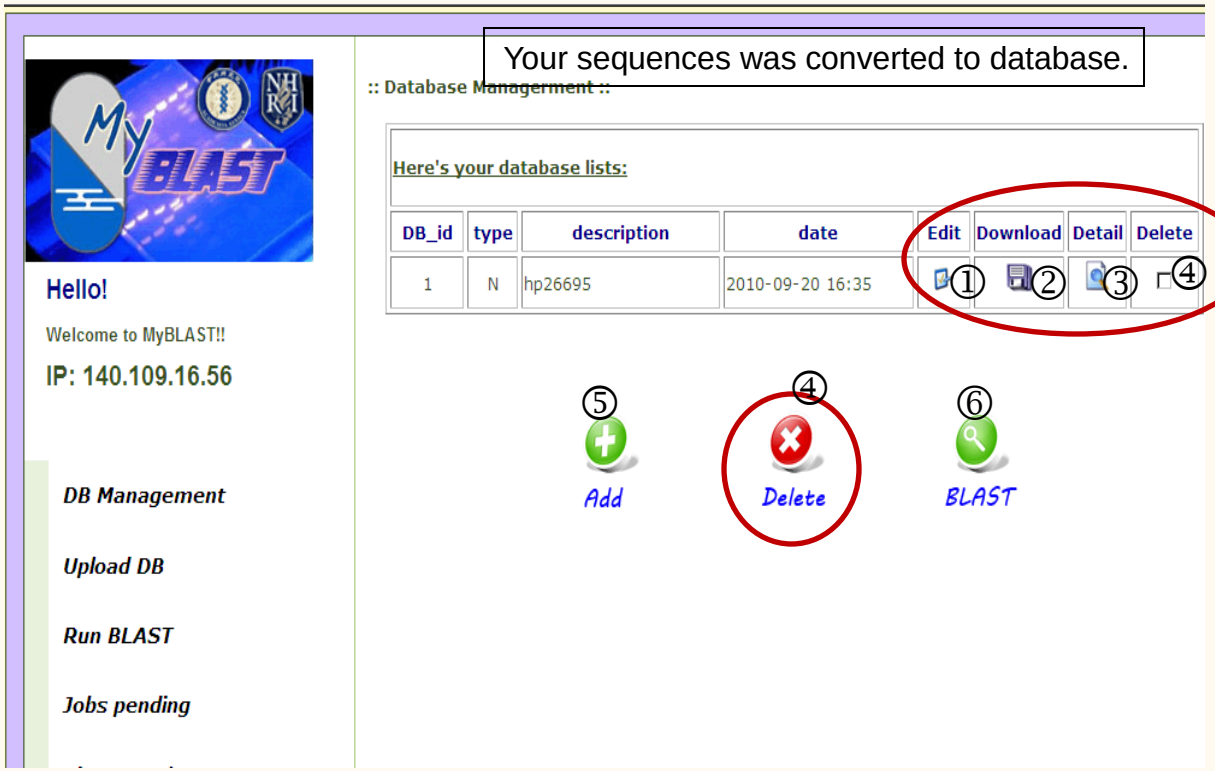
② Download icon: Download the particular uploaded DB FASTA format file.

③ Detail icon: view the BLAST command and database log.

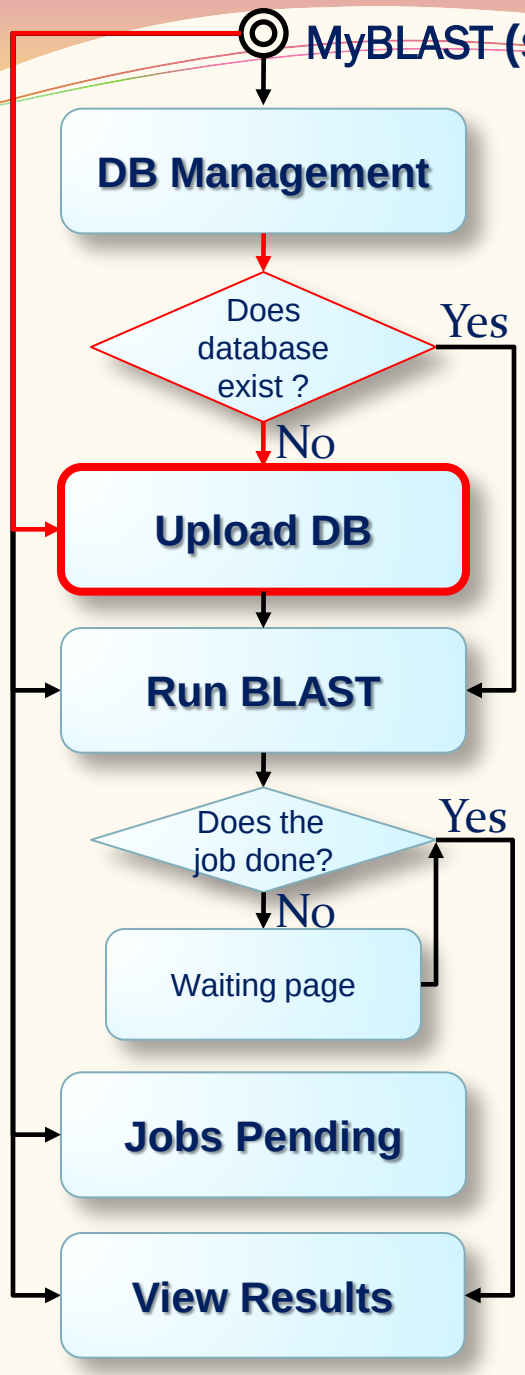
④ Delete checkbox: Check to delete the particular uploaded DB. Click the "Delete" button to start to do the DB deletion.

⑤ Link to Upload DB: Add a database.

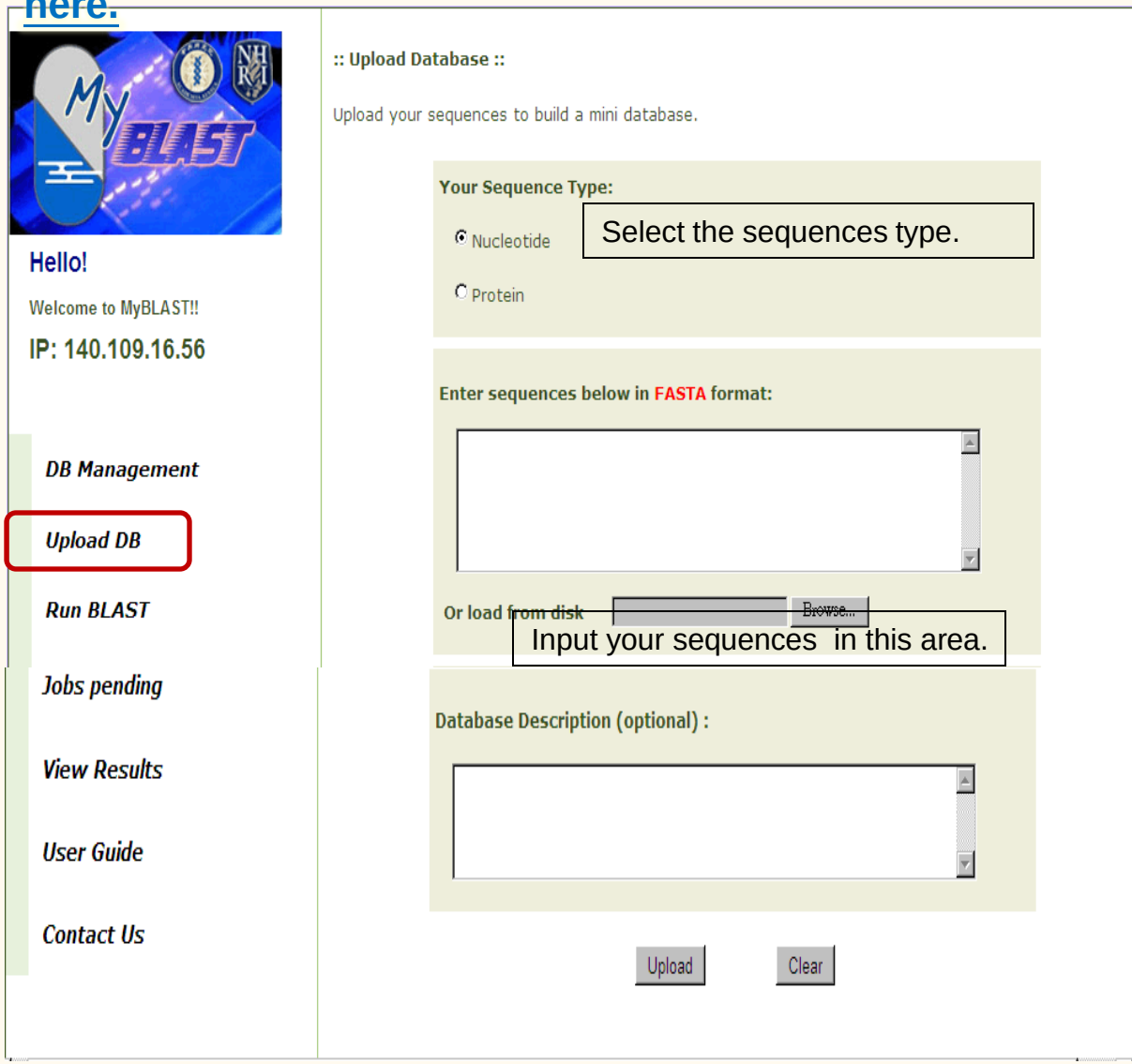
⑥ Link to Run BLAST: Run a BLAST program



MyBLAST (standalone)home page



User defined , BLAST searchable databases created here.



Upload DB

User defined , BLAST searchable databases created here.

①Sequence Type: Nucleotide or protein.

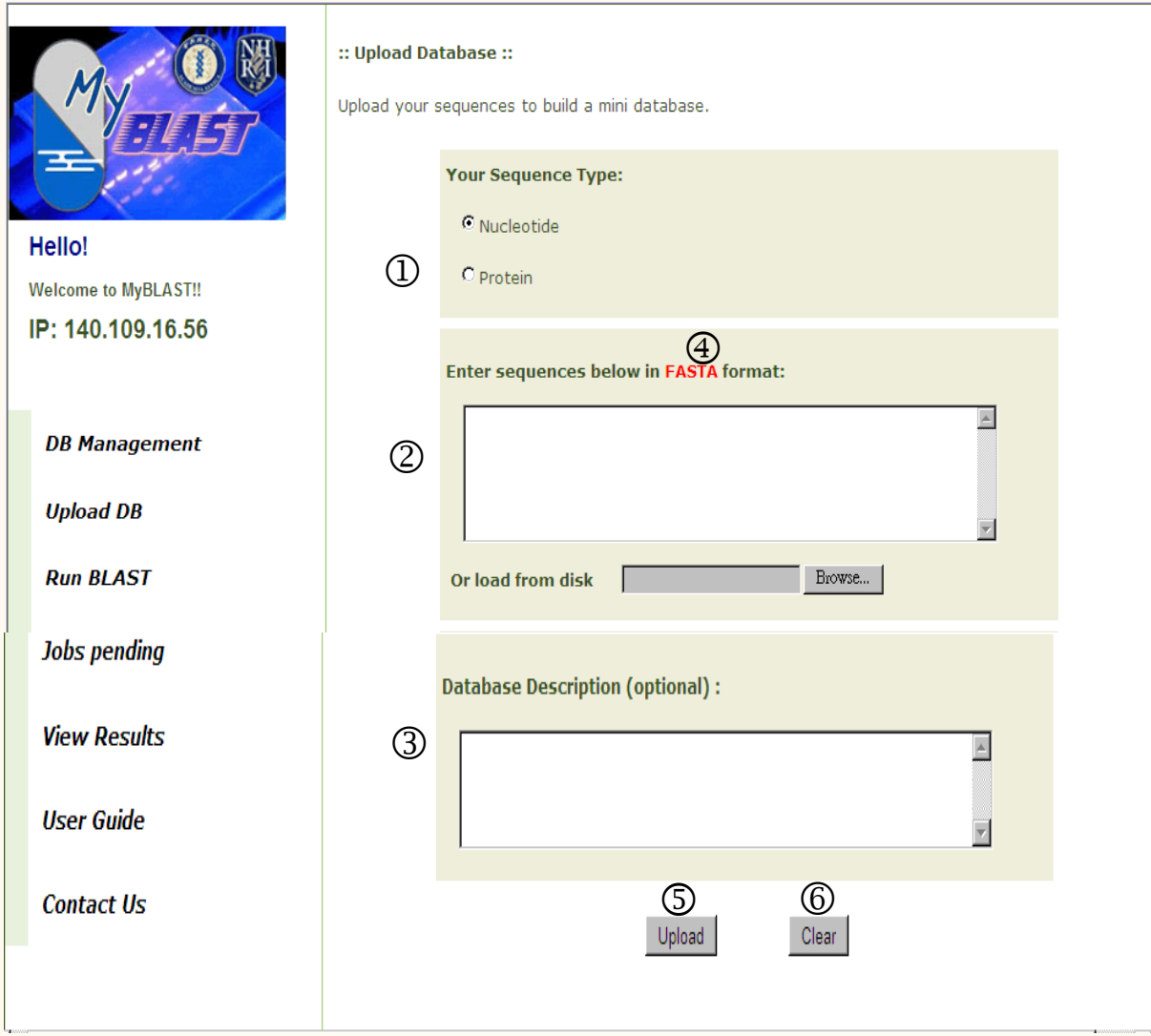
②Enter sequences:
Copy and paste the FASTA format to the text area.
Or enter the full file path and file name.

③Description: simple notes to this database.

④FASTA format description link

⑤Upload button: Click to send out the form.

⑥Clear button: clean up the form data.



Upload DB

User defined , BLAST searchable databases created here.

Step1. Fill in the form, then submit it.
Step2. Wait until the process is done.

Your Sequence Type:

- ○ Nucleotide
- ◉ Protein

Enter sequences below in FASTA format:

Or load from disk D:\seq\protein\hp26695.fa: Browse...

Database Description (optional) :

hp26695

Upload Clear

Click the Upload button

Waiting page

Your sequence is now converting, the results will be shown later....

A result will be shown on DATABASE RESULT PAGE after finishing.

Note: The job is running in batch mode, and it's ok to close this window.

Show the DB_id

[Back Home] [DB Managerment] [Upload DB] [Run BLAST] [View Results]

Database Conversion Finished !!

Your sequences was building a mini database now.

