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 <u>http://www.ncbi.nlm.nih.gov/genbank/submit/</u> 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.

Submitting Authors	٦
File Edit	
Submission Contact Authors Affiliation	
First Name M.I. Last Name Sfx	
Please include country code for non-U.S. phone numbers.	
Phone Fax	
Enail	
<< Prev Page Next Page >>	

Now fill out the author and affiliations tabs

Submitting Authors	Report of Automations States and second the or
File Edit	
Submission Co	mtact Authors Affiliation
First Name X X X X Consortum Consortum Consortum consortum is or publication authors may be consortum nam	M.I. Last Name Sfx M.I. L
Submitting Authors	

Submitting Authors				
File Edit				
Suba	ission Contact Authors Affiliation			
Institution				
Department				
Street Address				
City				
State/Province	Zip/Postal Code			
Country				
	Click here to export a template			
	<< Prev Page Next Form >>			

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				
O Uncultured Samples				
C rRNA-ITS-IGS sequences				
🔿 TSA				
Intergenic Spacer (IGS) sequences				
Microsatellite sequences				
O 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			
Single Sequence C Gapped Sequence Submission type Mutation Study C Phylogenetic Study Mutation Study C Environmental Samples Batch Submission C Transcriptome Shotgun Assembly			
Sequence data format (FASTA (no alignment) Alignment (FASTA+GAP, (TEXUS, PHT/LIP, etc.)			
© Original Submission Submission category			
<< 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.

Clinhoard	E Font E Paragraph
Organism and	Sequences
File Edit	
	Nucleotide Sequencing Method
	vour nucleotide alignment file.
	Click on 'Ouster Mignant Settings' if Servin
	has trouble reading your alignment file.
	· · · · ·
	Import Nucleotide Alignment Optional Alignment Settings
	Specify Molecule Specify Topology Vector Trim Tool
	5
	te
	F
	<< Prev FormNext Page >>

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.

Promat Painter Clinboard 5 Format Painter 5 Format Painter 5 Format Fo	Alignment reading summary File
Organism and Sequences File Edit Nucleotide Sequences, total len nucleotides Sequence 1 Length: 1108 nucleotides Sequence 2 Length: 1108 nucleotides mport Nucleotide Algoment Optional Algo Specty Molecule Specty Topology Ve	Congratulations, you have successfully created a sequin file; however, I had trouble reading part of your file. Please check your data carefully before submitting to be sure : were included correctly. If your file is incomplete, or contains incorrect sequences, p: to find the problem. ***** Error category 2 Missing organism information Found 83 sequences Found 0 organisms Gui_02341108 nucleotidesNo organism information Gui_07351108 nucleotidesNo organism information Gui_07351108 nucleotidesNo organism information Gui_07351108 nucleotidesNo organism information Gui_0801108 nucleotidesNo organism information Gui_041108 nucleotidesNo organism information Gui_041108 nucleotidesNo organism information Gui_041108 nucleotidesNo organism information Gui_04391108 nucleotidesNo organism information Gui_04391108 nucleotidesNo organism information CUWV93561108 nucleotidesNo organism information CUWV975501105 nucleotidesNo organism information
<< Pirev Form Next Page >>	Again, make sure that you have double checked ere are no mistakes (Stop codons) within your

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.

ſ	Organism and Sequences
L	File Edit
l	Nucleotide Sequencing Method
l	What methods were used to obtain these sequences? Sangerdideoxy sequencing 🔽 IonTorent
AT .	☐ 454
	F Helicos F SOLID
	🗌 Illumina 🧮 Other
	Are these sequence(s): raw sequence reads (not assembled) • assembled sequences
l	What program(s) did you use to assemble your sequences?- Assembly program (required): GeneStudo
l	Version or date (required): 22
tt	Add More Assembly Programs
l	Coverage (optional):
	Clear
	<< Prev Page Next Page >>

Click next page.

Linboard	Ed Font			
Organism and Sequences				
File Edit				
	Organism Proteins Annotation			
	Gui 0736, Gui 0801, Gui 0438, Gui 0439,			
	CUMV95968, CUMV97550, Gui_0537, Gui_0548,			
	Modifier Status			
	Organism All missing (83 sequences)			
	moltype All present, one unique value			
	topology All present, one unique value			
	genetic_code All present, one unique value			
	Add Organisms, Locations, and Genetic Codes			
	Import Source Modifiers Add Source Modifiers			
	Clear All Source Modifiers			
	A Prov Page Nett Page 22			
	(L			

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.

