Submission sequence to NCBI

At begin this book we mention the Database collect sequences or bio samples from researchers and database staffers. Now we explain how researchers submitted sequences to database (NCBI). Before started, we need to understand, what is bio samples? Why the researchers wanted submit their samples to database? In addition to some roles, we must follow it when we submitted sequence.

Bio sample that we can submitted it to NCBI

- ✓ DNA (Gene, genome, plasmid, metaDNA, tandem repeat and others).
- ✓ Protein (sequence of amino acid).
- ✓ RNA (all type of RNA).

Researchers submitted their sequence to proven their work, share their sample with other researchers and to largest the database to become contain a lot of information.

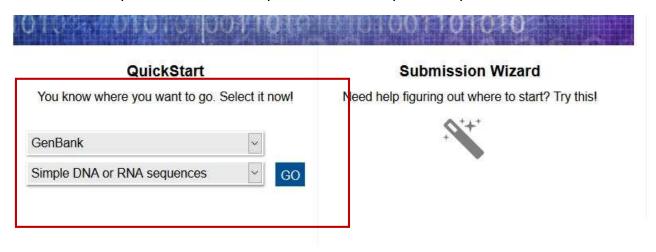
Some roles you must keep it in your mind, when you have DNA sample and submitted it to NCBI you cannot transcript or translate it to RNA or protein and submitted it too.

Now the following steps explain how you can submitted your sequence to NCBI

1. Visit NCBI website and choose submit.



Select the type of your data and type of your sequence, choose <GenBank> and <simple DNA or RNA sequence> as example. And press <GO>



3. Now choose <Banklt>

How to submit data to GenBank

The most important source of new data for GenBank® is direct submissions from scientists. GenBank depends on its contributors to help keep the database as comprehensive, current, and accurate as possible. NCBI provides timely and accurate processing and biological review of new entries and updates to existing entries, and is ready to assist authors who have new data to submit.

Receiving an Accession Number for your Manuscript

Most journals require DNA and amino acid sequences that are cited in articles be submitted to a public sequence repository (DDBJ/ENA/Genbank - INSDC) as part of the publication process. Data exchange between DDBJ, ENA and GenBank occurs daily so it is only necessary to submit the sequence to one database, whichever one is most convenient, without regard for where the sequence may be published. Sequence data submitted in advance of publication can be kept confidential if requested. GenBank will provide accession numbers for submitted sequences, usually within two working days. This accession number serves as an identifier for your submitted your data, and allows the community to retrieve the sequence upon reading the journal article. The accession number should be included in your manuscript, preferably in a footnote on the first page of the article, or as required by individual journal procedures.

Submissions to GenBank

ere are several options for submittil g data to GenBank:

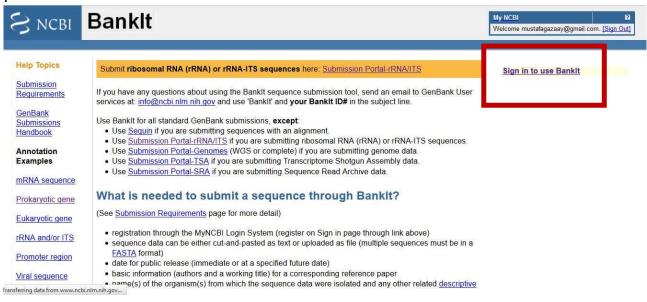
- Banklt, a WWW-based submission tool with wizards to guide the submission process
- tbl2asn, a command-line program, automates the creation of sequence records for submission to GenBank using many of the same functions as Sequence: a sequence and large batches of sequences and is available by FTP for use on MAC, PC and Unix platforms.
- <u>Submission Portal</u>, a unified system for multiple submission types. Currently only ribosomal RNA (rRNA), rRNA-ITS or Influenza sequences can be submitted with the GenBank component of this tool. This will be expanded in the future to