And here is the DB_id: 3

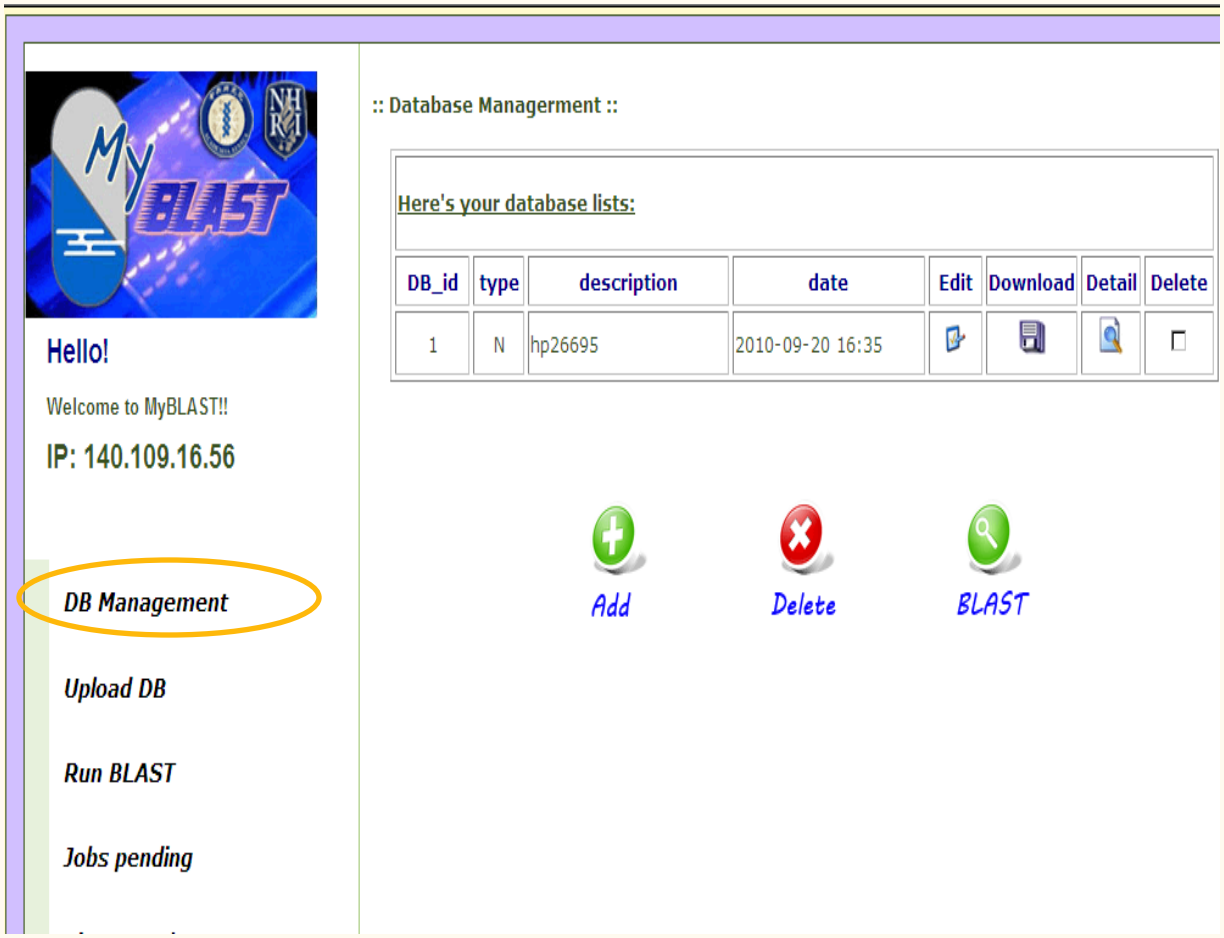
You can go here to see database result.

http://localhost/MainMenu

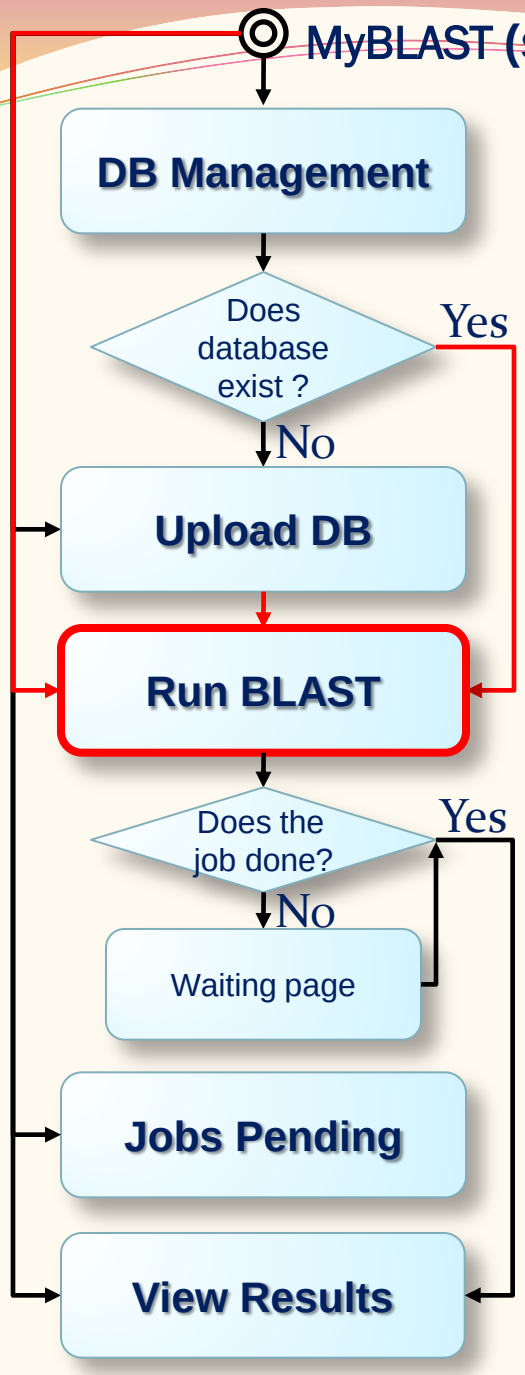
Or Run BLAST directly!

http://localhost/MainMenu

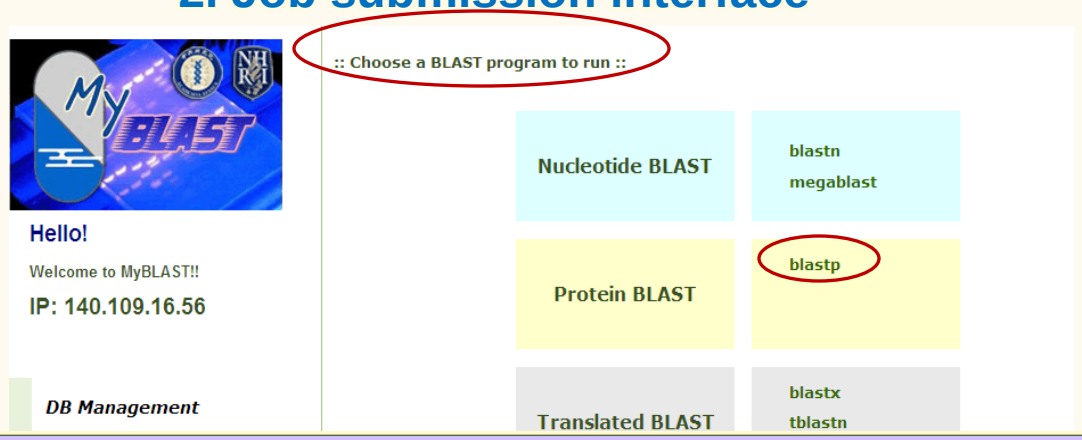
You may want to go to DB Management to view the database.

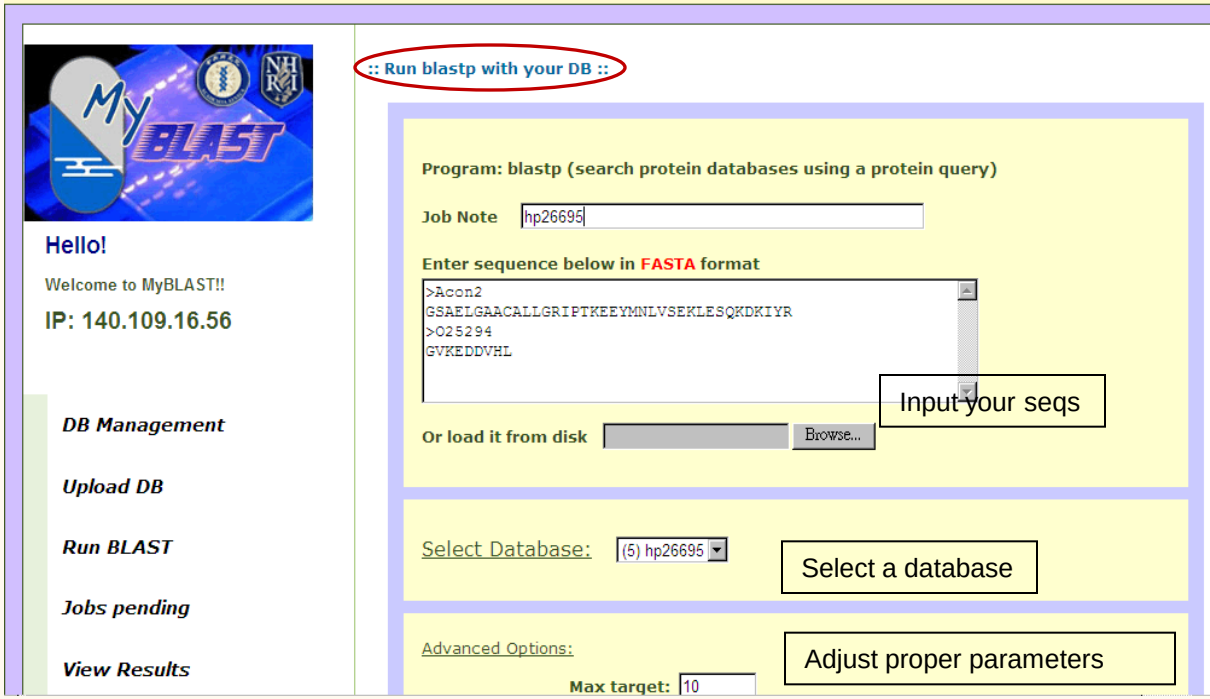


MyBLAST (standalone)home page



1. Select BLAST program
2. Job submission interface

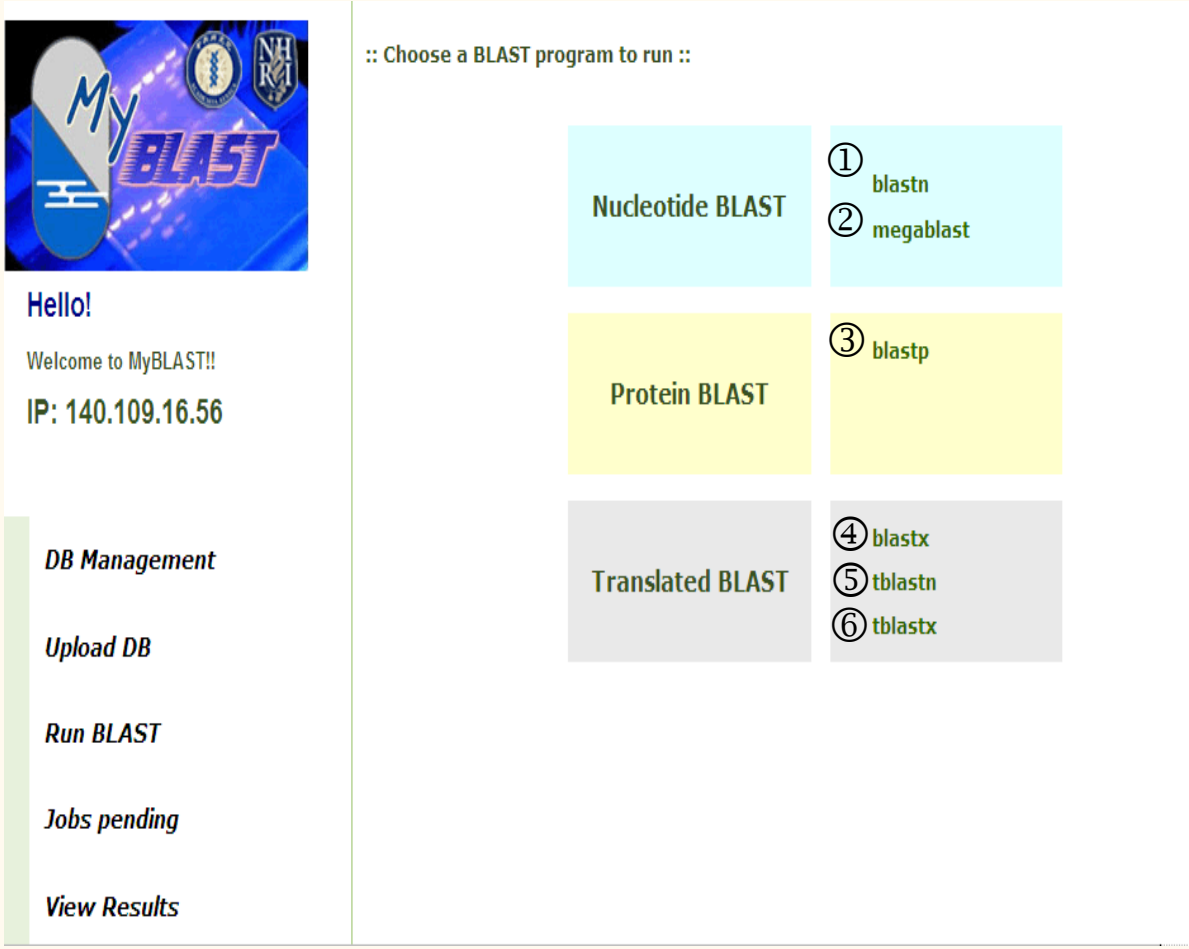




Run BLAST

1. *Select BLAST program*
2. *Job submission interface*

① blastn link
② megablast link
③ blastp link
④ blastx link
⑤ tblastn link
⑥ tblastx link

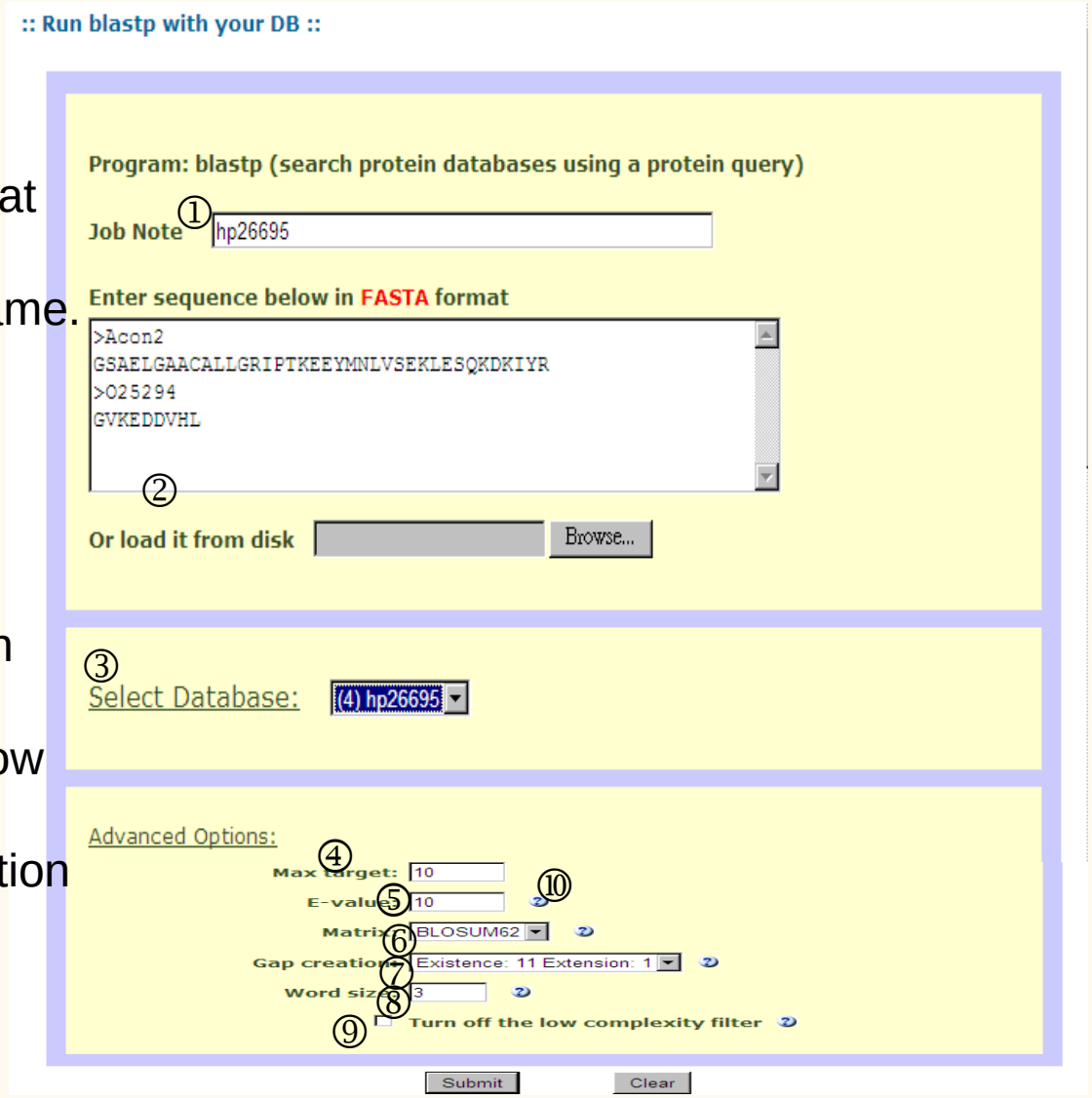


Run BLAST

1. *Select BLAST program*
2. *Job submission interface*

① Job note
② Copy and paste the FASTA format to the text area.
Or enter the full file path and file name.

③ Select Database
④ Advanced Options: Max target
⑤ Advanced Options: E-Value
⑥ Advanced Options: Matrix
⑦ Advanced Options: Gap Creation
⑧ Advanced Options: Word size
⑨Advanced Options: Turn off the low complexity filter
⑩ question icon: link to the description of the particular parameter.

