

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.

The screenshot shows the NCBI (National Center for Biotechnology Information) homepage. The top navigation bar includes links for 'Resources' and 'How To'. The main content area features a 'Welcome to NCBI' message and a grid of six primary actions: 'Submit', 'Download', 'Learn', 'Develop', 'Analyze', and 'Research'. The 'Submit' button, which includes the text 'Deposit data or manuscripts into NCBI databases' and an upward-pointing arrow icon, is highlighted with a red rectangular box. To the left of the main content is a 'Resource List (A-Z)' sidebar with categories like 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', and 'Training & Tutorials'. To the right, there are sections for 'Popular Resources' (listing PubMed, Bookshelf, PubMed Central, PubMed Health, BLAST, Nucleotide, Genome, SNP, Gene, Protein, and PubChem) and 'NCBI News & Blog' (mentioning PubMed Labs and a date of 03 Oct 2017).

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>



The screenshot shows the GenBank submission interface. The 'QuickStart' section is highlighted with a red box. It contains two dropdown menus: the first is set to 'GenBank' and the second is set to 'Simple DNA or RNA sequences'. A blue 'GO' button is located to the right of the second dropdown. The 'Submission Wizard' section is also visible, with the text 'Need help figuring out where to start? Try this!' and an icon of a pencil and stars.

3. Now choose <BankIt>

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

There are several options for submitting data to GenBank:

- [BankIt](#), 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 Sequin. It is used primarily for submission of complete genomes and large batches of sequences and is available by FTP for use on MAC, PC and Unix platforms.
- [Submission Portal](#), 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 BankIt to complete the process.

5. Now press on <new submission>.

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

First Name	<input type="text" value="Mustafa"/>
Last Name	<input type="text" value="Chazaay"/>
Department	<input type="text" value="genetic"/>
Institution	<input type="text" value="biotech"/>
Street Address	<input type="text" value="1010"/>
City	<input type="text" value="najaf"/>
State/Province	<input type="text"/>
ZIP/Postal Code	<input type="text" value="20820"/>
Country	<input type="text" value="iraq"/>
Phone	<input type="text" value="7812940687"/> <input type="text"/>
Fax	<input type="text"/> Example: 001-202-000-0000 (International), 202-000-0000 (U.S.A)
Email	<input type="text" value="mustafagazaay@gmail.com"/>
Please provide an alternate email address to ensure that messages are received.	
Alternate Email	<input type="text"/>
User profile update	<input checked="" type="checkbox"/> Retain changes to Contact information for all future BankIt submissions. (Uncheck if changes apply *only* to this su
<input type="button" value="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.

[Contact](#)
[Reference](#)
[Sequencing Technology](#)
[Nucleotide](#)
[Submission Category](#)
[Source Modifiers](#)
[Features](#)
[Review and Correct](#)

Submission # 2052331

Sequence Authors

First Name	Middle Initial(s)	Last Name	Suffix	Remove
musta	C	sarhan		X

[Add](#) more sequence authors.

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.

Reference Information #1

Please provide the title and relevant publication details (volume, issue, etc.) of a paper that discusses this submission.

PUBLICATION STATUS

☒ Unpublished
 ☐ In-Press
 ☐ Published

Reference Title

REFERENCE AUTHORS

☒ Same As Sequence Authors
☐ Specify New Authors

[Add Another Reference](#)

[Continue](#)

PUBLICATION STATUS

☐ Unpublished
 ☒ In-Press
 ☐ Published

Reference Title

Journal Title

Year
 Volume
 Issue
 Pages from to

REFERENCE AUTHORS

☒ Same As Sequence Authors
☐ Specify New Authors

[Add Another Reference](#)

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

☐ Ion Torrent

☐ 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.

Submission # 2052331

Submission Release Date

When may we release your sequence record?

☒ Immediately After Processing

☐ Release Date: Date format is 'DD-Mon-YYYY' (example: 20-Feb-2004)

16S rRNA submissions

Are the sequences in this submission ONLY 16S ribosomal RNA data? ☐ Yes ☒ No

Sequence(s) and Definition Line(s)

GenBank Submissions

[Contact](#) [Reference](#)

Submission

Submission

When may we re

☐ Immediately Af

☒ Release Date: 02-Nov-2018 Date format is 'DD-Mon-YYYY' (example: 20-Feb-2004)

16S rRNA submissions

Are the sequences in this submission ONLY 16S ribosomal RNA data? ☐ Yes ☒ No

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)

Molecule Type: genomic DNA

Topology: Linear

Are you submitting the complete sequence of an organelle genome, virus, viral segment, viroid, plasmid, or cloning vector? ☐ Yes ☒ No

Nucleotide Sequence(s) and Definition Lines

Sequences must be entered in the [FASTA](#) format, whether you are submitting a single sequence or multiple sequences. [Definition Lines](#) which are used to describe each sequence, should be included in the FASTA format.

How many nucleotide 1

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

Nucleotide Sequence(s) and Definition Lines

Sequences must be entered in the [FASTA](#) format, whether you are submitting a single sequence or multiple sequences. [Definition Lines](#) which are used to describe each sequence, should be included in the FASTA format.

