gmosDraw, version 1.0

Mirjana Domazet-Lošo

March 11, 2015

Table of Contents

1 Overview

gmosDraw is a program for visualization of the genome mosaic structure based on the output of the program **gmos** (1). The expected input to gmosDraw is the standard output of gmos comprising:

(1) a file that contains alignments representing the mosaic structure of a genome in multi-fasta format (.fasta file; automatically generated using –o option)
(2) an accessory file that contains a list of regions representing the mosaic structure of a genome (.txt file generated using –o option)
(3) an accessory file that contains lengths of query and subject sequences (.len file; automatically generated using –o option)

As input to gmosDraw, a user only specifies the name of the .fasta file, while the other two files are expected to be located in the same directory as the .fasta file. All three files should have been generated using the –o option of gmos and assigned the same name, but with different extensions. For example, if the name of the first file is o.fasta, then the other two files are expected to be named o.txt and o.len.
If any of the three files is missing, the program will terminate with an error message.

gmosDraw is written in C# and runs under Windows. It is installed in a standard manner by simply double-clicking the setup file (setup.exe) and following the instructions during the installation process.

Please contact mirjana.domazet-loso@fer.hr in case of any problems with the program.

2 Getting Started

1. Install gmosDraw
   Unpack the zip file gmosDraw.zip and install gmosDraw by double-clicking the setup file setup.exe.
   Follow the instructions during the installation process.
   The icon of the program gmosDraw will be added to the Start menu and the program can be started by clicking the gmosDraw icon.

   In addition, the three test files named o.fasta, o.txt and o.len are located in the directory "Test Data", which is a part of the gmosDraw.zip file, and can be used for test purposes.

2. Uninstall gmosDraw
   gmosDraw can be uninstalled using the standard Windows procedure, e.g. select *Control Panel* → select *Programs* → select *Programs and Features* → select *Uninstall a program* → select gmos from the list of installed programs → click *Uninstall/Change* button.

3. Running gmosDraw
   Select the gmosDraw icon from the Start menu.

3 Program Options

The input to gmosDraw is the output of gmos consisting of three files generated using –o option. These files are:

(1) a file that contains alignments representing the mosaic structure of a genome in multi-fasta format (.fasta file; automatically generated using –o option)
(2) an accessory file that contains a list of regions representing the mosaic structure of a genome (.txt file generated using –o option)
(3) an accessory file that contains lengths of query and subject sequences (.len file; automatically generated using –o option)

All three files are expected to be in the same directory and assigned the same name, but with different extensions. For example, if the name of the first file is o.fasta, then the second file is expected to be named o.txt and the third file o.len.

Please note that the three input files named o.fasta, o.txt and o.len are located in the directory "Test Data", which is a part of the gmosDraw.zip file, and can be used for the test purposes in this section.

If any of the files is missing, the program will terminate with an error message.

The file with .fasta extension, e.g. o.fasta, should be selected using gmosDraw option: *File* → *Open Analysis Results* (Fig. 1) followed by a standard open dialog. The other two files (with extensions .txt and .len) are also expected to be found in the same directory, but the program finds them automatically after the .fasta file has been chosen.

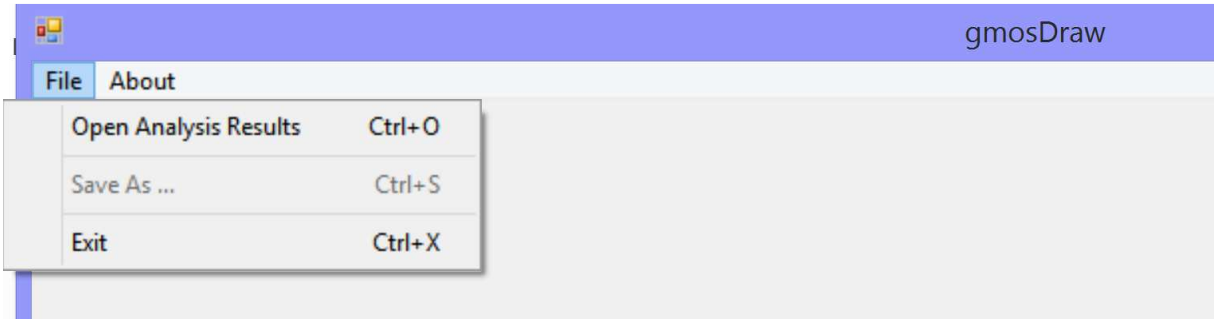

Fig. 1 Open gmos results file

When the .fasta file is opened, the results of gmos analysis are shown in a form: the short textual results are shown in the first tab (*Results of gmos analysis*; Fig. 2) and the graphical representation of the analysis is shown in the second tab (*Graphical results of gmos analysis*; Fig. 3).