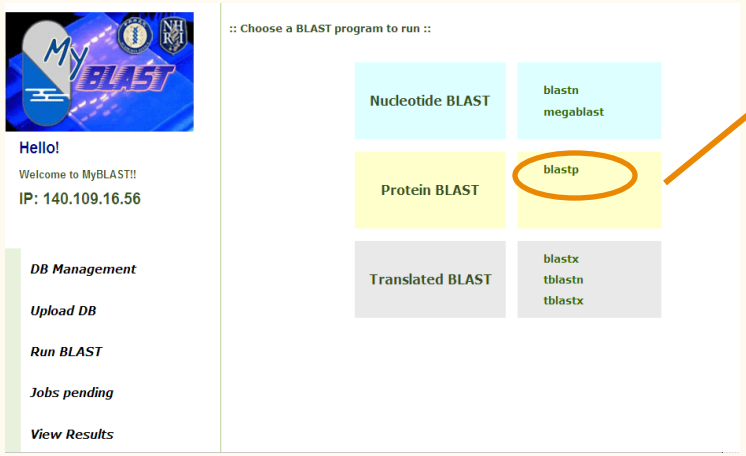


Run BLAST

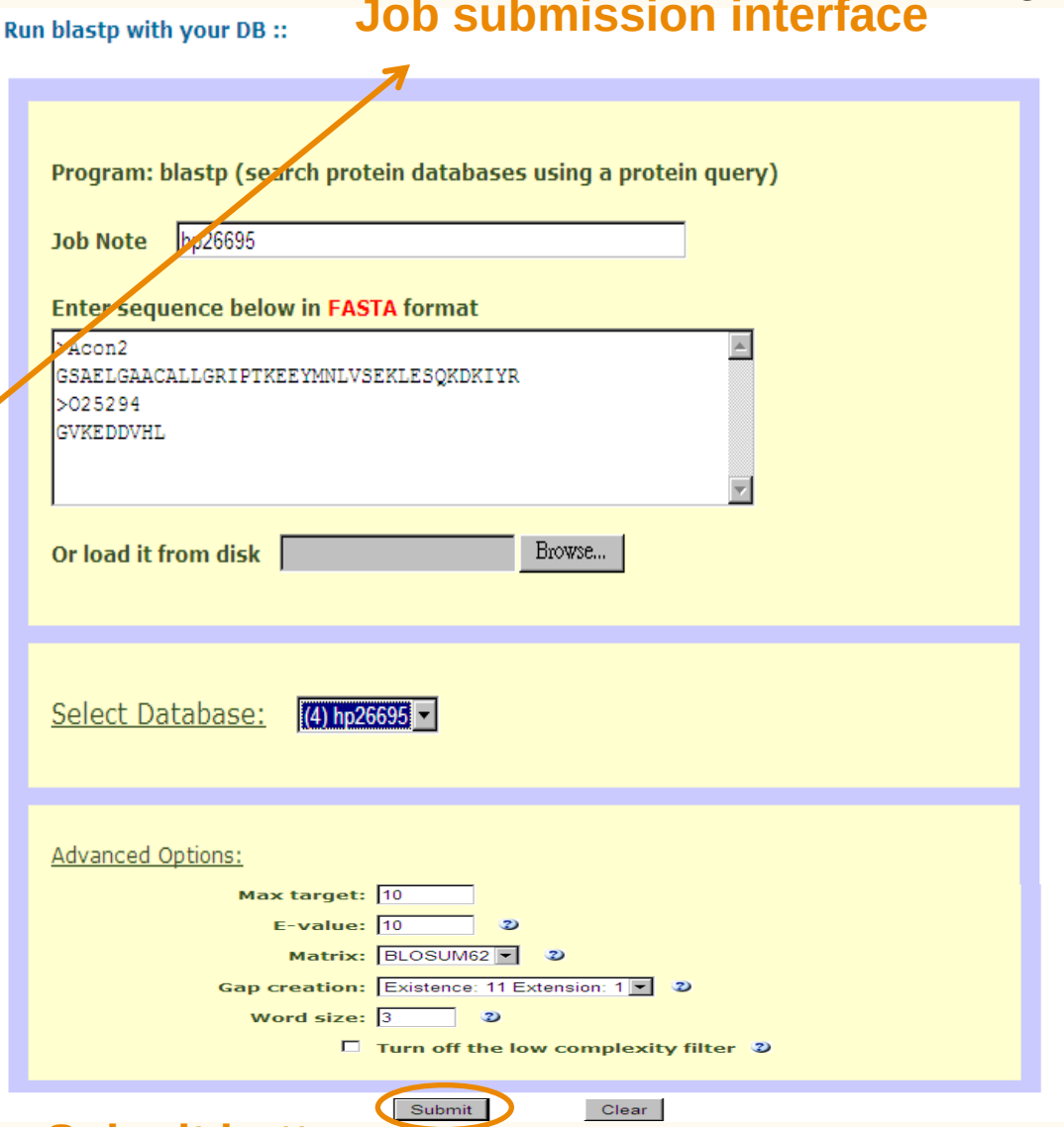
1. *Select BLAST program*
2. *Job submission interface*

Step1. Fill in the form, then submit it.
Step2. Wait until the process is done.

Select BLAST program



Job submission interface



Click the Submit button

Run BLAST

Step1. Fill in the form, then submit it.
Step2. Wait until the process is done.

Waiting page

Now is working, the BLAST results will be shown later....

The BLAST result will be shown on BLAST RESULT PAGE after finishing.

Note: The job is running in batch mode, and it's ok to close this window.

Show result

[Back Home] [Upload D

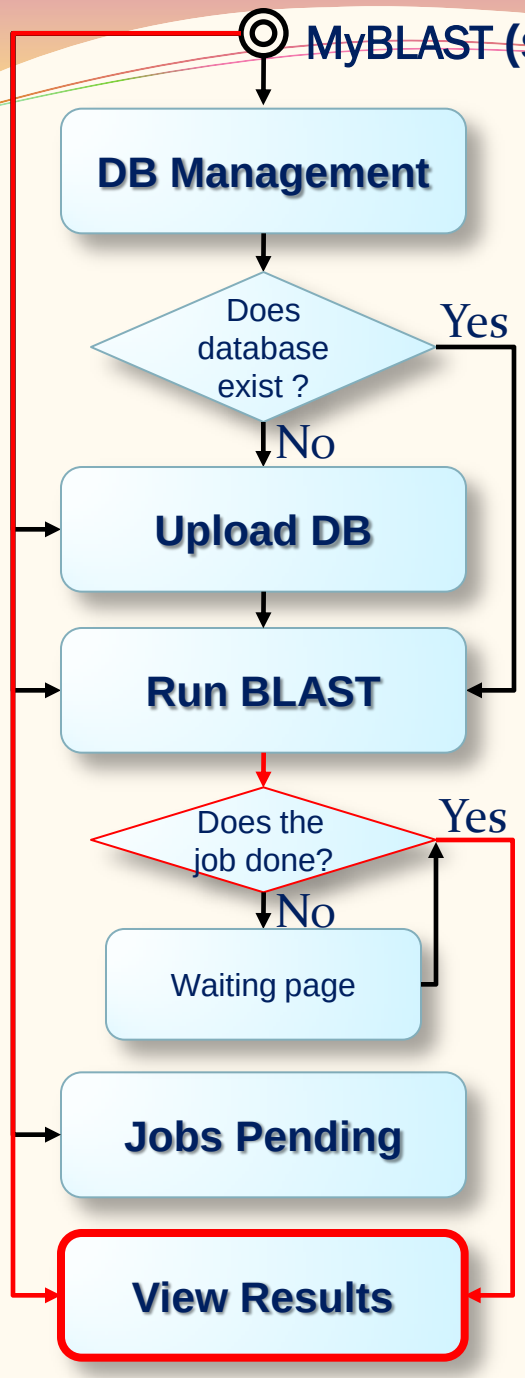
:: MyBLAST Results ::

DB description: (4) hp26695
Submit description: (3) hp26695
Download Output Files (text file) or (csv file)

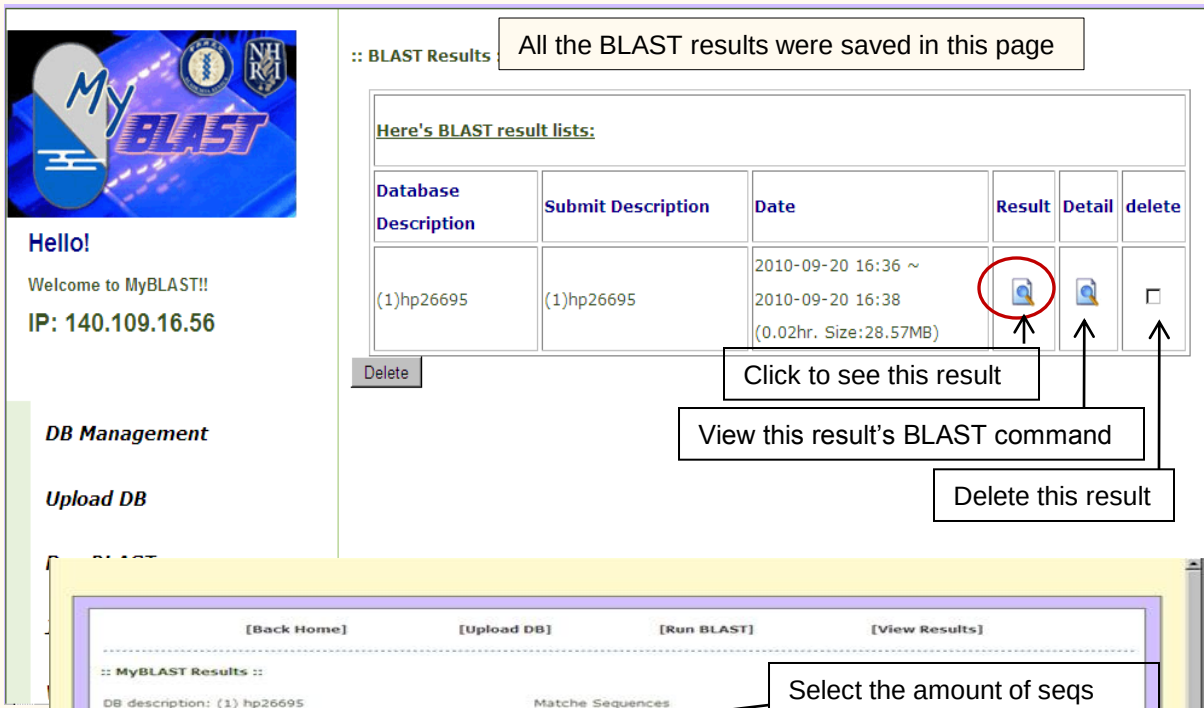
Matche Sequences: Top 3 | Go | Download

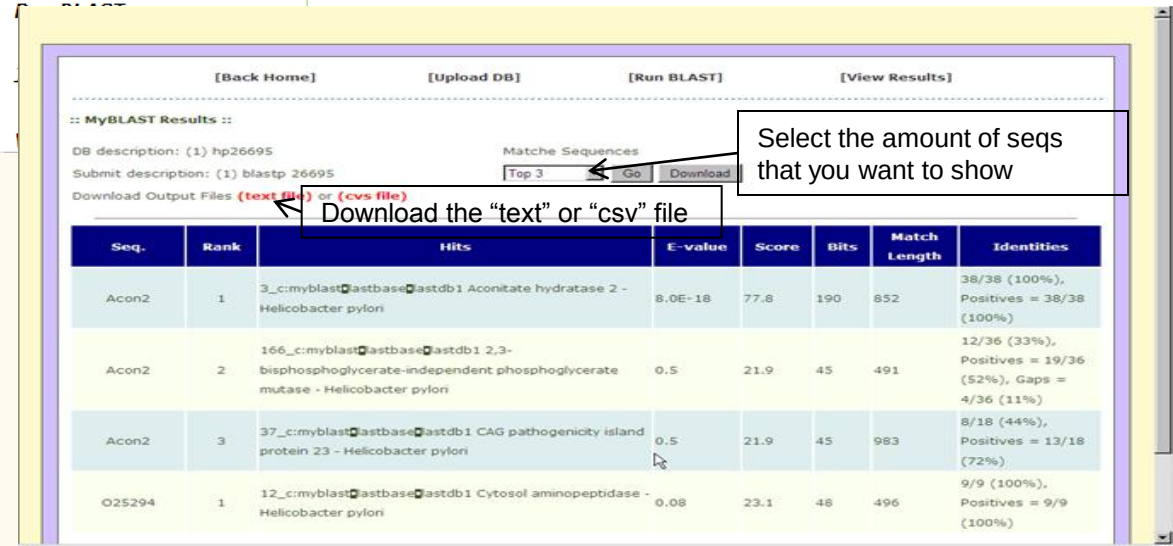
Seq.	Rank	Hits	E-value	Score	Bits	Match Length	Identities
Acon2	1	3_D:myblast□lastbase□lastdb4 Aconitate hydratase 2 - Helicobacter pylori	5.0E-17	77.8	190	852	38/38 (100%), Positives = 38/38 (100%)
Acon2	2	484_D:myblast□lastbase□lastdb4 UvrABC system protein C - Helicobacter pylori	3.1	21.9	45	594	9/19 (47%), Positives = 13/19 (68%)
Acon2	3	37_D:myblast□lastbase□lastdb4 CAG pathogenicity island protein 23 - Helicobacter pylori	3.1	21.9	45	983	8/18 (44%), Positives = 13/18 (72%)

