


# How to submit sequences to GenBank

using the BankIt website (other methods are available)

# What is GenBank?

 **National Library of Medicine**  
National Center for Biotechnology Information

rowandurrant@orcid

Nucleotide

Nucleotide rabies

Create alert Advanced

Help

Species

Animals (12,889)

Plants (3)

Fungi (13)

Protists (8)

Bacteria (662)

Viruses (29,170)

Customize ...

Molecule types

genomic DNA/RNA (34,392)

mRNA (11,253)

Customize ...

Source databases

INSDC (GenBank) (45,704)

RefSeq (639)

Customize ...

Sequence Type

Nucleotide (46,370)

EST (1)

Genetic compartments

Mitochondrion (3,580)

Plasmid (4)

Sequence length

Custom range...

Release date

Summary 20 per page Sort by Default order

Items: 1 to 20 of 46371

<< First < Prev Page 1 of 2319 Next > Last >>

☐ [Lyssavirus rabies isolate Ariana2.. partial genome](#)

1. 11,796 bp linear cRNA

Accession: MK981888.1 GI: 1693310585

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#)

☐ [Rabies lyssavirus isolate 18018LIB.. complete genome](#)

2. 11,923 bp linear cRNA

Accession: OK135148.1 GI: 2128399345

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

☐ [Rabies lyssavirus isolate 18009LIB.. complete genome](#)

3. 11,923 bp linear cRNA

Accession: OK135147.1 GI: 2128399339

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

☐ [Rabies lyssavirus isolate 18008LIB.. complete genome](#)

4. 11,885 bp linear cRNA

Accession: OK135146.1 GI: 2128399333

[Protein](#) [PubMed](#) [Taxonomy](#)

[GenBank](#) [FASTA](#) [Graphics](#) [PopSet](#)

Send to: Filters: [Manage Filters](#)

Results by taxon

Top Organisms [\[Tree\]](#)

[Lyssavirus rabies \(28293\)](#)

[Desmodus rotundus \(9227\)](#)

[synthetic construct \(1393\)](#)

[Vulpes lagopus \(680\)](#)

[Burkholderia pseudomallei \(598\)](#)

[All other taxa \(6180\)](#)

[More...](#)

Find related data

Database: [Select](#)


[Find items](#)

Search details

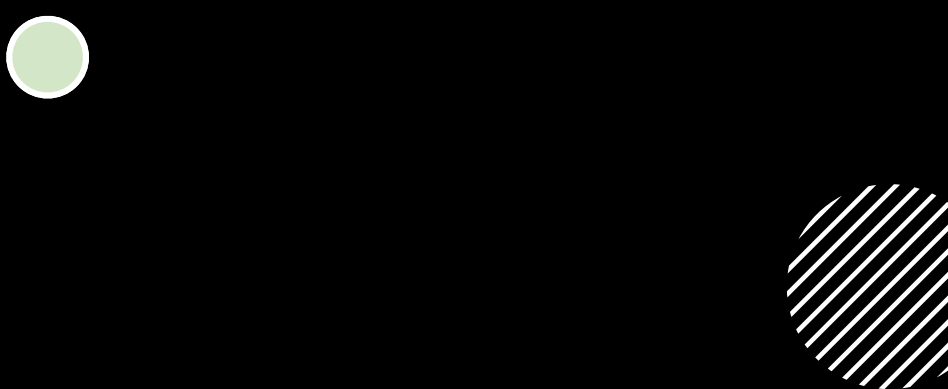
rabies[All Fields]

[Search](#)

[See more...](#)



# Why should I submit my sequences?



Get accession numbers




Most journals require it



Allows other researchers to use them in their work

# NCBI-compliant genome submissions: tips and tricks to save time and money

Walter Pirovano, Marten Boetzer, Martijn F L Derks, Sandra Smit 

*Briefings in Bioinformatics*, Volume 18, Issue 2, March 2017, Pages 179–182,  
<https://doi.org/10.1093/bib/bbv104>