Multiple Organ	ism Editor	and the second s	n	2
Ŷ	'ou can use	Acanthamoeba castellanii Acanthoscurria gomesiana Acetabularia acetabulum Acipenser sinensis Acipenser transmontanus Acorus americanus the Copy buttons to populate the	organism field fro	▲ ■ Dom the selector above.
SeqID	Сору	Organism	Location	Genetic Code
Gui_0234	->		genomic 💌	Standard 💌 🔺
G1u_0235	->		genomic 💌	Standard 👤
Gui_0735	->		genomic 💌	Standard
Gui_0736	->		genomic 💌	Standard
Gui_0801	->		genomic 💌	Standard 💌 🗸
	Scientific names should not be abbreviated. (use 'Drosophila melanogaster' instead of 'D. melanogaster') Accept Cancel			
Add Organisms, Locations, and Genetic Codes Dismiss Import Source Modifiers Add Source Modifiers Clear All Source Modifiers modifiers. This is the data that will go along v tific name, voucher number, locality, etc. We control to the set with this information.				

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

1	Multiple Organism E	ditor		<u> </u>	р. <u>р.ш. і</u>
l			Chiloglanis occidentalis		
a	candidate division TM7 single-cell isolate T Canis latrans Canis lupus Canis lupus familiaris Cannabis sativa Capra hircus ▼				
l	You c	an use	the Copy buttons to populate the org	anism field from	m the selector above.
l	SeqID	Сору	Organism	Location	Genetic Code
l	Gui_0736	•>	Chiloglanis occidentalis	genomic 💌	Vertebrate Mitochondrial
l	Gui_0801	->	Chiloglanis occidentalis	genomic 💌	Vertebrate Mitochondrial
l	Gui_0438	->	Chiloglanis occidentalis	genomic 💌	Vertebrate Mitochondrial
l	Gui_0439	->	Chiloglanis occidentalis	genomic 💌	Vertebrate Mitochondrial
l	CUMV95968	>	Chiloglanis occidentalis	genomic 💌	Vertebrate Mitochondrial
			Scientific names should not be abbre (use 'Drosophila melanogaster' inste Accept Cancel	eviated. ad of 'D. meland	ogaster')

Once entered you can hit accept.

Organism and Sequences				
File Edit				
Organism Proteins Annotation				
Add feature ac	ross full length of all sequences			
٠	CDS 🔿 rRNA 🔿 Gene 🔘 None			
[] Incor	mplete 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 *			
I	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.

