



Frequency Calculator

User Guide



<http://oligo.net>

Frequency Calculator is a sequence frequency table generator. It creates sequence frequency tables that are used by the Oligo software in the Eliminate Frequent Oligonucleotides sub search. Primers with low frequency numbers at the 3'-ends are preferred. Those frequent oligos may be 6-, 7- or 8-mers, and are generated from any number of sequences grouped in a single folder on your hard disk.

The list of sequence frequency tables is given in the Oligo 7 Manual in chapter 1.4.6, p. 11, and the Analyze – Sequence Frequency sub search is described in chapters 5.13, p. 75 and 6.1.3.11 on p. 96.

In the Search Parameters dialog in Oligo software there is “Frequency Table” item. The default sequence Frequency Table is GBPRI.FR6. This can be changed by clicking on the “Change” button located next to the file name mentioned above.

Oligo software comes with several frequency table files calculated from the DNA sequences of various organisms. There are 3 table types available. Tables containing all possible 6-mers end with extension FR6, and those with all possible 7-mers end with extension FR7. Oligo does not have default tables that list all 8-mers, but Frequency Calculator can generate those and Oligo 7 can use them. The file extension name for 8-mer tables is FR8.

To make your own frequent table file you need to:

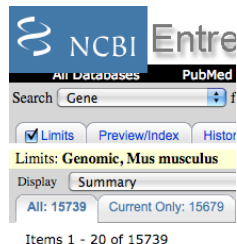
1. Assemble all your sequence files of interest on your local system in one folder.
2. Start Frequency Calculator and choose size of the database.
3. Select the Folder with sequences and click “Calculate” button.
4. Add your comments (optional) and save the sequence frequency table in a file.

The example below shows the details.

Example

1. Sequence retrieval.

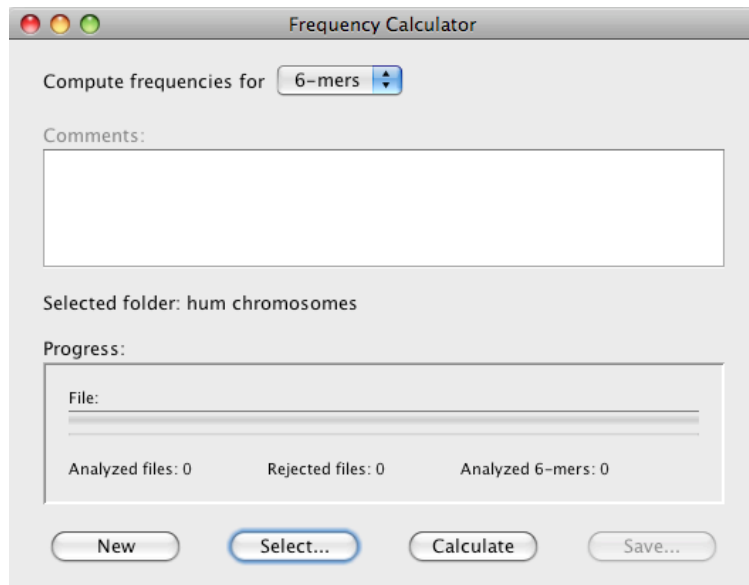
The first prerequisite is to have your sequence files on a local hard drive in one specific folder. You may download the sequences of interest from NCBI, <http://www.ncbi.nlm.nih.gov/>. They may be in GenBank or FASTA format. Files that do not contain sequences or are in a wrong format may be located in that specific folder and will be automatically ignored by the software. The extension of sequence files may be arbitrary (txt, seq, fasta are all fine). When retrieving sequences from NCBI using tab “Limits”



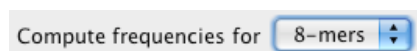
may be helpful. In our example, we created a folder “human chromosomes” on our computer desktop containing samples of human chromosome sequence files. Future versions of this software will have this step more automatic, but at this time, the retrieval of sequences is the most difficult step in this whole protocol.

2. Start Frequency Calculator and choose size of the database.

Once you have your folder containing sequences ready you may start Frequency Calculator and choose it. When you start the application the following windows appears:

