



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.



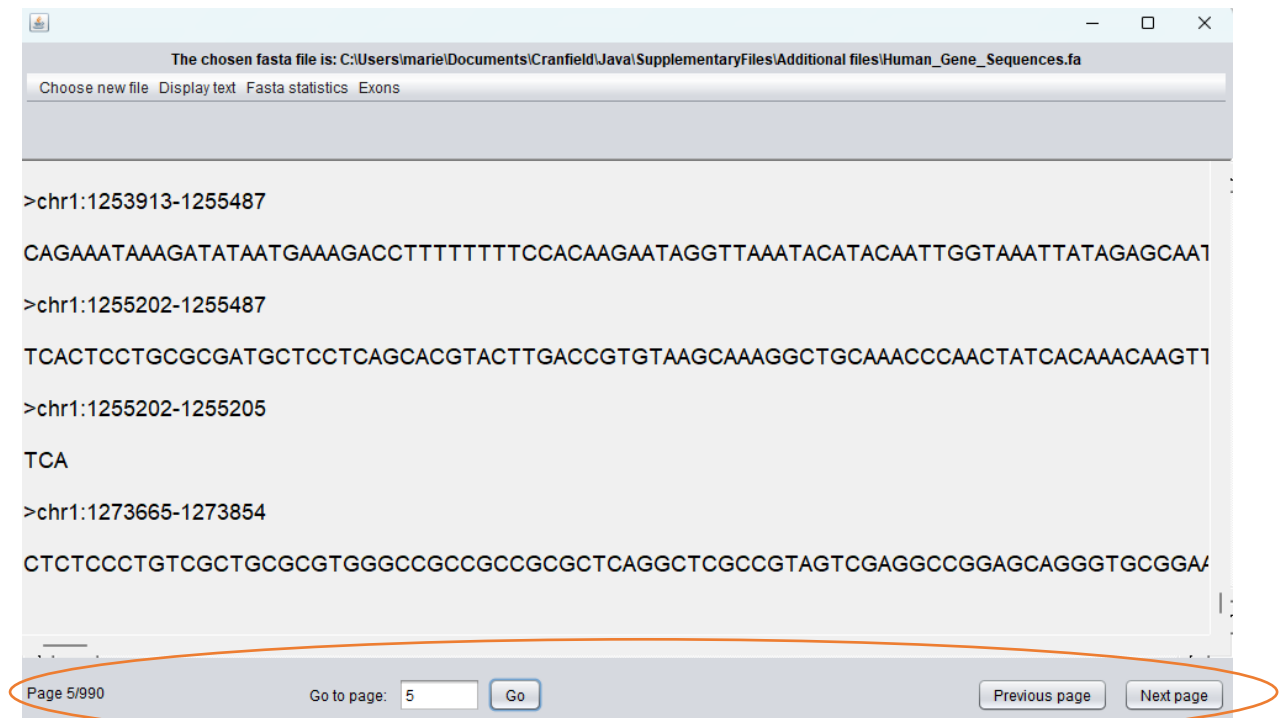
Once the file is chosen, click “Confirm” to continue. Please note that large files take time to process.

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).



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.

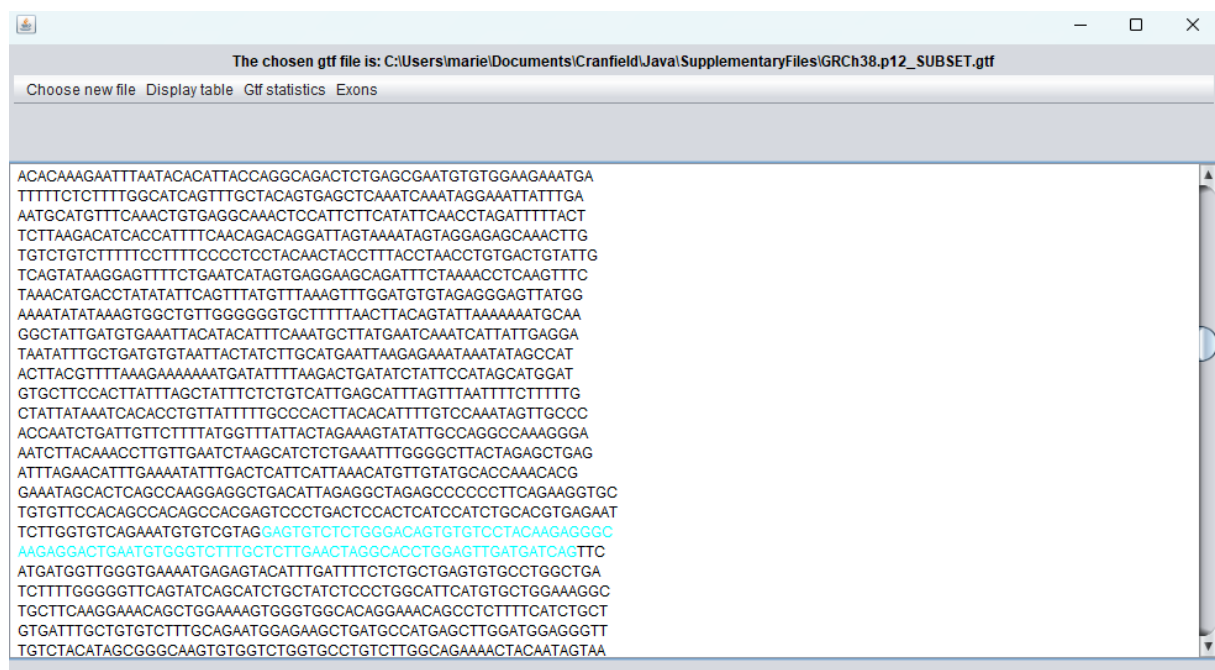


Figure 1 Exons textual display

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

Please pay attention to your terminal if graphic is not displayed: you might have uploaded an empty file or a multiple sequence fasta file. Message error will in that case be displayed on terminal.

[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.

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

Sequence...	Source	Feature	Start	End	Score	Strand	Frame	Attribute
chr1	ENSEMBL	exon	1311585	1311924	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	CDS	1311862	1311924	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	stop_codon	1311859	1311861	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	UTR	1324581	1324687	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	UTR	1323181	1323287	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	UTR	1321035	1321093	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	UTR	1311585	1311861	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000005455"
chr1	ENSEMBL	transcript	1311585	1324687	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1324581	1324687	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1324581	1324608	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	start_codon	1324606	1324608	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1323177	1323287	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1323177	1323287	.	-	2	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1315520	1315614	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1315520	1315614	.	-	2	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1315404	1315438	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1315404	1315438	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1314824	1314962	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1314824	1314962	.	-	1	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1314301	1314365	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1314301	1314365	.	-	0	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	exon	1313732	1313921	.	-	.	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"
chr1	ENSEMBL	CDS	1313732	1313921	.	-	1	gene_id "ENSG00000127054.22"; transcript_id "ENST0000006208"

Page 2/1996 Go to page: Go Previous page Next page

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.