MyBLAST (standalone)home page



1. BLAST results summary
2. BLAST result parser

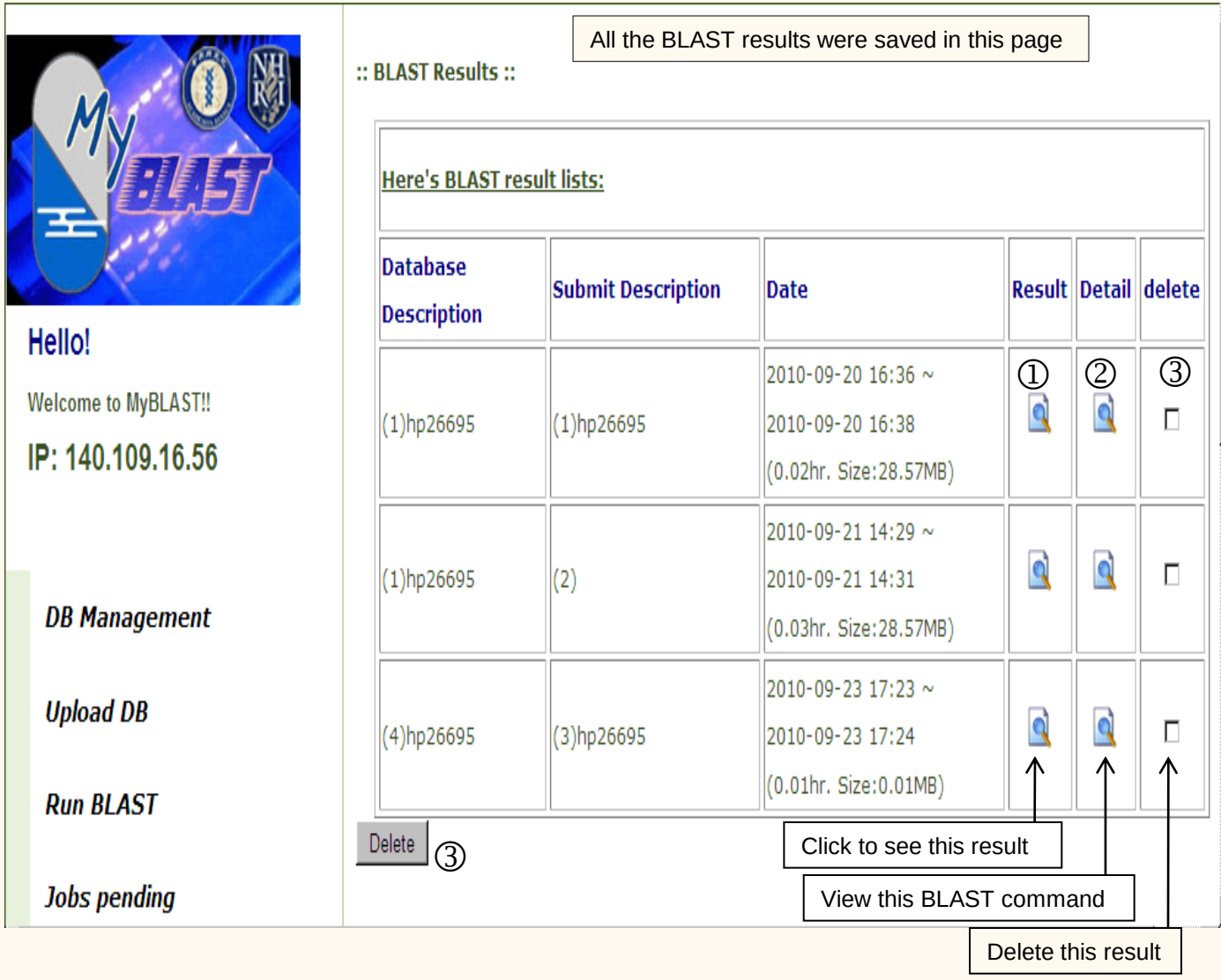




View Results

1. BLAST results summary
2. BLAST result parser

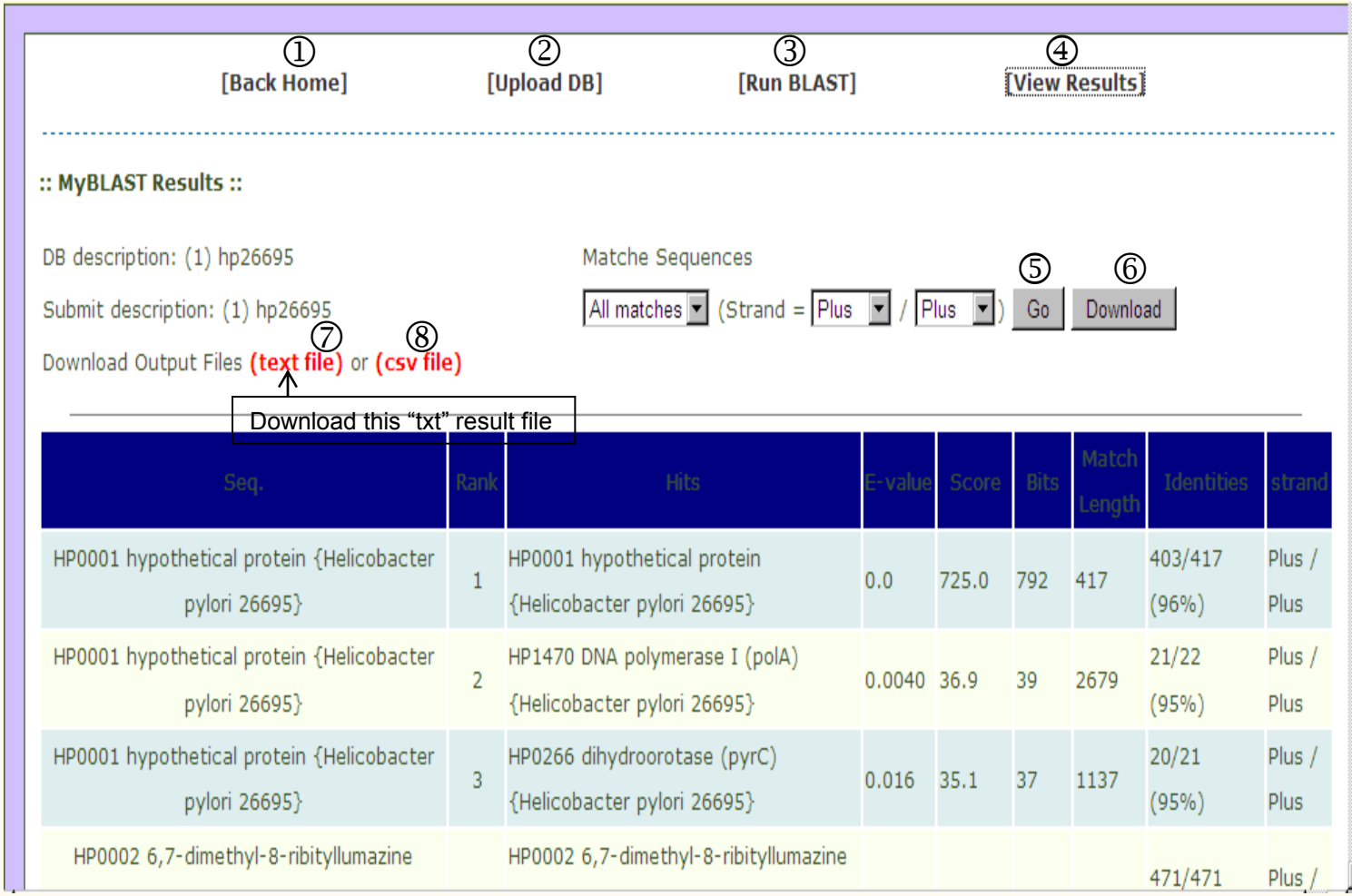
① Result parser link
② Log link.
③ Delete checkbox: Check to delete the particular result.
Click the “Delete” button to start to do the result deletion.



View Results

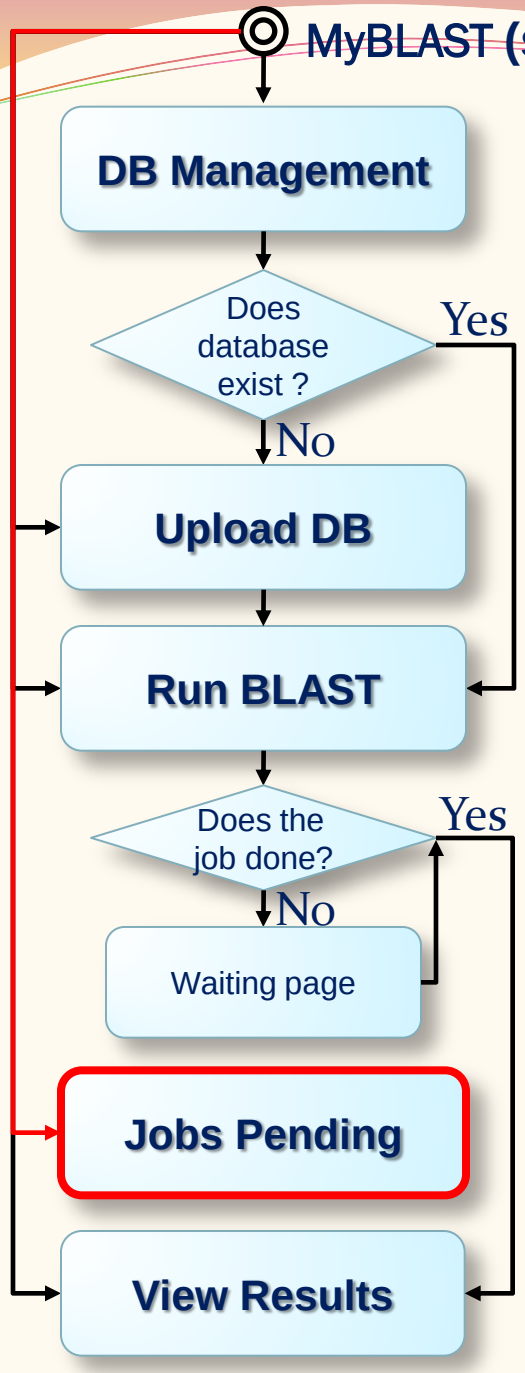
1. BLAST results summary
2. BLAST result parser

① Home link
② Upload DB link
③ Run BLAST link
④ Result Summary link
⑤ Filter
⑥ Filtered Result download
⑦ Output text file download
⑧ Output csv file download

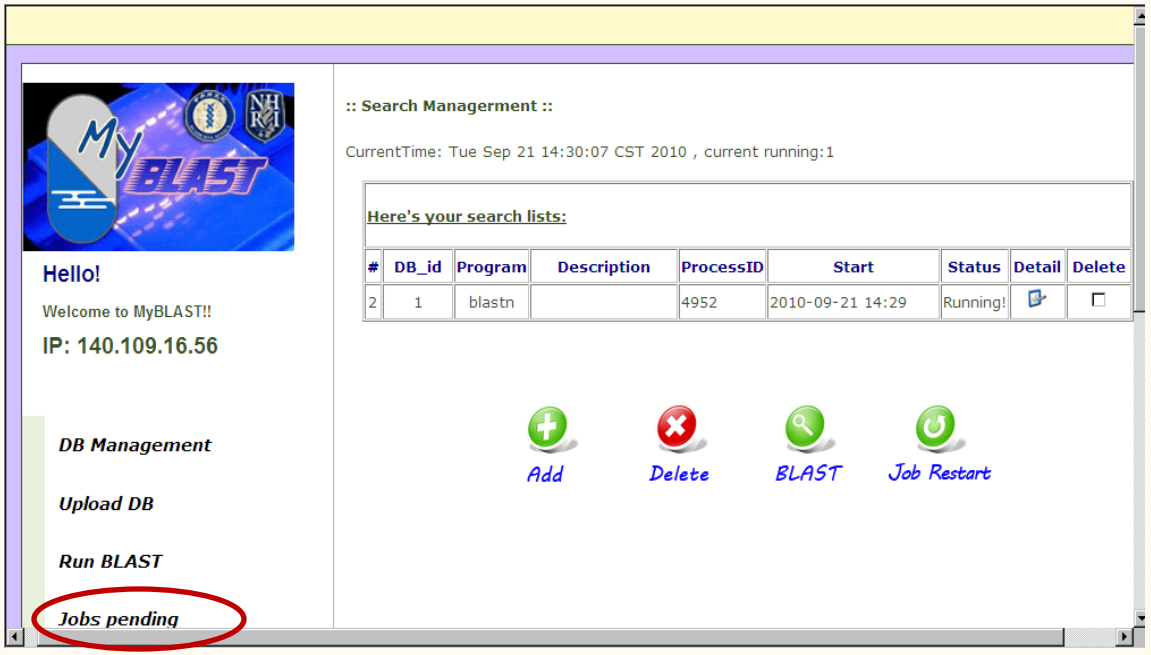


Seq.	Rank	Hits	E-value	Score	Bits	Match Length	Identities	strand
HP0001 hypothetical protein {Helicobacter pylori 26695}	1	HP0001 hypothetical protein {Helicobacter pylori 26695}	0.0	725.0	792	417	403/417 (96%)	Plus / Plus
HP0001 hypothetical protein {Helicobacter pylori 26695}	2	HP1470 DNA polymerase I (polA) {Helicobacter pylori 26695}	0.0040	36.9	39	2679	21/22 (95%)	Plus / Plus
HP0001 hypothetical protein {Helicobacter pylori 26695}	3	HP0266 dihydroorotase (pyrC) {Helicobacter pylori 26695}	0.016	35.1	37	1137	20/21 (95%)	Plus / Plus
HP0002 6,7-dimethyl-8-ribityllumazine		HP0002 6,7-dimethyl-8-ribityllumazine					471/471	Plus /

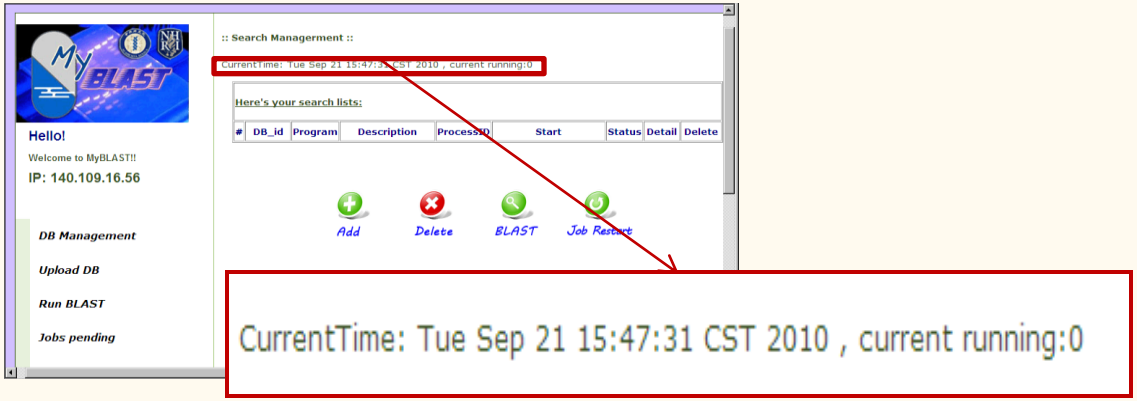
MyBLAST (standalone)home page



Current running job list.



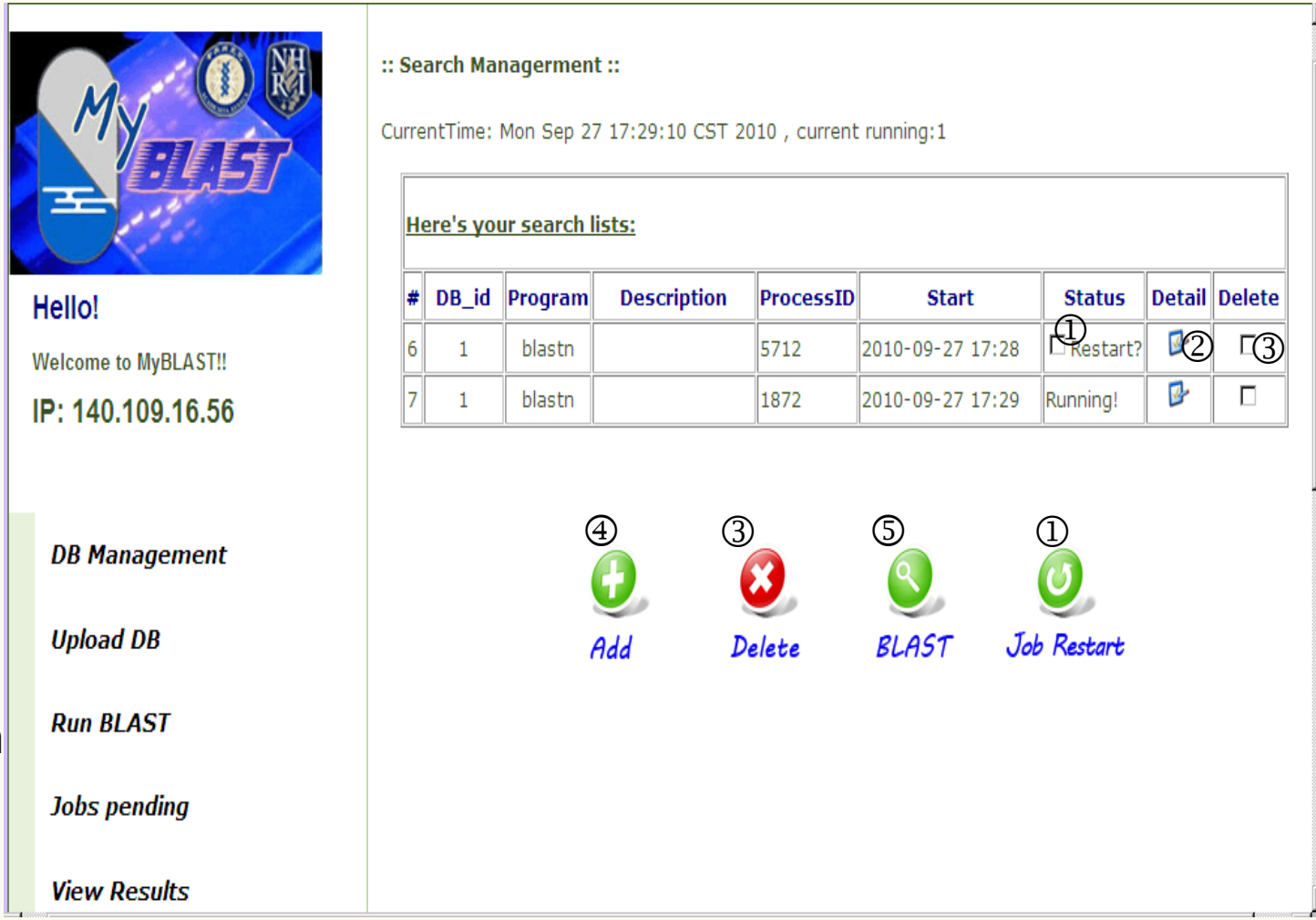
To get the newest running status , you'll need to click the browser's refresh button.



Jobs Pending

Current running job list.

