	FOOD SAFETY LAB / MILK QUALITY IMPROVEMENT PROGRAM *Standard Operating Procedure*

Title: User guide to assignment of *rpoB* allelic types for the order Bacillales			
SOP #: 8.4.6	Revision: 03	Revision Date: 11-1-23	Effective Date: 11/1/2023
Authors: Ahmed Gaballa / Steven Warchocki		Approved by:	

User guide to assignment of rpoB allelic types for the order Bacillales

FILE NAME: 8.4.6-rpoB_ATassignment.docx

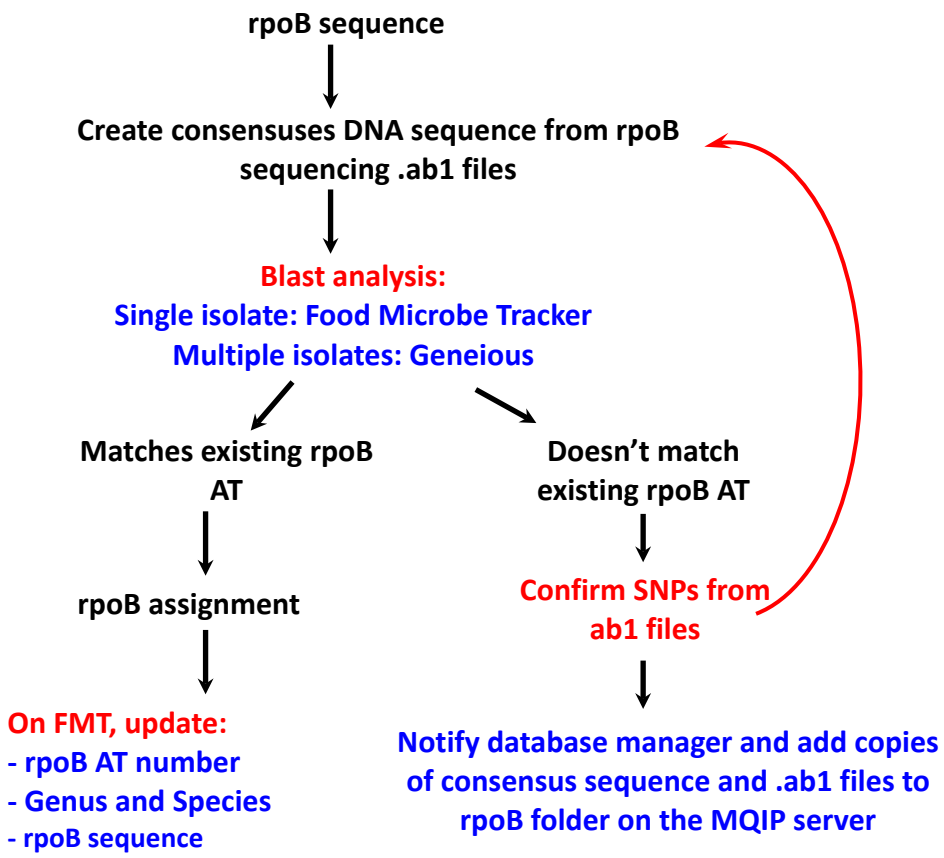


Figure 1: Flow Chart of steps needed to provide isolate with *rpoB* sequence allelic type number (AT number) and Genus/species identification.

TABLE OF CONTENTS

SECTION 1 INTRODUCTION

1.1 Purpose

To determine *rpoB* sequence allelic types (AT) of bacteria from the order Bacillales isolated in the Milk Quality Improvement Program (MQIP) and Food Safety Laboratory (FSL). AT assignment is achieved by comparing the sequence of an internal fragment of *rpoB* to an existing database.

1.2 Scope

This SOP applies to the MQIP and the FSL.

1.3 Definitions

AT: allelic type; defined as one specific DNA sequence of (or within) a gene, in this case a 632-nucleotide region of the *rpoB* gene in Bacillales.

bp: base pair

BLAST: Basic Local Alignment Sequence Tool

Consensus: a single sequence derived from a set of overlapping DNA segments originating from one genetic source

FMT: Food Microbe Tracker; WWW-based tool for information exchange on bacterial subtypes and strains, containing a large amount of bacterial gene information

PCR: Polymerase Chain Reaction, used to amplify a specific region within a DNA sequence.

Percentage sequence identity: proportion of identical nucleotides between two sequences multiplied by 100.