GenBank Resources

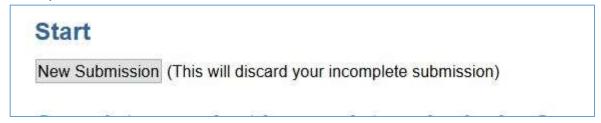
GenBank Home
Submission Types
Submission Tools
Search GenBank
Update GenBank Record



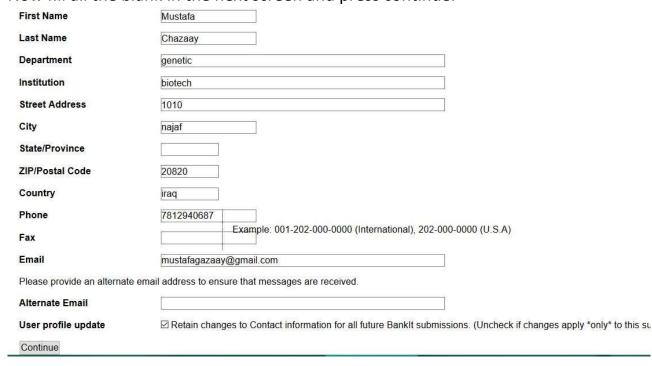
4. You will see the following screen. You must sign in to Banklt to complete the process.



5. Now press on <new submission>.



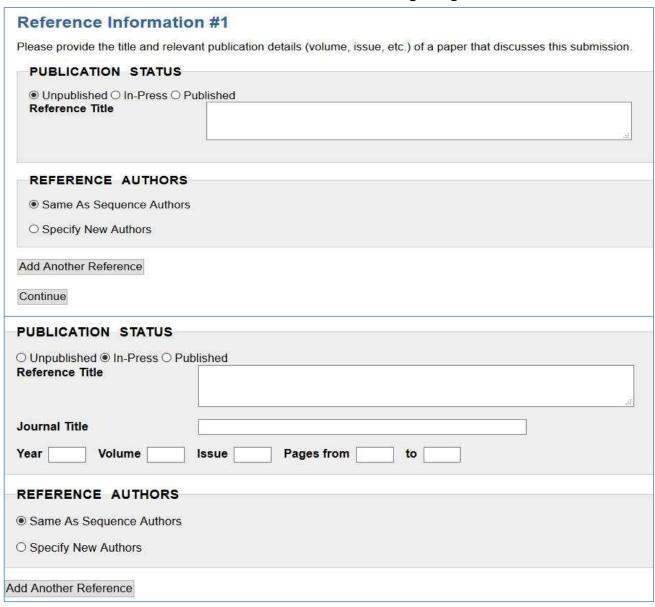
6. Now fill all the blank in the next screen and press continue.



7. Fill the blank, if you have more than one author click on <add> to mention other authors, in the middle name put just the first letter.

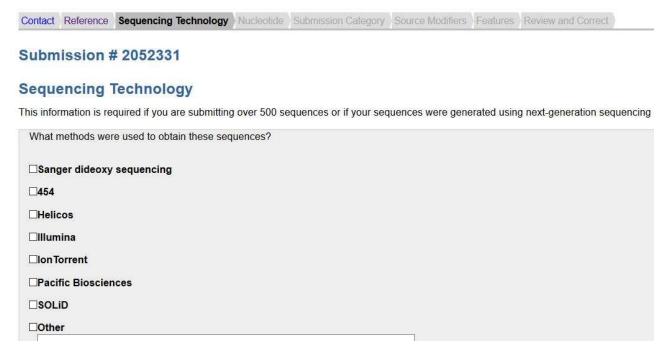


8. Reference information, you have three choices; unpublished, in-process and published. If the sequence of your search unpublished yet now you just write the title. If your search published, you must mention the journal and volume in addition to title and PubMed ID. Note the following images.

