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

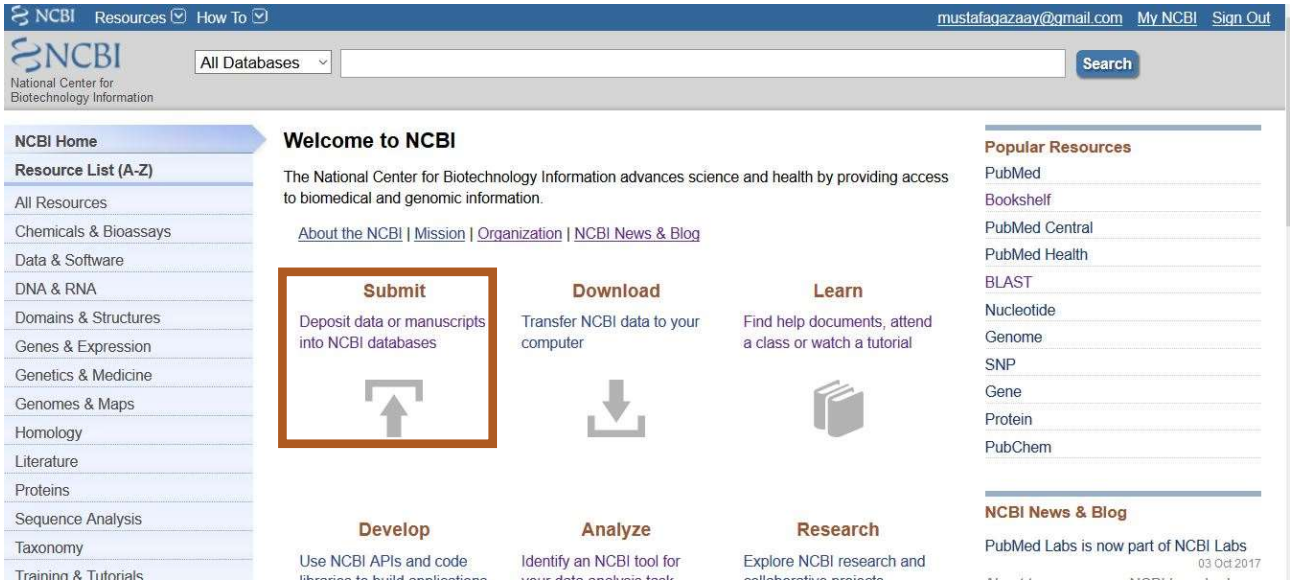
- ✓ DNA (Gene, genome, plasmid, metaDNA, tandem repeat and others).
- ✓ Protein (sequence of amino acid).
- ✓ RNA (all type of RNA).

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.

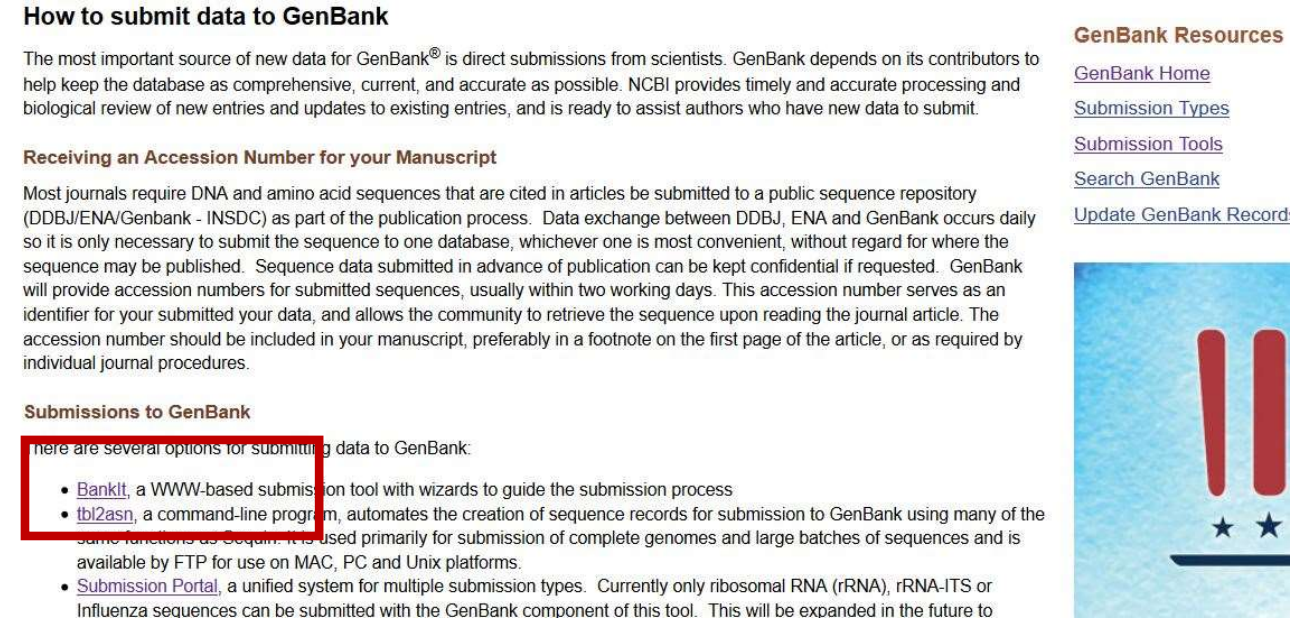
Now the following steps explain how you can submitted your sequence to NCBI

1. Visit NCBI website and choose submit.

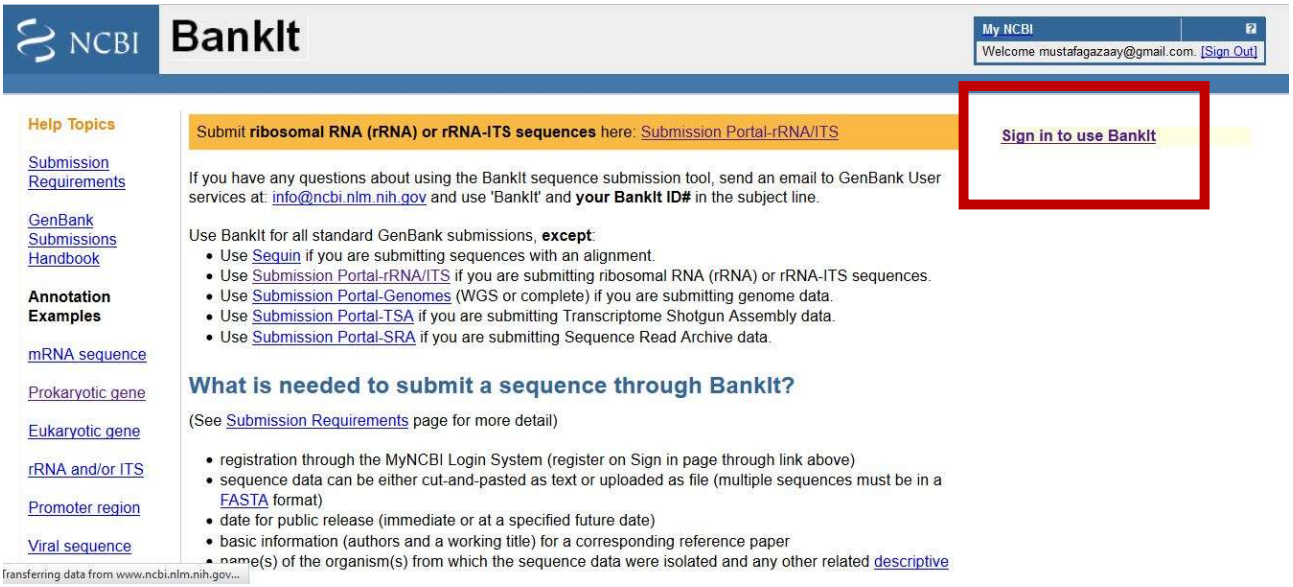


2. Select the type of your data and type of your sequence, choose <GenBank> and <simple DNA or RNA sequence> as example. And press <GO>