Phylogeny The evolutionary history of taxonomic groups.

***rpoB*:** RNA polymerase beta subunit

1.4 Required readings

https://pubmed.ncbi.nlm.nih.gov/34710512/

SECTION 2 MATERIALS

1. **Computer**

2. **Partial *rpoB* gene sequences:** From PCR products of isolates; sequences are obtained after PCR products are sent to the BRC facility (Biotechnology Research Center). Raw data files are in .ab1 format and edited consensus sequences are saved as .fas files.

3. **Internet Access:** For accessing Food Microbe Tracker.

4. ***rpoB* database via Food Microbe Tracker:** This can be accessed by logging into FMT, searching by DNA sequence, and selecting *rpoB* allelic typing.

5. **Local *rpoB* database**: This can found on the MQIP server at:

   \\cornell.edu\ag\FOOD\FOOD-MQIP\rpoB database\Current database file

   You will need to install Cisco AnyConnect software to connect to Cornell's VPN service and access the server off campus (https://it.cornell.edu/cuvpn).

6. **Geneious**: DNA and protein analysis software.

7. **Geneious workflow:** Geneious file that contains the analysis steps in chronological order. This file to be imported to Geneious.

SECTION 3 PROCEDURES

Note: If you need to analyze multiple sequences skip to section 3.2.

3.1 *rpoB* Allelic Type Assignment with Food Microbe Tracker (single isolate analysis)

1. Obtain edited (final) *rpoB* sequence data (Creating a consensus DNA sequence from ABI sequence data using Sequencher). These sequences should be at a minimum, 632 bp in length for *rpoB*, but are often longer.

2. Open website for Food Microbe Tracker: http://www.foodmicrobetracker.com

    a. Log-in, or request account for Log-in.

    b. In Food Microbe Tracker (FMT), on the left-hand side of the main page, under "Search By", click on "DNA Sequence".

    b. Once on this page use the pull-down menus to adjust your search parameters:

        i. "Number of Results": *default=10*, you may wish to increase/decrease this.

        ii. "Genus": *default=Unspecified*.

        iii. "Species": *default=Unspecified*.

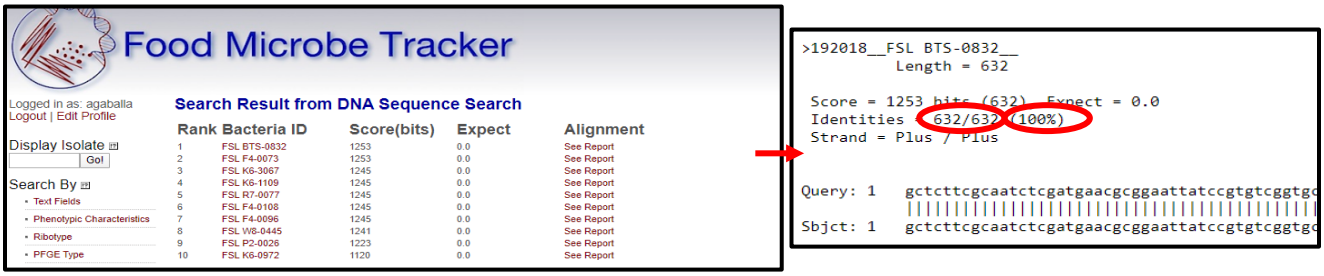
        iv. "Sequence Type": *default=Unspecified*, **this must be changed to "*rpoB* allelic typing"** in the pull-down menu (from the pull-down menu, make sure that you are using ***rpoB* allelic typing** and **not *rpoB***)

3. Open the *rpoB* consensus (final) sequence file (.fas) you wish to find an allelic type for. This can be done in either Notepad or Sequencher.

4. Copy and paste your *rpoB* sequence into the space labeled "Enter DNA sequence".

5. Click Submit. Once your results page ("Search Results from DNA Sequence Search") has loaded, choose the first Alignment file by clicking on "See Report" in red.

6. When the new page/tab has appeared in your web browser, click to view it and review some key details:



c. "Identities": this should read "632/632 (*100%*)" for a 100% allelic type match. Unique AT sequences will have less than a 100% AT match.

d. "Query": is the *rpoB* allelic type sequence you entered.
"Sbjct": is the database *rpoB* allelic type sequence and it should start at 1 and end at 632 (a few *rpoB* ATs have 829 bp or 635 bp).

e. If there are no *rpoB* ATs with a 100% match to the sequence queried (Identities do not equal "632/632 (100%)") you may have a new *rpoB* allelic type.
Check in Sequencher if the differences between your sequence and the sequence in the database are legitimate, i.e., they are not artifacts introduced during editing.