File Edit Search	Mirc Appotate			
rile Edit Search	Wise Amotale			
	ALL SEQUENCES			
Target Sequen	Ce Gui 0234_1			
	Glu 0235			
Format GenBank	▼ Mode Sequin ▼ Style Nomal ▼			
Pormate journame				
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				
SOURCE	Chiloglanis occidentalis			
ORGANISM	Chiloglanis occidentalis			
	Unclassified.			
REFERENCE	1 (bases 1 to 1108)			
AUTHORS	Smith, J.C.			
JOURNAL	my cicle here Unnublished			
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			
COMMENT	##Assembly-Data-START##			
	Assembly Method :: GeneStudio v. 2.2			
	Sequencing Technology :: Sanger dideoxy sequencing			
	##Assembly-Data-END##			
FEATURES	Location/Qualifiers			
000200	/organism="Chiloglanis occidentalis"			
	/mol_type="genomic DNA"			
gene	<11108			
CDS	/gene="Cyt b"			
022	/gene="Cyt b"			
	/codon_start=1			
	/transl_table=2			
	/product="Cytochrome b"			
	/translation="ILNDALIDLPAPSNISAWWNFGSLLLLCLIMQILTGLFLAMHYT			
	SDISTAFSSVVHICRDVNYGWVIRNLHANGASFFFICIYFHIGRGLYYGSYLYKETWN IGVVLLLLVMMTAFVGYVLPWGOMSFWGATVITNLLSAVPYVGDALVOWIWGGFSVDN			
	ATLTRFFTFHFLLPFVVIAATILHALFLHETGSNNPIGLNSDADKISFHPYFSYKDLL			
	GFFILLTALASLALFSPNLLGDPENFTPANPLVTPPHIKPEWYFLFAYAILRSIPNKL			
	FIIIGQIASVLYFMLFLILNPLIGIMENKALNRNC"			
BASE COUNT	318 a 292 c 152 g 346 t			
OKIGIN 1 atc	ttaaacg acgoggtaat cgaceteece geeeceteaa acatttetge etgatgaaac			
61 ttt	ggetece tectactett atgtettatt atacaaatet taacaggaet atttetagea			
121 atg	cactata catcogacat ctcaactgcc ttctcatcag ttgttcacat ctgccgagac			
181 gtt 241 tat	aattacg getgagttat tegaaatetg eaegeeaaeg gagetteatt titetteate atttaet tieatattog aegaggatta tattatoget eataettata taaagaaaea			
301 tga	aacattg gagtagtact tcttttacta gtaataataa ctgctttcgt aggatacgtc			
361 ctt	ccatgag gccaaatatc attctgaggt gccacagtaa ttacaaatct actctcagct			
421 gtc 481 gca	uuuauy tayyyyatyo titaytacaa tyaattiyay yäggottoto ogtägädääd acactaa caogattiit tacattocac tiooigotoo cattoqtaqt cattgoqqoo			
541 aca	attttac atgcactttt cctacacgaa acaggctcaa acaacccaat tggattaaac			
601 tcc	gatgcag ataaaatoto attocacoca tatttttoat acaaagacot attaggattt			
721 cca	arretar taacayeert ayeareeera geeetarret eteedaattt aetaggagae gaaaatt ttaeeeeege caaceeetta gtgaeteeee eteatattaa aeeagagtgg			
781 tac	tteetat ttgeetaege cateettega tetatteeea ataaactagg aggagtaeta			
841 gca	stattat titstattet agtasttata gitgigssat tistasasas sitesaasaa			
961 gcc	gyactaa ciiccogace ciicgeteaa ciactatti gaadadtaat tydayadytt attetaa eetgaategg gggeataeea gitgaaeate eatteateat tattggaeaa			
1021 atc	gceteag teetetaett catactatte ttaatettaa acceattaat tggtttaata			
1081 gaa	aataaag cootaaacog caactgoo			

🗈 Gui_0234				
File Edit Search Misc Annotate				
Target Sequenc	ALL SEQUENCES Gui 0234 Gui 0234 1 (Giu 0235			
Format GenPept 💌 Mode Sequin 💌 Style Nomal 💌				
BioSrc: Chiloglanis occidentalis				
LOCUS (DEFINITION (ACCESSION VERSION	Gui_0234_1 369 aa linear 27-SEP-2013 Cytochrome b, partial [Chiloglanis occidentalis].			
DBSOURCE U	UNKNOWN			
SOURCE ORGANISM	Chiloglanis occidentalis Chiloglanis occidentalis Unalanificad			
REFERENCE AUTHORS	Smith, J.C.			
JOURNAL	Unpublished			
REFERENCE	EFERENCE 2 (residues 1 to 369)			
TITLE	Direct Submission			
JOURNAL Submitted (27-SEP-2013) biology, my college, 2223 street, city, 1				
COMMENT 1	Method: conceptual translation.			
FEATURES	Location/Qualifiers			
source	1369			
Protein	<1369			
	/product="Cytochrome b"			
CDS	1369			
	/gene="Cyt b"			
/coded_py="Gu1_U234:<11106" /transl_table=2				
ORIGIN	, 020002_00020 2			
1 il:	ndalidlp apsnisawwn fgsllllcli mqiltglfla mhytsdista fssvvhicrd			
61 vn	ygwvirnl hangasfffi ciyfhigrgl yygsylyket wnigvvllll vmmtafvgyv			
121 lpt 181 +i	wgqmsiwg atvitniisa vpyvgdalvq wiwggisvdn atitriitin filpfvviaa Thalflhe tosnonioln sdadkisfhn vfsykdllof filltalasi alfennlod			
for orthogram of the starting and the st				

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.

sequin.exe			
Submission is now written. Please e-mail to gb-sub@ncbi.nlm.nih.gov.			
	ок		

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 <u>rschmidt@tulane.edu</u>