

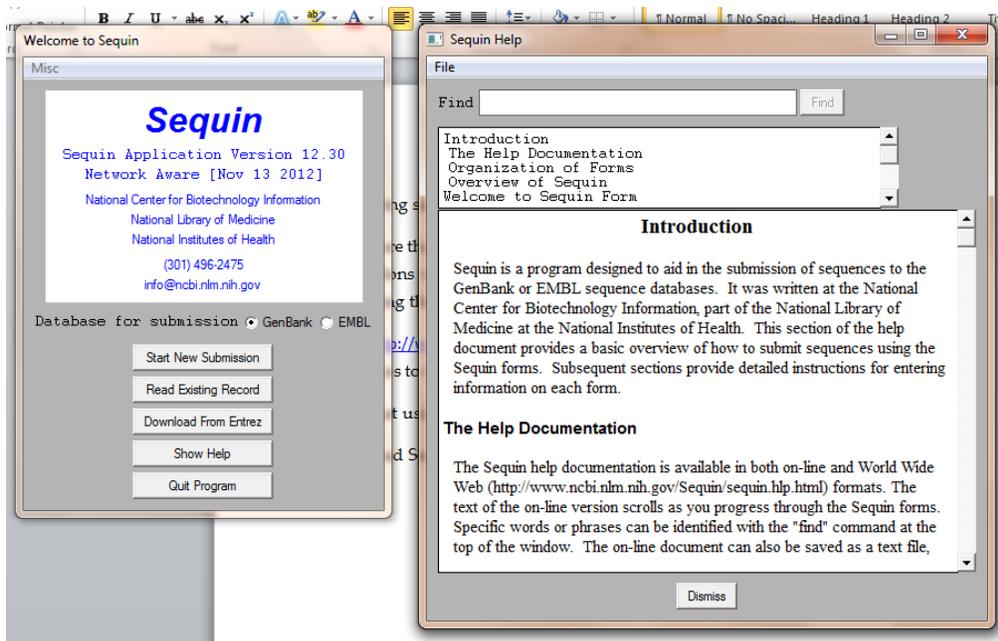
Submitting sequences to GenBank

Make sure that your alignment file is correct and also ensure that you don't have any stop codons within your sequences. Do this by setting the codon position and translating the sequence with the vertebrate mitochondrial code or standard code.

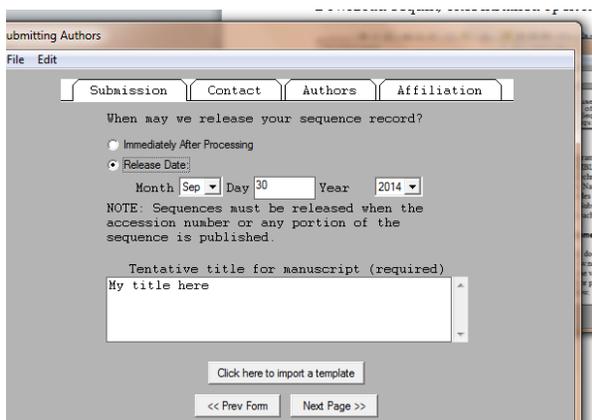
Go to <http://www.ncbi.nlm.nih.gov/genbank/submit/> to get complete details on submitting sequences to GenBank.

I find that using Sequin is fairly easy.

Download Sequin, once installed open it.



Click start new submission



Enter a tentative title and switch the release time from “immediately/ after processing” to “Release Date”, the default is one year from submission. Then click the contact tab and fill in this information.

The screenshot shows the 'Submitting Authors' form with the 'Contact' tab selected. The form has a menu bar with 'File' and 'Edit'. Below the menu bar are four tabs: 'Submission', 'Contact', 'Authors', and 'Affiliation'. The 'Contact' tab is active. The form contains the following fields: 'First Name', 'M. I.', 'Last Name', and 'Sfx' (a dropdown menu). Below these is a note: 'Please include country code for non-U.S. phone numbers.' There are also fields for 'Phone' and 'Fax'. At the bottom, there is an 'Email' field. Navigation buttons '<< Prev Page' and 'Next Page >>' are at the bottom center.