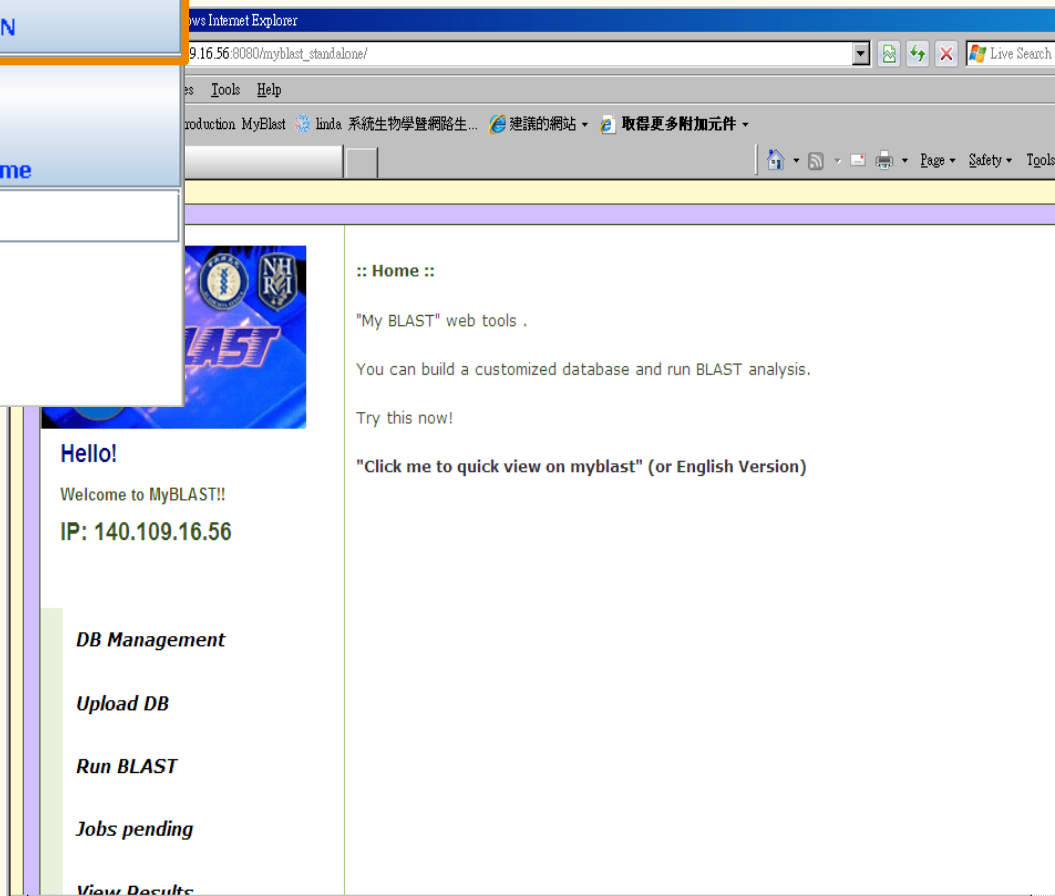
① Restart checkbox: Check to restart the particular job. Click the “Job Restart” button to start to do the BLAST again.
② Log link
③ Delete checkbox: Check to delete the particular job. Click the “Delete” button to start to stop the job and delete the search job.
④ Run BLAST link
⑤ View Results link



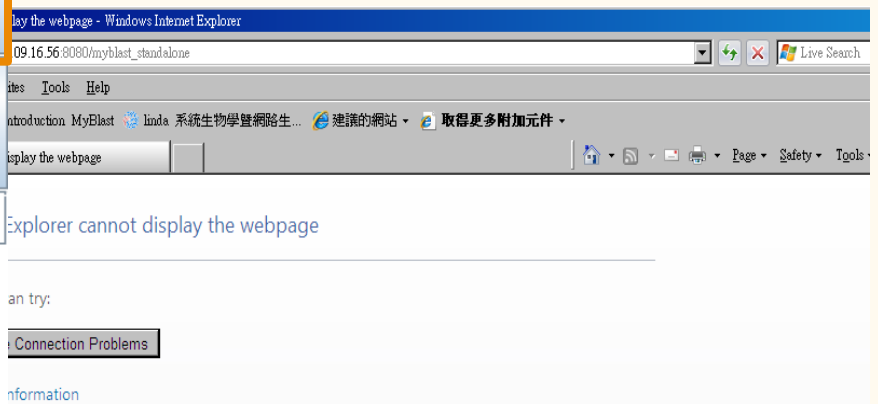
MyBLAST Control Panel

① Link to import Mysql database data and BLAST data page
② Link to backup Mysql database data and BLAST data page
③ Link to change port page
④ Link to help page
⑤ Shutdown or Restart button
⑥ MyBLAST Home page button
⑦ Link to lab home page
⑧ MySQL Community Server running status
⑨ Apache Tomcat Web Server running status
⑩ Message text area

- **Try to restart or shutdown Servers**

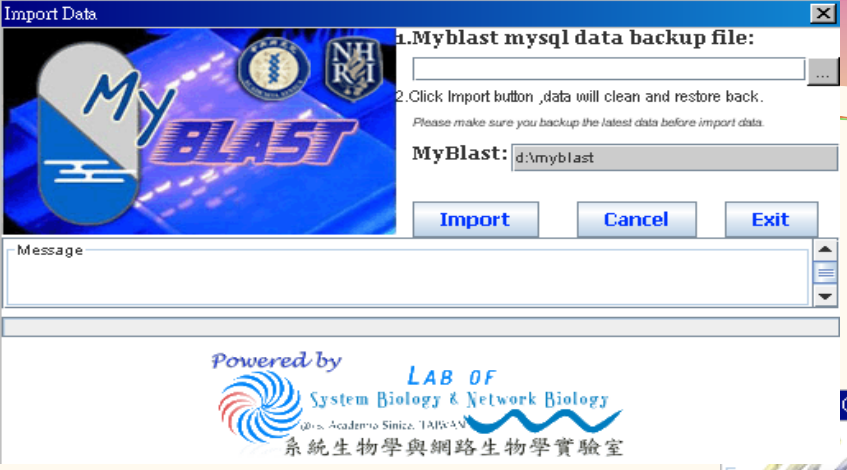


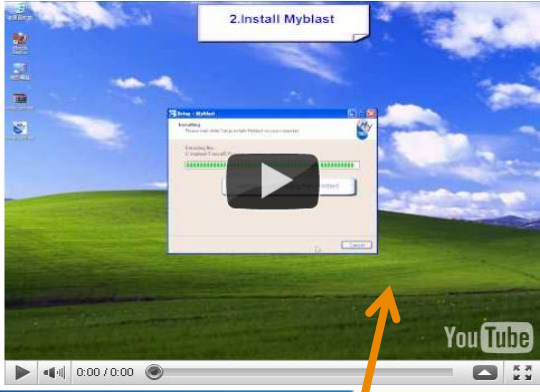
- **Try to restart or shutdown Servers**



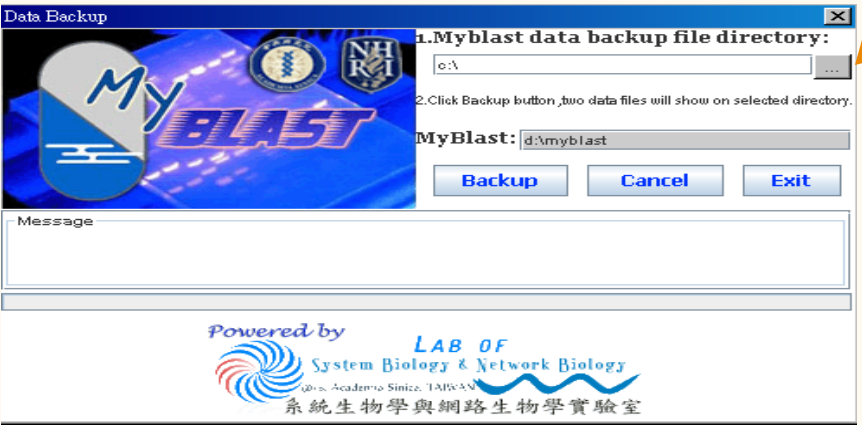
- **Close MyBLAST Control Panel**

Message


Import Data
1.Myblast mysql data backup file:
2.Click Import button ,data will clean and restore back.
Please make sure you backup the latest data before import data.
MyBlast: d:\myblast
Import
Cancel
Exit
Message
Powered by
LAB OF
System Biology & Network Biology
Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室


2.Install Myblast
YouTube
0:00 / 0:00

Control Panel
myBlast
MySQL Community Server
V.5.1.37
Apache Tomcat Web Server
V.5.5.27
Import
Backup
Tools
Help
SHUTDOWN
MyBlast Home
Powered by
LAB OF
System Biology & Network Biology
Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室


Data Backup
1.Myblast data backup file directory:
c:\
2.Click Backup button ,two data files will show on selected directory.
MyBlast: d:\myblast
Backup
Cancel
Exit
Message
Powered by
LAB OF
System Biology & Network Biology
Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室

Setting
Apache Tomcat Port
8080
(1~65536)
Mysql Port
3306
(1~65536)
view ports
Revert to defaults
Changing ports will require restarting MyBlast.
OK
Cancel
Message
Powered by
LAB OF
System Biology & Network Biology
Academia Sinica, TAIWAN
系統生物學與網路生物學實驗室

Example :

Get the FASTA data from After GenBank .

Upload DB, Run BLAST and View Results.

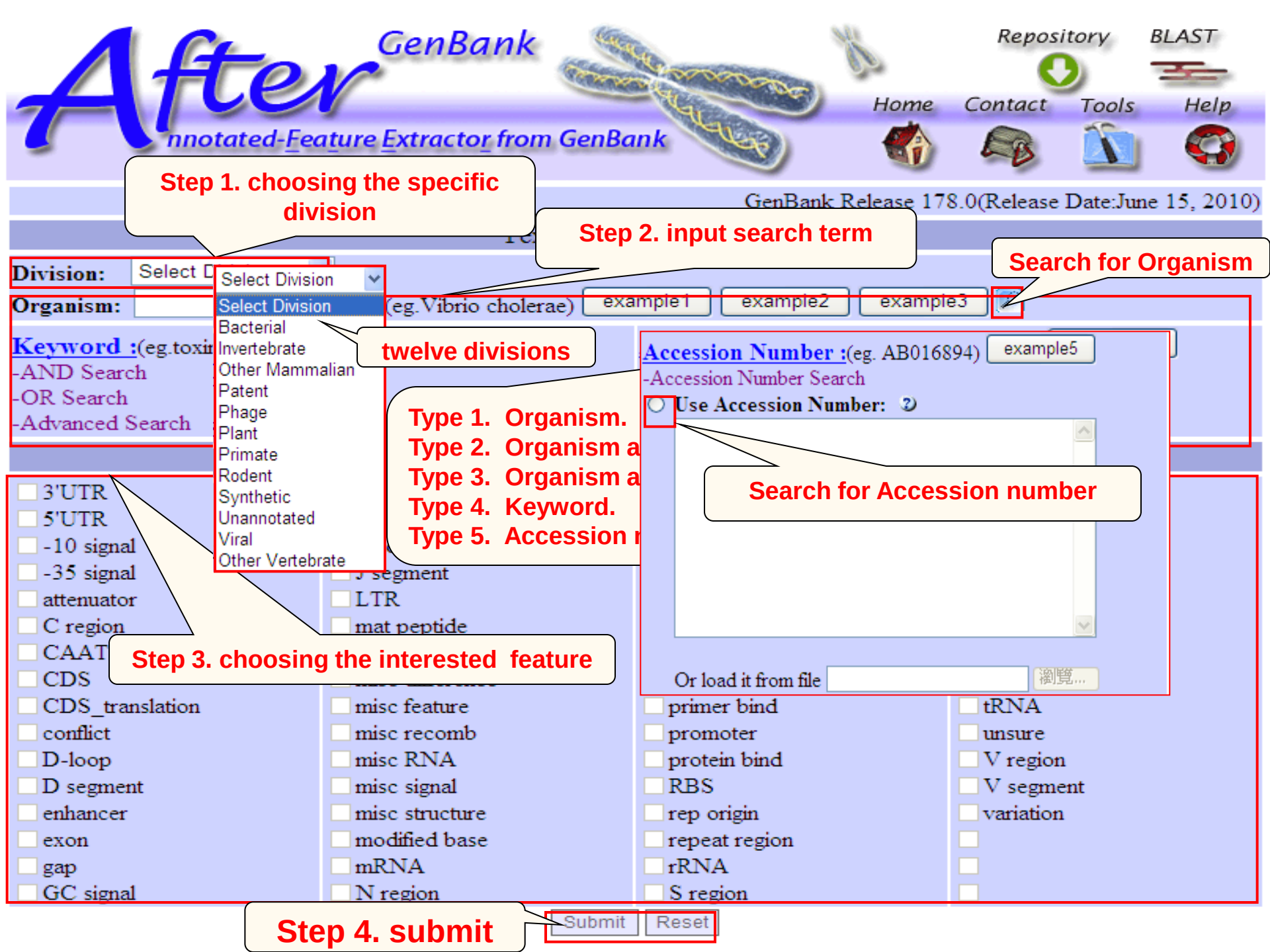
Motivation :: AfterGenBank

- A quick way to fetch the sequence features (5'UTR, 3' UTR, CDS, Translation) from GenBank only with mouse clicking not programming
- Provide most update collections for specific sequence features like 16s rRNA, ncRNA, etc which can be used to construct databases for search and blast (using myBLAST)
- Provide a friendly GUI to extract specific sequence features by submission of gene list.

Specificity of AfterGenbank

- Collect all the features existed in Genbank for 12 divisions
- Built-in BLAST with specific databases composed by selected features.
- Output in fasta and CSV can be used for futher analysis
- Access ：

http://aftergenbank.nhri.org.tw/


After GenBank
nnotated-Feature Extractor from GenBank
Repository
BLAST
Home
Contact
Tools
Help
Step 1. choosing the specific division
Step 2. input search term
Search for Organism
GenBank Release 178.0(Release Date:June 15, 2010)
Division:
Select Division
Organism:
(eg.Vibrio cholerae)
example1
example2
example3
Keyword :(eg.toxi
-AND Search
-OR Search
-Advanced Search
Select Division
Bacterial
Invertebrate
Other Mammalian
Patent
Phage
Plant
Primate
Rodent
Synthetic
Unannotated
Viral
Other Vertebrate
twelve divisions
Type 1. Organism.
Type 2. Organism a
Type 3. Organism a
Type 4. Keyword.
Type 5. Accession n
Accession Number :(eg. AB016894)
example5
-Accession Number Search
Use Accession Number:
Search for Accession number
Or load it from file
瀏覽...
3'UTR
5'UTR
-10 signal
-35 signal
attenuator
C region
CAAT
CDS
CDS_translation
conflict
D-loop
D segment
enhancer
exon
gap
GC signal
J segment
LTR
mat peptide
misc feature
misc recomb
misc RNA
misc signal
misc structure
modified base
mRNA
N region
primer bind
promoter
protein bind
RBS
rep origin
repeat region
rRNA
S region
tRNA
unsure
V region
V segment
variation
Step 3. choosing the interested feature
Step 4. submit
Submit
Reset

Search Result Page

Division : Bacterial Organism : Vibrio cholerae
Accession Number : EU523703 EU523704 EU523705 EU523706 ..... input_list
Feature : 5'UTR.

Search terms

7 feature(s) are found in this search.

Total result number

Page 1 of 1 page(s)

Show 20 features in p

Change page

Select Page 1 All/Null

Show not found Accession id

refresh 0 feature(s) are selected by u

Refresh the count

Download ALL (.fasta)

Back to Searc

ALL (.fasta)
ALL (.CSV)

download file

Result

Locus id

1.	EU523703	Vibrio cholerae strain NAG-V17 hemagglutinin protease regula mplete cds.	*Vibrio cholerae*
	Feature	5'UTR	
	Range	..542	
	gene		
2.	EU523704	erae strain TSI-4S hemagglutining/protease regulatory protein-like (hapR) gene, partial sequence.	*Vibrio cholerae*
	Feature	5'UTR	
	Range	1..542	
	gene	hapR	
3.	EU523705	Vibrio cholerae strain TSI-4P aggutining/protease regulatory protein-like (hapR) gene, partial sequence.	*Vibrio cholerae*
	Feature	5'UTR	
	Range	1..542	
	gene	hapR	
4.	EU523706	Vibrio cholerae strain CB1 hemagglutinin/protease regulatory protein (hapR) gene, complete cds.	*Vibrio cholerae*