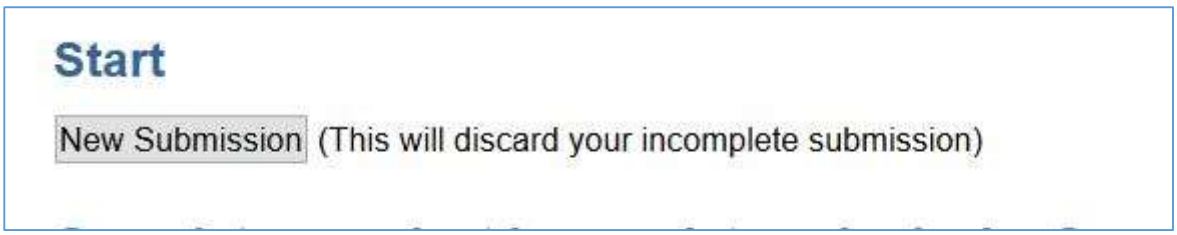
3. Now choose <BankIt>



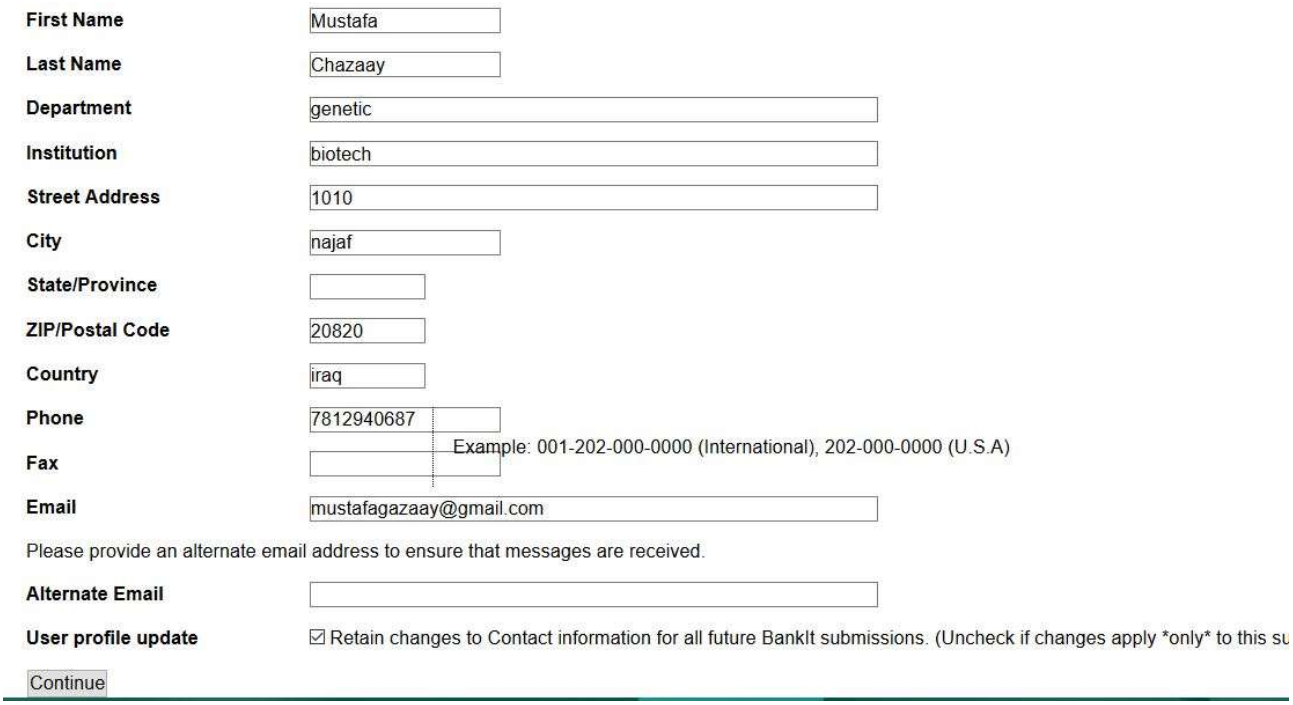
4. You will see the following screen. You must sign in to BankIt to complete the process.



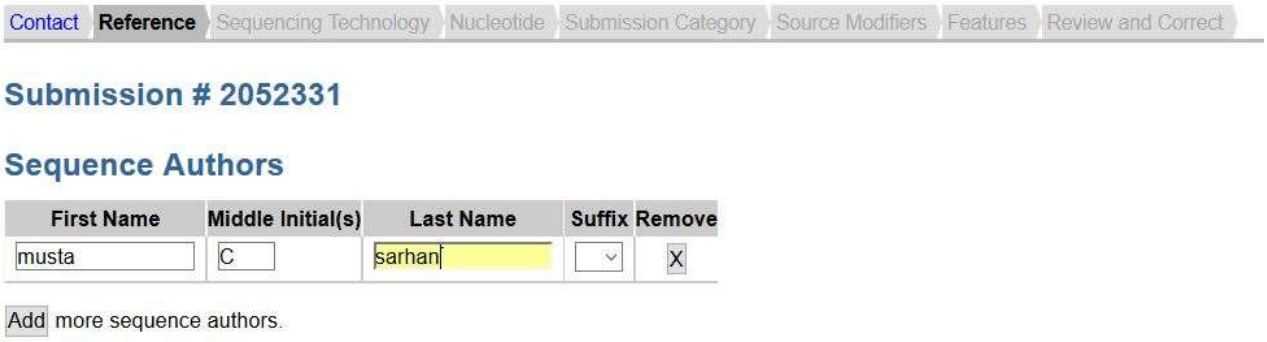
5. Now press on <new submission>.