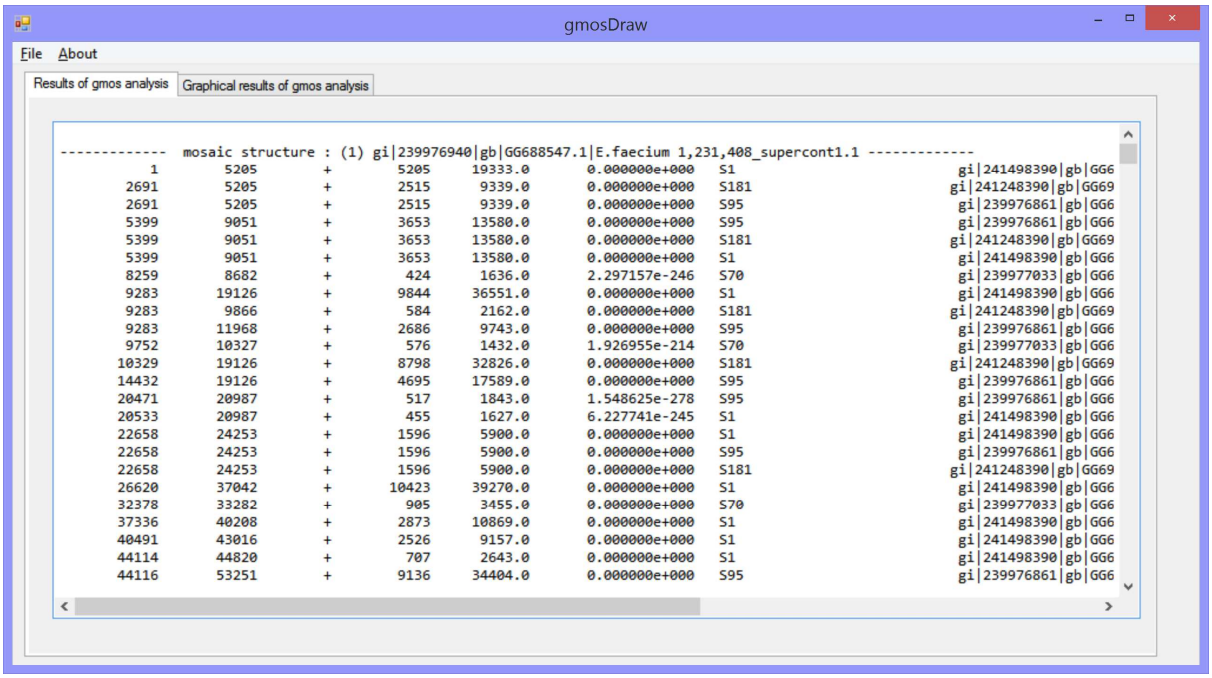

Fig. 2 Results of gmos analysis

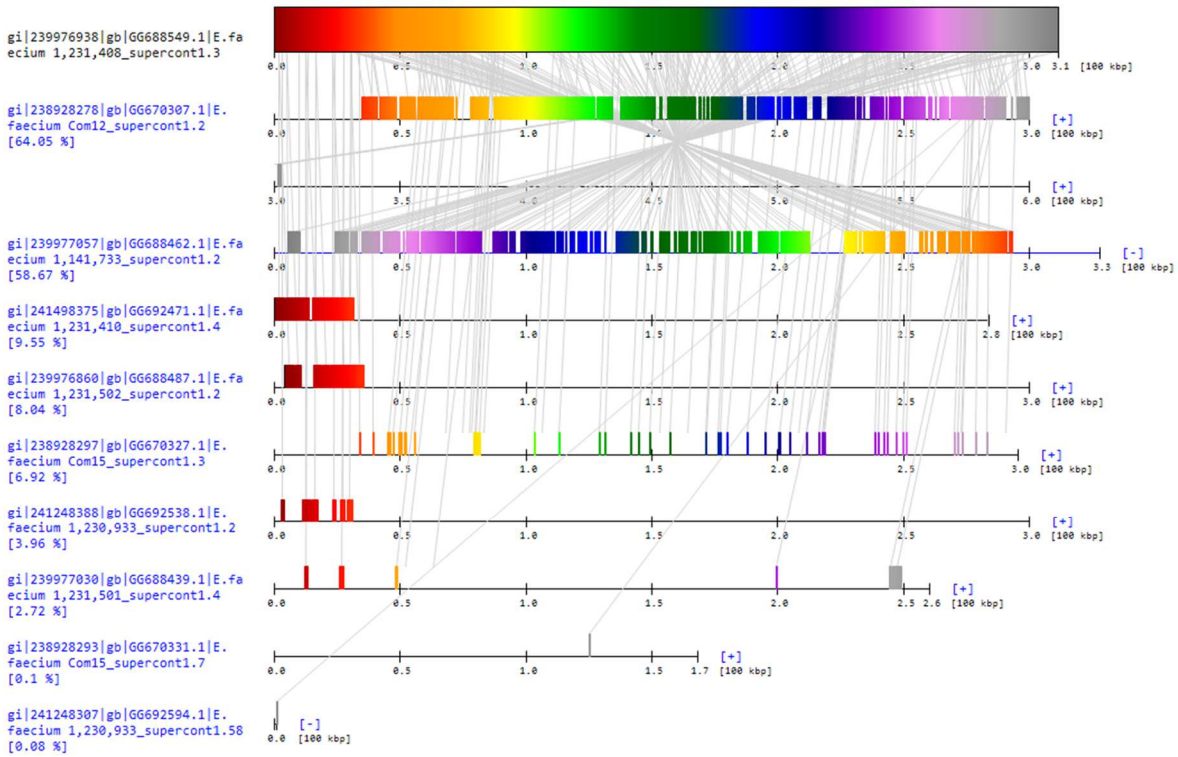

Fig. 3 Graphical results of gmos analysis

In the top row in Fig. 3 a query sequence is represented by its query name (on the left) and a rectangle colored with a range of colors (in Fig. 3 a query sequence is *E. faecium 1,231,408 supercontig 1.3*, whose length is 310 kbp). The query sequence is colored with a range of colors for easier visualization of query regions and the corresponding (most similar) subject regions.

Each query sequence is represented by its forward strand, while the subject sequences are represented by either their forward, reverse or both strands depending on which subject strand or strands are mapped to particular query regions. A subject's strand sign (+/- for the forward/reverse strand) is annotated at the end of a line.

Each subject sequence is annotated by its sequence name and the percentage of the query sequence to which the subject sequence maps. For example, in Fig. 3, a subject sequence *+E. faecium Com12 supercontig 1.2* is most similar to the query along 64.05% of the query's genome. In this case, query regions are mapped to the subject's forward strand. The next most similar subject sequence is *E. faecium 1,141,733 supercontig 1.2* (its reverse strand), which has been mapped to 58.67% of the query sequence.

At the top of the tab *Graphical results of gmos analysis* is a combo box comprising a list of query sequences. By default, the selected item is the first query in the list.

When a query name is selected in the combo box, the graphical representation of the gmos analysis of the query is shown below the combo box. In Fig. 3, the selected query is *E. faecium 1,231,408 supercontig 1.3*.

