

SEQUENCHER®

Tutorial for Windows and Macintosh

Local-BLAST

© 2016 Gene Codes Corporation

Gene Codes Corporation



Gene Codes Corporation
775 Technology Drive, Ann Arbor, MI 48108 USA
1.800.497.4939 (USA) +1.734.769.7249 (elsewhere)
+1.734.769.7074 (fax)
www.genecodes.com gcinfo@genecodes.com

Local-BLAST

About File Formats	3
Installing Local-BLAST	3
Creating a New BLAST Database	4
Invoking a Sequencher Connections Session.....	5
Running Local-BLAST	6
Viewing your Local-BLAST Results.....	7
Local-BLAST Options.....	8
Conclusion	10

Local-BLAST

Running BLAST searches locally can save time and provide you with flexibility in terms of which version of BLAST and database you choose to use. You can download pre-formatted BLAST databases from NCBI or create your own. With **Sequencher Connections'** simple interface for setting up and running BLAST searches, you'll find a system that meets your needs. Your BLAST results are shown in **Connections'** tabbed windows, which allow for easy viewing of the results in different formats.

In this tutorial, you will query the revised Cambridge Reference Sequence against a database of mitochondrial sequences. This tutorial also guides you through creating a Local-BLAST database from a FASTA file.

Information on using **Sequencher Connections'** other powerful features can be found in the **Sequencher Connections** tutorial and chapter in the **Sequencher** User Manual.

ABOUT FILE FORMATS

In this tutorial, you will download a FASTA file from which you will use one of the tools of the BLAST toolkit to create your own database. If you want to use your own data, you will need to provide your own query sequence file(s) in **FASTA** or **FASTQ** format and a BLAST database. If you have been working with a well-characterized genome, then you will probably be able to obtain pre-formatted BLAST databases from the NCBI BLAST website at <ftp://ftp.ncbi.nlm.nih.gov/blast/db/>. Note: If you are on a Mac and prompted for a name and password when trying to access any of the ftp sites in this tutorial, select the "Connect as: Guest" radio button and click "Connect." Once you have downloaded the .tar.gz archive, unzip it and copy the files to your Local-BLAST database location described in the instructions below for Creating a New BLAST Database.

INSTALLING LOCAL-BLAST

In this tutorial, you will need to install BLAST locally on your machine and download the mito.nt FASTA file from NCBI. We recommend using BLAST version 2.2.40 as this is the default version in **Sequencher**.

Windows and Mac installers for BLAST can be downloaded

<ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/>.

After running the Local-BLAST installer, you will need to set up the **BLASTDB** environment variable. There are different instructions for Mac and Windows operating systems.

For Mac OS systems:

- From the Finder, open a Terminal window by clicking on the **Terminal icon** in the **Applications/Utilities/** folder.
- In the terminal window that appears, enter the following command line (you can also copy and paste this command) into the Terminal window.

```
echo "[BLAST]" > ~/.ncbirc; echo "BLASTDB=/usr/local/ncbi/blast/db" >>
~/.ncbirc
```

This will create a hidden file called .ncbirc in your Home (~) directory. To ensure this change takes effect, you must log out and log back into your account.

For Windows operating systems:

- Select **Start > Control Panel > System > Advanced system settings**.
- Click on the **Advanced** tab and then on the **Environment Variables....** button.
- Under “User variables for ...” click on the **New...** button.

Note: This will enable only the current user to use Local-BLAST. You can, however, enable Local-BLAST for all users by creating a new System variable.

- For the Variable name, enter **BLASTDB**.
- For the Variable value, enter **C:\Program Files\NCBI\blast-2.2.28+\db**.
- Click **OK**.

CREATING A NEW BLAST DATABASE

Before you can run Local-BLAST, you will also need to have a database which has a compatible format installed on your computer. You can download segments of the databases that are used on the BLAST servers at NCBI or you can use the *makeblastdb* utility to create a database from a file containing FASTA sequences. To create a database from such a file, follow these steps:

- Download **mito.nt.gz** from <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/> and copy the file to the location where BLAST databases are stored (in this example, **C:\Program Files\NCBI\blast-2.2.28+\db** on Windows or **/usr/local/ncbi/blast/db** on Mac). You may need to create this directory.
- Unzip the file. A FASTA file named **mito.nt** will appear in that folder.

