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.

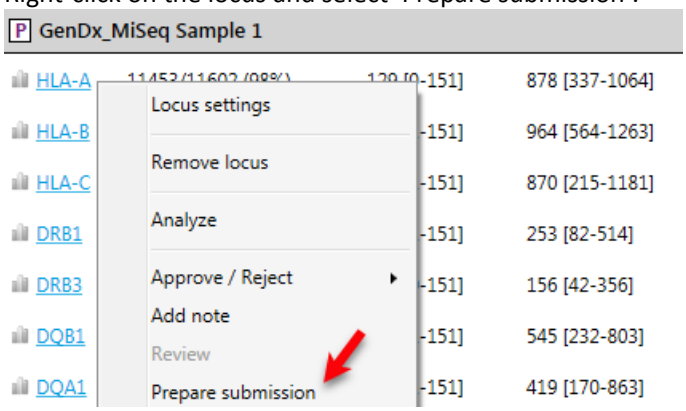
The allele name generated by NGSengine can be:
- [Sample]_[Locus]_[Allelename] (Sample01_HLA-A_B4050) when no new nucleotides are involved
- [Sample]_[Locus]_[Allelename]var (Sample01_HLA-B_B4050var) in case of exon mismatches
- [Sample]_[Locus]_[Allelename]extended (Sample01_HLA-B_B4050extended) in 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.

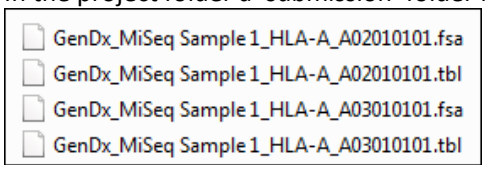
Part 1: Generate the submission files

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.