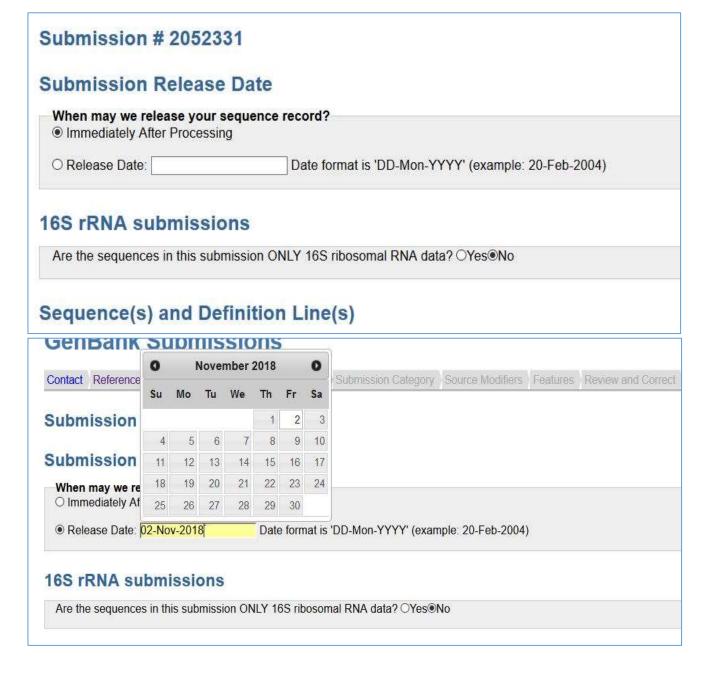


PUBLICATION STATUS		
○ Unpublished ○ In-Press ● Reference Title	Published	
Journal Title		
Year Volume	Issue Pages from to	
OR		
PubMed ID		
REFERENCE AUTHORS		
Same As Sequence Authors	S ₁	
O Specify New Authors		
Add Another Reference		

9. After filled all click on <same As Sequence Authors> and press continue. Then you will see the following screen, this screen specialized if you want to submit many sequence or if you use next generation sequence. So we now training on simple sequence, directly go down and press continue.



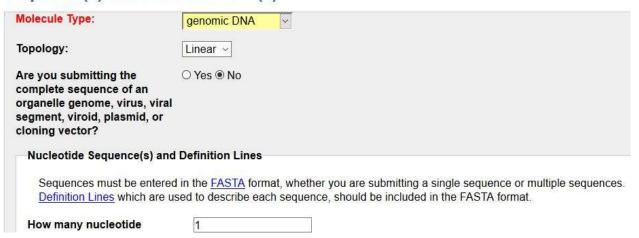
10. Submission release data you have two choice; immediately after process and release data (the balk to select the data). The first choice, your sequence directly become available for all researchers after processed by NCBI staff. Second, you must select the data in future.



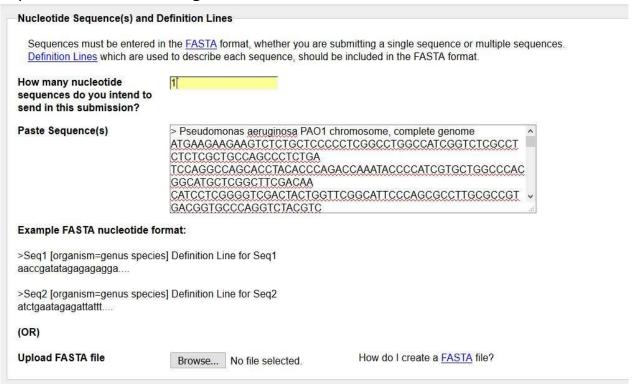
11.Select type of sequence from <molecular type> menu. Select topology if your sample liner or circular from <topology> menu. Write how many sequence

you will submitted it.