Now fill out the author and affiliations tabs

The screenshot shows the 'Submitting Authors' form with the 'Authors' tab selected. The form has a menu bar with 'File' and 'Edit'. Below the menu bar are four tabs: 'Submission', 'Contact', 'Authors', and 'Affiliation'. The 'Authors' tab is active. The form contains a table with columns 'First Name', 'M. I.', 'Last Name', and 'Sfx'. There are three rows, each with an 'X' in a small box to the left of the 'First Name' column. Below the table is a button labeled 'Insert First Author'. There is a checkbox labeled 'Consortium' followed by a text field. Below this is a note: 'The consortium field should be used when a consortium is responsible for the sequencing or publication of the data. Individual authors may be listed along with a consortium name.' Navigation buttons '<< Prev Page' and 'Next Page >>' are at the bottom center.

The screenshot shows the 'Submitting Authors' form with the 'Affiliation' tab selected. The form has a menu bar with 'File' and 'Edit'. Below the menu bar are four tabs: 'Submission', 'Contact', 'Authors', and 'Affiliation'. The 'Affiliation' tab is active. The form contains the following fields: 'Institution', 'Department', 'Street Address', 'City', 'State/Province', 'Zip/Postal Code', and 'Country'. There is a button labeled 'Click here to export a template'. Navigation buttons '<< Prev Page' and 'Next Form >>' are at the bottom center.

Once completed hit next form

Preparing the Sequences

How do you want to prepare your submission?

Use the normal submission dialog

Use a Submission Wizard:

Viruses

Uncultured Samples

rRNA-ITS-IGS sequences

TSA

Intergenic Spacer (IGS) sequences

Microsatellite sequences

D-loops and control regions

Back Next

Select which way you want to prepare the submission. If using Cytochrome b use the normal submissions dialog. Click next

Sequence Format

File

Submission type

Single Sequence Gapped Sequence

Population Study Phylogenetic Study

Mutation Study Environmental Samples

Batch Submission Transcriptome Shotgun Assembly

Sequence data format

FASTA (no alignment)

Alignment (FASTA+GAP, NEXUS, PHYLIP, etc.)