The definition of Locus id

Feature

Organism

More annotation

AfterGenBank

- **website**

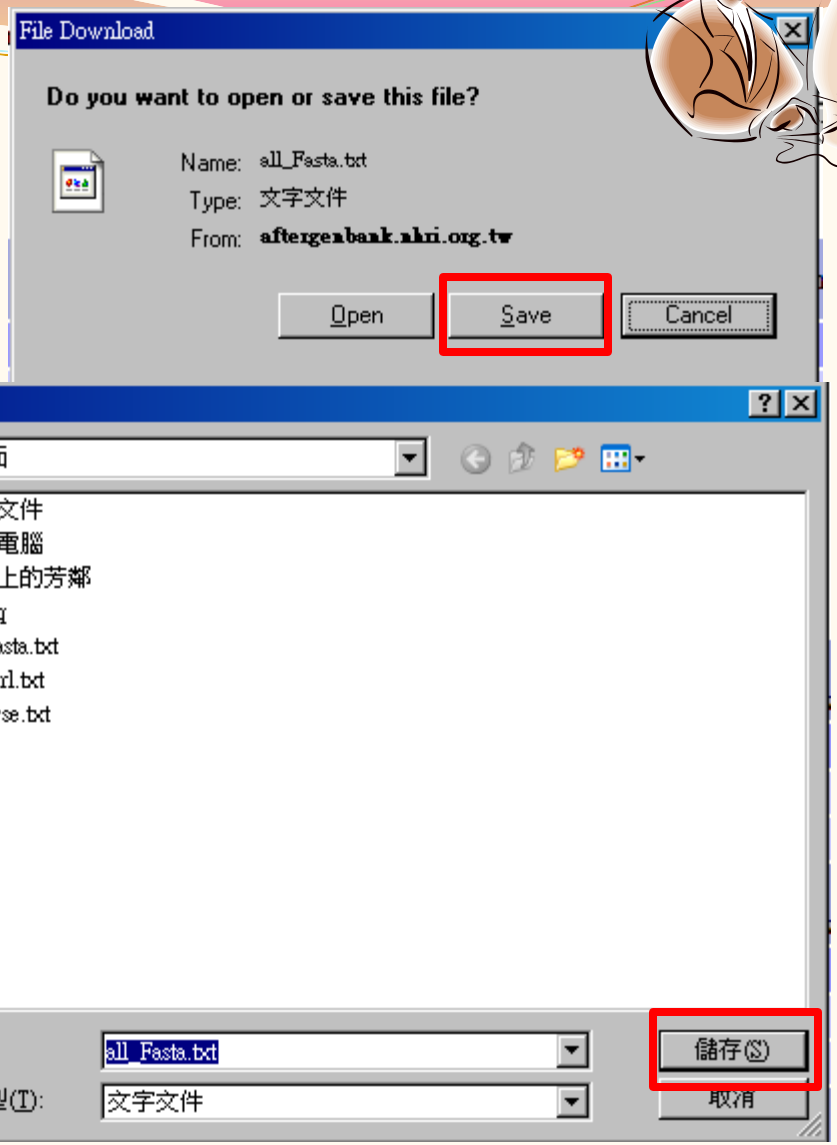
http://aftergenbank.nhri.org.tw/

- **Let's try to click example1**

- **Wait the result.**

- **Click Download.**

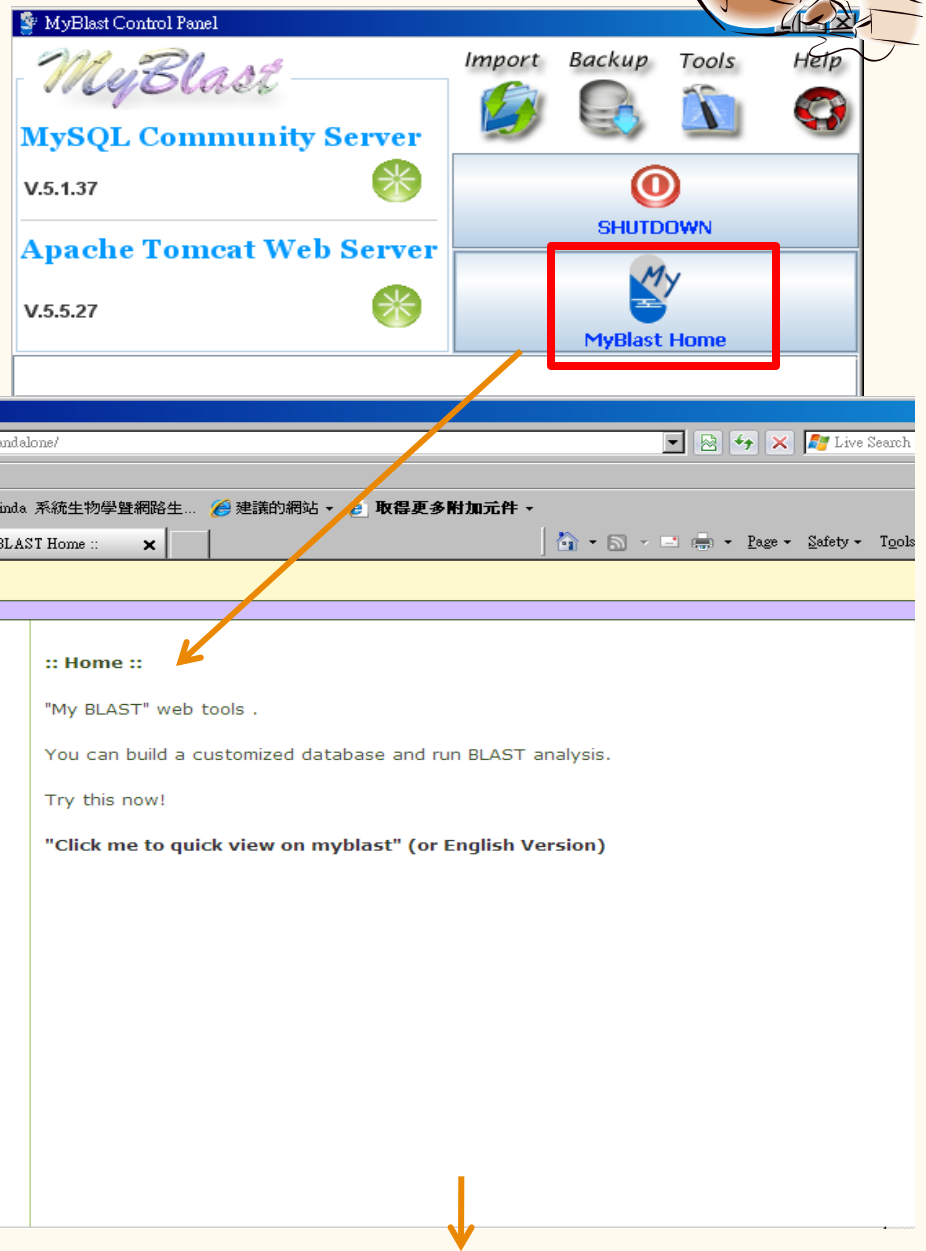
- **Save to the computer (Ex. desktop).**



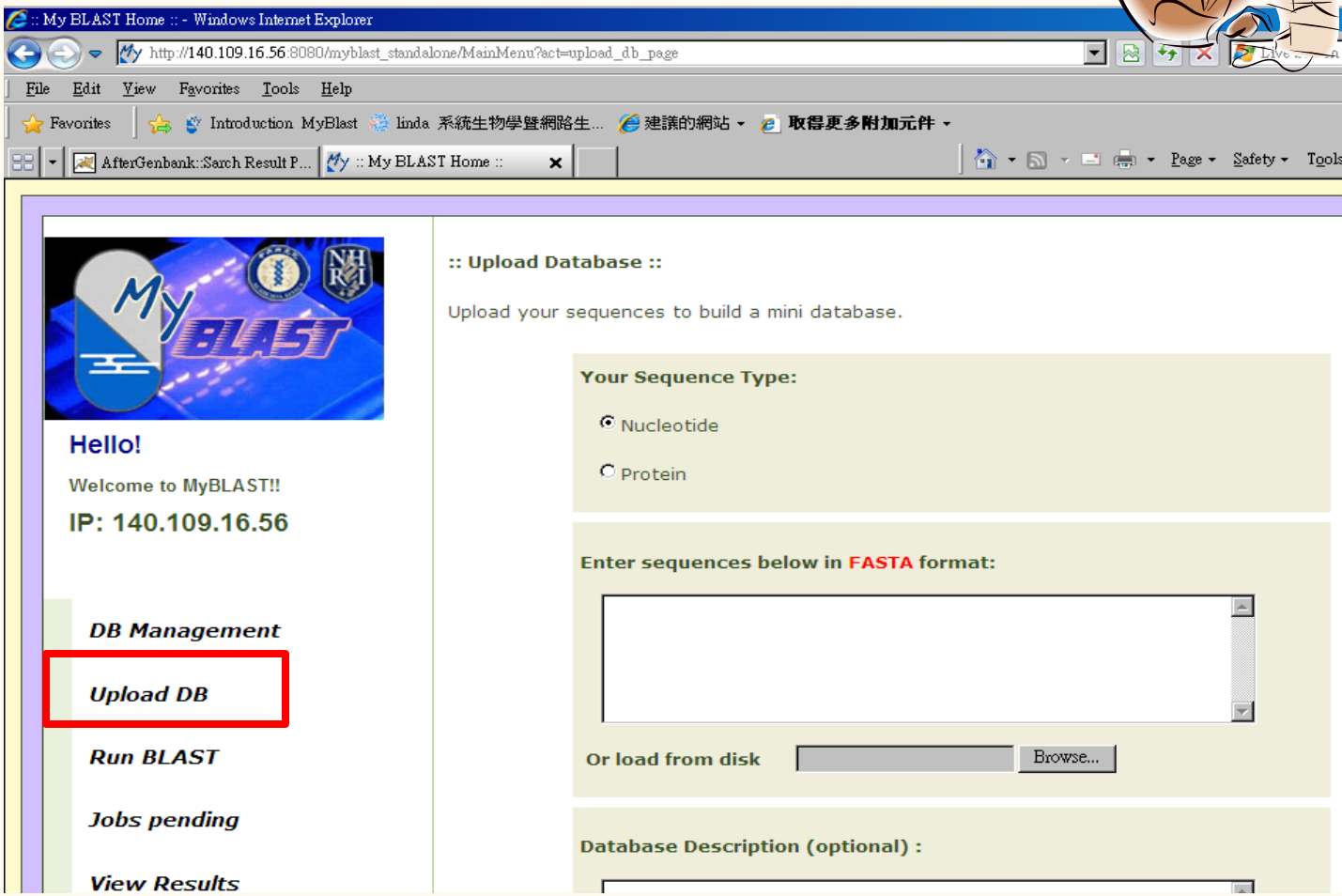
- **MyBLAST website**

  Open MyBlast ControlPanel.

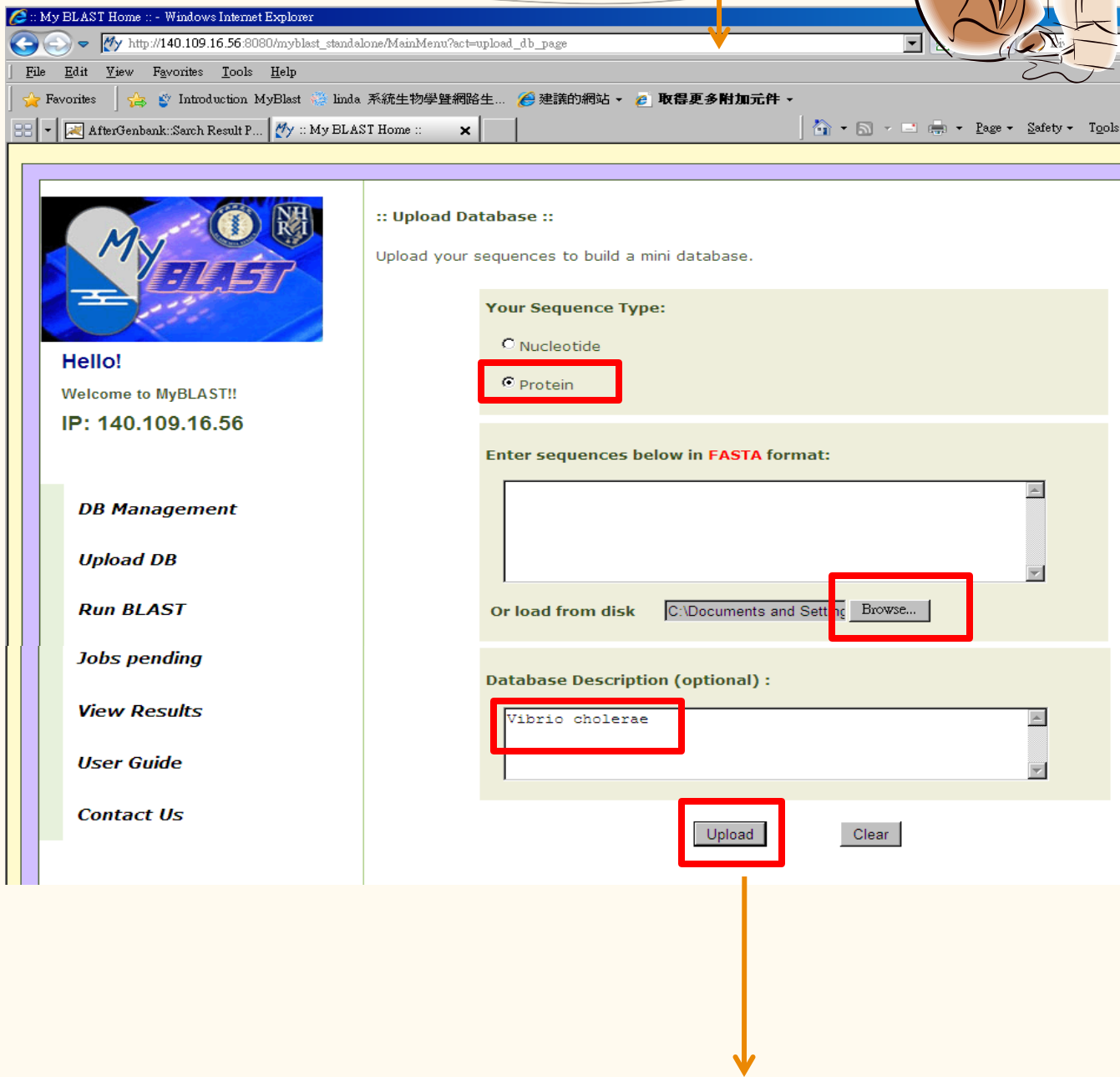
  Click Myblast Home button.



