User manual,

Java assignment

MSc Applied Bioinformatics,

Marie Schmit

1- Set-up and run the program

Open the project in a Java IDE. Open the files stored in source packages. Clean and build the program, then click on the "run" icon.

If you do not have any IDE installed, open a command prompt window. Go in the repertory where the project is saved and write in the terminal "javac mainFrame.java" to compile the code, then "java mainframe.java" to run it.

2- Functionalities of the program

The first window allows you to choose a file to analyse, either in gtf (.gtf) or in fasta (.fa) format. You can write the path of the file, or its name, in the search bar and chose "browse file" to search it, or directly open the file browser by clicking "browse file" button. If the format of the chosen file is not fasta or gtf, a message error appears.
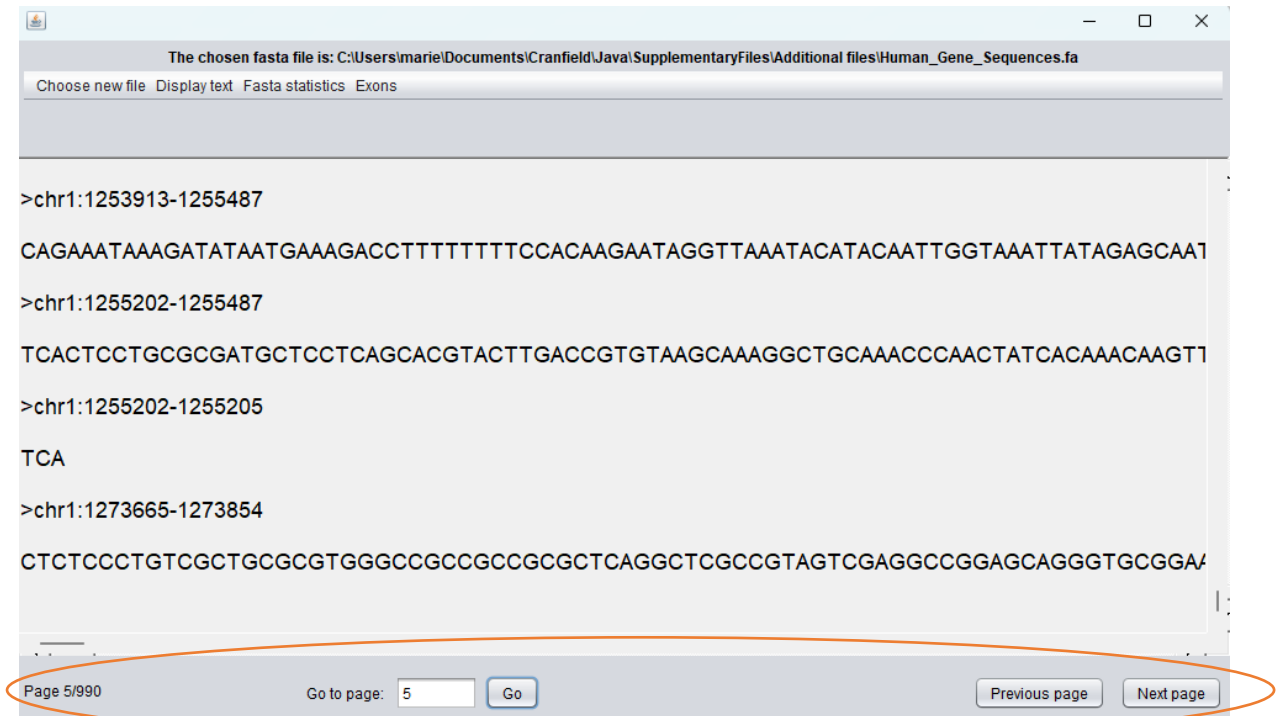

Once the file is chosen, click "Confirm" to continue. Please note that large files take time to process.

The next page displays a menu, depending on the type of file previously chosen. Click again on "Choose new file" to go back to the previous step.

Fasta files

Display Text

This command displays the sequence(s) contained in the file. If the file is large, it's text can be displayed on multiple pages. Use the buttons at the bottom to choose a particular page (Go to page) or navigate between pages (next, previous).

