


What do you want to submit?

Enter a few words about your sequence data.

BankIt

This submission option is for genomic DNA (e.g. protein-coding gene, regulatory element), transcripts (e.g. mRNA, ncRNA), or small genomes (organelle, plasmid, and phage and other viral) from any organism. Artificial sequences (cloning/expression vector) as well as annotated or assembled third party sequences can also be submitted here.

Prepare the following information:

1. General: your contact details, authors, publication, data release date
2. Submission type:
 - Original or third-party assembly/annotation
 - Set designation (if applicable) for multiple sequences of the same locus
 - Molecule type
3. Nucleotide sequences in FASTA or alignment format
4. Organism name(s)
5. Source metadata, such as: isolate, strain, collection date, country
6. Feature annotation, such as CDS (coding region), tRNA, ncRNA, gene

You may submit sequences in one of two formats:

- 1) FASTA, which is acceptable for one or more sequences. Please use the FASTA format that starts with a definition line, followed with a hard return and the sequence. The simplest definition line requires the ">" symbol and a sequence_ID.

Example:

```
>Seq1 [organism=Mus emesi]
CCTTTAT...
>Seq2 [organism=Mus bufo]
GGTAGGT...
```

- 2) The alignment format, which is acceptable for multiple sequences from the same locus or same genomic region. Accepted alignment formats include FASTA+GAP, Nexus, Phylip, and Clustal(w).

All sequence files must be in plain text using ASCII characters only. Use IUPAC codes for your sequences.

Source modifiers will be requested as part of submission and use a controlled vocabulary to describe how, when, and where you obtained your samples. You can also uniquely identify your samples from the same organism with source modifier such as isolate, clone, strain or specimen voucher.

You will be asked to provide values for certain source modifiers based on your organism information. Additional modifiers will be available to add.

Source modifiers can be provided through the web form or through a tab delimited table.

Prepare to annotate features on your sequence(s):

- For simple annotation (e.g. same feature for all sequences), follow the web form's instructions. For complex annotation, prepare a tab-delimited, five-column feature table to upload. Provide feature intervals based on the sequence(s) you are submitting. For protein-coding sequences, annotate the coding regions (CDS) on your sequence(s), whether they are partial or complete.
 - If you submitted an alignment, you will have an option to 'Propagate features' from a single sequence (longest sequence recommended) to the other sequences in your submission. You will have the option to manually edit or remove features after propagation.
 - Not providing complete feature annotation will delay accession number assignment and processing.
-

Login to NCBI (can use ORCID but must link the accounts first)

Click on one of the buttons below to start a new submission

[Start SARS-CoV-2 Submission](#)

[Start rRNA or rRNA/ITS Submission](#)

[Start Metazoan COX1 Submission](#)

[Start Influenza, Norovirus, Dengue submission](#)

[Start BankIt Submission](#)

Contact Information

First Name	<input type="text" value="Carla"/>
Last Name	<input type="text" value="Cicero"/>
Department	<input type="text" value="Museum of Vertebrate Zoology"/>
Institution	<input type="text" value="University of California, Berkeley"/>
Street Address	<input type="text" value="3101 Valley Life Sciences Building"/>
City	<input type="text" value="Berkeley"/>
State/Province	<input type="text" value="CA"/>
ZIP/Postal Code	<input type="text" value="94720-3160"/>
Country	<input type="text" value="United States"/>
Phone	<input type="text" value="510-642-7868"/>
Fax	<input type="text" value="510-643-8238"/>
Email	<input type="text" value="ccicero@berkeley.edu"/>

Please provide an alternative email address to ensure that messages are received.

Alternative Email	<input type="text" value="carlacic@gmail.com"/>
-------------------	---

User profile update Retain changes to Contact information for all future BankIt submissions. (Uncheck if changes apply *only* to this submission).

Sequence Authors

First Name	Middle Initial(s)	Last Name	Suffix	Remove
<input type="text" value="Carla"/>	<input type="text"/>	<input type="text" value="Cicero"/>	<input type="text" value="v"/>	<input type="button" value="X"/>

more sequence authors.

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

Sequencing Technology

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

What methods were used to obtain these sequences?

Sanger dideoxy sequencing

454

Helicos

Illumina

IonTorrent

PacBio

SOLiD

Other

Are these sequence(s):

unassembled sequence reads

assembled sequences (consisting of two or more sequence reads)

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).

Release date must be after 6 months from today and before 10 years from today.

Sequence(s) and Definition Line(s)

Molecule Type:

Select the type of molecule that was isolated and sequenced.

Topology:

Are you submitting the complete sequence of an organelle genome, virus, viral segment, viroid, plasmid, or cloning vector?

Yes No

Nucleotide Sequence Format

Sequence data format:

FASTA sequences (not an alignment, most common data format, [FASTA help](#))

For example:

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

```
aaccgatatagagagagga....
```

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

```
atctgaatagattattt....
```

[Definition Lines](#) which are used to describe each sequence, should be included with your sequence data.