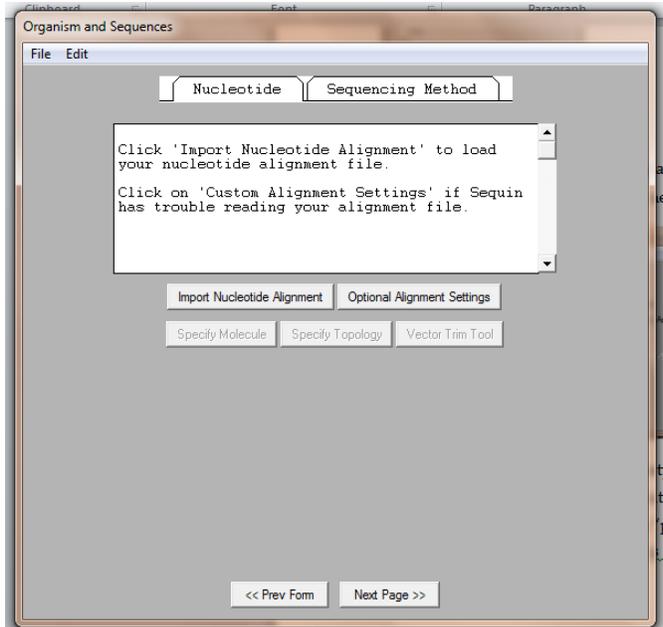
Submission category

Original Submission

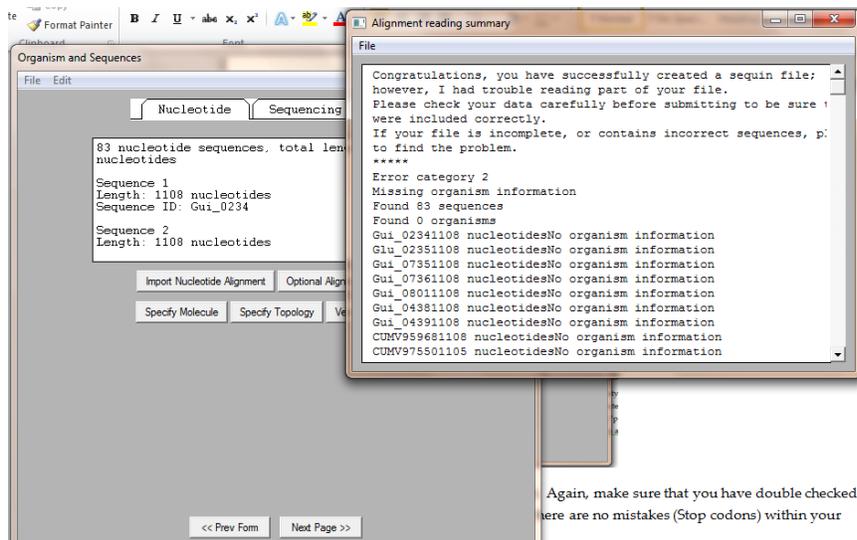
Third Party Annotation

<< Prev Form Next Form >>

We can submit sequences in a variety of ways: a single sequence, population studies, etc. Select the one that is appropriate for you. If you want to submit many sequences of similar or the same species choose “population study” or “phylogenetic study”. Once selected you can also select that it is an alignment. Check original submission, then click next form.

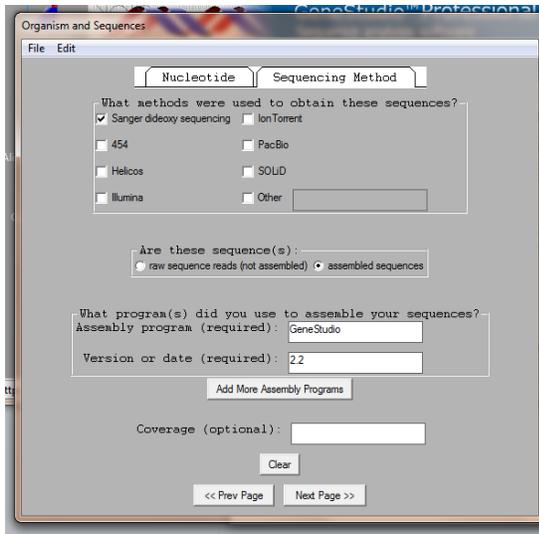


Click Import nucleotide alignment. Again, make sure that you have double checked your alignment to make sure that there are no mistakes (Stop codons) within your alignments. Find your alignment and import it.

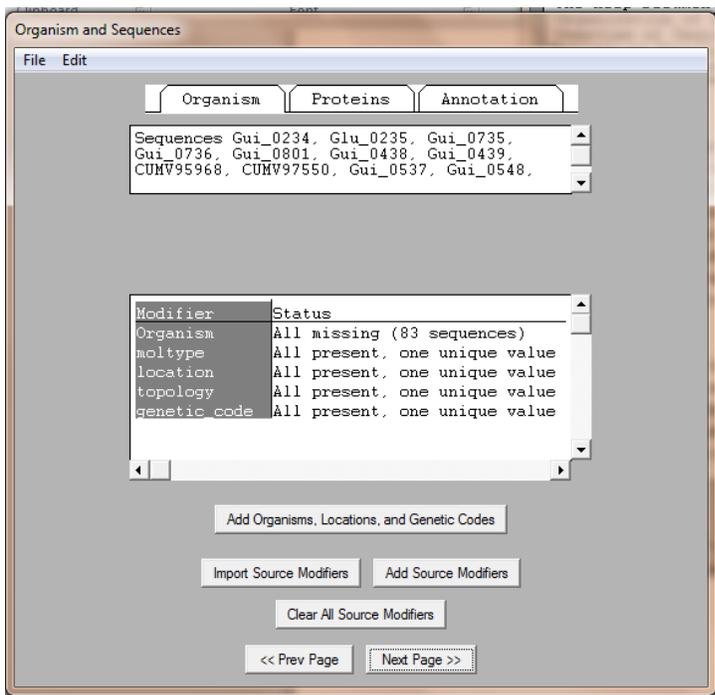


Again, make sure that you have double checked your alignment to make sure that there are no mistakes (Stop codons) within your alignments.