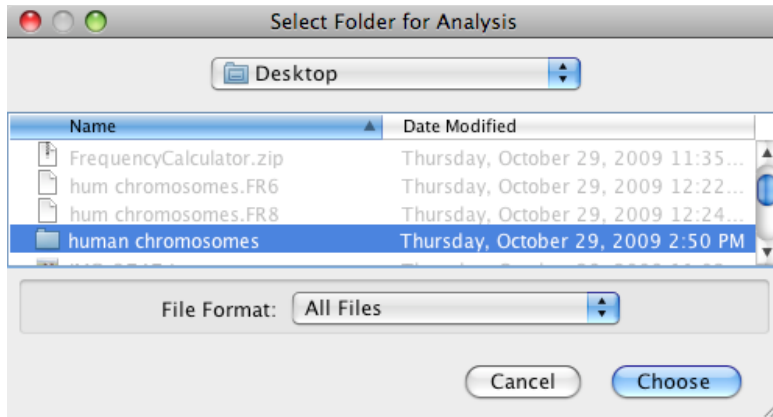


Choose the oligo size to calculate their frequencies using the top pop-up menu. If you choose “6-mers” the table of all possible 6-mer DNA combinations would have 4,096 entries. “7-mers” and “8-mers” would give 16,384 and 65,536, respectively. When you generate FR8 files, make sure that your sequence database is large enough to represent all possible 8-mers. For example when analyzing about 300Mb – long human DNA sequences you would come up with only about 6 eight-mers CGCGTACG and GTACGCGA (most rare) and about 110,000 repeats of A and T octamers. In our example, select “8-mers”.

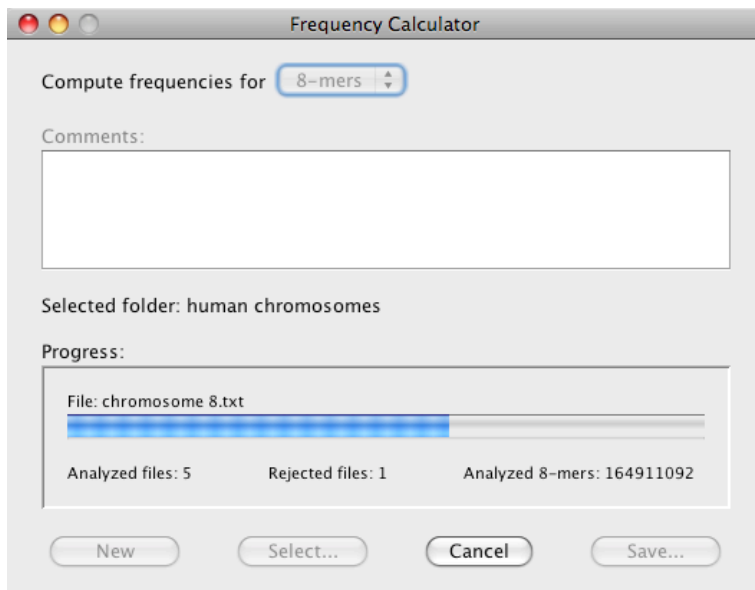


3. Select the Folder with sequences.

When you click the “Select” button the following window appears



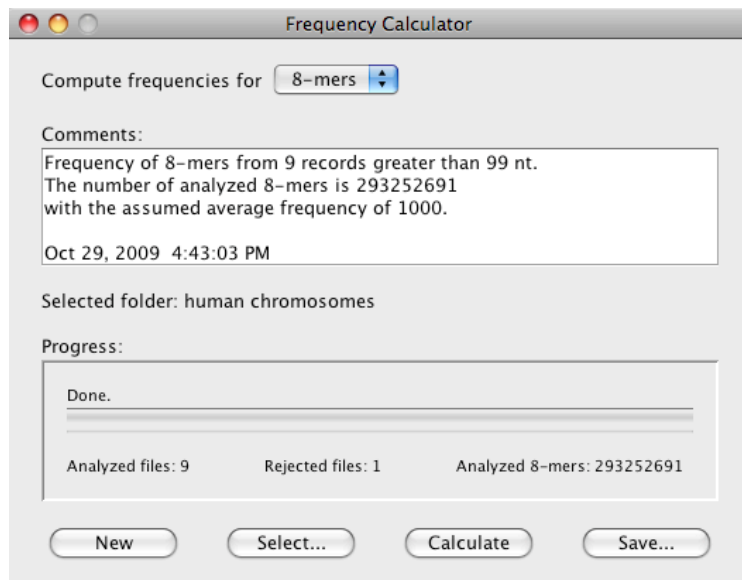
That will let you select a folder with your sequences. When you highlight the correct folder click the “Choose” button. Click the “Calculate” button when the folder is chosen.



At this point Frequency Calculator would start the calculations and display the progress bar and the statistics: how many files have been analyzed, how many rejected (files containing smaller sequences than 100 nt. are rejected), and how many x-mers have been analyzed.

4. Save the table in a file.

When the calculation is completed the Frequency Calculator displays comments it generated, as shown on the figure below.



A sequence with assumed average frequency below the 1000 number is more rare than 50% of all oligos and with above 1000 is more common than the other 6, 7 or 8-mers. You may write additional comments in the Comments field.

Button "Save" is for saving the table file. If you'd like to generate another frequency table, for example for 7-mers, choose "7-mers" and click "Calculate" button again. If you've created another folder with sequences, click "New" and start the process beginning from step 3 at this point.