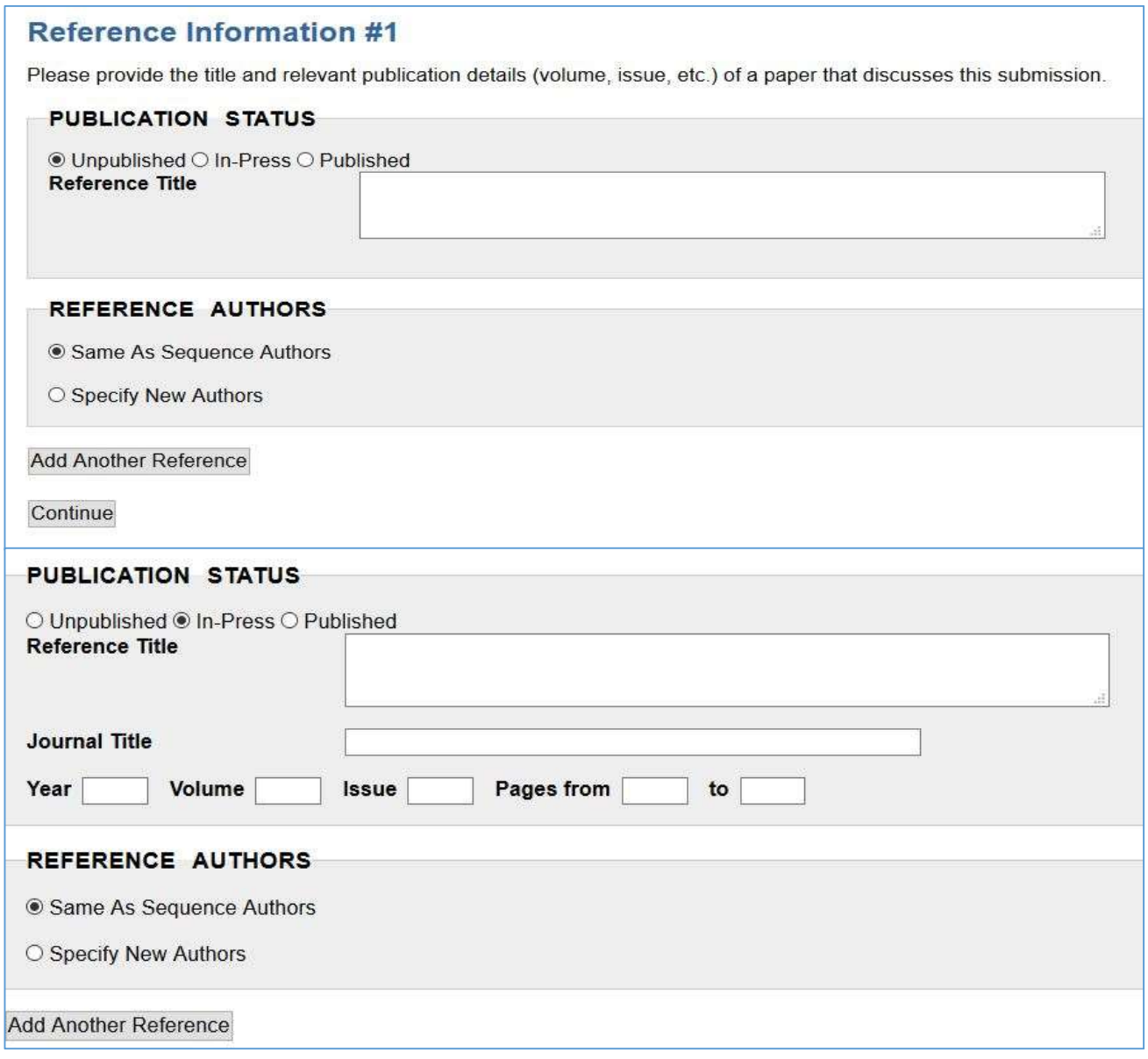
6. Now fill all the blank in the next screen and press continue.



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 ◉ Published

Reference Title

Journal Title

Year **Volume** **Issue** **Pages from** **to**

OR

PubMed ID

REFERENCE AUTHORS

◉ Same As Sequence Authors

○ 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.

Contact Reference **Sequencing Technology** Nucleotide Submission Category Source Modifiers Features Review and Correct

Submission # 2052331

Sequencing Technology

This information is required if you are submitting over 500 sequences or if your sequences were generated using next-generation sequencing

What methods were used to obtain these sequences?

