

## Submission of allele sequences to GenBank

When you encounter new or extended alleles, the sequences can be submitted to the IMGT/HLA database. Before doing this, sequences first need to be submitted to GenBank, which is a general database of DNA sequences. NGSengine can be used to generate files describing the allele. Two files will be generated: the .fsa file (a fasta format of the DNA sequence), and the .tbl file (a feature table that describes the features of the sequence, such as exon and intron location, and protein sequence).

The submission of alleles is only possible when for class I exon 2, intron 2 and exon 3, and for class II exon 2 are fully covered and phased. The submission involves the largest phased region containing these core regions. If e.g. phasing breaks before exon 7, then exon 7 is not included in the submission.

The allele name generated by NGSengine can be:

- [Sample]\_[Locus]\_[Allelename] (Sample01\_HLA-A\_B4050) when no new nucleotides are involved
- [Sample]\_[Locus]\_[Allelename]var (Sample01\_HLA-B\_B4050var) in case of exon mismatches
- [Sample]\_[Locus]\_[Allelename]extended (Sample01\_HLA-B\_B4050extended) in case of allele extension

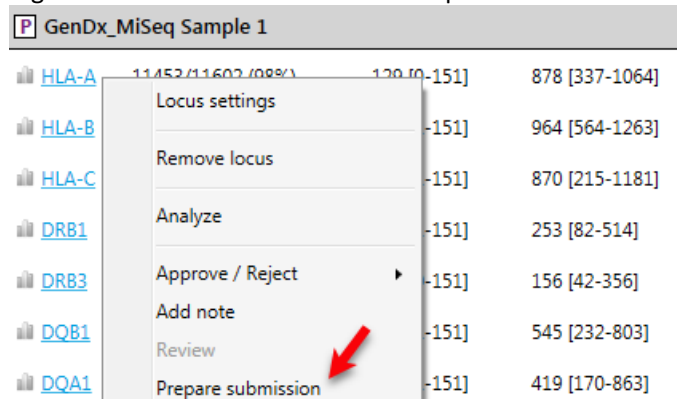
This document describes how a submission can be done. Note that this is a guideline, and the exact submission is partly dependent on the sequence submitted, and the submission website.

### Part 1: Generate the submission files

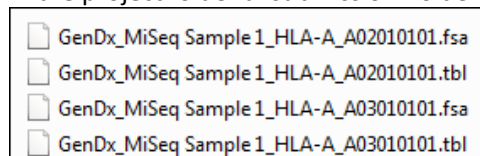
You can use NGSengine to generate the files to be submitted.

Open NGSengine.

Right-click on the locus and select 'Prepare submission'.



In the project folder a 'submission' folder is created with four files, two for each allele:



This is all you need from NGSengine to prepare the submission files.

## Part 2: Upload to GenBank using BankIt

BankIt is one of the submission tools of GenBank, which can be accessed via the following link:  
<https://www.ncbi.nlm.nih.gov/genbank/submit/>

Select 'Submit' > BankIt

On the top right corner click 'Sign in to use BankIt'

Sign in. If you do not have an account yet, register for an account.

After you signed in, an overview is shown of the last submissions, which may include partial submissions.

Click 'New submission'

The contact information is shown.

If complete, click 'Continue'

Fill in the form and click 'Continue'

Fill in the sequencing information

Click 'Continue'

### GenBank Submissions

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

Submission # 2099087

#### 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
- Ion Torrent
- Pacific Biosciences
- SOLID
- Other

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
NGSengine	2.9.0.10412	<input type="checkbox"/>
		<input type="checkbox"/>

Add additional assembly programs.

Assembly Name:

Coverage:

Fill in the form. Click 'Choose File' and choose the .fsa file generated by NGSengine:

## 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)

Molecule Type:

Topology:

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 sequences do you intend to send in this submission?

Paste Sequence(s)

### Example FASTA nucleotide format:

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

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

(OR)

Upload FASTA file

C1-114-F-51...0101var.fsa

[How do I create a FASTA file?](#)

Click 'Continue'

In case you choose to submit a single sequence, BankIt gives a warning, and you have to click 'Continue' again.

Select 'Original' and click 'Continue'.

<input checked="" type="radio"/> Original	Directly sequenced by submitter.
<input type="radio"/> Third Party Annotation	Derived from other primary sequence data.

Click 'Continue' again.

# GENDX

Select 'Add features by uploading five column feature table file'.  
Click 'Choose File' and choose the .tbl file generated by NGSengine.  
Click 'Upload file'


**GenBank Submissions**

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

**Submission # 2099087**

**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

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

Upload Features Table File  C1-114-F-512...0101var.tbl  
[How do I create a \[features table\]\(#\) file?](#)

Added Features for editing/removal  
No features entered yet

**Remove All Features** clicked.

BankIt will now combine the fasta sequence and feature table. In the current version the output will start with warnings, something like:

**Submission # 2099087**

**Features (Overview)**

**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.  
**Warning:** gene is not a valid qualifier for this feature. Converting to note.