>chr1:1253913-1255487

CAGAAATAAAGATATAATGAAAGACCTTTTTTTTTTCCACAAGAATAGGTTAAATACATACAATTGGTAAATTATAGAGCAAT

>chr1:1255202-1255487

TCACTCCTGCGCGATGCTCCTCAGCACGTACTTGACCGTGTAAGCAAAGGCTGCAAACCCAACTATCACAAACAAGTT

>chr1:1255202-1255205

TCA

>chr1:1273665-1273854

CTCTCCCTGTCGCTGCGCGTGGGCCGCCGCCGCGCTCAGGCTCGCCGTAGTCGAGGCCGGAGCAGGGTGCGGAA

Page 5/990   Go to page: 5   [Go]   [Previous page]   [Next page]

Fasta statistics

This item displays the length of the sequence and its GC content value (or their average values for multiple sequences files). All option displays both statistics.

Exons

For graphical or textual display of exons, a new window opens, inviting you to choose a second file. This should be the complementary gtf file of the fasta, containing information on the same gene.
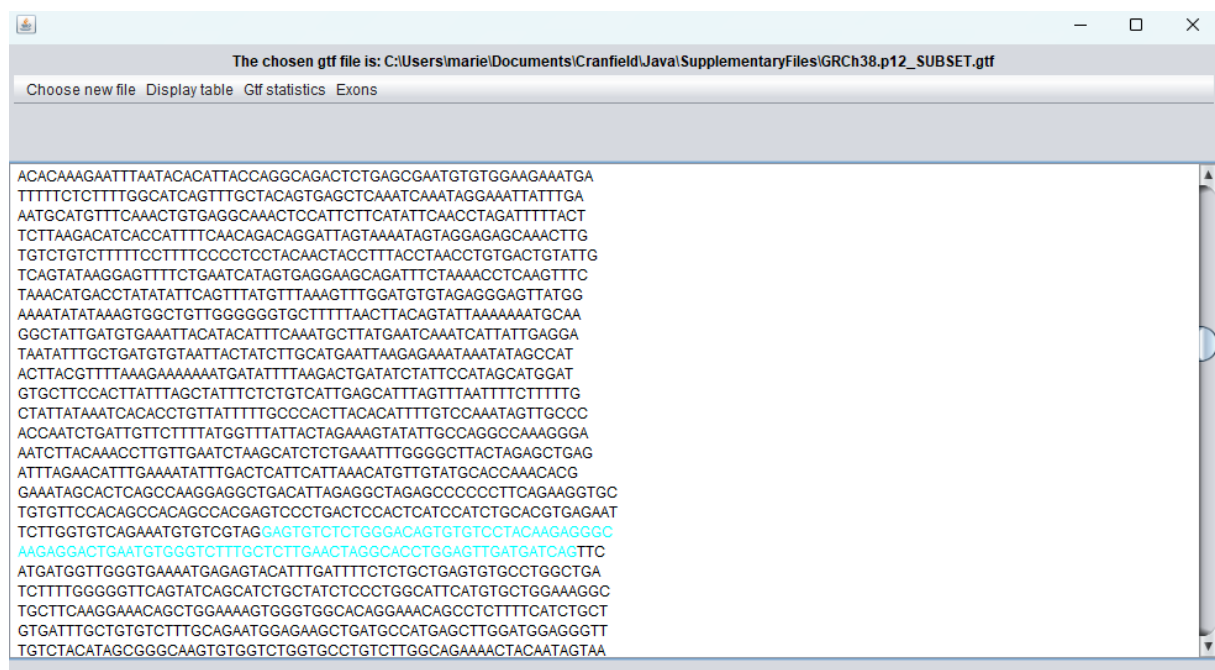
ACACAAAGAATTTAATACACATTACCAGGCAGACTCTGAGCGAATGTGTGGAAGAAATGA
TTTTTCTCTTTTGGCATCAGTTTGCTACAGTGAGCTCAAATCAAATAGGAAATTATTTGA
AATGCATGTTTCAAACTGTGAGGCAAACTCCATTCTTCATATTCAACCTAGATTTTTACT
TCTTAAGACATCACCATTTTCAACAGACAGGATTAGTAAAATAGTAGGAGAGCAAACTTG
TGTCTGTCTTTTTCCTTTTCCCCTCCTACAACTACCTTTACCTAACCTGTGACTGTATTG
TCAGTATAAGGAGTTTTCTGAATCATAGTGAGGAAGCAGATTTCTAAAACCTCAAGTTTC