☐**Sanger dideoxy sequencing**

☐**454**

☐**Helicos**

☐**Illumina**

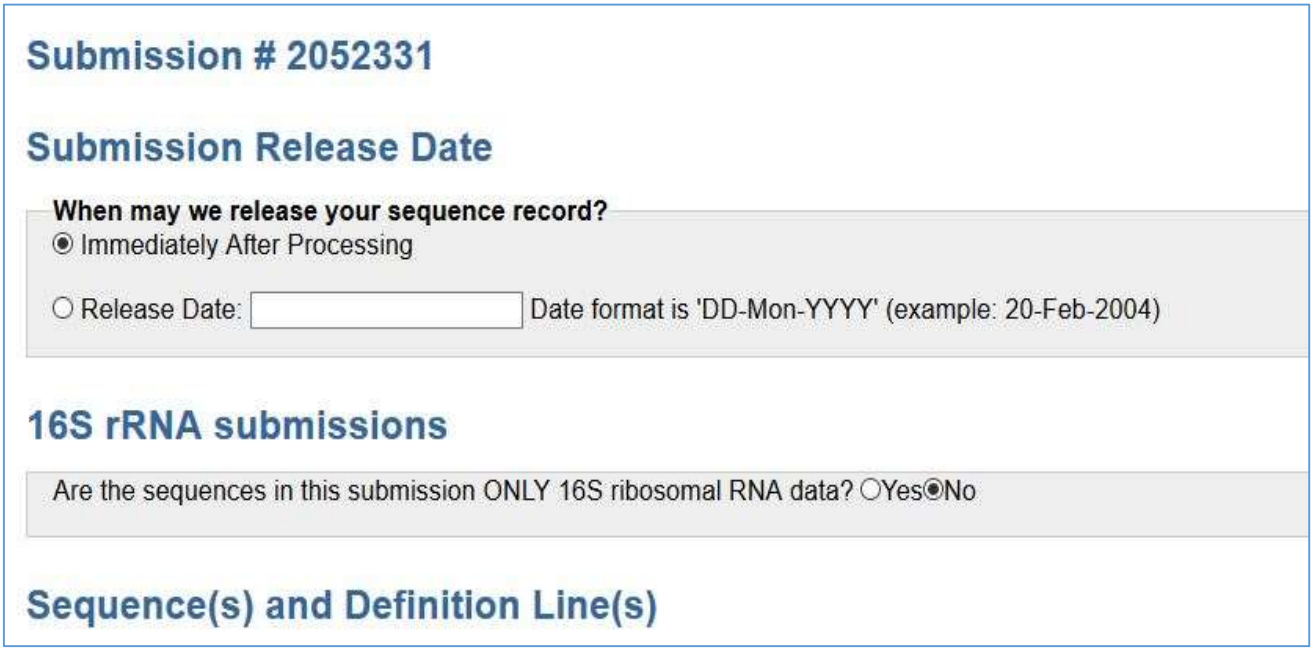
☐**IonTorrent**

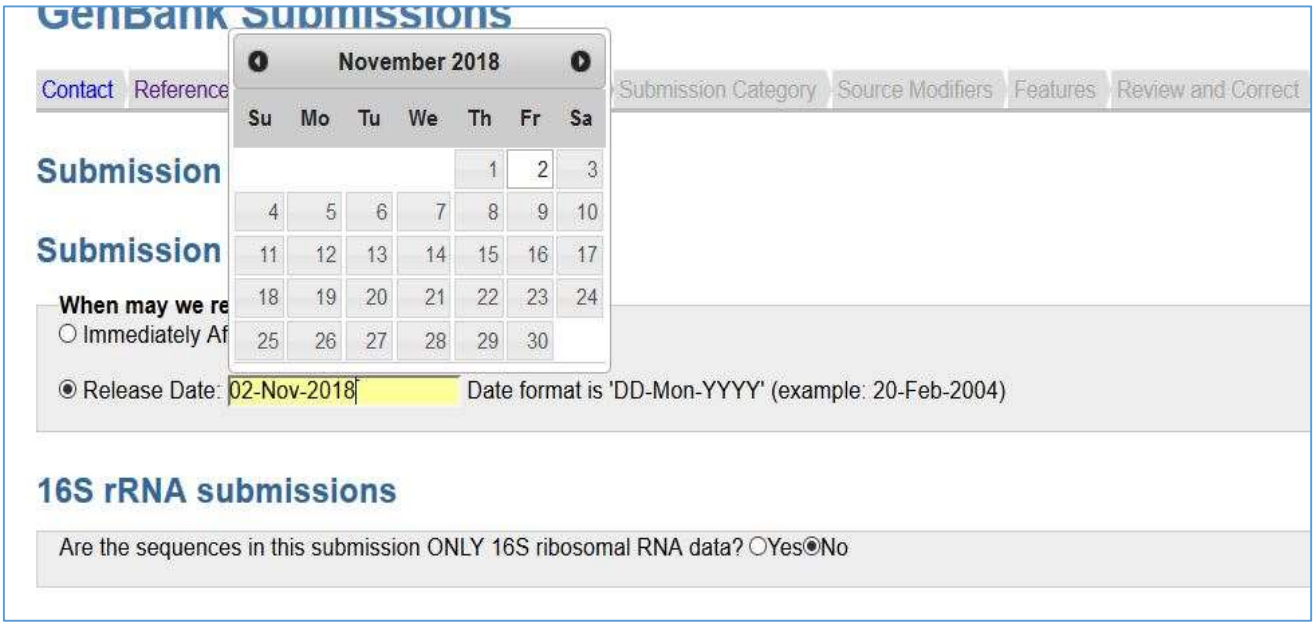
☐**Pacific Biosciences**

☐**SOLiD**

☐**Other**

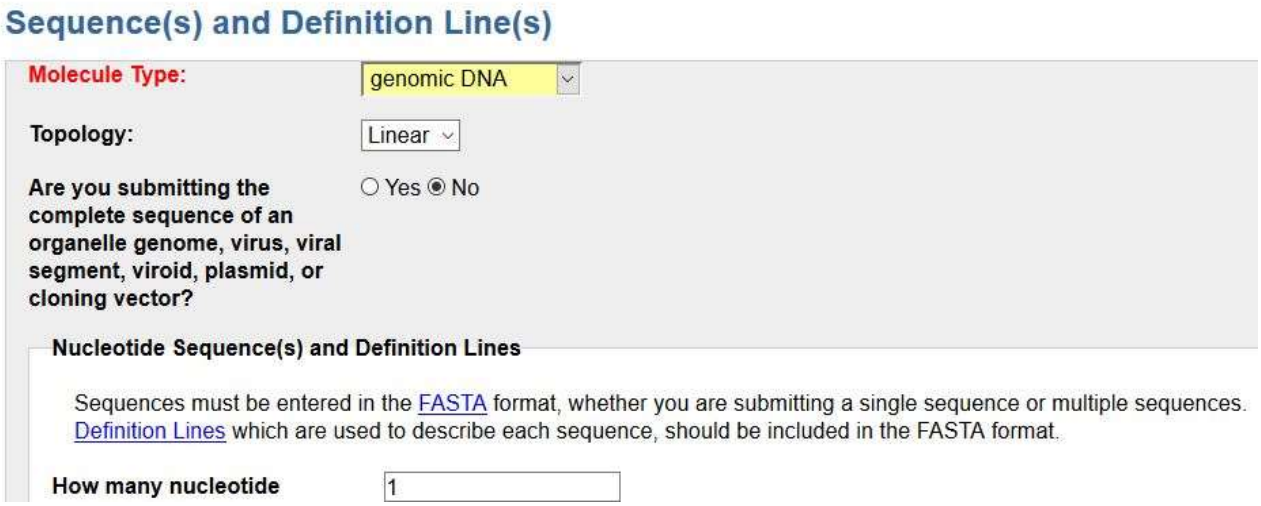
10. Submission release data you have two choice; immediately after process and release data (the balk to select the data). The first choice, your sequence directly become available for all researchers after processed by NCBI staff. Second, you must select the data in future.