GENDX

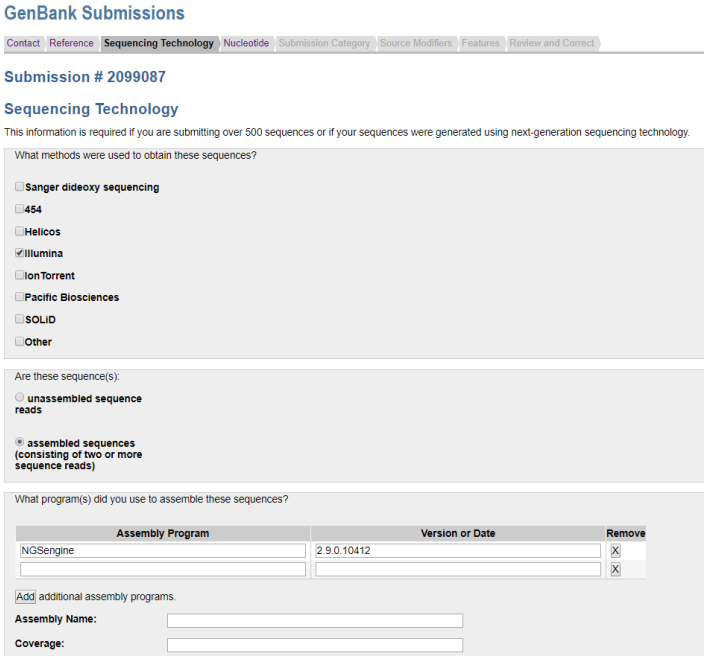
Part 2: Upload to GenBank using BankIt

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/

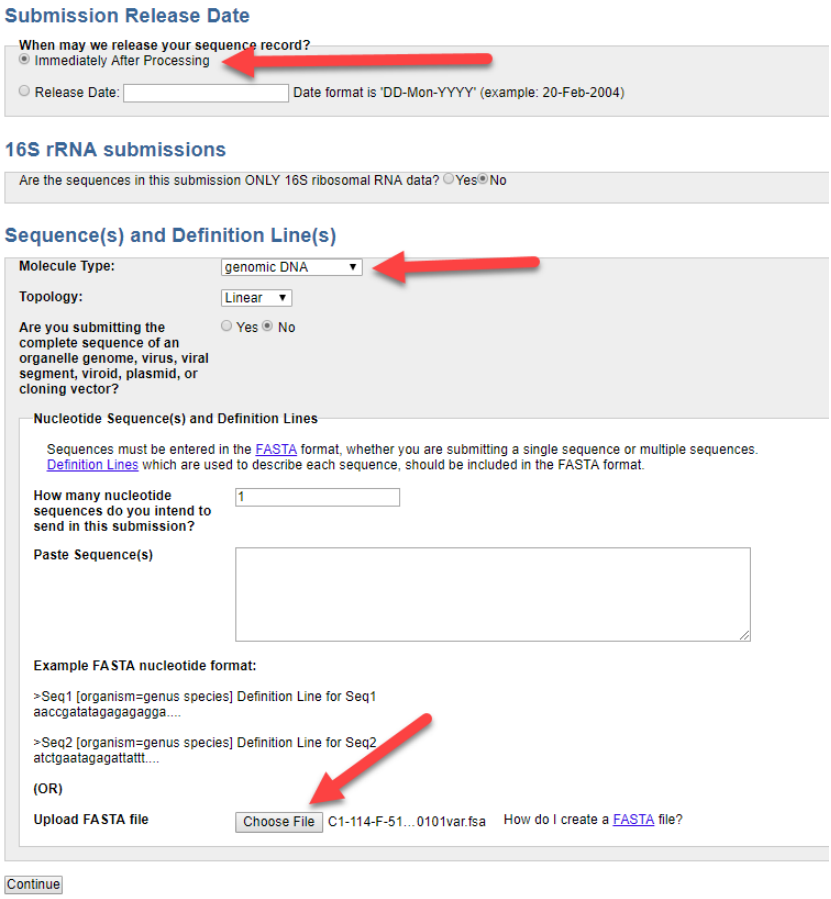
Select ‘Submit’ > BankIt
On the top right corner click ‘Sign in to use BankIt’
Sign in. If you do not have an account yet, register for an account.
After you signed in, an overview is shown of the last submissions, which may include partial submissions.

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’

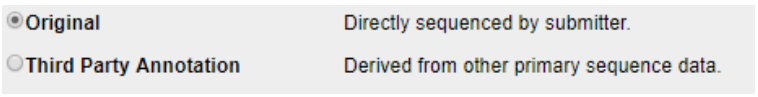


Fill in the form. Click ‘Choose File’ and choose the .fsa file generated by NGSengine:



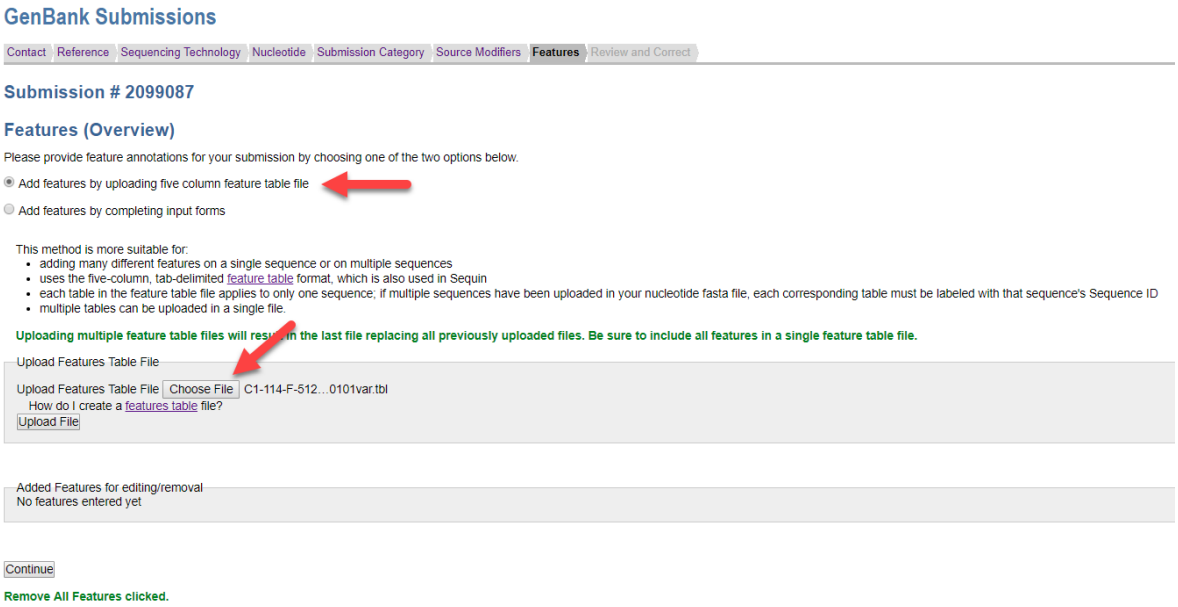
Click ‘Continue’
In case you choose to submit a single sequence, BankIt gives a warning, and you have to click ‘Continue’ again.