If the SNPs are not due to mis-analysis of the sequence to the database manager, currently Ahmed Gaballa (ag67@cornell.edu).

Additionally, add .ab1 files and consensus *rpoB* sequence (final) data to \\cornell.edu\ag\FOOD\FOOD-MQIP\rpoB database\Possible new ATs

If more than one new *rpoB* ATs are found, a list or file of consensus sequences may be compiled to facilitate easy data transfer.

7. If you have confirmed a 100% identity match between an *rpoB* allelic typing sequence currently listed in FMT and your query sequence, return to the search results page and click on the "Bacterial ID" in red. When directed to the isolate information page scroll down to "Additional Characteristics" and look for the "*rpoB* allelic type" number in bright red font. This number is the allelic type of the sequence you have searched.

8. The *rpoB* allelic type number may be entered into FMT for your isolate by going to the information page for your isolate, scrolling down to "Additional Characteristics" and clicking "Edit" and entering the correct AT number in the box for "*rpoB* allelic type". For hits with 100% identity matches, Genus/species information from the type strain should also be added to your query strain's FMT information. Update "Basis of species identification" entry on FMT using "rpoB sequencing".

9. At this time *rpoB* sequence data (raw and consensus) must also be added to Food Microbe tracker (see step 10 below).
NOTE: *Only sequences with both "rpoB allelic type" and "rpoB allelic typing" sequences are representative rpoB type strains. This means you will never enter "rpoB allelic typing sequence" data; this is only done by the database manager.*

10. In order to add sequence data for an isolate, open its page, scroll down to "DNA sequences" and click "Add". On the "Add DNA Sequence page", under "Type" select "*rpoB*". Where "Sequence" is listed, use the "Browse" button to upload your consensus

(final) sequence (.fas). Under "Raw data files" add the corresponding two raw files (.ab1AB1, for both Forward and Reverse). Update PCR and sequencing primers entries. Click "Submit".

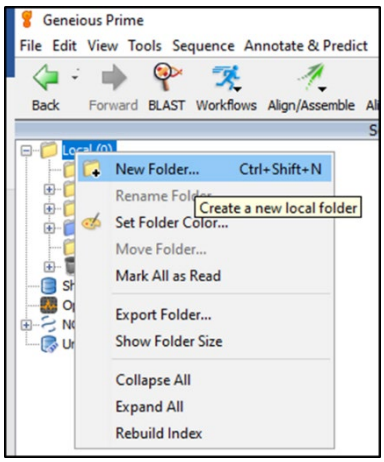
11. As a check, search FMT for all entries with the same *rpoB* AT you have found for your *rpoB* sequence (You can do this using the advanced search option, under "additional characteristics" and using quotation marks around the AT number). If all Genus/species identifications match, no action is required for your new entry; if there are discrepancies, email the database manager with the *rpoB* AT number.

3.2 *rpoB* Allelic Type Assignment with Geneious (multiple isolates analysis)

1. Copy the *rpoB* current database and "rpoB_AT_Workflow" files from: \\cornell.edu\ag\FOOD\FOOD-MQIP\rpoB database\Current database file to a local folder on your computer.

   When starting a new analysis, always copy a new database file from the MQIP server to guarantee that you are using the current database version.