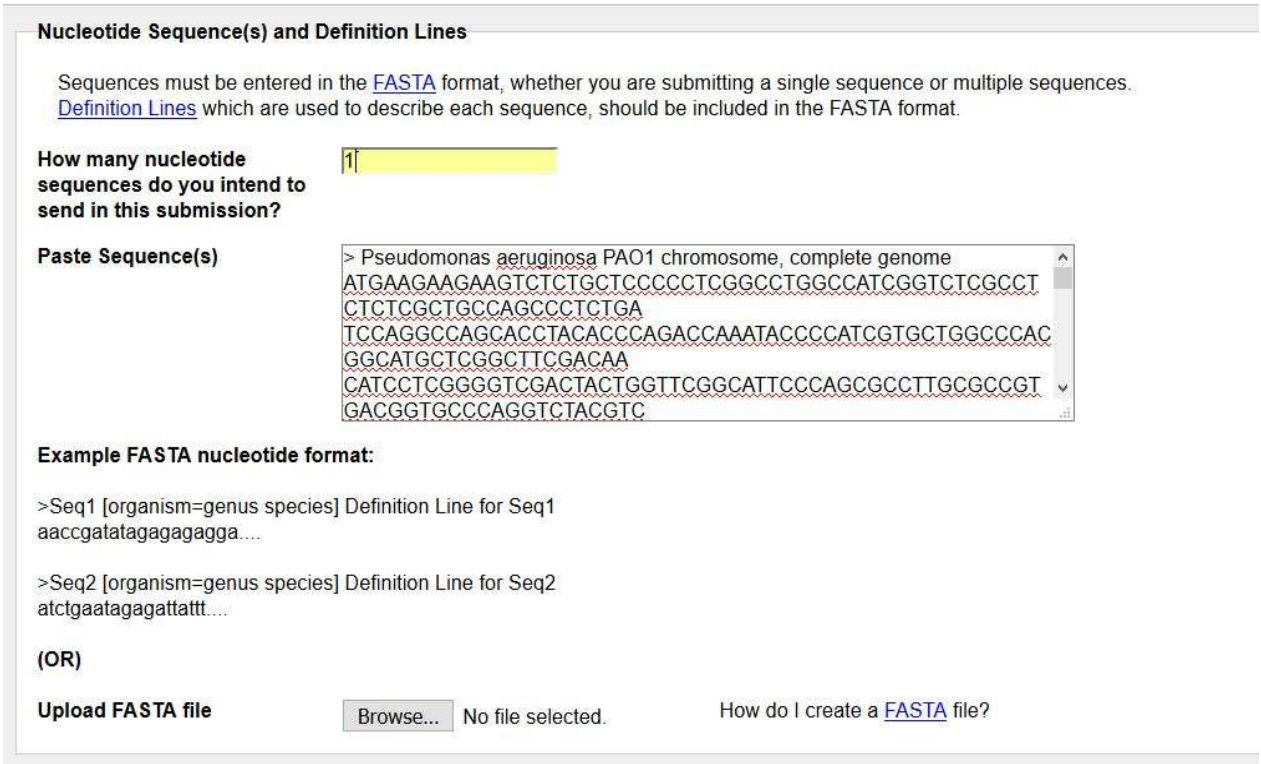


11. Select type of sequence from <molecular type> menu. Select topology if your sample liner or circular from <topology> menu. Write how many sequence

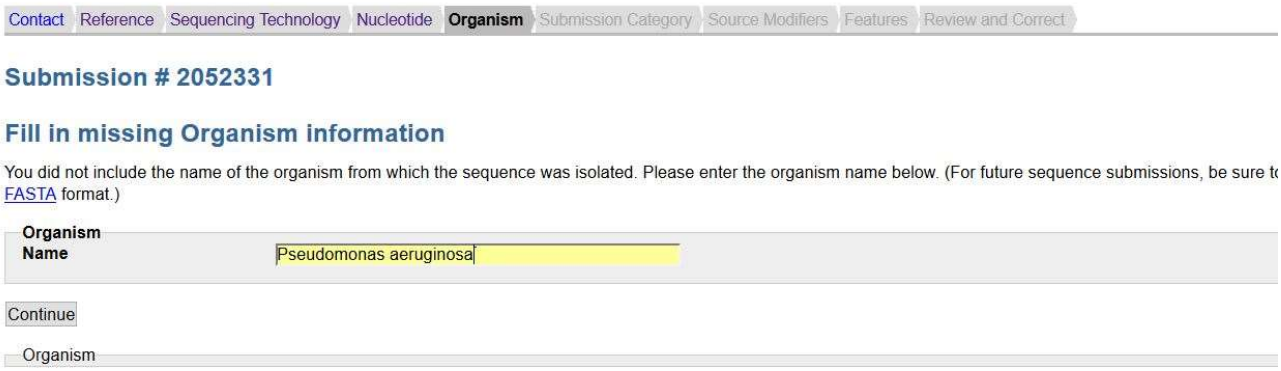
you will submitted it.



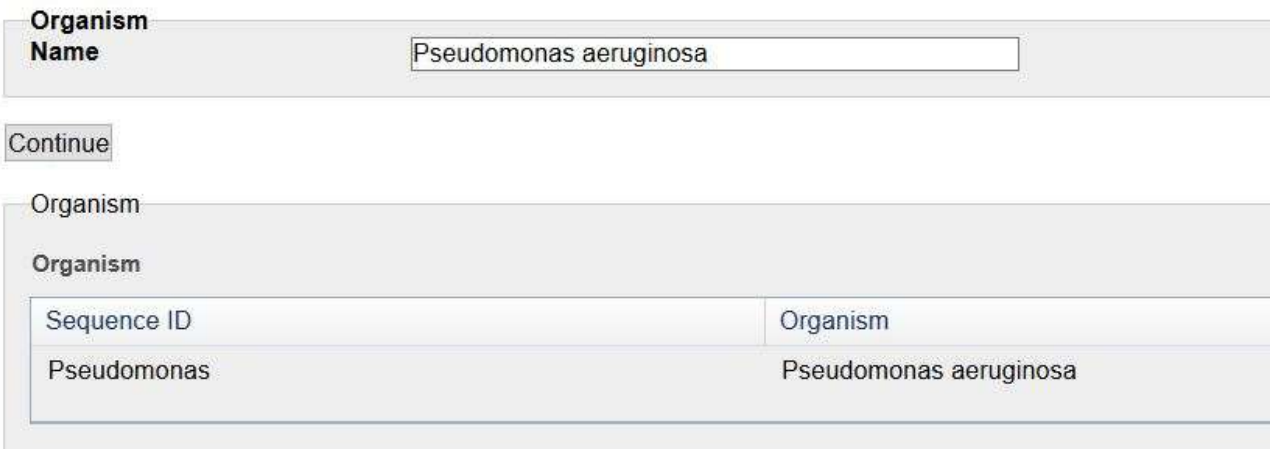
12. Copy your sequence from Fasta form and past it in the following blank, which explained in the next image. And click continue.



