ADDITIONAL FILES

Additional file 1 – UserGuideforB2C2.pdf

This text provides step by step instruction for installation and use of the B2C2 software.

USER GUIDE FOR B2C2:

Windows® based chimera checking software

For Windows®

Version:1.0

Research and Testing Laboratory of the South Plains

2010

INTRODUCTION

B2C2 is a stand-alone application designed to be able to evaluate sequences and determine the likelihood of the reads being chimeras. B2C2 is a fast and accurate algorithm for evaluating sequences in FASTA format, particularly, but not limited to, next generation data.

With B2C2, the user is able to specify the stringency of the analysis and customizing the end result. Not only are the user's input files organized into four subcategories: Definite Chimeras, Possible Chimeras, Not Chimeras, and Unclassified, the user is able to remove definite and/or possible chimeras from the original input to result in a Final file to be further processed.

This software provides the ease of use, speed and accuracy necessary in 16S bacterial research, in particular next generation sequencing output.

INSTALLING B2C2

Installing on Windows®

To install B2C2 on Windows, launch **setup.exe or the .msi files** and then follow the on-screen instructions.

By default, **B2C2.exe** will be installed to the following directory depending on the bit version:

**C:\Program Files\ResearchAndTestingLaboratory\B2C2Setup_x64**

or

C:\Program Files\ResearchAndTestingLaboratory\B2C2Setup_x32

The installation will also place a B2C2 icon on your Desktop and in your Start Menu/All Programs tab.
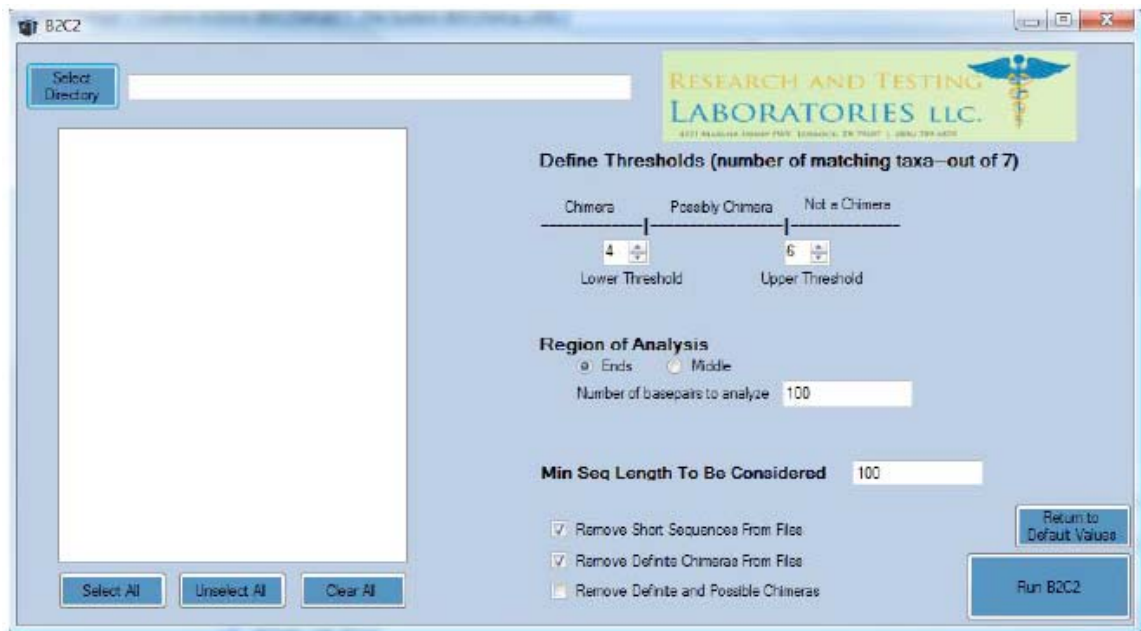
HOW TO USE B2C2

B2C2 is a windows application that is run with a user interface.