2. Open Geneious and create new folder.

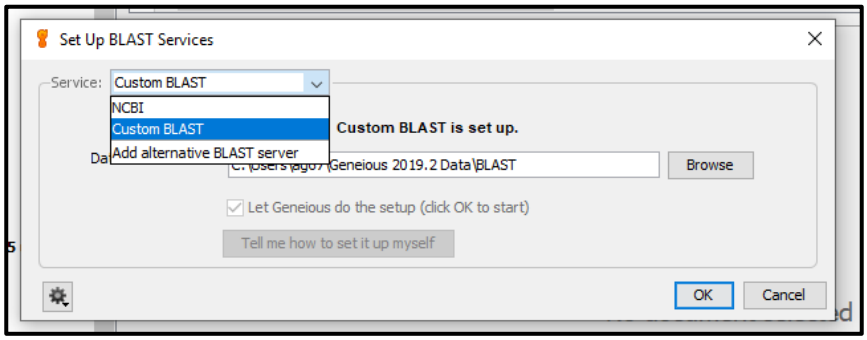


3. Drag and drop or import the *rpoB* current database file (File → import → from file: keyboard shortcut: Ctrl+I).

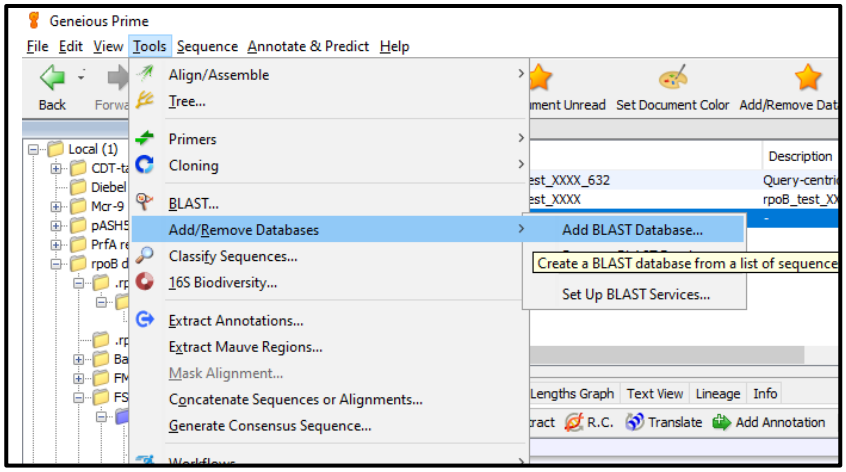
4. Create local Blast database in Geneious:

   a. Set Up Blast Services: This step must be done only once on the first time of using local Blast in Geneious.
   i. In Geneious menu go to Tools→ Set Up Blast Services.

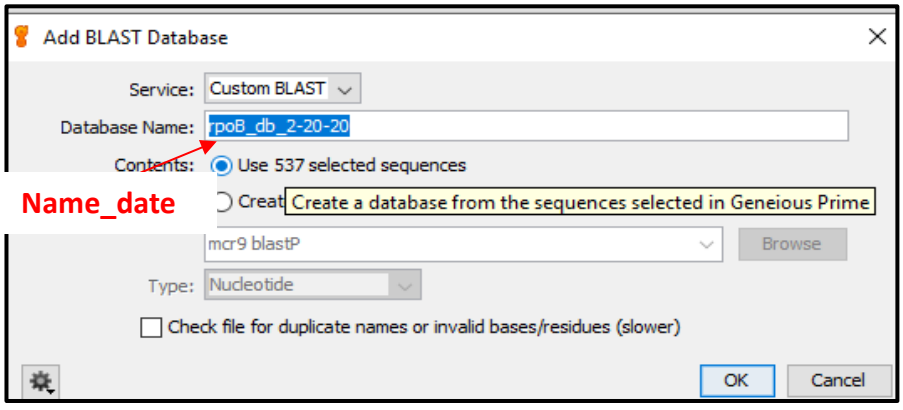
ii. In the popup window:
Service: select "Custom Blast". You can leave Folder location as suggested or browse to change where you want to save your files. Hit OK and wait until the setup is complete.



b. In Geneious, select the *rpoB* current database file (single left click) →Tools→ Add/remove databases → Add BLAST database.

