A DATABASE FOR TAXONOMIC AND PHYLOGENETIC IDENTIFICATION OF THE GENUS BRADYRHIZOBIUM USING MULTILOCUS SEQUENCE ANALYSIS

1 TUTORIAL

In this tutorial we'll show you how to navigate through some of the available options.

2 ABOUT

On the main page of our database, you will find an overview of the work that led to this site.

Biological nitrogen fixation, with an emphasis on the legume-rhizobia symbiosis, is a key process for agriculture and the environment, allowing the replacement of nitrogen fertilizers, reducing water pollution by nitratE as well as emission of greenhouse gases. Soils contain numerous strains belonging to the bacterial genus Bradyrhizobium, which establish symbioses with a variety of legumes. However, due to the high conservation of Bradyrhizobium 16S rRNA genes-considered as the backbone of the taxonomy of prokaryotes-few species have been delineated. The multilocus sequence analysis (MLSA) methodology, which includes analysis of housekeeping genes, has been shown to be promising and powerful for defining bacterial species, and, in this study, it was applied to Bradyrhizobium species, increasing our understanding of the diversity of nitrogen-fixing bacteria.

Classification of bacteria of agronomic importance is relevant to biodiversity, as well as to biotechnological manipulation to improve agricultural productivity. We propose construction of an on-line database that will provide information and tools using MLSA to improve phylogenetic and taxonomic characterization of Bradyrhizobium, allowing the comparison of genomic sequences with those of type and representative strains of each species.

A database for the taxonomic and phylogenetic identification of the *Bradyrhizobium* genus, using MLSA, will facilitate the use of biological data available through an intuitive web interface. Sequences stored in the on-line database can be compared with multiple sequences of other strains with simplicity and agility through multiple alignment algorithms and computational routines integrated into the database. The proposed database and software tools can be used, free of charge, by researchers worldwide to classify *Bradyrhizobium* strains; the database and software can be applied to replicate the experiments presented in this study as well as to generate new experiments. The next step will be expansion of the database to include other rhizobial species.

2.1 URL, AND CONTACT

Clicking on the About tab will show you the Contact sub-menu, which contains contact data of the main researcher.

2.1.1 **URL:** http://mlsa.cnpso.embrapa.br

2.1.2 **Important:** This website and its applications are best viewed in **Firefox browser** version 30 or later.

2.1.3 **Contact:** Contact for information about the website and database.

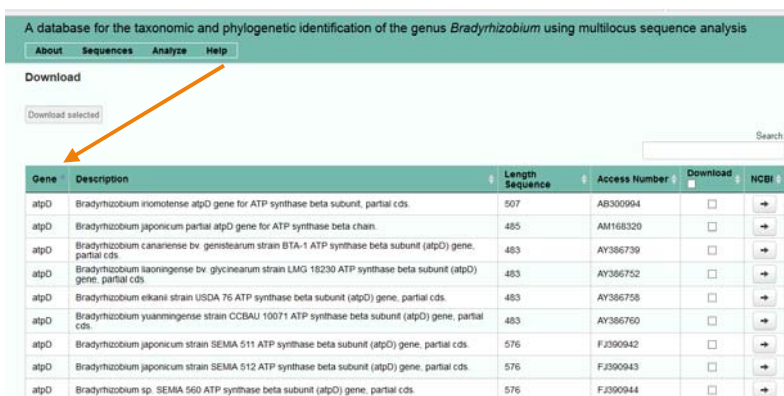**Name:** Fabrício Martins Lopes
E-mail: fabricio@utfpr.edu.br
URL: http://www.utfpr.edu.br/cornelioprocopio

3 SEQUENCES DOWNLOAD

Clicking on the Sequences tab will show you the Download sub-menu, which will link you to the Data Table component, with all sequences stored in our database.

The database currently has 286 entries, distributed as follows:

- For all (57) strains stored, we provide four genes (atpD, dnaK, glnII and recA);
- For 30 strains stored, we provide five different genes (atpD, dnaK, glnII, recA and gyrB);
- For 28 strains stored, we provide four different genes (atpD, dnaK, glnII, recA, gyrB and rpoB).

3.1.1 **Ordering can be changed by clicking on the arrows that appear to the right of each field in the table's first row.**

3.1.2   You can also search for specific organisms using keywords in the Search field on the right.

3.1.3   By clicking on the download check box (on the right of each strain) will give file in Fasta format, for one or more sequences.

3.1.4   By clicking the download check box button will give file in Fasta Format, for all sequences displayed.

3.1.5   The Next Step is selected click download button to download the desired sequences into a single file.

3.1.6   Clicking on the NCBI button (on the right of each strain) will give you access to the desired sequence in the NCBI.