**Published:** 10 December 2015    **Article history** ▼



PDF

Split View

Cite



Permissions



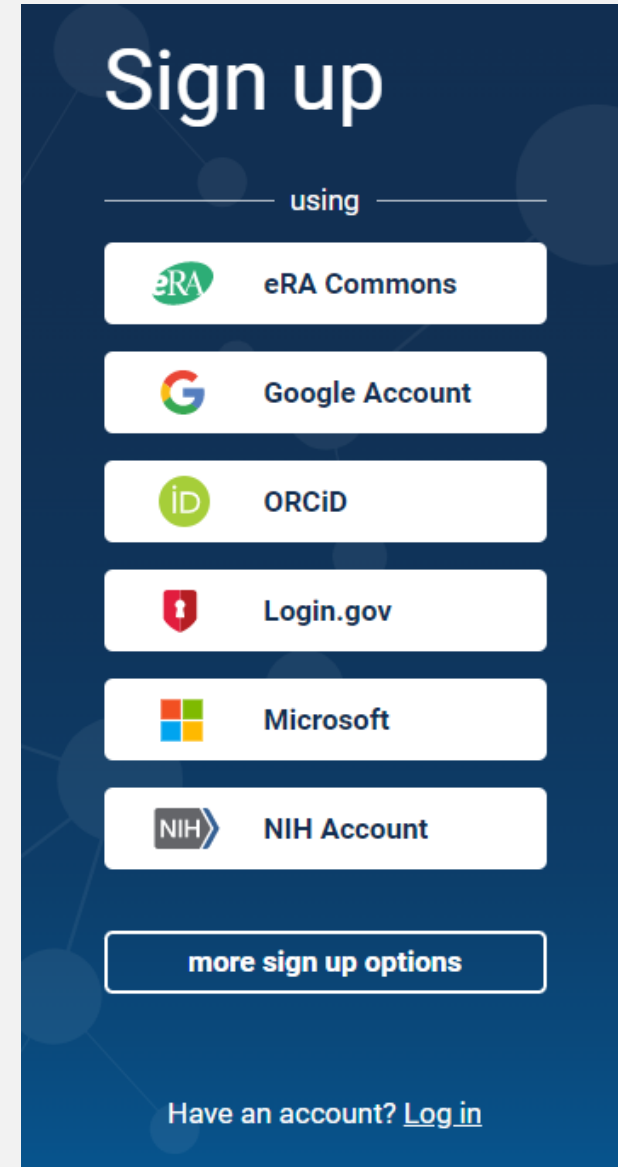
Share ▼

## Abstract

Genome sequences nowadays play a central role in molecular biology and bioinformatics. These sequences are shared with the scientific community through sequence databases. The sequence repositories of the International Nucleotide Sequence Database Collaboration (INSDC, comprising GenBank, ENA and DDBJ) are the largest in the world. Preparing an annotated sequence in such a way that it will be accepted by the database is challenging because many validation criteria apply. In our opinion, it is an undesirable situation that researchers who want to submit their sequence need either a lot of experience or help from partners to get the job done. To save valuable time and money, we list a number of recommendations for people who want to submit an annotated genome to a sequence database, as well as for tool developers, who could help to ease the process.

# How do I submit my sequences?

- Online is easiest:  
[www.ncbi.nlm.nih.gov/WebSub/](http://www.ncbi.nlm.nih.gov/WebSub/)
- You will need to make an account
- You can log in through ORCID, a Google account, Microsoft etc



# Step 1: What kind of sequence do you have?

## Submit new sequences to GenBank

### What type of sequence data do you have?

- ☐ SARS-CoV-2 ?
- ☐ Ribosomal RNA (rRNA) or rRNA-ITS
- ☐ Metazoan (multicellular animal) COX1
- ☐ Eukaryotic nuclear mRNA ? **NEW**
- ☐ Influenza virus
- ☐ Norovirus
- ☐ Dengue virus
- ☐ Eukaryotic and Prokaryotic Genomes (WGS or Complete)
- ☐ Transcriptome Shotgun Assembly (TSA)
- ☐ Unassembled sequence reads (SRA)
- ☒ Sequence data not listed above (through BankIt): genomic DNA, organelle, ncRNA, plasmids, other viruses, phages, other mRNA, synthetic constructs

#### Submission requirements:

- Sequence data in [FASTA](#) or [alignment](#) format
- Name(s) of the organism(s) where sequence data were isolated and any other [descriptive data](#)
- Sequence features (Examples: CDS, gene, tRNA, with nucleotide intervals and product names)

Start

## Step 2: Contact details

Contact Reference Sequencing Technology Nucleotide Organism Set/Batch Submission Category Source Modifiers Features Review and Correct

### Submission # 2626608

#### Contact Information

**First Name**

**Last Name**

**Department**

**Institution**

**Street Address**

**City**

**State/Province**

**ZIP/Postal Code**

**Country**

**Phone**

**Fax**

**Email**

Example: 001-202-000-0000 (International), 202-000-0000 (U.S.A)