If you are working on Windows using a Cygwin or DOS cmd command prompt or Terminal window on Mac, you now need to navigate to the db folder.

On Mac operating systems:

- From the Finder, open a Terminal window by clicking on the **Terminal icon** in the **Applications/Utilities/** folder.
- At the command prompt in the Terminal window, enter this command:
`cd /usr/local/ncbi/blast/db`

On Windows operating systems:

- Click on the **Start** menu and, in the search bar, type **cmd** and press **enter**.
- Right-click **Command Prompt** and select **Run as administrator**.
- At the command prompt in the cmd window, enter this command: **cd "C:\Program Files\NCBI\blast-2.2.28+\db"**

You now need to verify that you are in the same folder as the *mito.nt* file by running the “list” command (Mac: *ls*, Windows: *dir*) at the command prompt.

- Run the **makeblastdb** tool on the FASTA file:
`makeblastdb -in mito.nt -input_type fasta -dbtype nucl`
- Note: The Mac installer may not update the PATH environment variable with the path to BLAST’s executables. As a result, you may need to run the above command as:

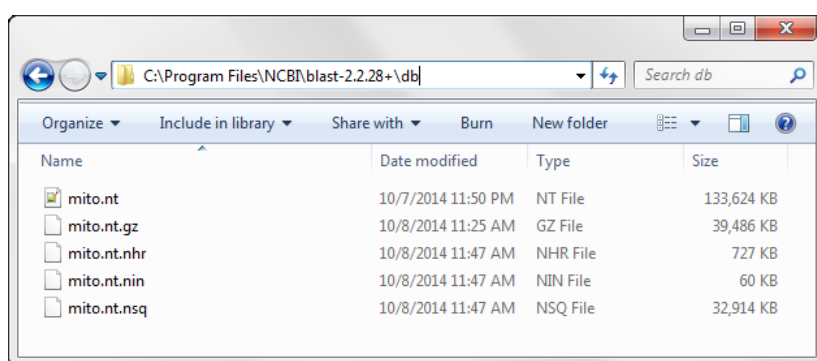
```
/usr/local/ncbi/blast/bin/makeblastdb -in mito.nt -input_type fasta -dbtype nucl
```

```
Administrator: C:\Windows\system32\cmd.exe
C:\Program Files\NCBI\blast-2.2.28+db> makeblastdb -in mito.nt -input_type fasta -dbtype nucl

Building a new DB, current time: 10/08/2014 11:51:32
New DB name:   mito.nt
New DB title:  mito.nt
Sequence type: Nucleotide
Deleted existing BLAST database with identical name.
Keep Linkouts: T
Keep MBits: T
Maximum file size: 1000000000B
Adding sequences from FASTA; added 5087 sequences in 2.48967 seconds.
C:\Program Files\NCBI\blast-2.2.28+db>
```

The BLAST database files created are mito.nt.nhr, mito.nt.nin and mito.nt.nsq.

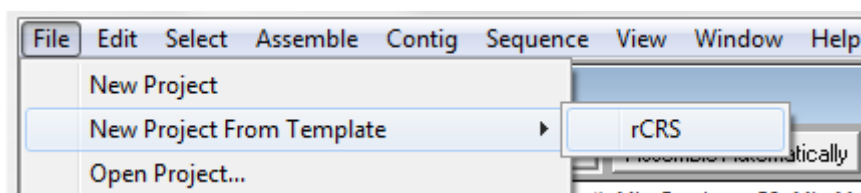
- Quit the DOS command prompt or Terminal session.