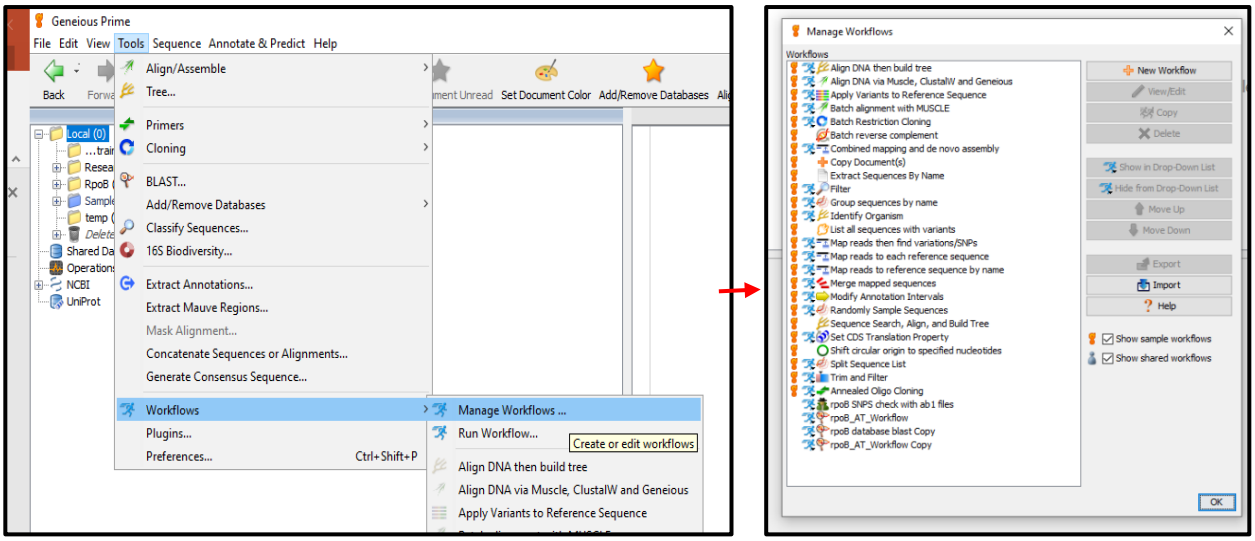


i. In the popup window:
Service: select "Custom BLAST"
Database Name: "*rpoB* database DATE"
Click on: Use "number" selected sequences
Click: OK



5. In Geneious, import the consensus *rpoB* sequences:
   Create a new folder; drag and drop all consensus *rpoB* sequences (in fasta format) into the folder. Geneious might ask to either keep the sequences separate or in a list: choose separate.

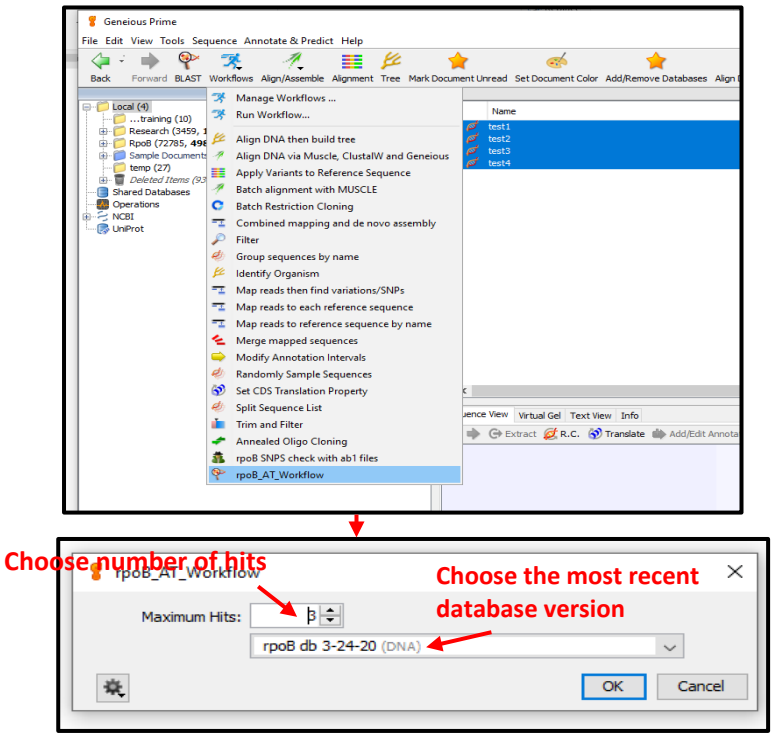
6. Import "rpoB_AT_Workflow":
   In Geneious: go to Tools → Workflows → Manage workflows
   in the popup window: click Import (on the right) and upload the "rpoB_AT_Workflow".



7. Select all *rpoB* sequences to be analyzed (depending on your computer's memory, you might have to analyze 50 to 100 sequences at a time).