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

**Alternative Email**

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

## Step 3: Authors & Reference

Submission # 2626608

### Sequence Authors

First Name	Middle Initial(s)	Last Name	Suffix	Remove
<input type="text" value="Gurdeep"/>	<input type="text"/>	<input type="text" value="Jaswant"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Kirstyn"/>	<input type="text"/>	<input type="text" value="Brunker"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Katie"/>	<input type="text"/>	<input type="text" value="Hampson"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="S"/>	<input type="text" value="M"/>	<input type="text" value="Thumbi"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Julius"/>	<input type="text" value="O"/>	<input type="text" value="Oyugi"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Charles"/>	<input type="text"/>	<input type="text" value="Kayuki"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Davis"/>	<input type="text"/>	<input type="text" value="Kuchaka"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Marco"/>	<input type="text"/>	<input type="text" value="van Zwetselaar"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Kathryn"/>	<input type="text"/>	<input type="text" value="Campbell"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Rowan"/>	<input type="text"/>	<input type="text" value="Durrant"/>	<input type="text" value="v"/>	<input type="button" value="X"/>
<input type="text" value="Marieke"/>	<input type="text"/>	<input type="text" value="Dekker"/>	<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



## Step 4: Sequencing Technology

### 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

Oxford Nanopore MinION

Are these sequence(s):

☐ unassembled sequence  
reads

☒ assembled sequences  
(consisting of two or more  
sequence reads)

What program(s) did you use to assemble these sequences?

Assembly Program	Version or Date	Remove
Guppy		X
Minimap2		X

additional assembly programs.

Assembly Name:

# Step 5: Nucleotide

## 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  
aaccgatagagagagga....  
>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

No file chosen


**Warning:** Terminal ends of the following sequence(s) are low quality (too many ambiguous bases) and have been trimmed. These are your options:


- Press Continue below to continue with the trimmed sequence(s)
- or, trim the sequence(s) yourself and upload the new corrected sequence(s)
- or, click [here](#) to undo **all** trimming, and then click Continue to submit the original untrimmed sequence(s)


Sequence ID	Original Length	New Trimmed Length
2021153276	11926	11698
HB002	11923	11695
MR001	11923	11695
MR005	11923	11695
MR006	11926	11698
MR008	11923	11695
MR009	11923	11695
RB001	11923	11695
RB002	11923	11695
RB003	11923	11695
RB004	11923	11695
SD703	11923	11620
SD738	11923	11695
SD741	11923	11695
SD743	11923	11694


**Warning:** The sequences listed below contain more than 50% N's (unknown nucleotides) and are of low quality, which we cannot accept. Please trim or resequence these sequences to provide better quality data and upload the corrected data.


Sequence ID
Z0826751

 RowanDurrant / BankIt\_Checker Public

 Code

 Issues

 Pull requests

 Actions

# Step 6: Organism

## 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**

No file chosen

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

[Continue](#)

Organism

Organism

Sequence ID	Organism
Z0083197	Lyssavirus rabies
Z0083198	Lyssavirus rabies
Z0083201	Lyssavirus rabies
Z0083205	Lyssavirus rabies
Z0083214	Lyssavirus rabies
Z0083219	Lyssavirus rabies

## Step 7: Set/ Batch

### 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:

- |                               |  |
|-------------------------------|--|
| <input type="radio"/> Pop set | <b>Population study:</b> a set of sequences that were derived by sequencing the same gene from different isolates of the same organism.              |
| <input type="radio"/> Phy set | <b>Phylogenetic study:</b> a set of sequences that were derived by sequencing the same gene from different organisms.                                |
| <input type="radio"/> Mut set | <b>Mutation study:</b> a set of sequences that were derived by sequencing multiple mutations of a single gene.                                       |
| <input type="radio"/> Env set | <b>Environmental study:</b> 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.

- |  |   |
|--|---|
| <input checked="" type="radio"/> Batch | Multiple, related nucleotide sequences that are not from the same gene, but may be from the same study or organism. |
|--|---|

## Step 8: Submission Category

### 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](#)

## Step 9: Source Modifiers

### 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

HumanRabies\_GenomeBatch\_dates.txt

No file chosen

Set one value for all sequences (each chosen modifier/value will be applied to all sequences)

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

- ☐ Choose source modifier

List of source modifiers and formats can be found here: <https://www.ncbi.nlm.nih.gov/WebSub/html/help/genbank-source-table.html>

	A	B	C	D	E
1	Sequence_ID	Isolation_	Country	Host	Collection_date
2	02196160414A		Kenya		2018
3	2010818843	Brain	Kenya: Nairobi	Human	12-Mar-13
4	2011100075	Brain	Kenya: Nthangu	Bovine	13-Jul-18
5	2011100176	Brain	Kenya: Kwa DC	Bovine	11-Jul-18
6	2011100186	Brain	Kenya	Domestic dog	12-Sep-17
7	AR02		Tanzania: Arusha	Domestic dog	01-Mar-22
8	AR03		Tanzania: Arusha	Domestic dog	01-Mar-22
-	-----	-		..	