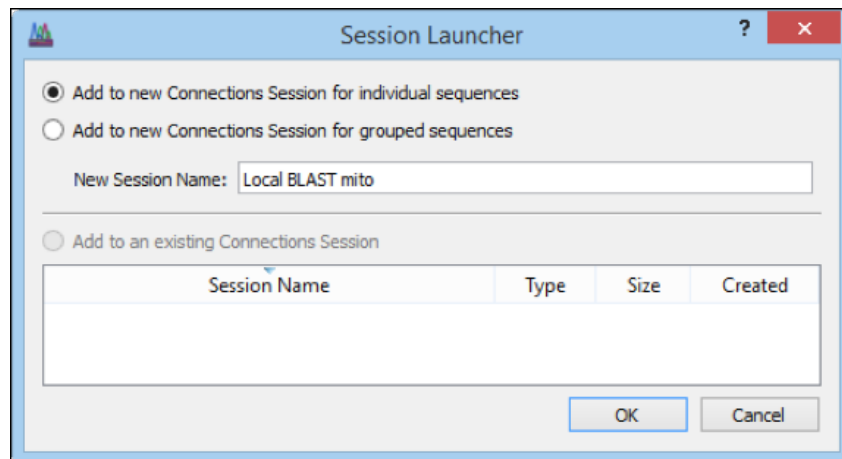
INVOKING A SEQUENCHER CONNECTIONS SESSION

Now that you have Local-BLAST installed and have created a compatible database you are ready to run some searches. First you need to choose the sequence or sequences you will be using as queries and create a new session with them in **Sequencer** Connections.

- Launch **Sequencer**.
- Go to the **File** menu and select **New Project From Template>rCRS**.



- Select the sequence named **rCRS**. Go the **Window** menu and select **Add to Connections Session...**
- You will then see the **Session Launcher** dialog.
- Select the **Add to new Connections Session for individual** sequences radio button.

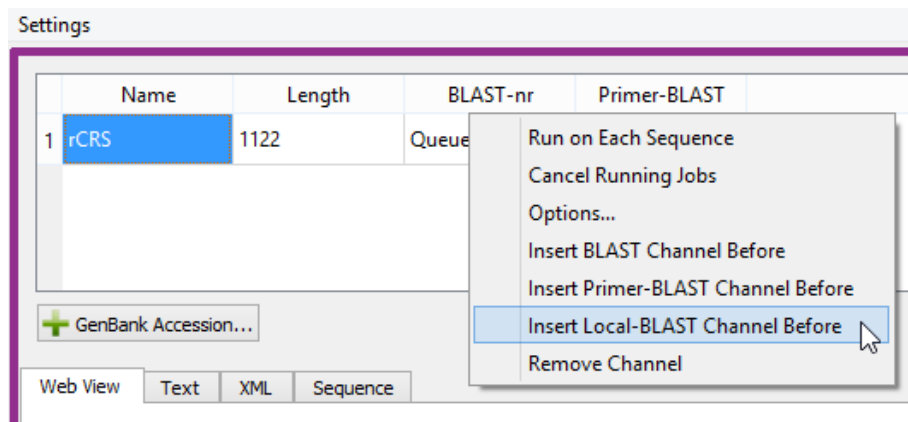


- Enter **“Local BLAST mito”** into the **New Session Name** input field.
- Click on the **OK** button.
- A **Sequencer Connections** session window will appear.

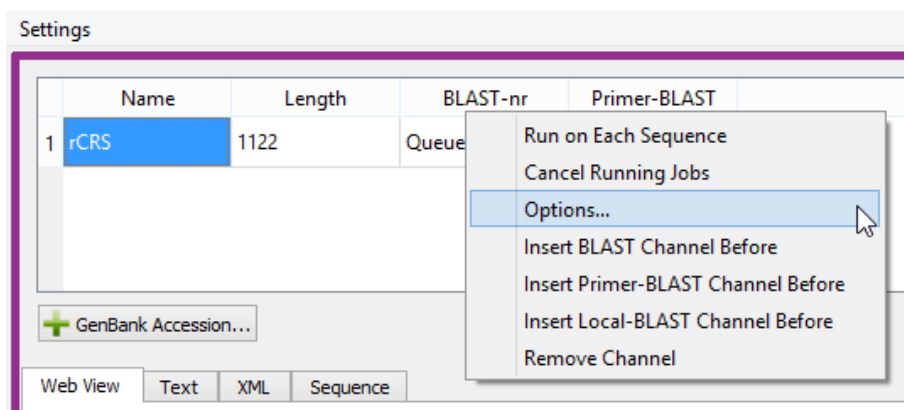
RUNNING LOCAL-BLAST

When you launch a new **Connections** session for individual sequences, two default channels are created. These channels are BLAST (at NCBI) and Primer BLAST (at NCBI). You need to create a channel for Local-BLAST.