Alignment (FASTA+GAP, NEXUS, PHYLIP, CLUSTAL(W), [alignment help](#))

Nucleotide Sequence(s)

Paste sequence data

(OR)

Upload file

Choose File

No file chosen

File used for submission:

[Sequences_MVZonly.fasta+\(trimmed\)](#)

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 the [FASTA](#) format.)

Organism

Name

Input same organism name for all sequences.

Upload File

Choose File

No file chosen

How do I create an [Organism](#) file?

Organism

Sequence ID	Organism
MX_Chiapas_C1056_MVZ188144	Cyanocitta stelleri
GUA_Huehuetenango_C1057_MVZ189266	Cyanocitta stelleri
GUA_Huehuetenango_C1058_MVZ189267	Cyanocitta stelleri
1_Canada_VancouverId_C25_MVZ177865	Cyanocitta stelleri
1_Canada_VancouverId_C26_MVZ177866	Cyanocitta stelleri
1_Canada_VancouverId_C733_MVZ177867	Cyanocitta stelleri
1_Canada_VancouverId_C734_MVZ177868	Cyanocitta stelleri
1_Canada_VancouverId_C735_MVZ177869	Cyanocitta stelleri
1_Canada_VancouverId_C736_MVZ177870	Cyanocitta stelleri
1_Canada_VancouverId_C737_MVZ177871	Cyanocitta stelleri
2_WA_Clallam_C98_MVZ181877	Cyanocitta stelleri
2_WA_Clallam_C99_MVZ181878	Cyanocitta stelleri
2_WA_Clallam_C104_MVZ181879	Cyanocitta stelleri
2_WA_Clallam_C678_MVZ181880	Cyanocitta stelleri
2_WA_Clallam_C679_MVZ181881	Cyanocitta stelleri
2_WA_Clallam_C680_MVZ181882	Cyanocitta stelleri
2_WA_Clallam_C681_MVZ181883	Cyanocitta stelleri

Submission Set/Batch

All sequences in a set must be from the same gene/locus and are expected to be released at the same time.

Please select a set type:

- Pop set **Population study:** a set of sequences that were derived by sequencing the same gene from different isolates of the same organism.
- Phy set **Phylogenetic study:** a set of sequences that were derived by sequencing the same gene from different organisms.
- Mut set **Mutation study:** a set of sequences that were derived by sequencing multiple mutations of a single gene.
- Env set **Environmental study:** a set of sequences that were derived by sequencing the same gene from a population of unclassified or unknown organisms.

If your sequences are NOT all from the same gene/locus and NOT intended to be released at the same time, then choose 'Batch' below.

- Batch Multiple, related nucleotide sequences that are not from the same gene, but may be from the same study or organism.

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.

Source Modifiers

Source Organelle/Location Information

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

Organelle/Location

mitochondrion

Source Modifiers

- For multiple sequences: use only a table file that contains all the source modifiers you want to add.
- Do **not** add source modifiers using both a table file and the source modifier menus below.
- A reloaded source modifier table file will overwrite a previously loaded table file.

Set multiple values for sequences

- Upload source modifiers Table File
How do I create a [source modifiers table](#) file?

Source Modifier File

SourceFile_MVZonly3.txt

Choose File SourceFile_...nly_NCBI.txt

Any other source related information? [Add](#)

For example, Taxonomic Classification or Lineage for unknown/unrecognized organisms. You should create a [Source Modifier Table](#) for notes or other information that is unique for individual sequences; use 'note' as the header label. **Do not** enter individual source modifiers as plain text here. Choose them from the drop down list in the Source Modifier section above.

If you want to include PCR primers in your submission, check here:

Features (Overview)

Provide feature annotations for your submission using one of the options below: [INSDC feature definitions](#)

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

This method is more suitable for:

- adding many different features on a single sequence or on multiple sequences
- uses the five-column, tab-delimited [feature table](#) format, which is also used in Sequin
- each table in the feature table file applies to only one sequence; if multiple sequences have been uploaded in your nucleotide fasta file, each corresponding table must be labeled with that sequence's Sequence ID
- multiple tables can be uploaded in a single file.

Uploading multiple feature table files will result in the last file replacing all previously uploaded files. Be sure to include all features in a single feature table file.

Upload Features Table File

1. No file chosen [How do I create a features table file?](#)
2.

Features (Overview)

Provide feature annotations for your submission using one of the options below: [INSDC feature definitions](#)

- Add features by uploading five column feature table file
- Add features by completing input forms
 - This method is suitable for annotating a single feature or a few features
 - Features may be annotated across an entire sequence or using specific intervals within a sequence

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

Features (Detail)

Adding Feature 'CDS'

Information on Coding Sequences

Strand? + -

Partial? 5' 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)

Sequences that have this feature: All Specific

Available Sequences		Sequences for this feature
51_ID_BearLake_C964_MVZ186068	>	26_CA_Plumas_C119_MVZ179196
10_CA_ContraCosta_C745_MVZ177791		26_CA_Plumas_C2_MVZ179195
26_CA_Plumas_C1_MVZ179194	<	25_CA_Lassen_C762_MVZ182217
		25_CA_Lassen_C761_MVZ182216
		25_CA_Lassen_C760_MVZ182215

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