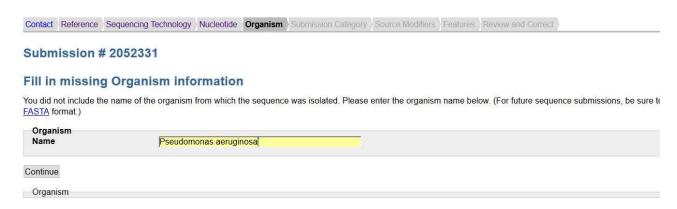
Sequence(s) and Definition Line(s)



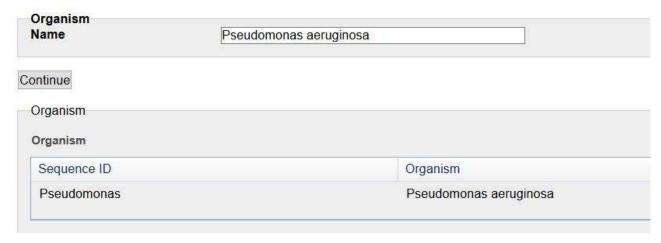
12. Copy your sequence from Fasta form and past it in the following blank, which explained in the next image. And click continue.



13. Organism name, write organism name and wait to appear the name spontaneously and click on it and press continue.



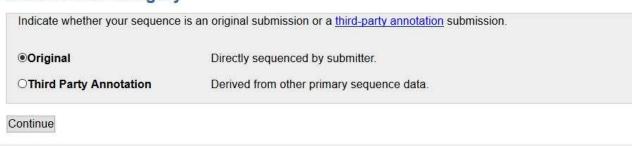
14. You will see the following screen. Press continue.



15. Click original and press continue.

Submission # 2052331

Submission Category



16. Source modifiers, you have three choices; first pure culture.

Submission # 2052331

Warning: Bacterial/Archaeal sequences require additional information about how they were obtained. Please indicate how they were obtained choice.

Source Modifiers

Bacterial/Archaeal Sequences: How were they obtained?

●pure-cultured strain(s): contains only one microbial species (axenic culture)

Oenrichment culture: sequenced directly from a mixed culture (non-axenic). Do not choose this option for purified strains. ○

Ouncultured, bulk environmental DNA:

PCR-amplified directly from environmental sample or host; samples were not grown in culture.

Please provide Isolation Source (if applicable) and Strain identifier in the Source Modifier section below.

Carres Organistical acadian Information

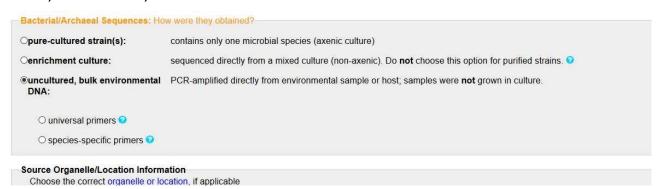
Second enrichment culture.

Source Modifiers

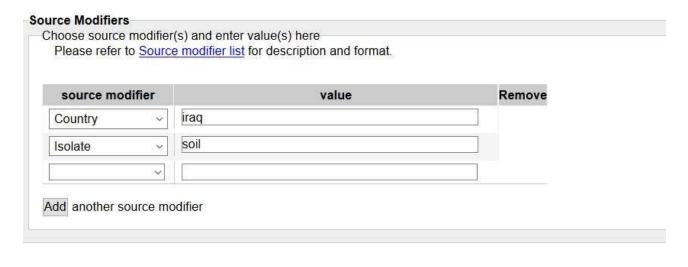
Bacterial/Archaeal Sequences:	How were they obtained?
Opure-cultured strain(s):	contains only one microbial species (axenic culture)
enrichment culture:	sequenced directly from a mixed culture (non-axenic). Do not choose this option for purified strains.
Did you further purify individ	ual strains from the enrichment culture and sequence from the purified strain(s)?
O Yes, sequences are from	purified strains isolated from an enrichment culture.
O No, sequences are directly	y from the mixed enrichment culture.
Ouncultured, bulk environmen DNA:	PCR-amplified directly from environmental sample or host; samples were not grown in culture.

