Subitting sequences to GenBank

FILE NAME: Submitting sequences to GenBank.doc

Authored by: Emily Wright
Last Modified on: 23OCT12
Approved by: Martin Wiedmann

EFFECTIVE DATE: Date of Approval

APPROVED BY:

_____   _____
Dr. Martin Wiedmann   (date)

AUTHORED BY:

_____   _____
(Name)   (date)

TABLE OF CONTENTS

SECTION 1   INTRODUCTION

1.1   Purpose

The purpose of this protocol is to outline how to submit sequences to GenBank.

1.2   Scope

This protocl applies to the Food Safety Lab, including the Laboratory for Food Microbiology and Pathogenesis of Foodborne Diseases.

1.3   Definitions

1.4   Safety

N/A

SECTION 2   MATERIALS

N/A

SECTION 3   PROCEDURES

Submitting Sequences to GenBank

There are ten steps to submit you sequences to GenBank (Contact, Reference, Nucleotide, Organism, Submission category, Sequencing Technology, Source Modifiers, Primers, features, and Review and Contact). Each is listed as a top on the top of the website. The steps below will help walk you through you submission.
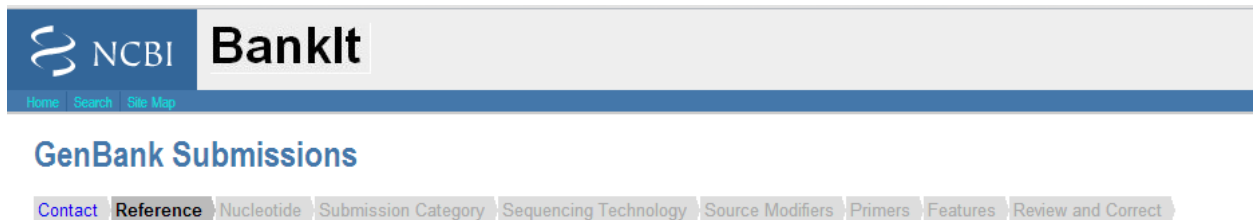
1. Go to the NCBI website: http://www.ncbi.nlm.nih.gov/
2. Under the "Get Started" menu, select "Submissions"
3. Under the Sequence Data heading, select GenBank
4. Under the GenBank Resources header, select 'Submission Tools"
5. Under the Submissions to GenBank header, now select "BankIt"
   a. Use BankIt if
      i. you have a single sequence, a simple set of sequences (for example:16S rRNA, matK, ITS/rRNA, amoE, tefB, cytb, or COI sets), or a small batch of different sequences
      ii. you prefer to use a web-based submission tool
      iii. the feature annotation for your sequences is not complicated
      iv. you do not require advanced sequence analysis tools
   b. Use Sequin if:
      i. you prefer to work on your submission off-line
      ii. you have a sequence or sequences that are complex
      iii. you would like graphical viewing and editing options, including an alignment editor
      iv. you would like the option to have network access to related analytical tools
6. On the far right, click 'Sign in to use BankIt'. If you do not have an account, it is relatively quick to create on.
7. Click on New Submission to start
8. In the next window, fill out all submission and contact information: coauthors, reference information, and publication status.
9. Chose an appropriate release date for your sequences. You may choose to embargo your sequences for anytime in the future.
10. For this example we will add 16S rDNA sequences by selecting 16S

11. You will then be asked whether you used a chimeric checking tool to check your sequences. Select no. All your isolates should be pure cultures.
12. Fill in Molecular Type (genomic DNA), Topology (linear), and Genomic completeness (partial).
13. Enter who many sequences you have
14. If you have just one sequences, you can paste the sequence in the box below OR if you have multiple sequences, you can upload a FASTA file containing all your sequences
15. Enter organism information
16. Enter Sequencing technology. At Cornell BRC, we use Sanger dideoxy sequencing.
17. Select assembled sequences and list what program was used to assemble them (Sequencher v5.0.1, etc)
18. Select '2x' coverage; for forward and reverse
19. Fill in Organism and Source Mofofiers table. List 'culture' (ie single organism), isolation source (ie. Water, farm, etc), country, and any other descriptive information you wish to provide. If you have multiple strains, you can create this table in Excel and upload it.
20. If you are submitting 16S sequences, you do not have to list primers. But do list primers if they are unique.
21. On the Featurs page, make sure your sequence is the correct length, etc.
22. Review Submission and click "Finish Submission"
23. Copy down your submission number that is generated
24. It may take up to 72 hours for you to receive your Accession numbers via email.

SECTION 4   TROUBLESHOOTING

The NCBI GenBank website has a very detailed manual on how to submit sequences that can be found at:
http://www.ncbi.nlm.nih.gov/books/NBK51157/

Requirements for GenBank Submissions through BankIt

- Contact Information
  - Name, address, phone number, fax number and email address of the submitter must be entered when registering and submitting for the first time
  - Subsequent BankIt submissions will retain this information and display it once the submitter logs in

- Release date information
  - Immediately after it is processed at NCBI OR
  - On a date the submitter specifies

- Reference information
  - Sequence authors: names of the researchers who are credited with the sequence
  - Publication information: Unpublished, In-Press, or Published; and applicable citation information (paper's title, authors, journal title, volume, issue, year, pages)

- Submission Category and Type
  - Original sequencing or Third Party Annotation
  - Single sequence, sequence set (phylogenetic, population, environmental, etc), or batch

- Nucleotide sequence(s)
  - Input (cut-and-paste) single or multiple sequences OR
  - Upload them as a FASTA file; FASTA files should include organisms in their definition lines
  - Sequences must be at least 200 nucleotides long (unless they are complete exons, non-coding RNAs (ncRNAs), microsatellites or ancient DNA)
  - Molecule type: what was sequenced? (genomic DNA, mRNA, genomic RNA, cRNA, etc)
  - Topology: linear or circular (circular must be complete, such as a complete plasmid)

- Organism name, applicable source modifiers, location
  - Genus and species names (if not previously provided in FASTA file)
  - If name is new or unrecognized, provide best known taxonomic lineage
  - If genus and/or species names are not known, provide most specific name known (for example:Bacillus sp., Uncultured bacterium, Uncultured archaeon)
  - Most complete name for any synthetic vector (for example: Cloning vector pAB234, Transfer vector p789Abc)
  - Source modifiers include: strain, clone, isolate, specimen-voucher, isolation-source, country
  - Location: organelle (mitochondrion, chloroplast, etc); map and/or chromosome

- Features of the sequence
  - Upload files or use input forms to add all applicable features (for example: CDS, gene, rRNA, tRNA, microsatellite, exon, intron)

SECTION 5   REFERENCES

http://www.ncbi.nlm.nih.gov/