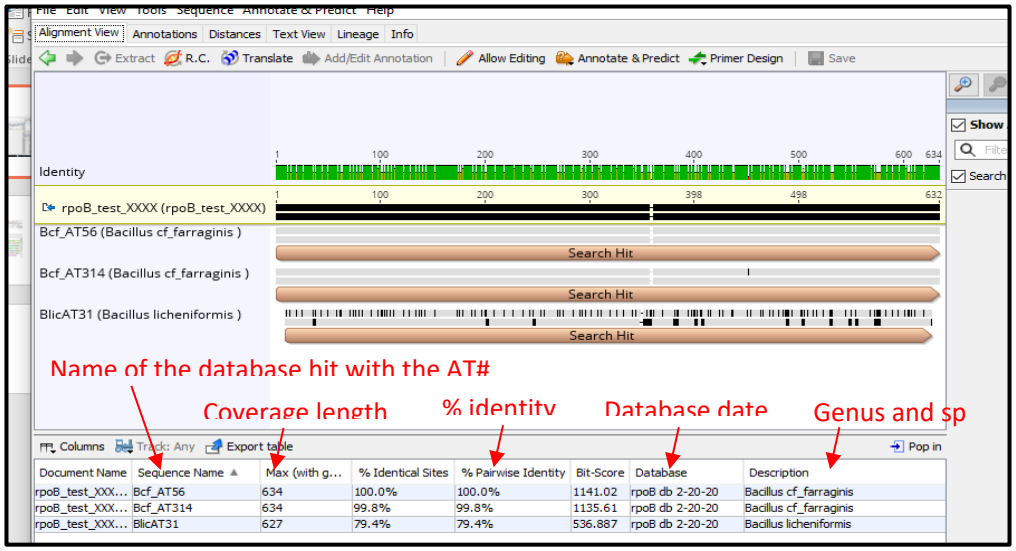
8. Run the workflow:
   Tools → Workflows → click on "rpoB_AT_Workflow".
   Select the number of hits that you want to see for each sequence and the *rpoB* database

(make sure to use the database with most recent date). Hit OK to run workflow.



9. The workflow will analyze each sequence separately and generate an alignment file for each sequence.

10. Click on the file and the alignment will appear in the right-bottom window. Click on "Annotations" in Geneious right-bottom window toolbar to show details of the database hit, percent identity, description of the database hit, which includes genus, species, AT number etc.



Hint: if you do not see all the columns in the toolbar of Geneious right-bottom window, click on columns, select "show all"

11. **To assign rpoB AT, Genus and species follow steps 5-10 from 3.1 section.**

Optional: you can export results to an Excel file:

1. In Geneious, select all Blast generated files from the right-top window.

   Hint: sort files by "Description" or "created date"; left-click on the first file name, hold shift, scroll down, and left-click on the last file name.

2. A large table should appear in the right-bottom window.

   Hint: if a large number of sequences were analyzed, a warning might appear in the right-bottom window of Geneious as "Large Documents" click "View Documents" underneath the warning.

3. In the toolbar of Geneious' right-bottom window, click on columns, select "show all"

4. In the toolbar of Geneious' right-bottom window, click on "Export table" and save the Comma Separated Values file (csv).

5. Open the csv file in Excel. This file should contain all the information you need to analyze your sequences as described in section 3.1 (points 5-10) including:

   a. Column "Document Name".

   b. Column "Sequence Name": Database Hit name.

   c. Column "% Pairwise Identity" shows the percent identity between the query and the database hit.

   d. Column "Length (with gaps)" shows the length of alignment.

   e. Column "Description" Genus, species, AT number of the database Hit.

SECTION 4 TROUBLESHOOTING

4.1 If "Identities" in your results read anything but out of 632 (e.g. 630/630 or 626/631), the BLAST algorithm has somehow trimmed your sequence. First check the second-best match, if that one isn't out of 632 either, then it is best to pull out *rpoB* sequences from each isolate's FMT page and align them (ClustalW or Mesquite can do this) and see if the complete length matches for 632bp. Another possibility is that Sequencher trimmed part of the rpoB allelic typing region from either ends. Go back to the consensus Sequencher and try to call the missing bases manually.

4.2 Matches showing 635/635 usually indicate a match with a *Staphylococcus* sp. Close attention should be paid to these isolates. Record percentage identities to the best *rpoB* AT match. Report to database manager, or Martin Wiedmann. These sequences may be added to the database so that these Genera can be identified, however if you find many of these in your project there may be a breakdown in laboratory methods that may need to be investigated.

SECTION 5 METHOD REVIEWS & CHANGES

VERSION	DATE	EDITOR	COMMENTS
Version 1 (SOP 8.1.1.1.12)	11/17/2014	Steven Warchocki	Original SOP. Assignment rpoB and sigB ATs
Version 2 (SOP 8.4.6)	04/22/2020	Ahmed Gaballa	Specific SOP for rpoB allelic typing.
Version 3	11/01/2023	Ahmed Gaballa	Additon of 2021 rpoB database paper as required reading.