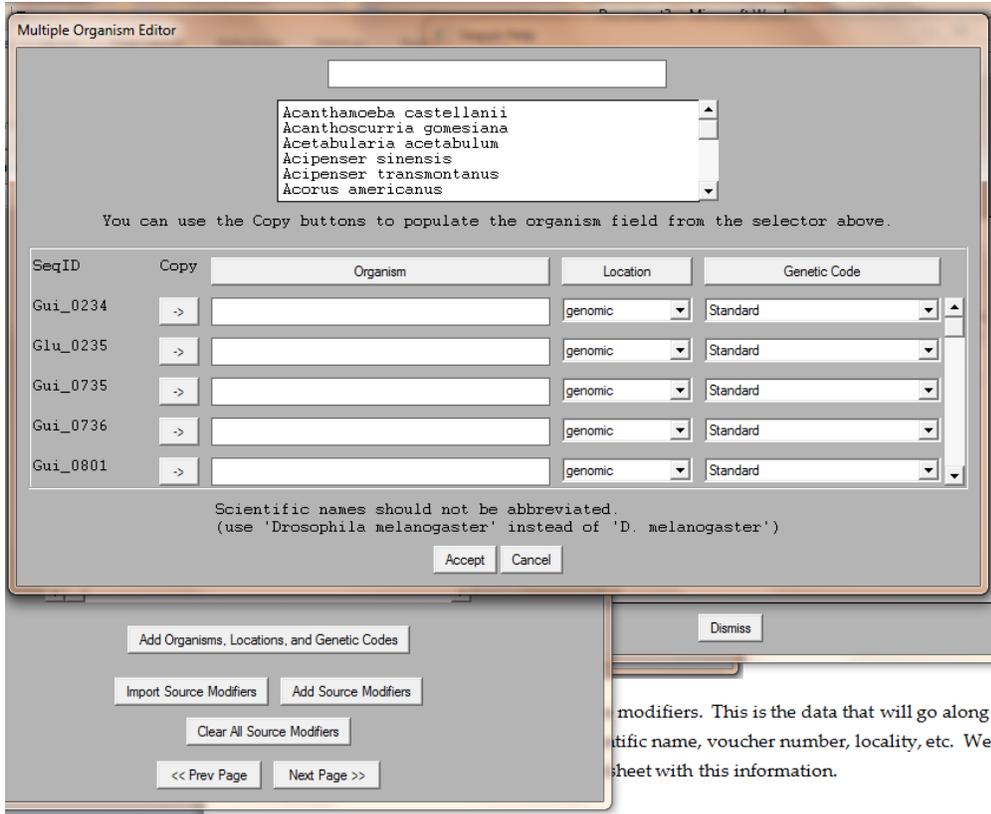
We successfully imported our sequence however it says that we are missing organism information. We can add this later. We can click next page. On the next page we need to say how these sequences were produced.



Click next page.

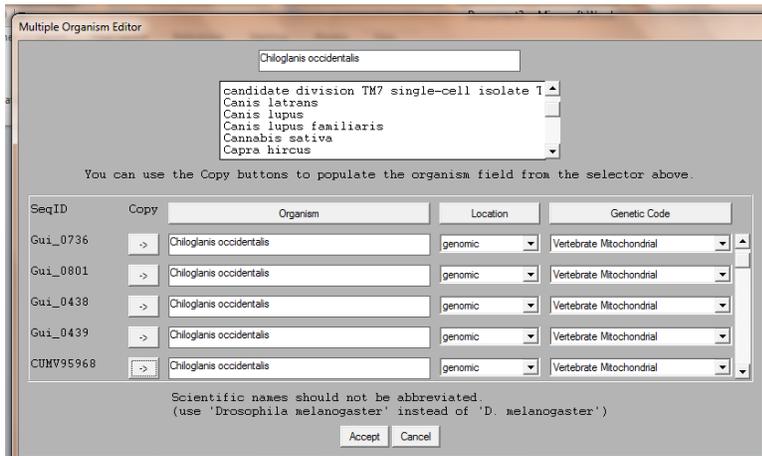


Now we can import or enter source modifiers. This is the data that will go along with your sequences. For example, scientific name, voucher number, locality, etc. We can do this one at a time or input a spreadsheet with this information. If you don't have very many you can click Add organisms, Locations and genetic codes.



modifiers. This is the data that will go along with the scientific name, voucher number, locality, etc. We can include this information on a separate sheet with this information.

Enter your species names and change the genetic code to vertebrate mitochondrial if from the mitochondria.



Once entered you can hit accept.

Organism and Sequences

File Edit

Organism Proteins Annotation

Add feature across full length of all sequences

CDS rRNA Gene None

Incomplete at 5' end Incomplete at 3' end

Plus strand Minus strand

Protein Name: Cytochrome b

Protein Description:

Gene Symbol: Cyt b

Comment:

Add title to all sequences if not in definition line

Title: cytochrome b (cytb) gene, partial cds; mitochondrial

Prefix title with organism name

<< Prev Page Next Form >>

You can skip the protein tab and proceed to the annotation tab. Put in your protein name and gene symbol. You can get this information for genbank. Also input a title that will be prefixed by your organism name. Again, look at other submissions in genbank to get a good idea of what appropriate titles are. We also want to click whether or not the sequences are complete are the 5' and 3' end. Do you have a sequence for the start and end? If so then leave both un-checked. Our sequences here are incomplete at the 5' end so I am going to check that. Click next form.