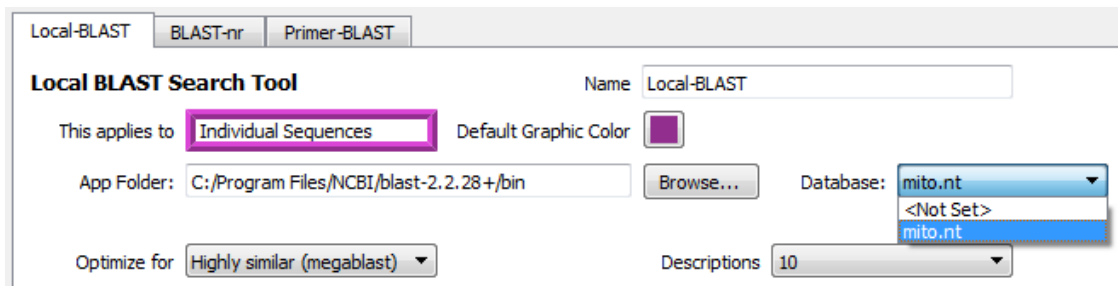
- Right-click on the column **BLAST-nr** and select **Insert Local-BLAST Channel Before**.



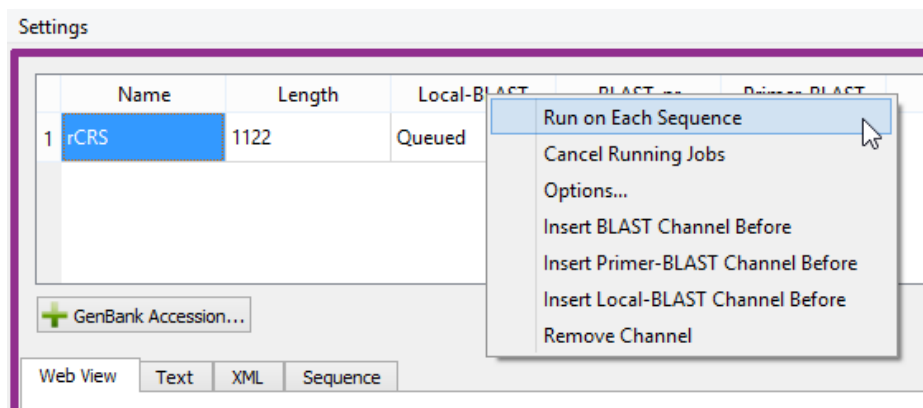
- Right-click on the column **Local-BLAST** and select **Options...**



- From the **Channel Options** dialog, select **mito.nt** from the Database drop-down menu.



- If you prefer, you can name this channel to something more memorable by typing the new name into the **Name** input field, otherwise its default name will be Local-BLAST-mito.nt.
- Click on the **OK** button to close the **Options** dialog.
- Right-click on the **Local-BLAST-mito.nt** column and select **Run on Each Sequence**. The status of the channel will go from Queued to Pending to Done.



Note that if you close and reopen a session that had earlier search results, the **Done** status changes to **Done**. Also, if a session had previous search results and a new search is run but then cancelled before the search is completed, the earlier results will be restored.

VIEWING YOUR LOCAL-BLAST RESULTS

The results of your session can be viewed in the tabs labeled Web View and Text. The Sequence tab contains the original query sequence.

If you used the mitochondrial sequence and database used in this tutorial, then your results will look like the following image.

- Click the cell in the **Local-BLAST-mito.nt** channel that has a status of **Done**.
- The results will appear in the Web View tab.
- Now click in the Text tab. Notice that there are no hyperlinks to other parts of the results in this view.