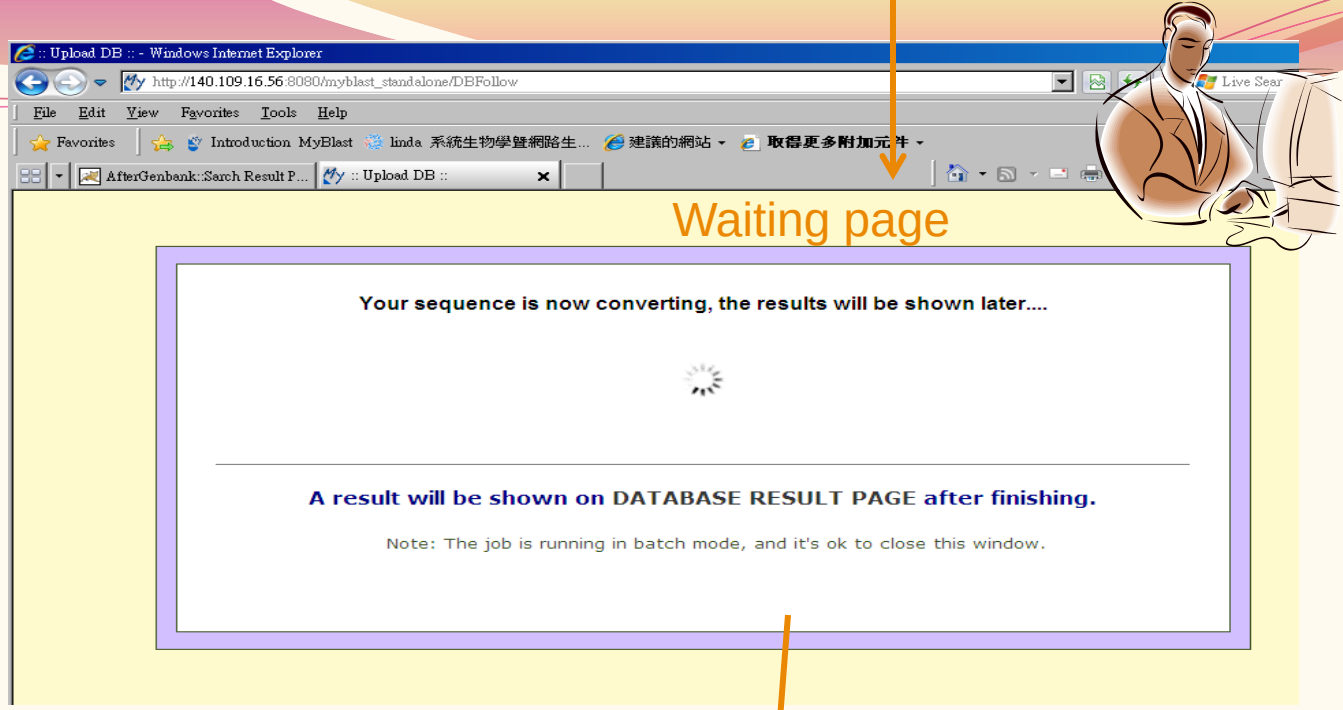
- **Upload DB**

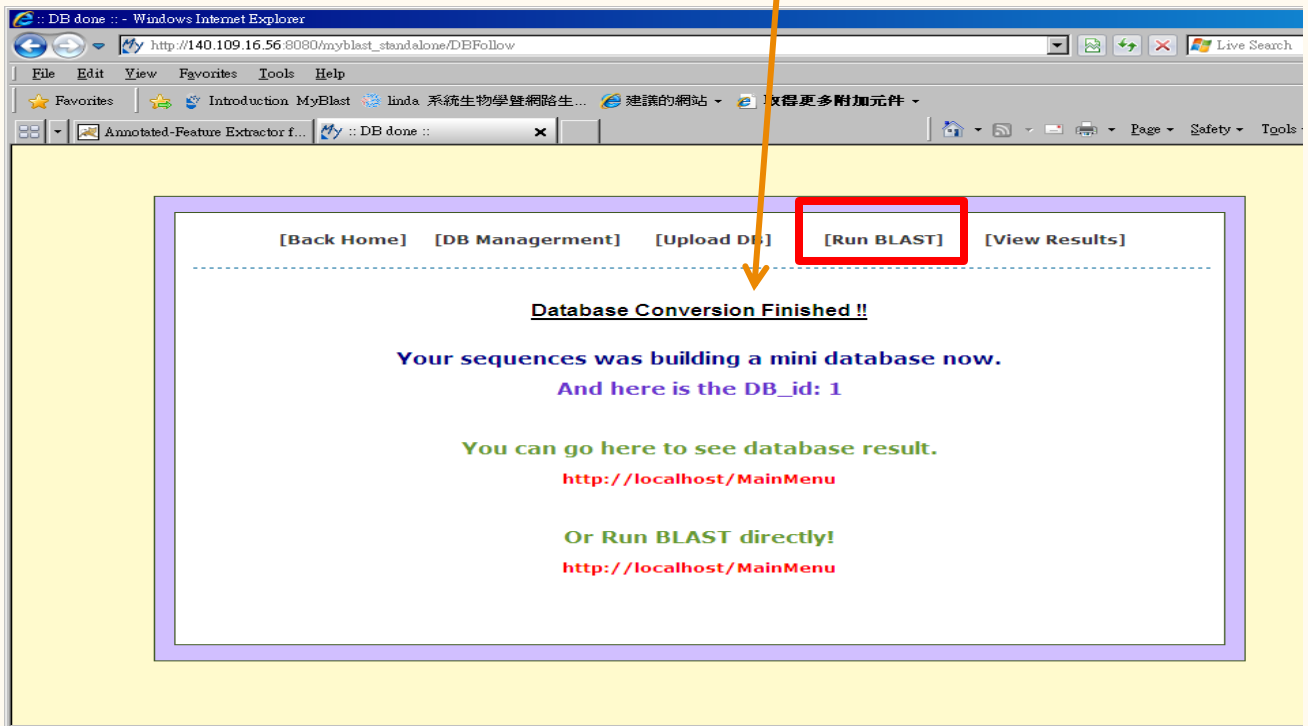

:: My BLAST Home :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/MainMenu?act=upload_db_page
File Edit View Favorites Tools Help
Favorites
Introduction MyBlast
linda 系統生物學暨網路生...
建議的網站
取得更多附加元件
AfterGenbank::Sarch Result P...
:: My BLAST Home ::
Page Safety Tools
Hello!
Welcome to MyBLAST!!
IP: 140.109.16.56
DB Management
Upload DB
Run BLAST
Jobs pending
View Results
:: Upload Database ::
Upload your sequences to build a mini database.
Your Sequence Type:
Nucleotide
Protein
Enter sequences below in FASTA format:
Or load from disk
Browse...
Database Description (optional) :

- **Upload DB**

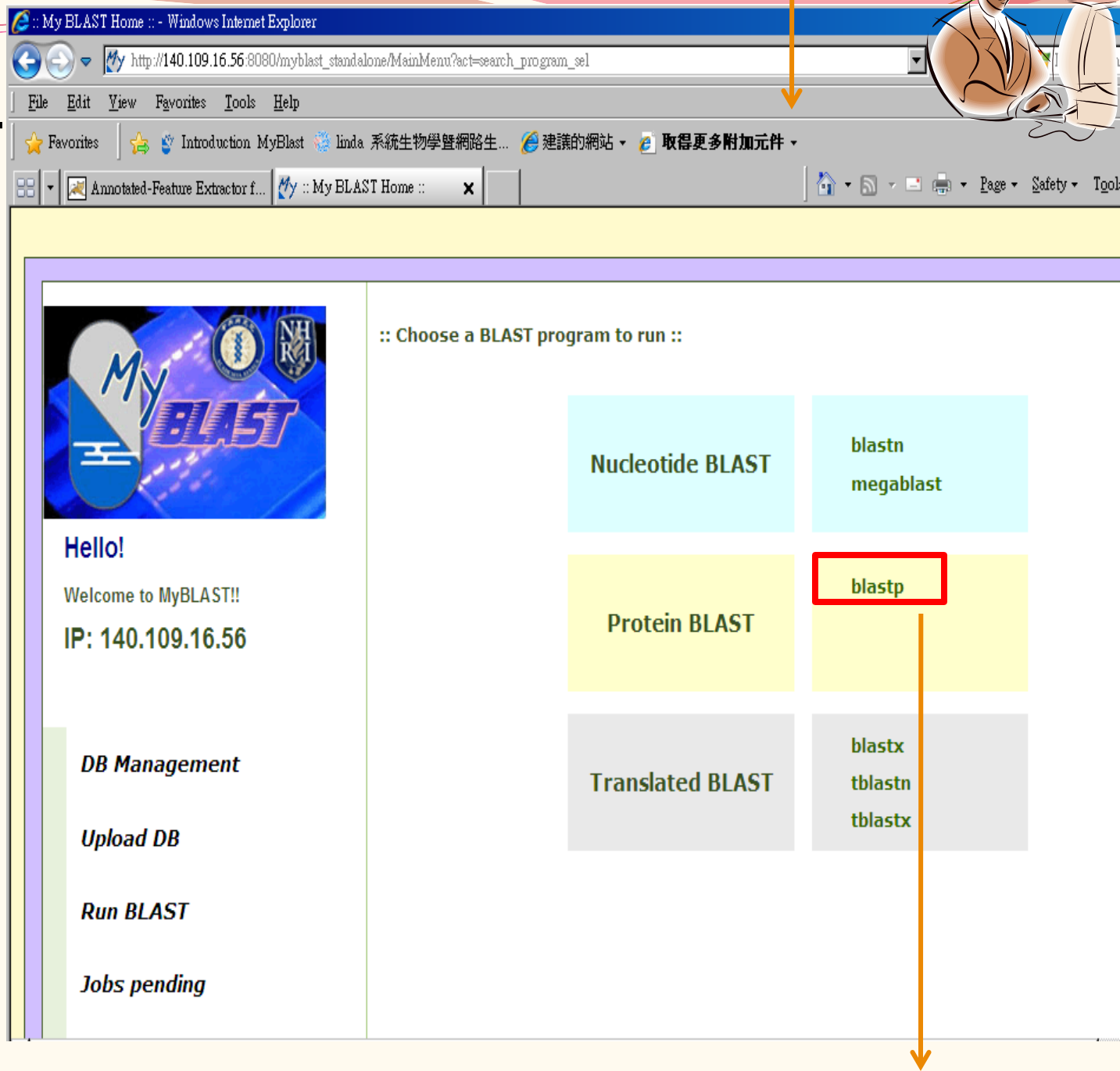

:: My BLAST Home :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/MainMenu?act=upload_db_page
File Edit View Favorites Tools Help
Favorites
Introduction MyBlast
linda 系統生物學暨網路生...
建議的網站
取得更多附加元件
AfterGenbank::Sarch Result P...
:: My BLAST Home ::
Page Safety Tools
Hello!
Welcome to MyBLAST!!
IP: 140.109.16.56
DB Management
Upload DB
Run BLAST
Jobs pending
View Results
User Guide
Contact Us
:: Upload Database ::
Upload your sequences to build a mini database.
Your Sequence Type:
Nucleotide
Protein
Enter sequences below in FASTA format:
Or load from disk
C:\Documents and Setting
Browse...
Database Description (optional) :
Vibrio cholerae
Upload
Clear

- **Upload DB**

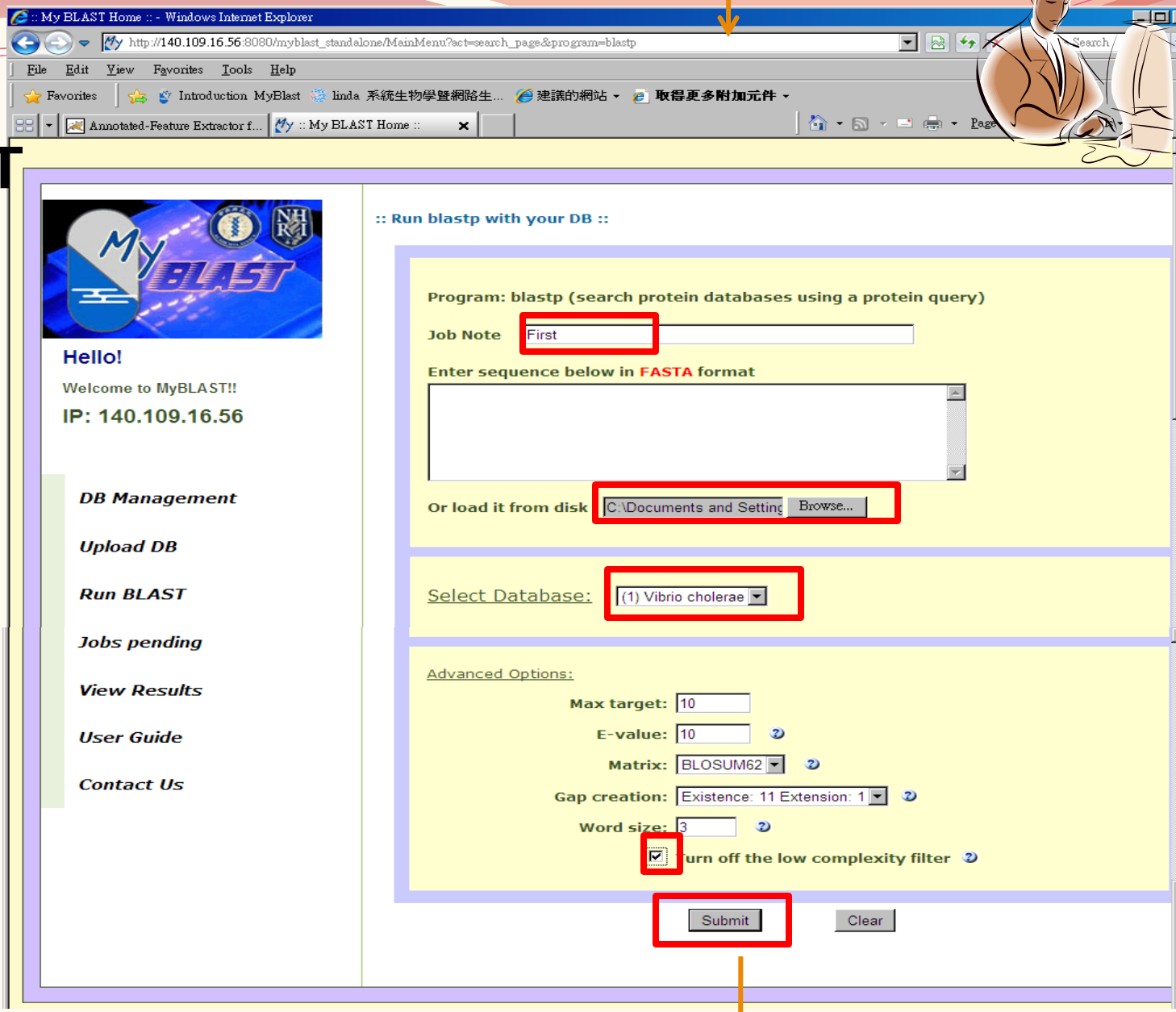

:: Upload DB :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/DBFollow
File Edit View Favorites Tools Help
Favorites Introduction MyBlast linda 系統生物學暨網路生... 建議的網站 取得更多附加元件
AfterGenbank::Sarch Result P... :: Upload DB ::
Waiting page
Your sequence is now converting, the results will be shown later....
A result will be shown on DATABASE RESULT PAGE after finishing.
Note: The job is running in batch mode, and it's ok to close this window.


:: DB done :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/DBFollow
File Edit View Favorites Tools Help
Favorites Introduction MyBlast linda 系統生物學暨網路生... 建議的網站 取得更多附加元件
Annotated-Feature Extractor f... :: DB done ::
[Back Home] [DB Managerment] [Upload DB] [Run BLAST] [View Results]
Database Conversion Finished !!
Your sequences was building a mini database now.
And here is the DB_id: 1
You can go here to see database result.
http://localhost/MainMenu
Or Run BLAST directly!
http://localhost/MainMenu