TAAACATGACCTATATATTCAGTTTATGTTTAAAGTTTGGATGTGTAGAGGGAGTTATGG
AAAATATATAAAGTGGCTGTTGGGGGGTGCTTTTTAACTTACAGTATTAAAAAAATGCAA
GGCTATTGATGTGAAATTACATACATTTCAAATGCTTATGAATCAAATCATTATTGAGGA
TAATATTTGCTGATGTGTAATTACTATCTTGCATGAATTAAGAGAAATAAATATAGCCAT
ACTTACGTTTTAAAGAAAAAATGATATTTTAAGACTGATATCTATTCCATAGCATGGAT
GTGCTTCCACTTATTTAGCTATTTCTCTGTCATTGAGCATTTAGTTTAATTTTCTTTTTG
CTATTATAAATCACACCTGTTATTTTTGCCCACTTACACATTTTGTCCAAATAGTTGCCC
ACCAATCTGATTGTTCTTTTATGGTTTATTACTAGAAAGTATATTGCCAGGCCAAAGGGA
AATCTTACAAACCTTGTTGAATCTAAGCATCTCTGAAATTTGGGGCTTACTAGAGCTGAG
ATTTAGAACATTTGAAAATATTTGACTCATTCATTAAACATGTTGTATGCACCAAACACG
GAAATAGCACTCAGCCAAGGAGGCTGACATTAGAGGCTAGAGCCCCCCCTTCAGAAGGTGC
TGTGTTCCACAGCCACAGCCACGAGTCCCTGACTCCACTCATCCATCTGCACGTGAGAAT
TCTTGGTGTCAGAAATGTGTCGTAGGAGTGTCTCTGGGACAGTGTGTCCTACAAGAGGGC
AAGAGGACTGAATGTGGGTCTTTGCTCTTGAACTAGGCACCTGGAGTTGATGATCAGTTC
ATGATGGTTGGGTGAAAATGAGAGTACATTTGATTTTCTCTGCTGAGTGTGCCTGGCTGA
TCTTTTGGGGGTTCAGTATCAGCATCTGCTATCTCCCTGGCATTCATGTGCTGGAAAGGC
TGCTTCAAGGAAACAGCTGGAAAGTGGGTGGCACAGGAAACAGCCTCTTTTCATCTGCT
GTGATTTGCTGTGTCTTTGCAGAATGGAGAAGCTGATGCCATGAGCTTGGATGGAGGGTT
TGTCTACATAGCGGGCAAGTGTGGTCTGGTGCCTGTCTTGGCAGAAAACTACAATAGTAA