Look at the window below. For each organism it shows the accession file for the sequence and it also created one for the protein translation. Your original sequence name is the sequence and the name_1 is the protein translation.

Gui_0234

File Edit Search Misc Annotate

Target Sequence

Format Mode Style

Double click on an item to launch the appropriate editor.

```

LOCUS      Gui_0234                1108 bp    DNA        linear      27-SEP-2013
DEFINITION Chiloglanis occidentalis cytochrome b (cytb) gene, partial cds;
            mitochondrial.
ACCESSION
VERSION
KEYWORDS
SOURCE     Chiloglanis occidentalis
ORGANISM   Chiloglanis occidentalis
            Unclassified.
REFERENCE  1 (bases 1 to 1108)
AUTHORS    Smith, J.C.
TITLE      My title here
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1108)
AUTHORS    Smith, J.C.
TITLE      Direct Submission
JOURNAL    Submitted (27-SEP-2013) biology, my college, 2223 street, city, LA
            67367, ANY
COMMENT    ##Assembly-Data-START##
            Assembly Method      :: GeneStudio v. 2.2
            Sequencing Technology :: Sanger dideoxy sequencing
            ##Assembly-Data-END##
FEATURES   Location/Qualifiers
     source          1..1108
                    /organism="Chiloglanis occidentalis"
                    /mol_type="genomic DNA"
     gene            <1..1108
                    /gene="Cyt b"
     CDS             <1..1108
                    /gene="Cyt b"
                    /codon_start=1
                    /transl_table=2
                    /product="Cytochrome b"

```

```

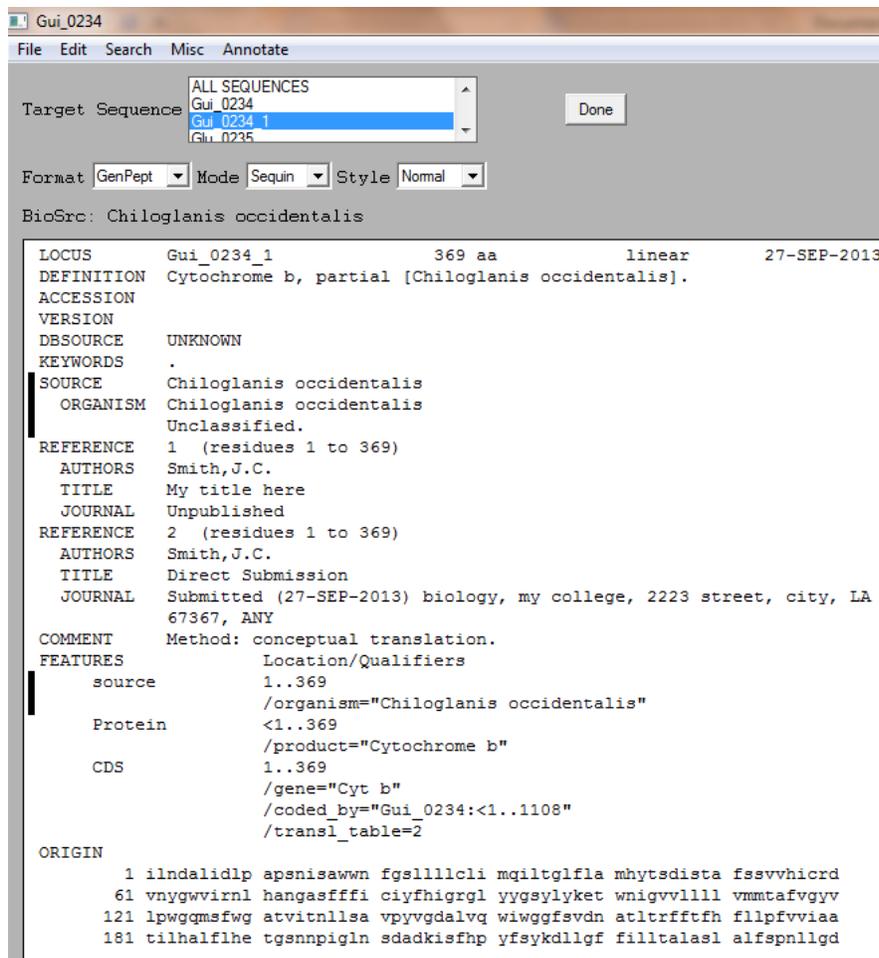
/translation="ILNDALIDLPAFNSISAWNNFGLSLLLLCLIMQILTGLFLAMHYT
SDISTAFSSVVHICRDVNYGWVIRNLHANGASFFFCIYFHFHIGRGLYGSYLYKETWN
IGVLLLLLMMTAFVGYVLPWGQMSFWGATVITNLLSAPVYVGDALVQWIWGGFSVDN
ATLTRFTFHFLPFVVAATILHALFLHETGSNNPIGLNSDADKISFHPYFSYKDLL
GFFILLTALASLALFSPNLLGDPENFTPANPLVTPPHIKPEWYFLFAYAILRSIPNKL
GVLALLFSILVLMVVPFLHTSKQQLTFRPFAQLLFWTLIADVAILTIGWGMPEVHE
FIIIGQIASVLYFMLFLILNPLIGLMENKALNRNC"