Select ‘Original’ and click ‘Continue’.

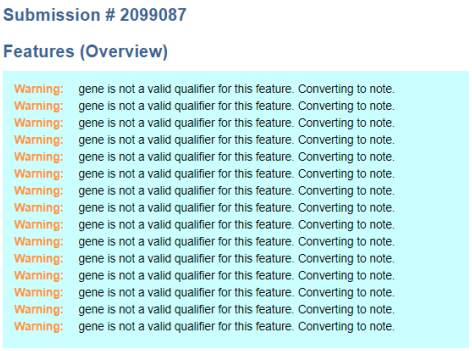


Click ‘Continue’ again.

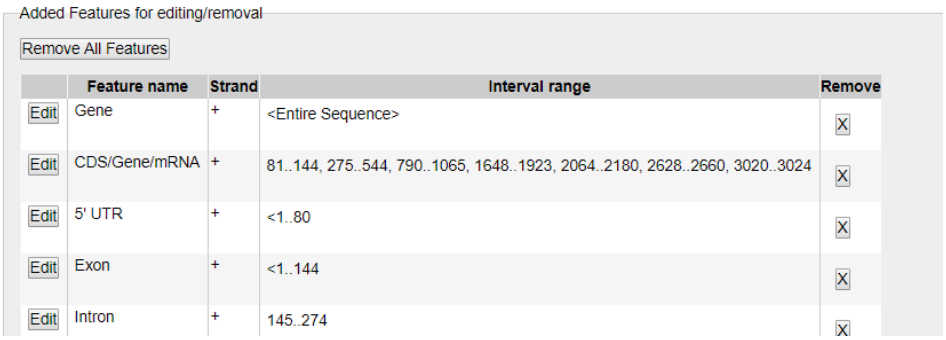
Select ‘Add features by uploading five column feature table file’.
Click ‘Choose File’ and choose the .tbl file generated by NGSengine.
Click ‘Upload file’