Figure 1 Exons textual display

The textual display shows sequence or multiple sequences of the fasta file, with blue highlight on exons indicated in the gtf file.

Graphical representation shows exons of a single sequence from fasta file, indicated by gtf file. Please note that this option is only available for single sequence. If the sequence is too long to be displayed on the width of the page, it is drawn on the next line.
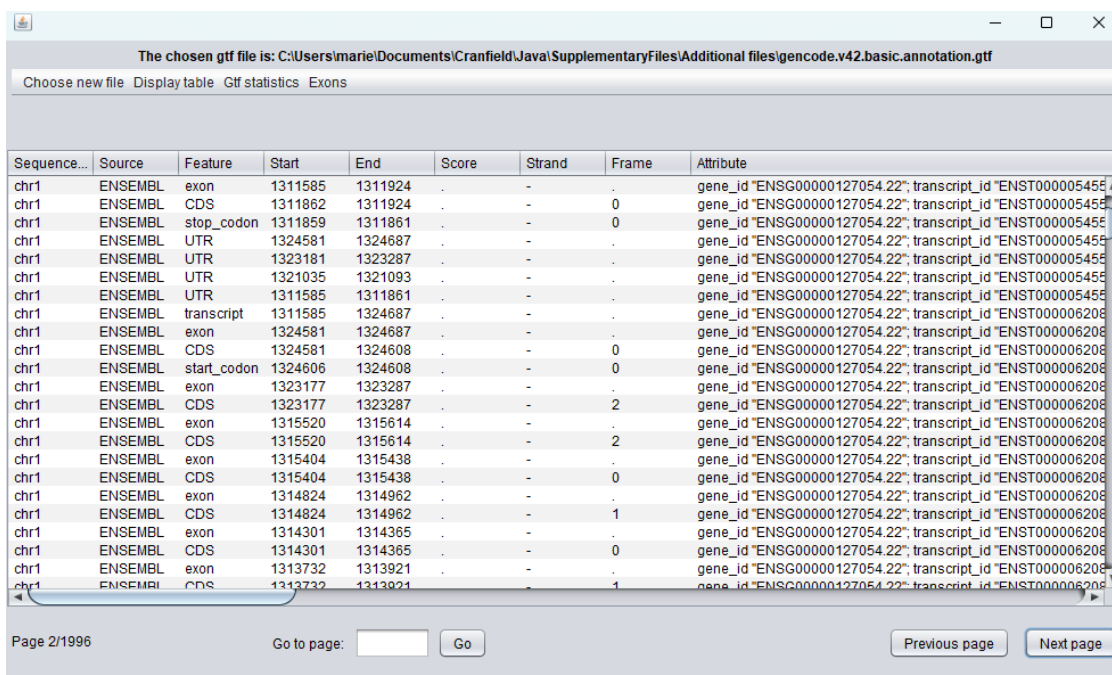

Figure 2 Exons graphical display, GRch38

Gtf file

Display table
Display table option for gtf files is like display text of fasta. A page navigator is also available for large files.

Figure 3 Display table with page navigation

Gtf statistics

Similarly to fasta, those statistics (average number of exons, shortest and longest genes, average length gene), are displayed on the text area.

The chosen gtf file is: C:\Users\marie\Documents\Cranfield\Java\SupplementaryFiles\Additional files\gencode.v42.basic.annotation.gtf

Choose new file Display table Gtf statistics Exons

The average number of exons for this file is: 13.593052188337374.

The shortest gene is: ENSG00000223997.1 with a length of: 7
The longest gene is: ENSG00000078328.23 with a length of: 2473538

The average length gene is: 1016.0
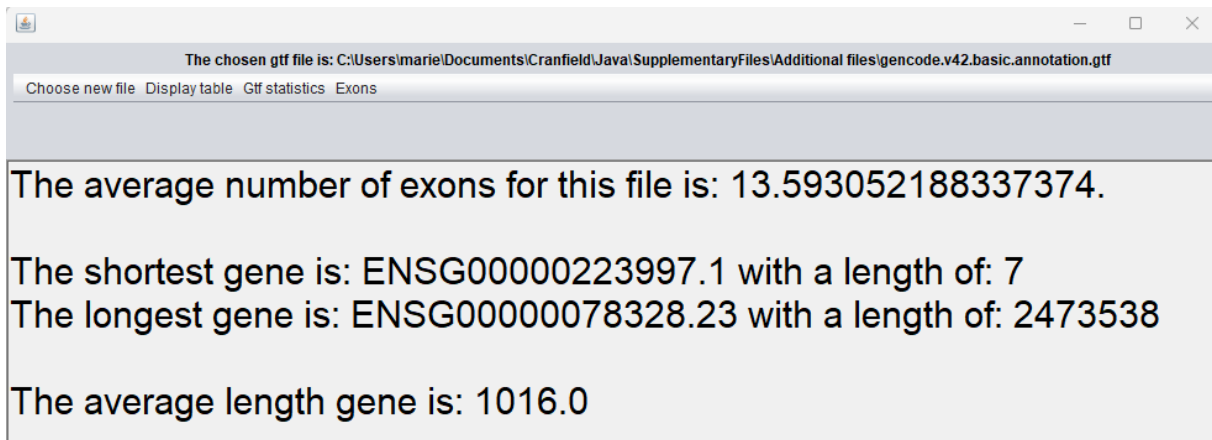
Figure 4 gtf statistics, all statistics

Exons

Both exons panels works like fasta's. The second file chosen should be of fasta format.