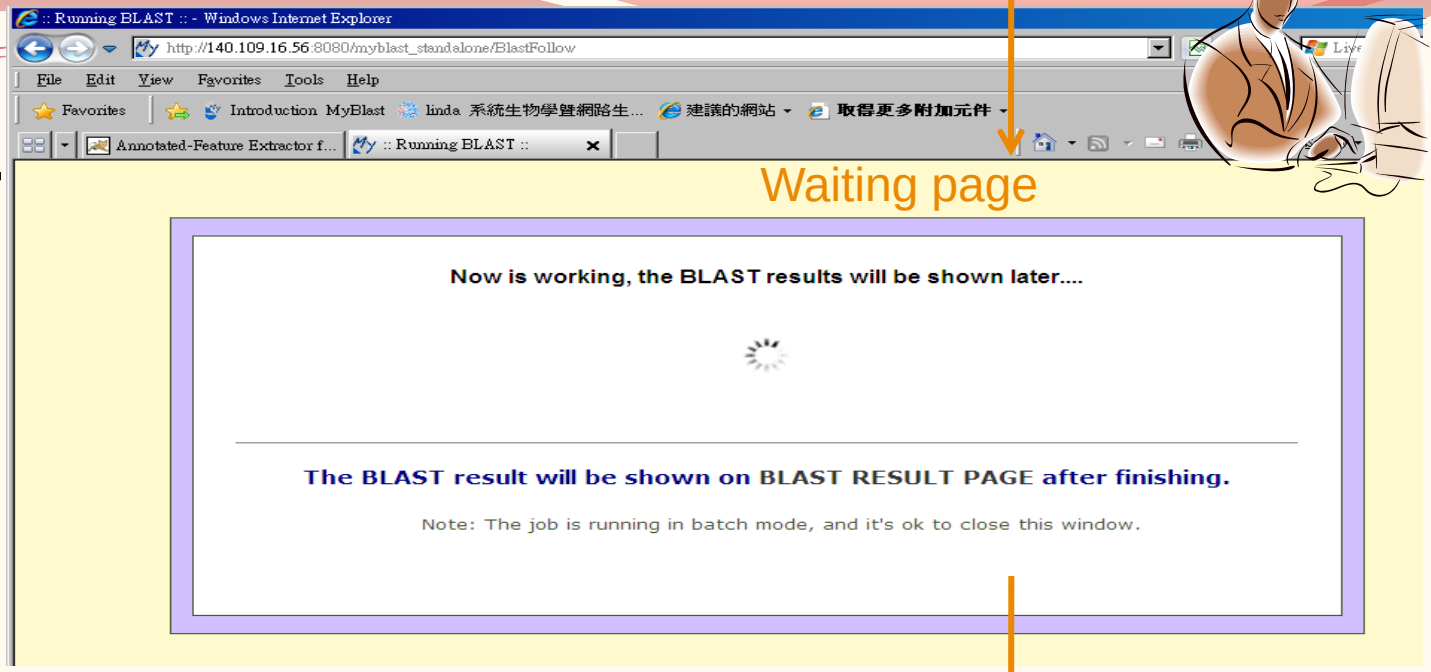
- **Run BLAST**

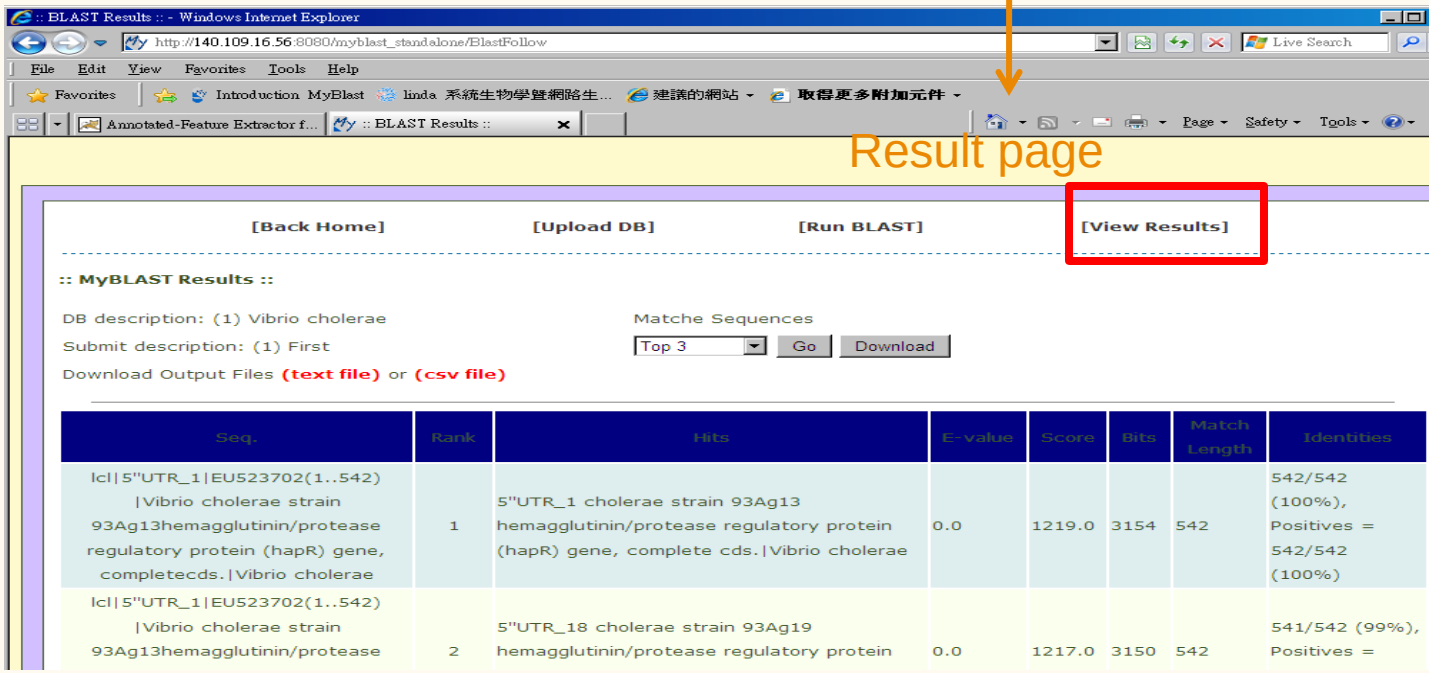

:: My BLAST Home :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/MainMenu?act=search_program_sel
File Edit View Favorites Tools Help
Favorites
Introduction MyBlast
linda 系統生物學暨網路生...
建議的網站
取得更多附加元件
Annotated-Feature Extractor f...
:: My BLAST Home ::
Page
Safety
Tools
Hello!
Welcome to MyBLAST!!
IP: 140.109.16.56
DB Management
Upload DB
Run BLAST
Jobs pending
:: Choose a BLAST program to run ::
Nucleotide BLAST
blastn
megablast
Protein BLAST
blastp
Translated BLAST
blastx
tblastn
tblastx

- **Run BLAST**

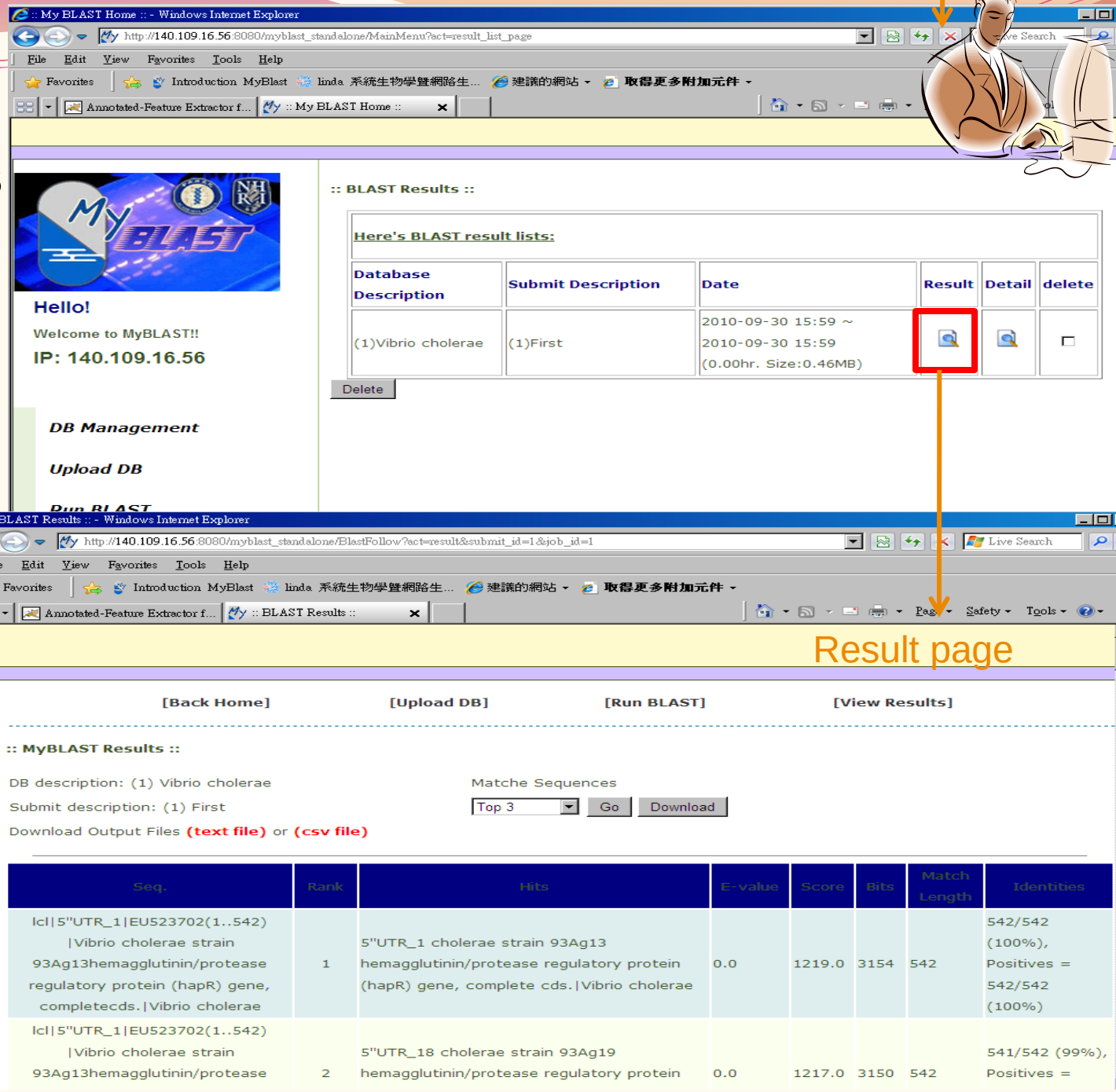

:: My BLAST Home :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/MainMenu?act=search_page&program=blastp
File Edit View Favorites Tools Help
Favorites
Introduction MyBlast
linda 系統生物學暨網路生...
建議的網站
取得更多附加元件
Annotated-Feature Extractor f...
:: My BLAST Home ::
Hello!
Welcome to MyBLAST!!
IP: 140.109.16.56
DB Management
Upload DB
Run BLAST
Jobs pending
View Results
User Guide
Contact Us
:: Run blastp with your DB ::
Program: blastp (search protein databases using a protein query)
Job Note
First
Enter sequence below in FASTA format
Or load it from disk
C:\Documents and Setting
Browse...
Select Database:
(1) Vibrio cholerae
Advanced Options:
Max target:
10
E-value:
10
Matrix:
BLOSUM62
Gap creation:
Existence: 11 Extension: 1
Word size:
3
Turn off the low complexity filter
Submit
Clear

- **Run BLAST**

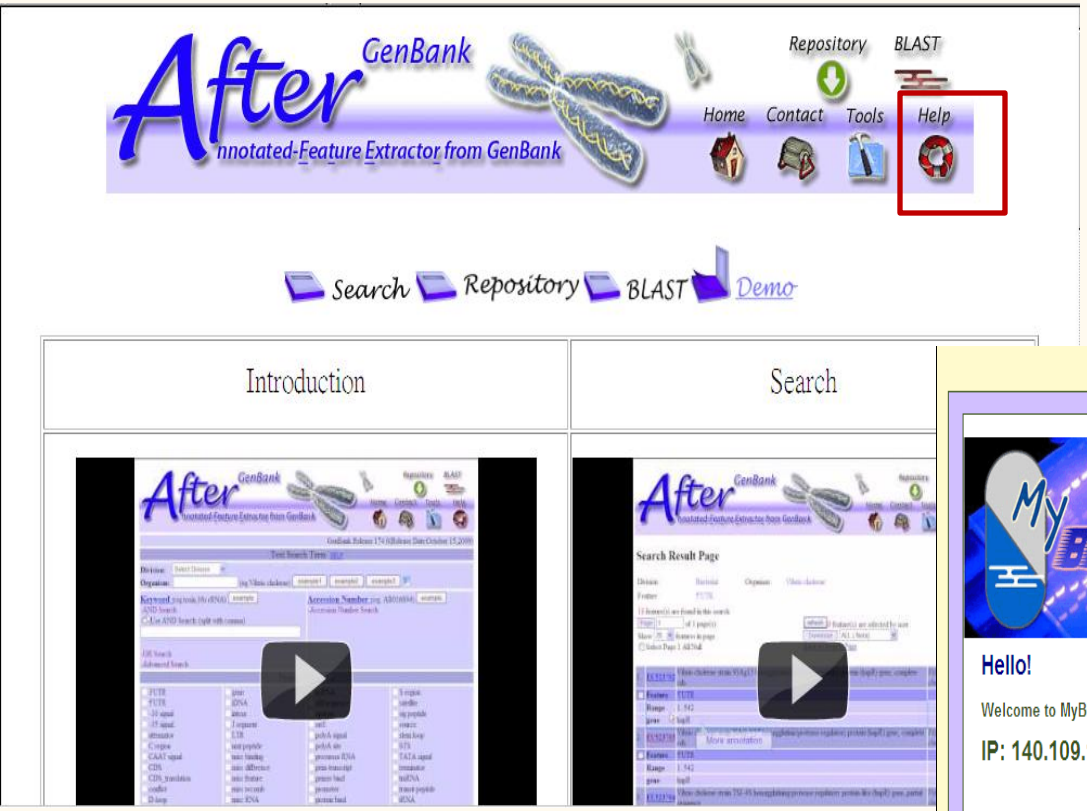

:: Running BLAST :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/BlastFollow
File Edit View Favorites Tools Help
Favorites Introduction MyBlast linda 系統生物學暨網路生... 建議的網站 取得更多附加元件
Annotated-Feature Extractor f... :: Running BLAST ::
Waiting page
Now is working, the BLAST results will be shown later....
The BLAST result will be shown on BLAST RESULT PAGE after finishing.
Note: The job is running in batch mode, and it's ok to close this window.

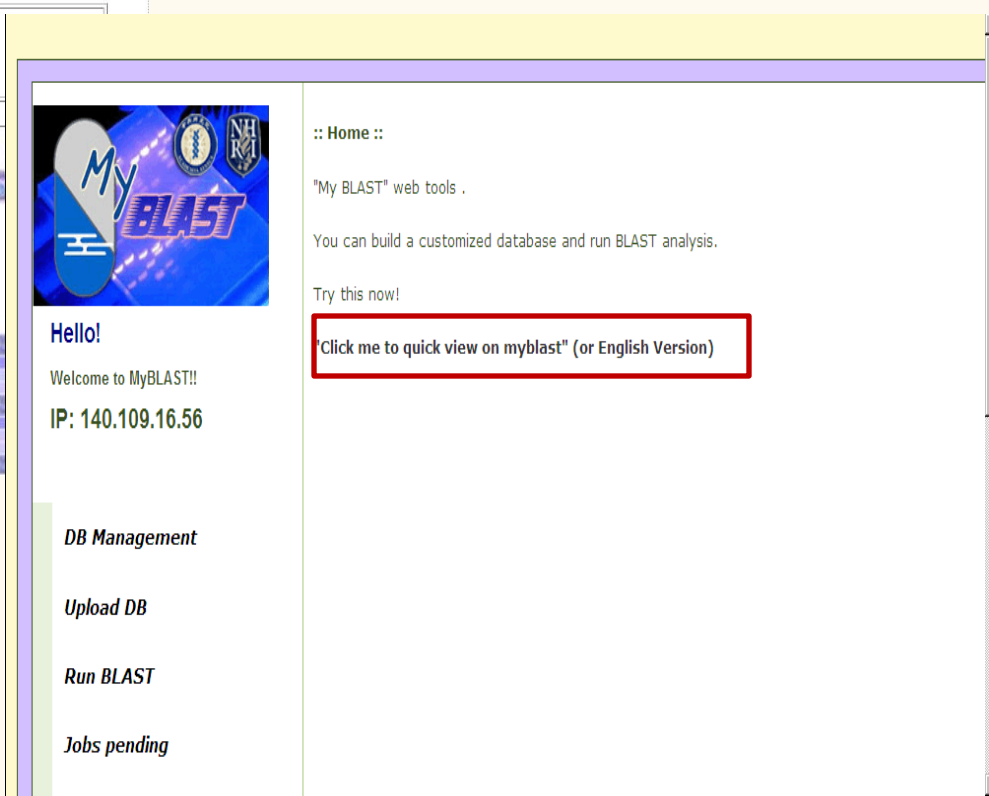

:: BLAST Results :: - Windows Internet Explorer
http://140.109.16.56:8080/myblast_standalone/BlastFollow
File Edit View Favorites Tools Help
Favorites Introduction MyBlast linda 系統生物學暨網路生... 建議的網站 取得更多附加元件
Annotated-Feature Extractor f... :: BLAST Results ::
Result page
[Back Home] [Upload DB] [Run BLAST] [View Results]
:: MyBLAST Results ::
DB description: (1) Vibrio cholerae
Submit description: (1) First
Download Output Files (text file) or (csv file)
Matche Sequences
Top 3 Go Download
Seq. Rank Hits E-value Score Bits Match Length Identities
lcl|5"UTR_1|EU523702(1..542) |Vibrio cholerae strain 93Ag13hemagglutinin/protease regulatory protein (hapR) gene, completecds.|Vibrio cholerae 1 5"UTR_1 cholerae strain 93Ag13 hemagglutinin/protease regulatory protein (hapR) gene, complete cds.|Vibrio cholerae 0.0 1219.0 3154 542 542/542 (100%), Positives = 542/542 (100%)
lcl|5"UTR_1|EU523702(1..542) |Vibrio cholerae strain 93Ag13hemagglutinin/protease 2 5"UTR_18 cholerae strain 93Ag19 hemagglutinin/protease regulatory protein 0.0 1217.0 3150 542 541/542 (99%), Positives =

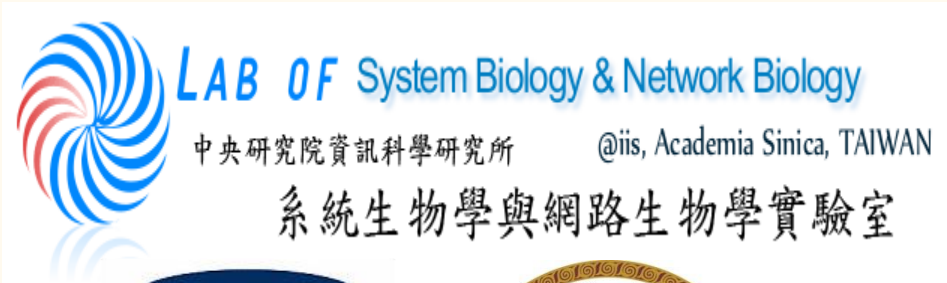
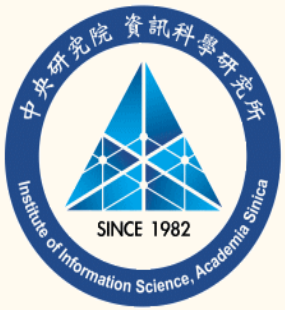
- **View Results**



Flash Demo


After GenBank
nnotated-Feature Extractor from GenBank
Repository
BLAST
Home
Contact
Tools
Help
Search
Repository
BLAST
Demo
Introduction
Search


:: Home ::
"My BLAST" web tools .
You can build a customized database and run BLAST analysis.
Try this now!
"Click me to quick view on myblast" (or English Version)
Hello!
Welcome to MyBLAST!!
IP: 140.109.16.56
DB Management
Upload DB
Run BLAST
Jobs pending

Authorships:

- Tsai, Ming-Hsin(蔡鳴興) Tang, Yueh-Hsia(唐月霞) Lin,Chieh-Hua(林介華) Wang,Shih- Hai(汪詩海) Hsiung, Chao(熊昭)
- Shu-Hwa Chen(陳淑華), I-Hsuan Lu(呂怡萱), Kuei-Chuan Huang(黃桂絹), Chung-Yen Lin(林仲彥)