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



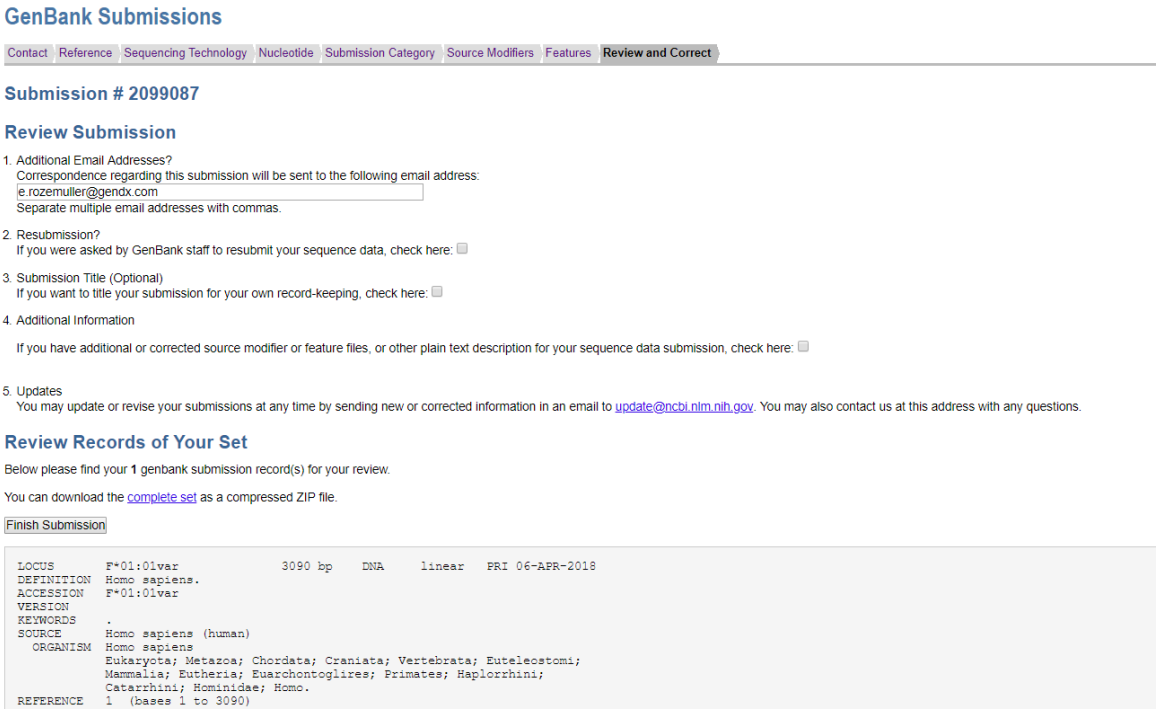
This shows a difference in definition between the NGSengine output and BankIt. This can be ignored, since it does not influence processing of the sequence.
Scrolling down will show a possibility to manually edit the features. Generally this is not necessary.



Scrolling down even further shows the submission definition:

Click ‘Continue’

GENDX



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:
   e.rozemuller@gendx.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.

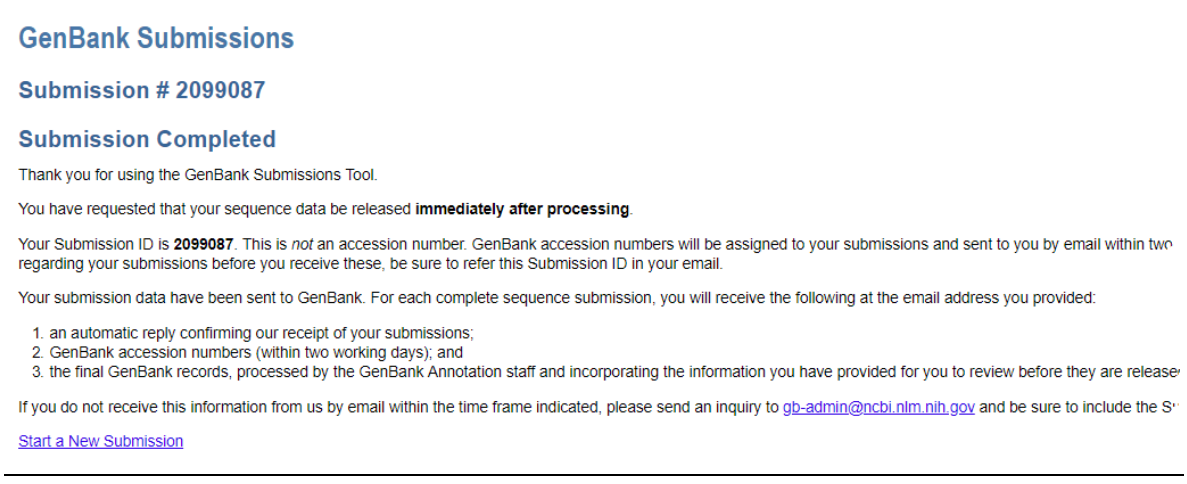
You can download the complete set as a compressed ZIP file.

Finish Submission

```
LOCUS       F*01:01var             3090 bp    DNA     linear   PRI 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
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 release

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 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.