```

```

BASE COUNT      318 a      292 c      152 g      346 t
ORIGIN
1 atcttaaacg acgcgctaat cgacotcccc gccccctcaa acattttcgc ctgatgaaac
61 tttggctccc tcctactctt atgtcttatt atacaaatct taacaggact atttctagca
121 atgcactata catccgacat ctcaactgcc ttctcatcag ttgttcacat ctgccgagac
181 gttaattacg gctgagttat togaaatctg caogccaaog gagcttcatt tttcttcac
241 tgtatttact ttcattattg acgaggatta tattatggct catacttata taaagaaaca
301 tgaaacattg gagtactact tcttttacta gtaataataa ctgctttcgt aggatacgtc
361 ctccatgag gccaaatata attctgaggt gccacagtaa ttacaaatct actctcagct
421 tccocctacg taggggatgc tttagtacaa tgaatttgag gaggcttctc cgtagacaac
481 gcaacactaa caagattttt tacattccac ttctctctcc cattcgtagt cattgctggc
541 acaattttac atgcactttt cctacacgaa acaggctcaa acaaccaaat tggattaaac
601 tccgatgcag ataaaatctc attccacca tatttttcat acaaagacct attaggattt
661 tttattctat taacagcctt agcatcccta gccctattct ctccaaatct actaggagac
721 ccagaaaatt ttacccccgc caaccctta gtgactcccc ctcatattaa accagagtgg
781 tacttctcat ttgcctacgc catccttoga tctattccca ataaactagg aggagtacta
841 gcactattat tttctattct agtacttata gttgtgcoat ttctacacac ctctaataca
901 caaggactaa ctttccgacc tttcgtctca ctactatttt gaacactaat tgcagacggt
961 gccattctaa cctgaatcgg gggcatacca gttgaacatc cattcatcat tattggacaa
1021 atgcctcag tctctactt cactactatc ttaactctaa acccattaat tggtttaata
1081 gaaataaag ccctaaaccg caactgcc

```

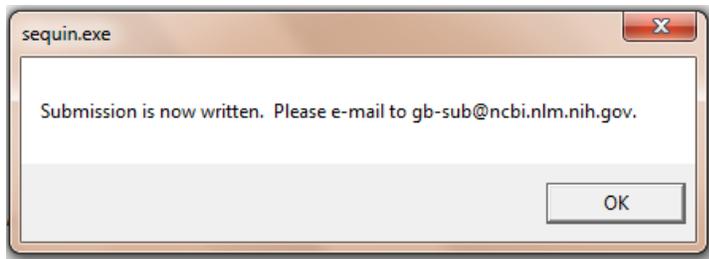


Go through all of the records to ensure that the information is correct. You can click on certain fields to edit them if needed (Organism)

Once you have reviewed all records you can hit Done

Sequin will review your errors. Hopefully you don't have many if any errors. You need to correct errors before you go forward. Click proceed and save this file to a folder of your choice.

Once saved you can email it to Genbank at the following address.



You can compose an email to this address with your submission file attached. I usually just write the following in the email:

Hello,

Please see the attached submission,

Best,

Your name

Institution

If you have any questions you can send me an email at rschmidt@tulane.edu