To launch the program click on the B2C2 icon on your desktop or in your Start Menu Program Tab. You may also go to **C:\Program Files\ResearchAndTestingLaboratory\B2C2Setup_x--** where the "--" indicates the version appropriate for your system. In this folder you may double click the B2C2.exe icon.
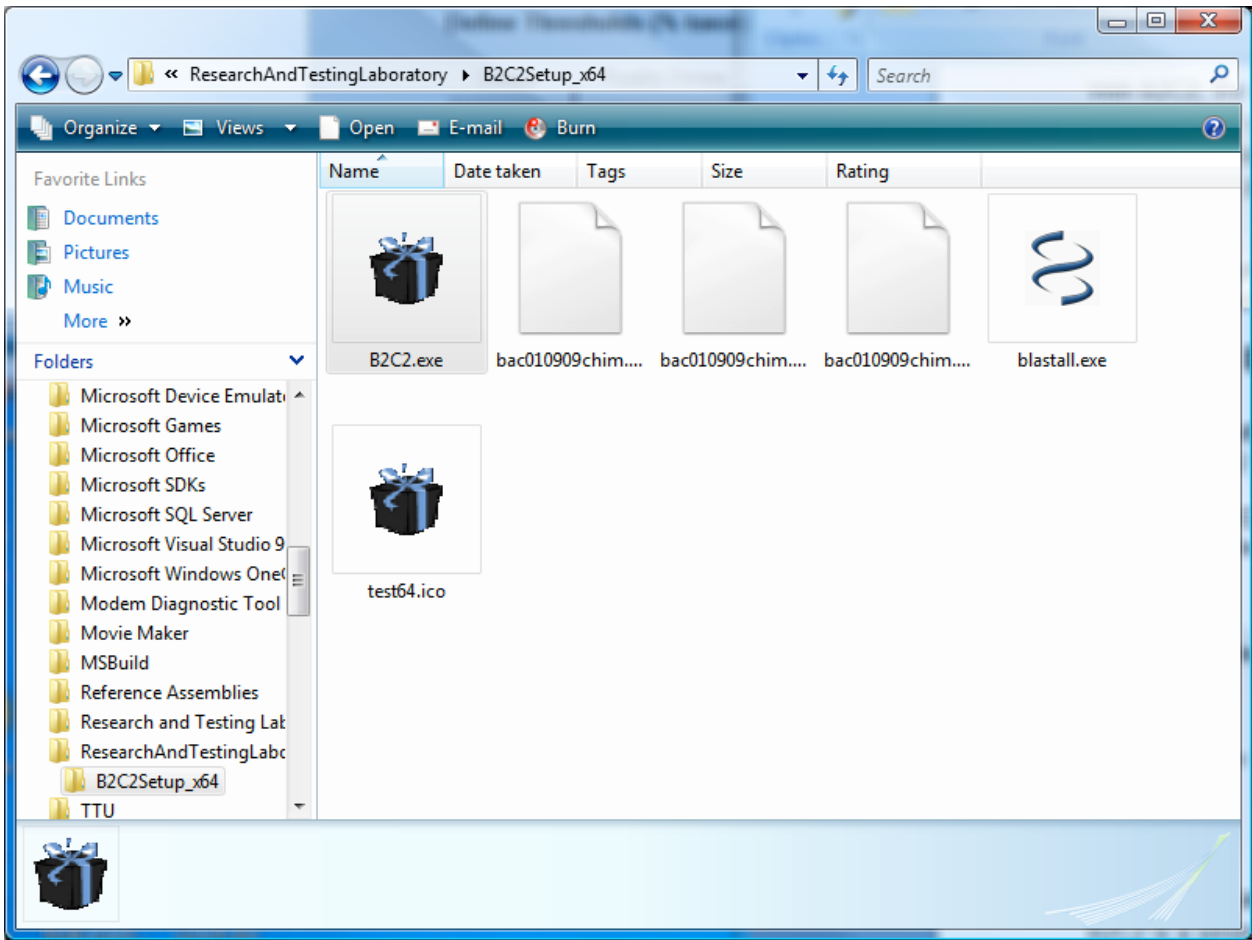
Data Input and Analysis