3.1.7   On the upper corner of this page you find the possible format options for the GenBank.

3.1.8   There are two ways to transfer the data to your device:

3.1.8.1   By selecting and copying the sequence data right clicking the mouse;
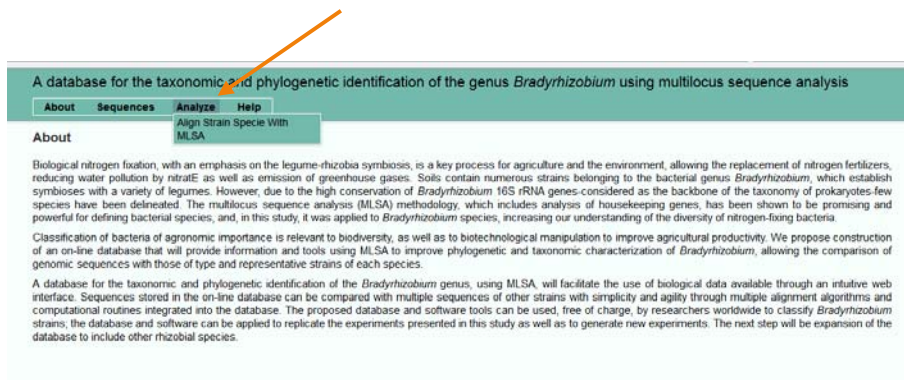
3.1.8.2   Or by the Send dialog box in the right upper corner of the screen and selecting "complete record" and "file".
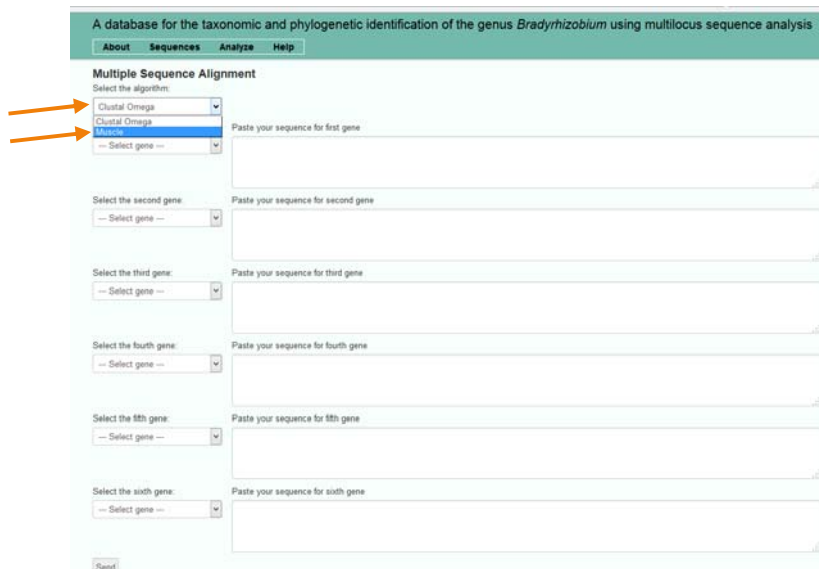

4   ANALYZE

In this part of tutorial, we will show you the process of multilocus sequence analysis.

4.1.1   Clicking the context menu "analysis" will show you the option "align with strain MLSA".

4.1.2   In the first dialogue box on your left, clicking on the arrow will give you the choice between Clusta Omega or Muscle.

4.1.3   The next six dialogue boxes on left allow you to select each of the one to six genes to be used for analysis.

4.1.4   In this example we must select 3 different genes and paste the sequences for each of them in the "field " paste your sequence for your first (second / third) gene.

4.1.5   When selecting a particular gene, information on the smallest size of the sequence stored: min xxx bp (example: min 369 bp) which means that the smallest sequence for gene dnaK in the databank has 369 base pairs.

4.1.6   After filling in your sequences, you can click the send button and wait for results.

4.1.7   A screen informs the procedure being performed.

4.1.8   The results will be shown in the following order (Alignment, Identity Matrix and Parameters for the Generation of the Phylogenetic Tree), and include download option to alignments, the identity matrix and the newick tree.