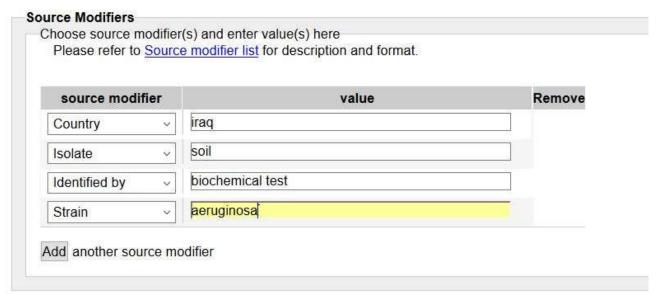
Source Organelle/Location Information

Third, un cultured, bulk environmental DNA.

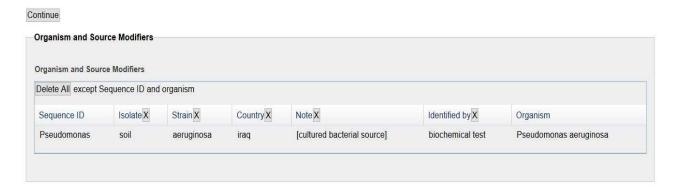


17.From <source modifier> menu select the approach and write your comment in <value> blank. More source modifier information make you sequence more clear and understood.





18. Click continue to see the following screen, click <continue> to continue.



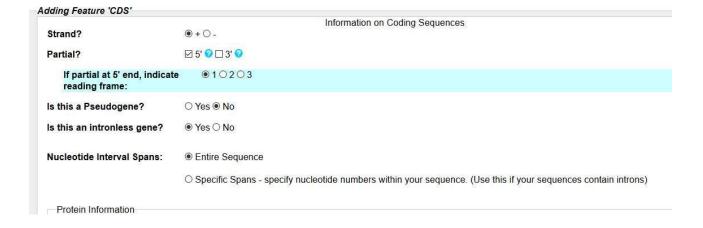
19. Now add feature to your sample (more information).

Add features by completing input forms	
This method is best suitable for:	
a single feature or a few features applied on a single sequence	
• the same single feature or the same few features applied to all sequences in a set or batch submission (for example: COI or 16S rRNA sequences)	ces)
features can be added across an entire sequence or by specific intervals within a sequence	
• one or more qualifiers can be chosen to apply to each feature; if it is not clear what qualifier is correct for a feature, use the 'note' qualifier	
To add a feature, select feature category and feature type within that category then click 'Add'. Coding Region (CDS) / Gene / mRNA if your sequence encodes a protein, choose this option	
○ RNA (rRNA, tRNA, non-coding RNA, misc_RNA, etc)	
O Repeat region (for sequence repeats, mobile elements and satellites)	
○ Regulatory feature (promoter, TATA_signal, RBS etc.)	
O Other	
O Regulatory feature (promoter, TATA_signal, RBS etc.)	
Add	

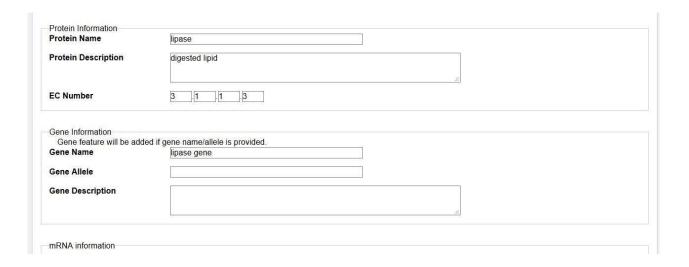
20. If your sequence coded to protein, write the information of protein look to the following screen.