1. Click the "Select Directory" button to find the folder containing the FASTA formatted files you wish to have chimera checked and click "OK". This action will load files with the extensions "*.fas", "*.fna" or "*.fasta".
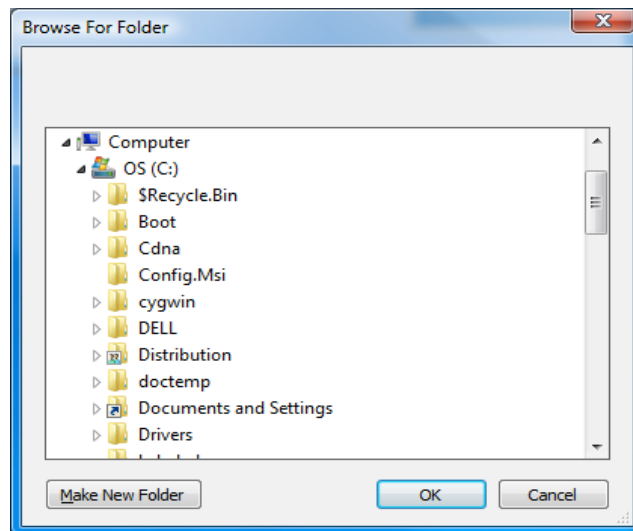
Click on the files you wish to analyze. This action will produce a check placed in the box in front of the file name you selected. Clicking once more on the file name will uncheck the box and the file will not be analyzed by B2C2.
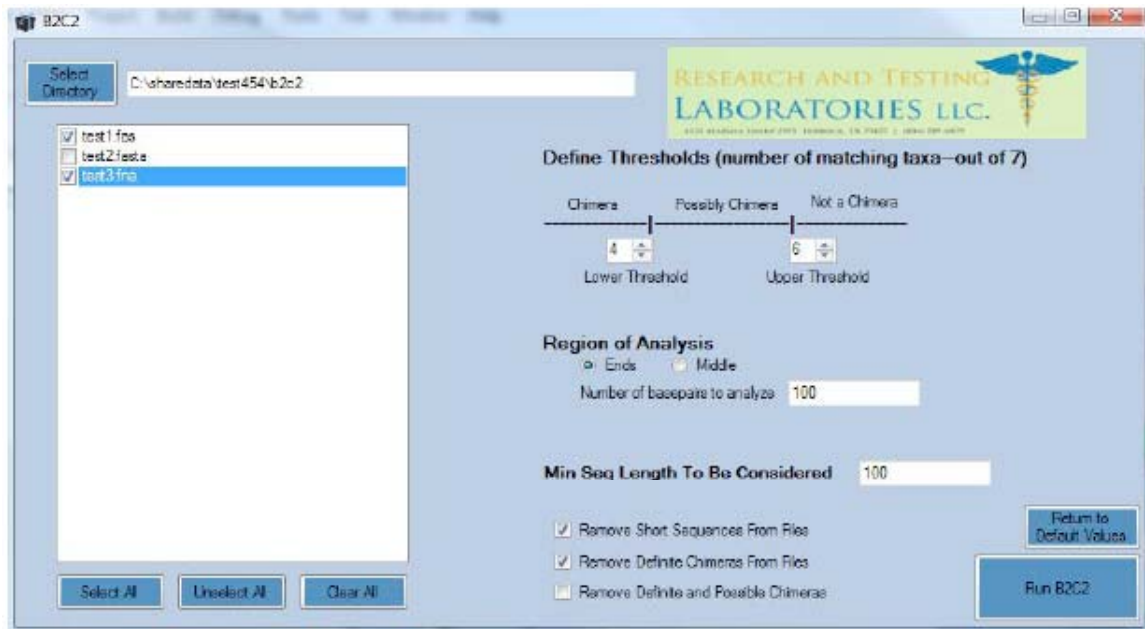
The "Select All", "Unselect All" and "Clear All" buttons may be used to check all files listed in the list box, uncheck all items, or restart the process by reselecting a folder and loading the files into the list box.

2. Define stringency thresholds. The analysis is performed on the seven major taxonomic levels: kingdom, phylum, class, order, family, genus, and species. When the start and end regions of a user's sequence are identified as the same bacteria, 7 of 7 taxonomic levels match, similarly if the regions align to two species of bacteria that fall into the same genus, 6 out of 7 names match.

Sequences whose names match at a lower percentage than the "Lower Threshold" defined, will be categorized as "YourFileNameDefiniteChimera.fas". Those whose percentage identity is between the lower and upper thresholds will be organized into the "YourFileNamePossibleChimera.fas " file and sequences above the "Upper Threshold" will be labeled as "YourFileNameNotChimera.fas ". Users sequences which were not able to be classified, for example if the start region or end region were not able to be identified, will be put into the "YourFileNameNotClassified.fas " file.

3. Select Region of Analysis and the number of basepairs to align. If the "Ends" are selected, the defined number of basepairs from the start and end of each sequence are selected to be aligned against a database of sequences. If the regions don't align to the same bacteria, the sequence in question could be chimeric. Similarly, the middle region could be selected for analysis. In this case, the areas to the left and right of the midpoint of a sequence will be aligned.

4. Select a minimum sequence length to be analyzed. Sequences below this minimum basepairs number will not be analyzed or included in the "Final" output file (select if you want short sequences removed).

If all sequences should be considered in the analysis, the minimum sequence length should be defined as 0 (zero).

5. Select if you would like definite chimeras, or both definite and possible chimeras to be removed from your original input files. The remaining, good quality sequences will be written to a YourFileNameFinal.fas file in the directory where the original file resides.

6. Finally, click the "Run B2C2" button.

Do Not close the black window that will appear during the chimera checking process.

When the process is complete, a notification box will appear.
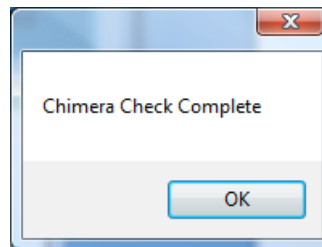
The files containing your results will be ready for your inspection in the folder with your original input file(s).