To the right of the combo box is a text field, *Subject threshold [%]*, which denotes the number (percentage) from the interval [0, 100]. For example, if the *Subject threshold* is set to 10%, then only those subjects, which are mapped to at least 10% of the query genome, are shown in an image below. In Fig. 3, all subjects mapped to some query regions are shown, since the *Subject threshold* is set to its default value: 0%.

The graphical representation of a query mosaic structure can be saved in different image formats (JPG, BMP, GIF, PNG) by selecting the menu item *File → Save As …* and then selecting the image format and the file name. This option is enabled only when the selected tab is *Graphical results of gmos analysis*.

Finally, by mouse-clicking a subject region mapped to a query region of the same color range, a simple text editor is opened and the corresponding local alignment in fasta format is shown. A query and a subject name are shown together with the nucleotide positions stated in the square brackets and the sign +/- representing the forward/reverse strand (Fig. 4). The local alignment can be saved in a fasta file using the menu item *File → Save As* of the text editor.

Fig. 4 Local sequence alignment between a query segment (*gi|239976940|gb|GG688547.1|E.faecium 1,231,408_supercont1.1* from 209657 to 219112 bp on the forward strand) and a subject segment *E. faecium 1,231,410_supercont1.1* from 226896 to 205417 bp on the reverse strand.

4 References

1. Domazet-Lošo M. Rapid detection of genome mosaicism over short evolutionary distances (submitted). 2015;