4.1.8.1   ALIGNMENT

4.1.8.2 IDENTITY MATRIX

4.1.8.3 PARAMETERS FOR GENERATION OF THE PHYLOGENETIC TREE.

4.1.9 After downloading, you can use programs for data processing.

4.1.10 Next step is an example alignment and phylogenetic tree analysis using MEGA software version 6.
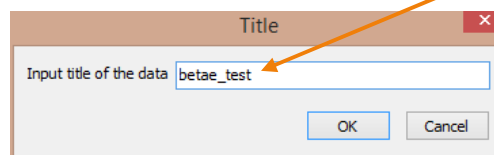
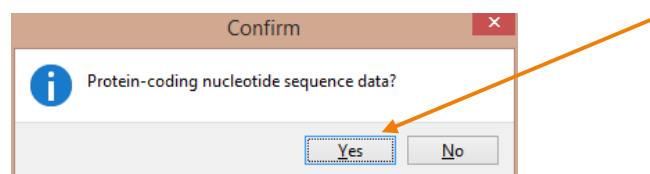### 4.1.11 Use option DATA > Export Alignment > MEGA Format
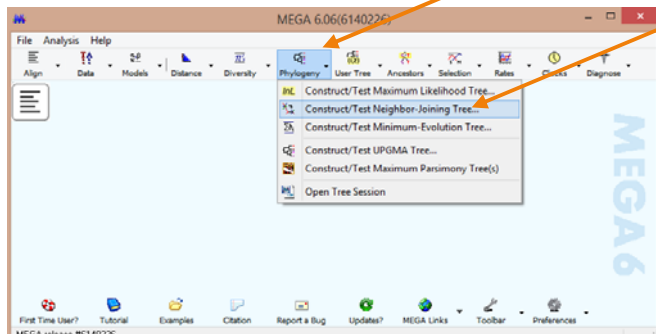

4.1.12 Save as alignment.meg

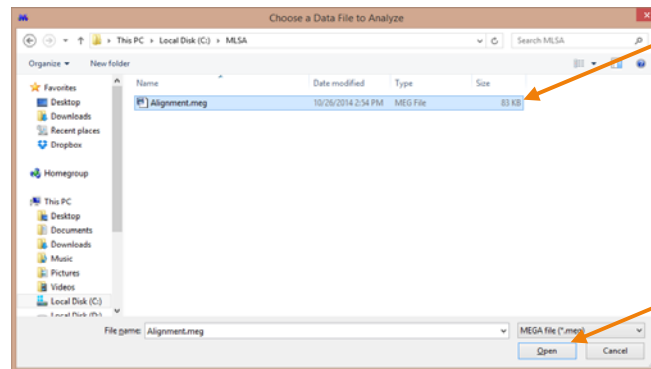### 4.1.13 Input title of the data: "betae_test"

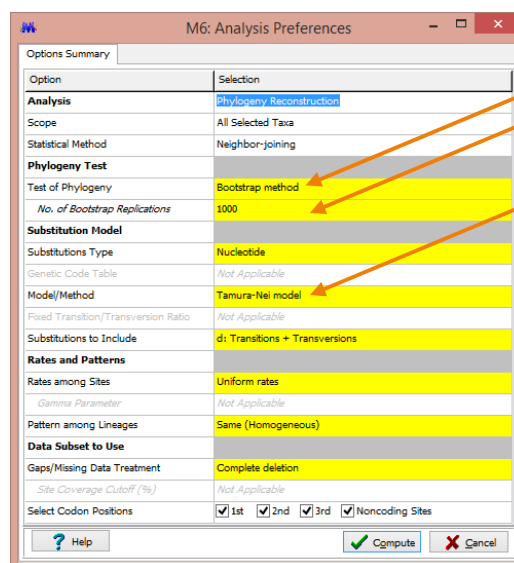### 4.1.14 Protein-coding nucleotide sequence data? "Yes"