How many nucleotide sequences do you intend to send in this submission? 1

Paste Sequence(s)

```
> Pseudomonas aeruginosa PAO1 chromosome, complete genome
ATGAAGAAGAAGTCTCTGCTCCCCCTCGGCCTGGCCATCGGTCTCGCCT
CTCTCGCTGCCAGCCCTCTGA
TCCAGGCCAGCACCTACACCCAGACCAAATACCCCATCGTGCTGGCCCAC
GGCATGCTCGGCTTCGACAA
CATCCTCGGGGTCGACTACTGGTTCGGCATTCCCAGCGCCTTGCGCCGT
GACGGTGCCCAGGTCTACGTC
```

Example FASTA nucleotide format:

```
>Seq1 [organism=genus species] Definition Line for Seq1
aaccgatatagagagga....

>Seq2 [organism=genus species] Definition Line for Seq2
atctgaatagattatt....
```

(OR)

Upload FASTA file Browse... No file selected. [How do I create a FASTA file?](#)

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

[Contact](#) [Reference](#) [Sequencing Technology](#) [Nucleotide](#) **Organism** [Submission Category](#) [Source Modifiers](#) [Features](#) [Review and Correct](#)

Submission # 2052331

Fill in missing Organism information

You did not include the name of the organism from which the sequence was isolated. Please enter the organism name below. (For future sequence submissions, be sure to use [FASTA](#) format.)

Organism Name

Continue

Organism

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

Organism Name

Continue

Organism

Sequence ID	Organism
Pseudomonas	Pseudomonas aeruginosa

15. Click original and press continue.

Submission # 2052331

Submission Category

Indicate whether your sequence is an original submission or a [third-party annotation](#) submission.

☒ **Original** Directly sequenced by submitter.

☐ **Third Party Annotation** Derived from other primary sequence data.

Continue

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 by making a choice.

Source Modifiers

Bacterial/Archaeal Sequences: How were they obtained?

- ☒ **pure-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. ?
- ☐ **uncultured, 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.

Source Organelle/Location Information

Second enrichment culture.

Source Modifiers

Bacterial/Archaeal Sequences: How were they obtained?

- ☐ **pure-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 individual strains from the enrichment culture and sequence from the purified strain(s)?
- ☐ Yes, sequences are from purified strains isolated from an enrichment culture.
- ☐ No, sequences are directly from the mixed enrichment culture.
- ☐ **uncultured, bulk environmental 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.

Bacterial/Archaeal Sequences: How were they obtained?

- ☐ **pure-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. ?
- ☒ **uncultured, bulk environmental DNA:** PCR-amplified directly from environmental sample or host; samples were **not** grown in culture.
- ☐ universal primers ?
- ☐ species-specific primers ?

Source Organelle/Location Information

Choose the correct [organelle](#) or [location](#), if applicable

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.

Source Modifiers

Choose source modifier(s) and enter value(s) here

Please refer to [Source modifier list](#) for description and format.

source modifier	value	Remove
Country ▾	iraq	
Isolate ▾	soil	
▾		

 another source modifier**Source Modifiers**

Choose source modifier(s) and enter value(s) here

Please refer to [Source modifier list](#) for description and format.

source modifier	value	Remove
Country ▾	iraq	
Isolate ▾	soil	
Identified by ▾	biochemical test	
Strain ▾	aeruginosa	

 another source modifier

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

Organism and Source Modifiers

Organism and Source Modifiers

Delete All except Sequence ID and organism

Sequence ID	Isolate <input checked="" type="checkbox"/>	Strain <input checked="" type="checkbox"/>	Country <input checked="" type="checkbox"/>	Note <input checked="" type="checkbox"/>	Identified by <input checked="" type="checkbox"/>	Organism
Pseudomonas	soil	aeruginosa	iraq	[cultured bacterial source]	biochemical test	Pseudomonas aeruginosa

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)
- 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)
☐ Repeat region (for sequence repeats, mobile elements and satellites)
☐ Regulatory feature (promoter, TATA_signal, RBS etc.)
☐ Other

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'**.

- Coding Region (CDS) / Gene / mRNA -- if your sequence encodes a protein, choose this option

Add CDS by

- providing intervals ☐ providing protein sequence data

- ☐ RNA (rRNA, tRNA, non-coding RNA, misc_RNA, etc)
☐ Repeat region (for sequence repeats, mobile elements and satellites)
☐ Regulatory feature (promoter, TATA_signal, RBS etc.)
☐ Other

Add

21. Now select suitable choice.

Adding Feature 'CDS'

Information on Coding Sequences

Strand? ☒ + ☐ -

Partial? ☒ 5' ☐ 3'

If partial at 5' end, indicate reading frame: ☒ 1 ☐ 2 ☐ 3

Is this a Pseudogene? ☐ Yes ☒ No

Is this an intronless gene? ☒ Yes ☐ No

Nucleotide Interval Spans: ☒ Entire Sequence

☐ Specific Spans - specify nucleotide numbers within your sequence. (Use this if your sequences contain introns)

Protein Information

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

Protein Information

Protein Name

Protein Description

EC Number

Gene Information

Gene feature will be added if gene name/allele is provided.

Gene Name

Gene Allele

Gene Description

mRNA information

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.

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

Added Features for editing/removal

	Feature name	Strand	Interval range	Remove
<input type="button" value="Edit"/>	CDS/Gene/mRNA	+	<Entire Sequence	<input type="button" value="X"/>

Features

24. When press continue you will see GenBank format 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='MKKKSLLPLGLAIGLASLAASPLIQASTYTTQTKYPIVLAHGMLG
FDNILGVDYWFGIPSA LR RDGAQVYVTEVSQ LDTSEVRGEQLLQQVVEEIVALSGQPKV
NLIGHSHGGPTIRYVAAVRPD LIASATSVGAP HKGSDTADFLRQIPPGSAGEAVLSGL
VNSLGALISFLSSGSTGTQNSIGSLES LNSEGAARFNAKYPQGIPTSACGEGAYKVNG
VSYYSWSGSSPLTNFLDPSDAFLGASSLTFKNGTANDGLVGTCS SHLGMVIRDNYRMN
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:

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 PA01 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).