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

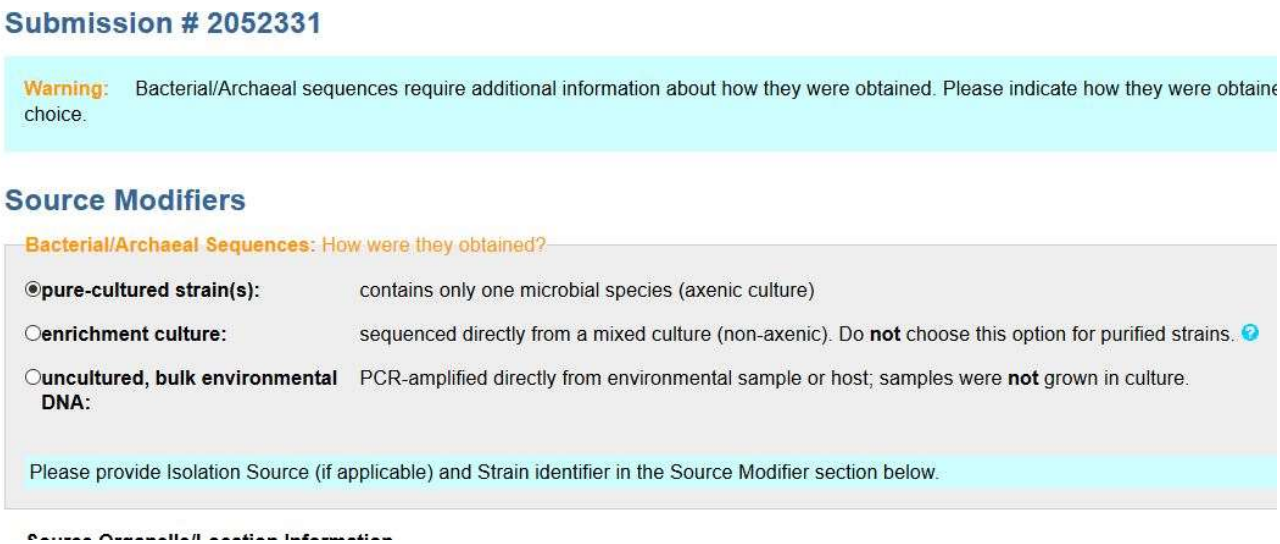


14. You will see the following screen. Press continue.

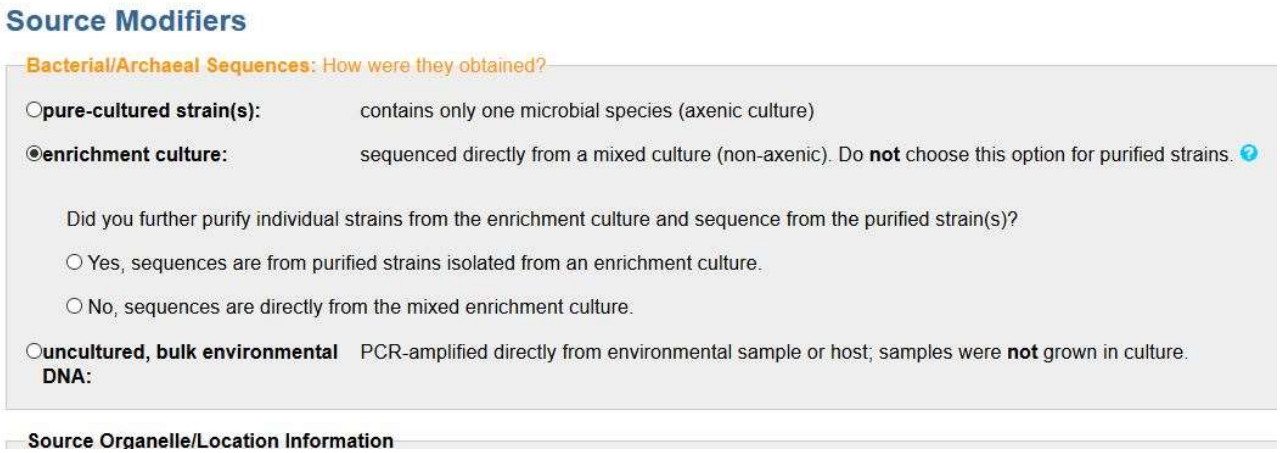


15.Click original and press continue.

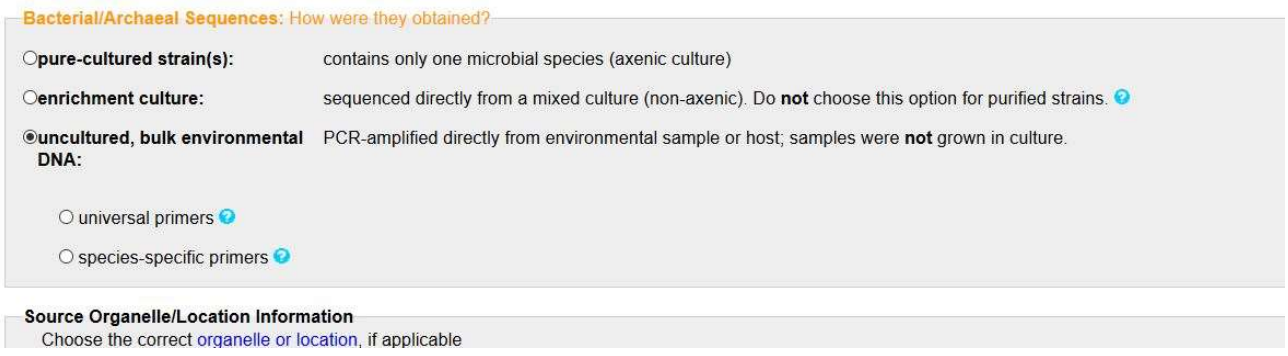
16. Source modifiers, you have three choices; first pure culture.



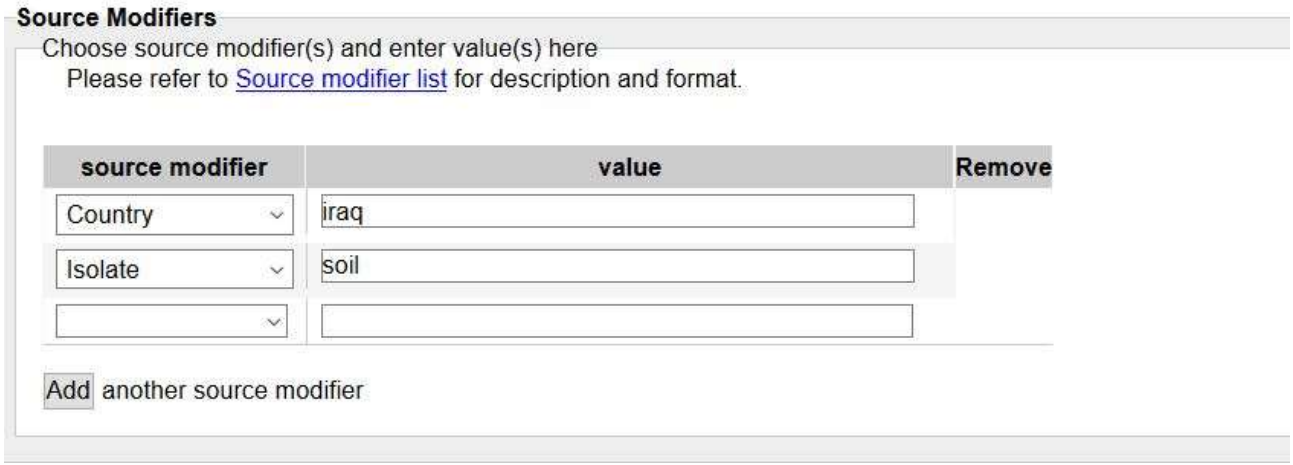
Second enrichment culture.