# Step 10: Feature table

## 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.

Added Features for editing/removal  
No features entered yet

The word “Feature” + the name of the sequence as it appears in the FASTA file

```
>Feature Seq3
1080 1210 CDS
1275 1315
product actin
note alternatively spliced
1055 1210 mRNA
1275 1340
product actin
1055 1340 gene
gene ACT
1055 1079 5'UTR
1316 1340 3'UTR
```

The word “Feature” + the name of the sequence as it appears in the FASTA file

>Feature Seq3				
1080	1210	CDS		
1275	1315			
			product	actin
			note	alternatively spliced
1055	1210	mRNA		
1275	1340			
			product	actin
1055	1340	gene		
			gene	ACT
1055	1079	5'UTR		
1316	1340	3'UTR		

Columns 1&2: Start and stop locations of the feature

The word “Feature” + the  
name of the sequence as  
it appears in the FASTA  
file

```
>Feature Seq3
1080 1210 CDS
1275 1315

product actin
note alternatively spliced

1055 1210 mRNA
1275 1340

product actin
gene ACT

1055 1340 gene
1055 1079 5'UTR
1316 1340 3'UTR
```

Columns 1&2: Start and  
stop locations of the  
feature

Column 3: Feature type

The word “Feature” + the name of the sequence as it appears in the FASTA file

>Feature Seq3			
1080	1210	CDS	
1275	1315		
			product actin
			note alternatively spliced
1055	1210	mRNA	
1275	1340		
			product actin
1055	1340	gene	
			gene ACT
1055	1079	5'UTR	
1316	1340	3'UTR	

Columns 1&2: Start and stop locations of the feature

Column 3: Feature type

On a new line,  
Column 4: Qualifier name

The word “Feature” + the name of the sequence as it appears in the FASTA file

>Feature Seq3				
1080	1210	CDS	product	actin
1275	1315			
			note	alternatively spliced
1055	1210	mRNA	product	actin
1275	1340			
1055	1340	gene	gene	ACT
1055	1079	5'UTR		
1316	1340	3'UTR		

Columns 1&2: Start and stop locations of the feature

Column 3: Feature type

On a new line,  
Column 4: Qualifier name      Column 5: Qualifier value

# This is the hard bit, especially if you have lots of sequences!

- GenBank seems to want you to do it by hand? Which takes ages
- More details on the ncbi site: <https://www.ncbi.nlm.nih.gov/WebSub/html/help/feature-table.html>
- BLAST might help
- I think there are some programs available to do this but I don't know how reliable they are

☐ Show results in a new window



Score	Expect	Identities	Gaps	Strand
21073 bits(11411)	0.0	11589/11695(99%)	0/11695(0%)	Plus/Plus
CDS: Putative 1	1	D K I V F K V N N Q V V S L K P E I I V		
Query	80	GACAAGATTGTATTCAAAGTCAATAACCAGGTGGTCTCTTTGAAGCCTGAGATTATCGTG		139
Sbjct	80	GACAAGATTGTATTCAAAGTCAATAACCAGGTGGTCTCTTTGAAGCCTGAGATTATCGTG		139
CDS:nucleoprotein [L	4	D K I V F K V N N Q V V S L K P E I I V		
CDS: Putative 1	21	D Q Y E Y K Y P A I K D L K K P C I T L		
Query	140	GATCAATATGAGTACAAGTACCCTGCTATCAAAGATTTGAAAAAGCCCTGCATAACCCTA		199
Sbjct	140	GATCAATATGAGTACAAGTACCCTGCTATCAAAGATTTGAAAAAGCCCTGTATAACCCTA		199
CDS:nucleoprotein [L	24	D Q Y E Y K Y P A I K D L K K P C I T L		

☐ Show results in a new window

product polymerase

4AGATTTTGAAAAAGCCCCGTGATAAACCCCTA 199  
K D L K K P C I T L

# Step 11: Review & Submit!

## 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](mailto: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 **first 5** genbank submission record(s) of your set for your review.

You can download the [complete set](#) as a compressed ZIP file.

LOCUS	Z0083197	11695 bp	cRNA	linear	VRL 24-SEP-2022
DEFINITION	Lyssavirus rabies.				
ACCESSION	Z0083197				
VERSION					



# Step 11: Review & Submit!

## 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](mailto: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 **first 5** genbank submission record(s) of your set for your review.

You can download the [complete set](#) as a compressed ZIP file.

LOCUS	Z0083197	11695 bp	cRNA	linear	VRL 24-SEP-2022
DEFINITION	Lyssavirus rabies.				
ACCESSION	Z0083197				
VERSION					