Web View Text XML Sequence

Query=
Length=1122

Sequences producing significant alignments:

	Score (Bits)	E Value
gi 5835121 ref NC_001643.1 Pan troglodytes mitochondrion, comp...	1173	0.0
gi 5835135 ref NC_001644.1 Pan paniscus mitochondrion, complet...	1151	0.0
gi 251831106 ref NC_012920.1 Homo sapiens mitochondrion, compl...	1064	0.0
gi 196123578 ref NC_011137.1 Homo sapiens neanderthalensis mit...	961	0.0
gi 292606408 ref NC_013993.1 Homo sp. Altai mitochondrion, com...	929	0.0
gi 568192363 ref NC_023100.1 Homo heidelbergensis mitochondrio...	843	0.0
gi 195952353 ref NC_011120.1 Gorilla gorilla gorilla mitochond...	667	0.0
gi 5835149 ref NC_001645.1 Gorilla gorilla mitochondrion, comp...	625	9e-178
gi 5835834 ref NC_002083.1 Pongo abelii mitochondrion, complet...	584	1e-165
gi 5835163 ref NC_001646.1 Pongo pygmaeus mitochondrion, compl...	556	3e-157

> gi|5835121|ref|NC_001643.1| Pan troglodytes mitochondrion, complete
genome
Length=16554

Score = 1173 bits (635), Expect = 0.0
Identities = 965/1124 (86%), Gaps = 24/1124 (2%)
Strand=Plus/Plus

Query	1	TTCTTTCATGGGGAAGCAGATTGGGTACCACCCAAGTATTGACTCACCCATCAACAACC	60
Sbjct	15442	TTCTTTCATGGGGAAGCAAATTTAGGTACCACCTAAGTACTGGCTCATTTCAT-TACAACC	15500
Query	61	GCTATGTATTTTCGTACATTACTGCCAGCCACCATGAATATTGTACGGTACCATAAATACT	120
Sbjct	15501	GCTATGTATTTTCGTACATTACTGCCAGCCACCATGAATATCGTACAGTACCAT-ATCACC	15559
Query	121	TGACCACCTGTAGTACATAAAAACCCAAT-CCACATCAAAACCCCTCCCATGCTTACA	179

From within the Text, XML, or Sequence tabs, you can save the results by right-clicking and selecting "Save as Text..." for Text and Sequence tabs and "Save XML..." for the XML tab.

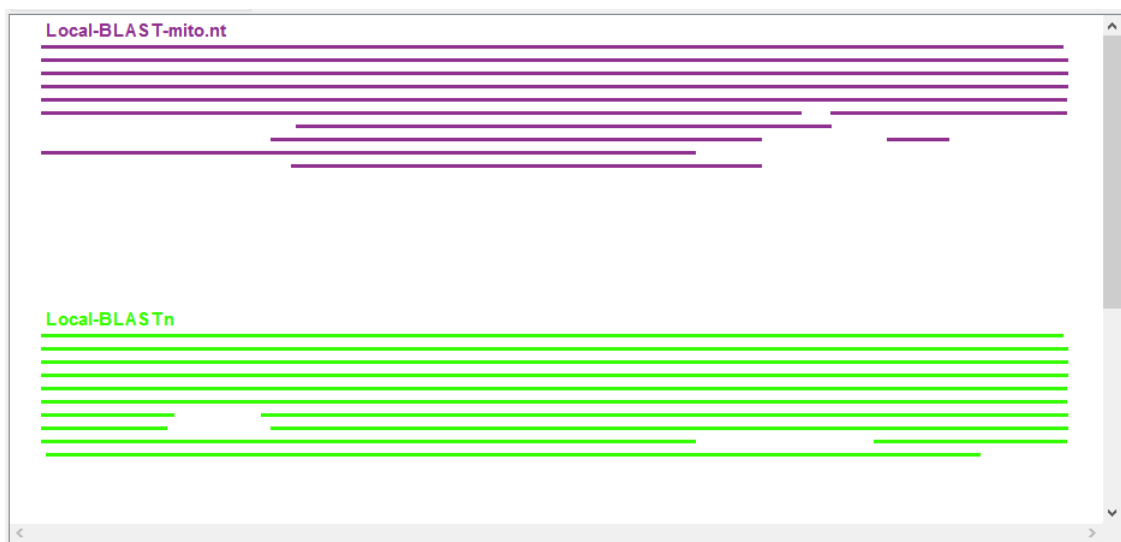
Local-BLAST Options

The majority of options for Local-BLAST and BLAST are the same. The most important Local-BLAST difference is the option to use any database that you are able to obtain or create. BLAST at NCBI is updated at regular intervals, with Local-BLAST you are in control of the version you use. For this tutorial, we have used version 2.2.28. For more details on using these options, please refer to the chapter called "**Sequencher Connections**" in the Sequencher User Manual.