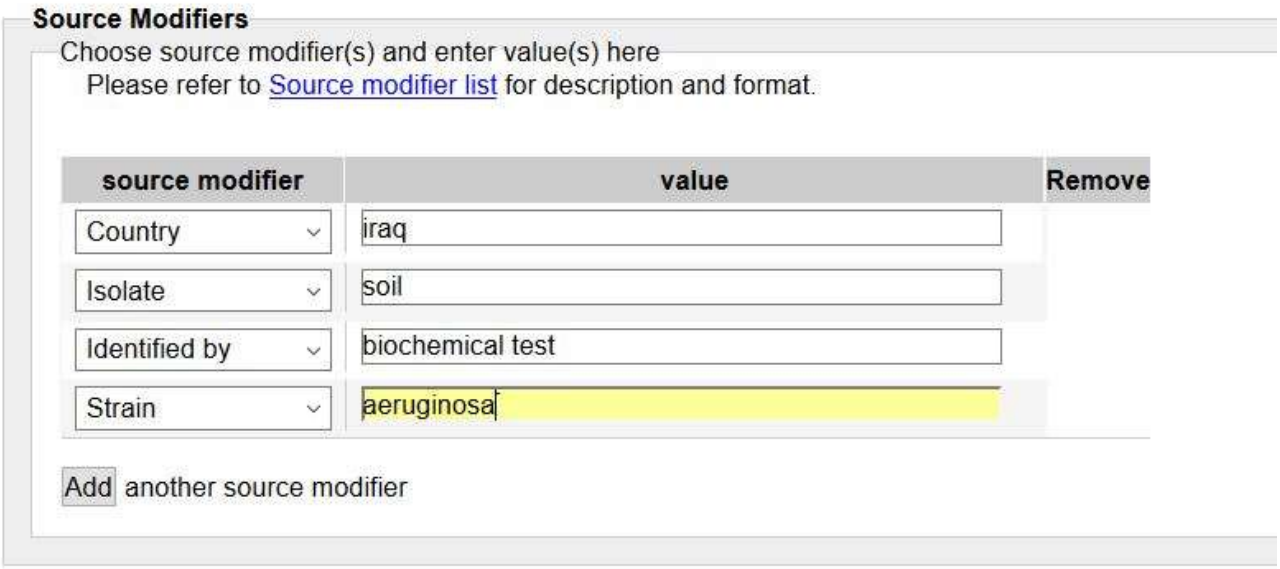


Third, un cultured, bulk environmental DNA.

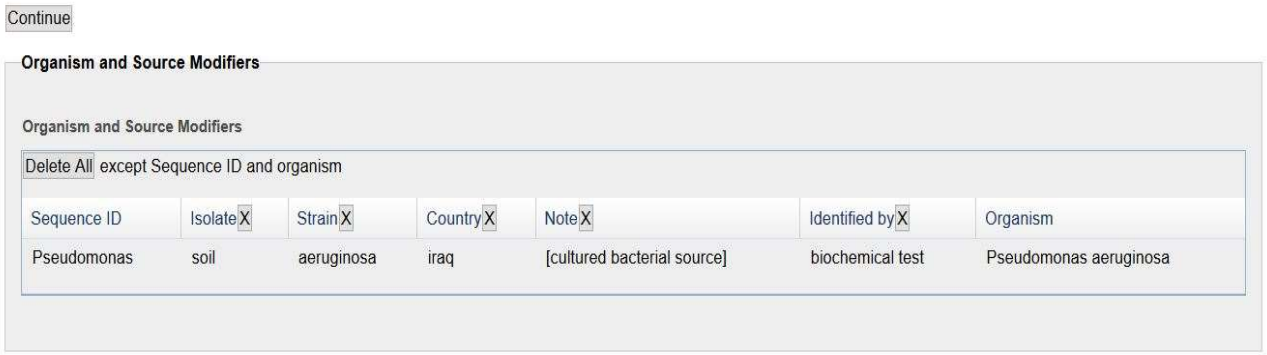


17. From <source modifier> menu select the approach and write your comment in <value> blank. More source modifier information make you sequence more clear and understood.

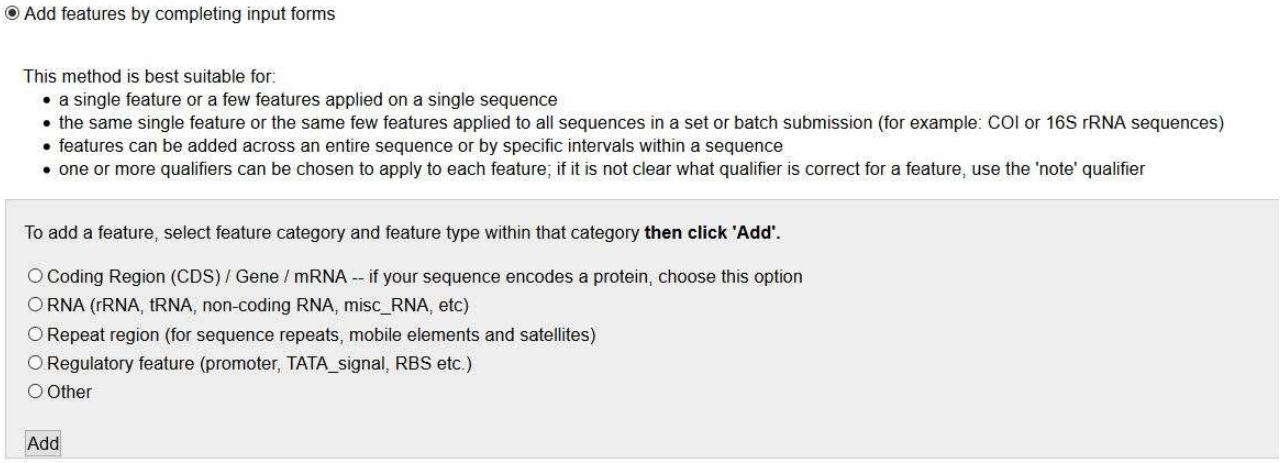




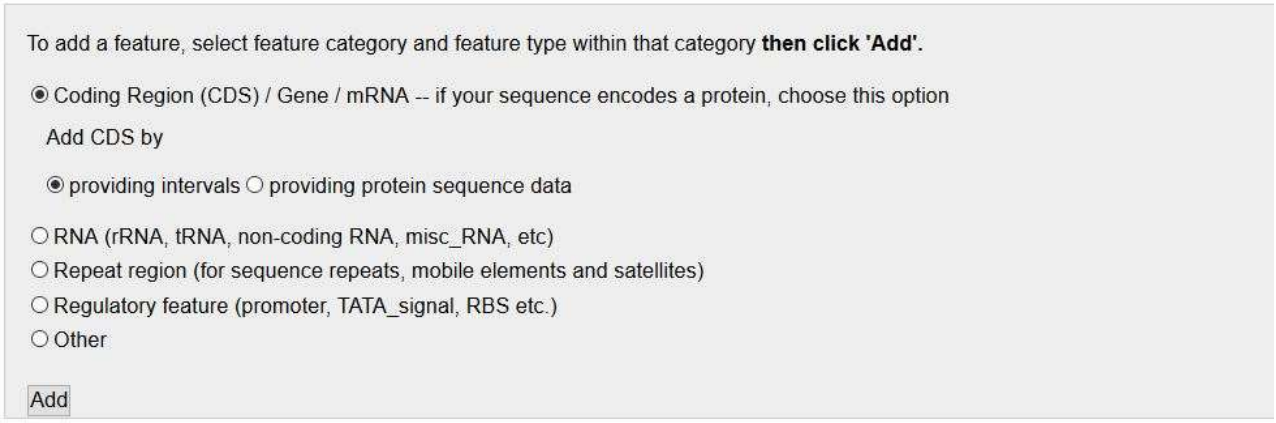
18. Click continue to see the following screen, click <continue> to continue.