This shows a difference in definition between the NGSengine output and BankIt. This can be ignored, since it does not influence processing of the sequence.  
Scrolling down will show a possibility to manually edit the features. Generally this is not necessary.


# GENDX

Added Features for editing/removal

[Remove All Features](#)

	Feature name	Strand	Interval range	Remove
<a href="#">Edit</a>	Gene	+	<Entire Sequence>	<input type="checkbox"/>
<a href="#">Edit</a>	CDS/Gene/mRNA	+	81..144, 275..544, 790..1065, 1648..1923, 2064..2180, 2628..2660, 3020..3024	<input type="checkbox"/>
<a href="#">Edit</a>	5' UTR	+	<1..80	<input type="checkbox"/>
<a href="#">Edit</a>	Exon	+	<1..144	<input type="checkbox"/>
<a href="#">Edit</a>	Intron	+	145..274	<input type="checkbox"/>

Scrolling down even further shows the submission definition:

[Continue](#) 

Upload File clicked.

Features

```
gene <1..>3090
      /gene='HLA-F'
exon <1..144
      /gene='HLA-F'
      /number=1
      /allele='F*01:01var'
5'UTR <1..80
      /gene='HLA-F'
      /allele='F*01:01var'
CDS join(81..144,275..544,790..1065,1648..1923,2064..2180,
2628..2660,3020..3024)
      /gene='HLA-F'
      /codon_start=1
      /allele='F*01:01var'
      /product='Non-classical MHC Class I sequence (HLA-F)'
      /translation='MAPRSLLLLSGALALDTWAGSHSLRYFSTAVSRPFGGEPFYI
AVEIVDTQFLREDSDAIIFRMEFRFPWVEQGFQIWEWTGIAKANAQTDRVALRNL
LRVYVQSEAGSHLQQMNSGDMGSPDRLLRGGVHQHAKGKDIYISLNEDRSGWALADV
AQITQRFYAEAEVVAEEFRTVLEGECLLRRLVLENGKETLQRADFPKAHVHHFISDH
EATLRWALGFVPAETTLTWQRDGEETQDTELVEVTRPAGDGTFGKNAAVVPSGEEQ
RYTCHVQHEGLPQLLILRWESQPQTIPIVGIAGLVVLGAVVTGAVVAAMVRKKS
DRNRGSYSQAAV*'
intron 145..274
      /gene='HLA-F'
      /number=1
      /allele='F*01:01var'
exon 275..544
      /gene='HLA-F'
      /number=2
      /allele='F*01:01var'
intron 545..789
      /gene='HLA-F'
      /number=2
      /allele='F*01:01var'
exon 790..1065
      /gene='HLA-F'
      /number=3
      /allele='F*01:01var'
intron 1066..1647
      /gene='HLA-F'
      /number=3
      /allele='F*01:01var'
exon 1648..1923
      /gene='HLA-F'
      /number=4
      /allele='F*01:01var'
intron 1924..2063
      /gene='HLA-F'
      /number=4
      /allele='F*01:01var'
```

Click 'Continue'

## GenBank Submissions

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

### Submission # 2099087

#### 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 1 genbank submission record(s) for your review.

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

#### Finish Submission

```
LOCUS      F*01:01var          3090 bp    DNA       linear   FRI 06-APR-2018
DEFINITION Homo sapiens.
ACCESSION  F*01:01var
VERSION
KEYWORDS
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
           Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3090)
```

Click 'Finish Submission'

Now the sequence has been submitted to GenBank.

A message appears:

## GenBank Submissions

### Submission # 2099087

#### Submission Completed

Thank you for using the GenBank Submissions Tool.

You have requested that your sequence data be released **immediately after processing**.

Your Submission ID is **2099087**. This is *not* an accession number. GenBank accession numbers will be assigned to your submissions and sent to you by email within two weeks regarding your submissions before you receive these, be sure to refer this Submission ID in your email.

Your submission data have been sent to GenBank. For each complete sequence submission, you will receive the following at the email address you provided:

1. an automatic reply confirming our receipt of your submissions;
2. GenBank accession numbers (within two working days); and
3. the final GenBank records, processed by the GenBank Annotation staff and incorporating the information you have provided for you to review before they are released.

If you do not receive this information from us by email within the time frame indicated, please send an inquiry to [gb-admin@ncbi.nlm.nih.gov](mailto:gb-admin@ncbi.nlm.nih.gov) and be sure to include the S

[Start a New Submission](#)

The result can be downloaded, which is a zip file containing the details of the submission.

After some time you will get an accession number which you can use to submit the allele to the IMGT/HLA nomenclature committee.