4.1.15   Now Open Phylogeny Tab in Software Mega.
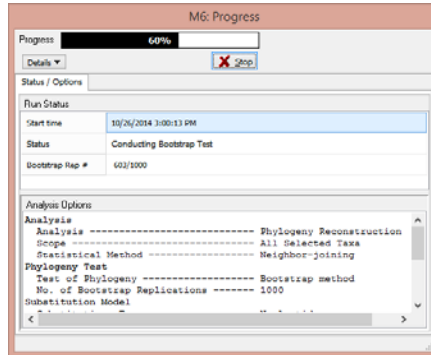

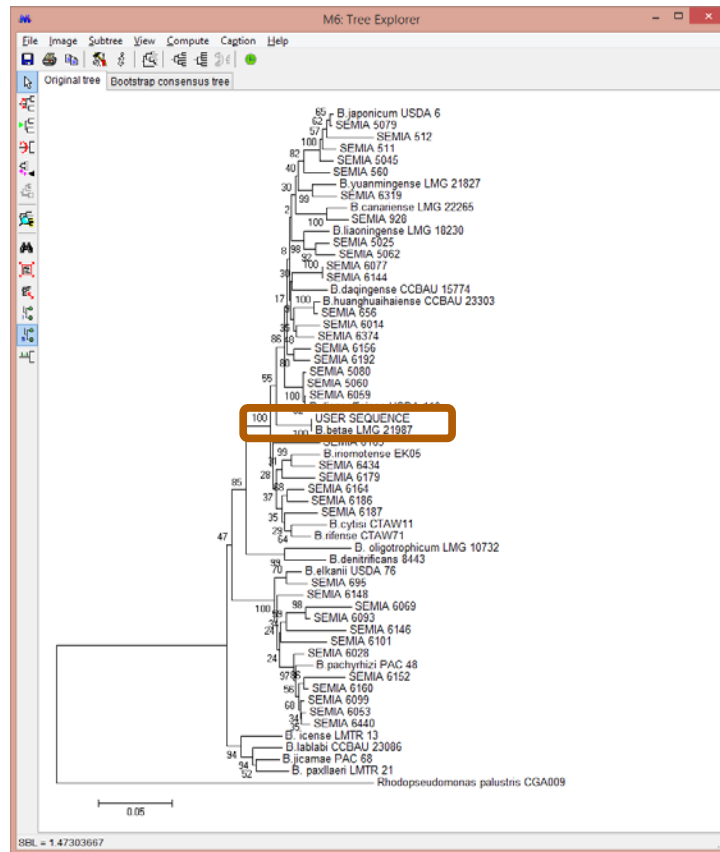4.1.16   Choose a Data File to Analyze: "alignment.meg"

4.1.17   In Options Summary, input 3 parameters different of default analysis preferences: "Test Phylogeny: Bootstrap method", "No. of Bootstrap Replications: 1000" and "Model/Method: Tamura-Nei model". Click in compute.
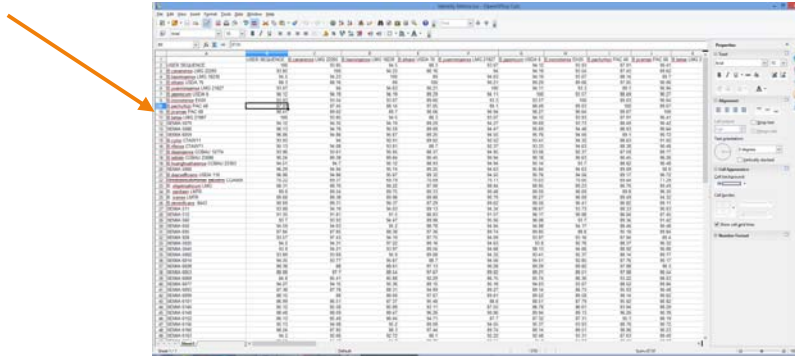
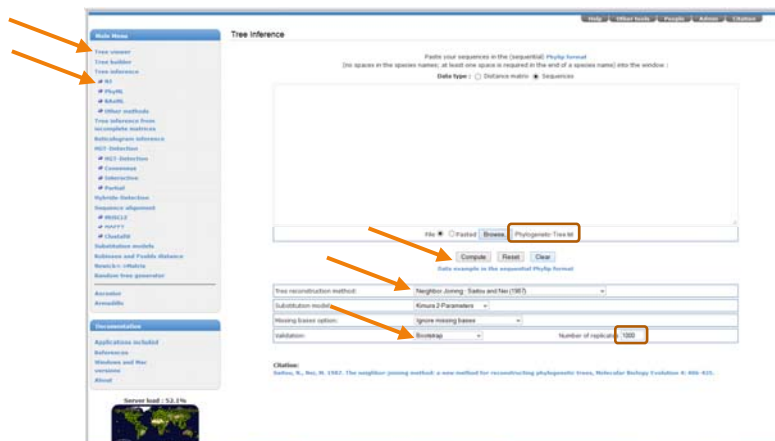### 4.1.18 A screen informs the procedure being performed

4.1.19 Result of analysis:

4.1.20 To make an analysis of Identity-Matrix.csv file, can be used a software Open Calc of Apache OpenOffice software available in: http://www.openoffice.org/

4.1.21 To the phylogenetic analysis of the file "Phylogenetic-Tree.txt" an option can be found in the services available on the website: http://www.trex.uqam.ca/

5 COMMENTS AND SUGGESTIONS

5.1 WE WOULD BE VERY HAPPY WITH YOUR COMMENTS OR SUGGESTIONS.

PLEASE CONTACT US.