- Right-click on the column BLAST-nr and select **Insert Local-BLAST Channel Before**.
- Right-click on the new **Local-BLAST** column and select **Options....**
- Change the channel **Name** to "**local blastn**."
- Choose a new Default Graphic Color.
- Choose **mito.nt** from the **Database** drop-down menu.
- Choose **blastn** from the **Optimize for** drop-down menu.
- You are warned that you need to change the **Word Size** to 11, you can type the number **11** directly over the existing value.
- Click on the **OK** button to dismiss the dialog.
- Right-click on the new **Local-BLAST-mito.nt** column and select **Run on Each Sequence**.

You now have results using two different Local-BLAST algorithms – blastn and megablast. You can compare the results using **Sequencher Connections Schematic** and you can look at the alignments. To compare the results using the **Schematic**, perform the following steps:

- Right-click on either the number next to the sequence whose schematic you wish to view or its name. In this case, right-click on **rCRS**.
- Choose **Show Schematic** from the menu. If you hold the cursor over a single alignment in the **Schematic**, you will see a tooltip showing information about that alignment.



- Click on the cell in the **Local-BLAST-mito.nt** channel that has the status of **Done**.
- Click on the cell in the **Local-BLASTn** channel that has a status of **Done**.
- Compare the two lists of results and view the actual alignments.

The top image shows the list of alignments obtained using the megablast algorithm while the second image shows the list of alignments obtained using the blastn algorithm with a Word size of 11. Note how the list of results differs. If you go on to explore the alignments, you will see that even where the same sequence is reported, it may have fewer aligned bases.

Local-BLAST using megablast

Sequences producing significant alignments:	Score (Bits)	E Value
gi 5835121 ref NC_001643.1 Pan troglodytes mitochondrion, comp...	1173	0.0
gi 5835135 ref NC_001644.1 Pan paniscus mitochondrion, complet...	1151	0.0
gi 251831106 ref NC_012920.1 Homo sapiens mitochondrion, compl...	1064	0.0
gi 196123578 ref NC_011137.1 Homo sapiens neanderthalensis mit...	961	0.0
gi 292606408 ref NC_013993.1 Homo sp. Altai mitochondrion, com...	929	0.0
gi 568192363 ref NC_023100.1 Homo heidelbergensis mitochondrio...	843	0.0
gi 195952353 ref NC_011120.1 Gorilla gorilla gorilla mitochond...	667	0.0
gi 5835149 ref NC_001645.1 Gorilla gorilla mitochondrion, comp...	625	9e-178
gi 5835834 ref NC_002083.1 Pongo abelii mitochondrion, complet...	584	1e-165
gi 5835163 ref NC_001646.1 Pongo pygmaeus mitochondrion, compl...	556	3e-157

Local-BLAST using blastn

Sequences producing significant alignments:	Score (Bits)	E Value
gi 5835121 ref NC_001643.1 Pan troglodytes mitochondrion, comp...	1260	0.0
gi 5835135 ref NC_001644.1 Pan paniscus mitochondrion, complet...	1242	0.0
gi 251831106 ref NC_012920.1 Homo sapiens mitochondrion, compl...	1040	0.0
gi 196123578 ref NC_011137.1 Homo sapiens neanderthalensis mit...	957	0.0
gi 292606408 ref NC_013993.1 Homo sp. Altai mitochondrion, com...	915	0.0
gi 568192363 ref NC_023100.1 Homo heidelbergensis mitochondrio...	845	0.0
gi 195952353 ref NC_011120.1 Gorilla gorilla gorilla mitochond...	722	0.0
gi 5835149 ref NC_001645.1 Gorilla gorilla mitochondrion, comp...	693	0.0
gi 5835834 ref NC_002083.1 Pongo abelii mitochondrion, complet...	648	0.0
gi 408772040 ref NC_018753.1 Nomascus gabriellae mitochondrion...	594	2e-168

CONCLUSION

In this tutorial, you have created your own Local-BLAST database using a FASTA file as the source of your sequences. You have learned how to query this database using locally installed version of BLAST. You have also learned how to create additional channels so that you can test the veracity of your results using a second algorithm and different BLAST options. Finally you have learned how to use the Schematic and compare results.