To add a feature, select feature category and feature type within that category then click 'Add'.	
Add CDS by	
● providing intervals ○ providing protein sequence data	
O RNA (rRNA, tRNA, non-coding RNA, misc_RNA, etc)	
O Repeat region (for sequence repeats, mobile elements and satellites)	
O Regulatory feature (promoter, TATA_signal, RBS etc.)	
O Other	
Add	

21. Now select suitable choice.



22. Write protein name, protein descript and enzyme commission number (if the protein have enzymatically properties. In addition to gene name, gene allel and gene description.



23. After click on continue you will see the following screen.

Features (Overview)

Please provide feature annotations for your submission by choosing one of the two options below.

- O Add features by uploading five column feature table file
- O Add features by completing input forms



24. When press continue you will see GenBank forma of lipase gene for amino acid sequence as the following screen.

Continue

```
Features
                       <1..936
                       /gene='lipase gene'
      CDS
                       <1..936
                       /gene='lipase gene'
                       /EC number='3.1.1.3'
                       /note='digested lipid; [intronless gene]'
                       /codon start=1
                       /transl_table=11
                       /product='lipase'
                       translation='MKKKSLLPLGLAIGLASLAASPLIQASTYTQTKYPIVLAHGMLG/
                       FDNILGVDYWFGIPSALRRDGAQVYVTEVSQLDTSEVRGEQLLQQVEEIVALSGQPKV
                       NLIGHSHGGPTIRYVAAVRPDLIASATSVGAPHKGSDTADFLROIPPGSAGEAVLSGL
                       VNSLGALISFLSSGSTGTQNSLGSLESLNSEGAARFNAKYPQGIPTSACGEGAYKVNG
                       VSYYSWSGSSPLTNFLDPSDAFLGASSLTFKNGTANDGLVGTCSSHLGMVIRDNYRMN
                       HLDEVNQVFGLTSLFETSPVSVYRQHANRLKNASL*'
```

25. Press continue to open review submission page you will see all information you write it and choose it.

Submission # 2052331

Review Submission

Additional Email Addresses?
 Correspondence regarding this submission will be sent to the following email address: mustafagazaay@gmail.com
 Separate multiple email addresses with commas.

2 Resultmission?

2. Resubmission?

If you were asked by GenBank staff to resubmit your sequence data, check here: \Box

3. Submission Title (Optional)

If you want to title your submission for your own record-keeping, check here: \Box

4. Additional Information

If you have additional or corrected source modifier or feature files, or other plain text description for your sequence data submission, check here: \Box

5. Updates

You may update or revise your submissions at any time by sending new or corrected information in an email to update@ncbi.nlm.nih.gov. You may also contact us at this address with any questions.

Review Records of Your Set

Below please find your 1 genbank submission record(s) for your review.

You can download the complete set as a compressed ZIP file

Finish Submission

```
936 bp
                                                         DNA
                                                                  linear BCT 04-OCT-2017
              Pseudomonas
DEFINITION
              aeruginosa PAO1 chromosome, complete genome.
Pseudomonas
ACCESSION
VERSION
KEYWORDS
              Pseudomonas aeruginosa
SOURCE
  ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
              Pseudomonadaceae; Pseudomonas.
                  (bases 1 to 936)
  AUTHORS sarhan, m.C.
TITLE test for submission sequence to NCBI
  JOURNAL
              Unpublished
2 (bases 1 to 936)
REFERENCE
  AUTHORS
              sarhan, m.C.
              Direct Submission
Submitted (04-OCT-2017) genetic, biotech, 1010, najaf 20820, iraq
Bankit Comment: TOTAL # OF SEQS:1.
  TITLE
COMMENT
FEATURES
                         Location/Qualifiers
      source
                          1..936
                         /organism='Pseudomonas aeruginosa'
```

When you click on finish submission, your sample and metadata will send to NCBI staff to see it and accepted to publish or not. When they accepted your sample you will receive email contain number of sequence (accession number).