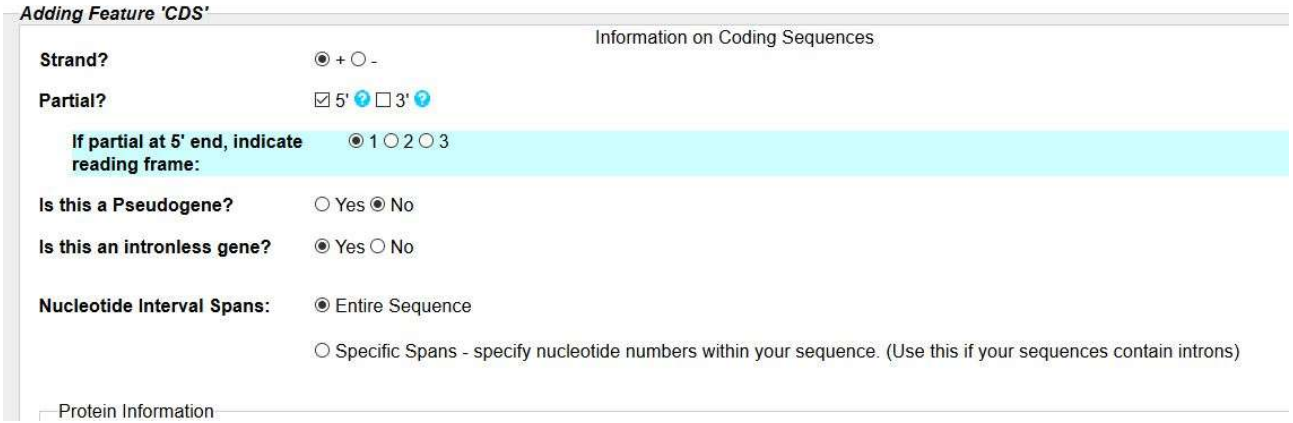
19. Now add feature to your sample (more information).



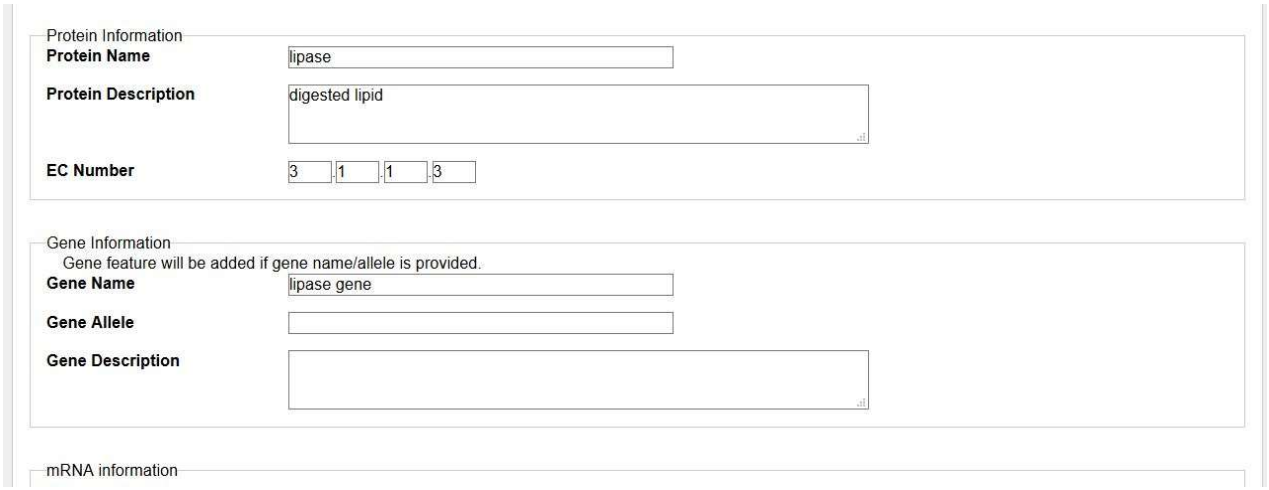
20. If your sequence coded to protein, write the information of protein look to the following screen.



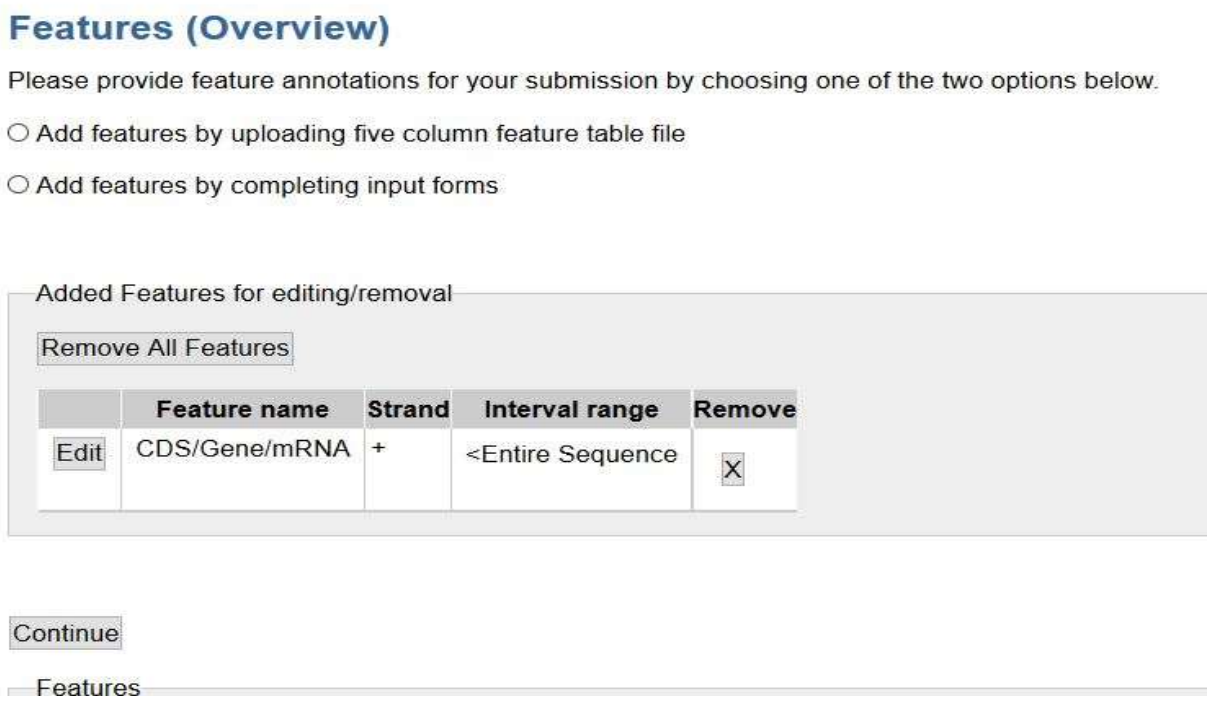
21. Now select suitable choice.



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.



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

Continue

Features

```
gene            <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
                NLIGHSHGGPTIRYVAAVRPDLIASATSVGAPHKGSDTADFLRQIPPGSAGEAVLSGL
                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

1. 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. Resubmission?
   If you were asked by GenBank staff to resubmit your sequence data, check here: ☐

3. Submission Title (Optional)
   If you want to title your submission for your own record-keeping, check here: ☐

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: ☐

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

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

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).