

Contents

1 Installation instructions	3
1.1 System requirements	3
1.2 Launching the program	3
2 Start Up User interface	4
3 Comparative Genomics Start Screen	5
3.1 Performing Alignment within MapOptics	6
3.1.1 Connecting to a server	7
3.1.2 Uploading Reference Genome Files	8
3.1.3 Uploading Query Genome Files	8
3.1.4 Choosing Digestion Enzyme	8
3.2 Uploading completed alignment to MapOptics	9
4 Viewing results in Comparative Genomics View	10
4.1 Navigation	10
4.2 Genome View	10
4.3 Chromosome View	11
4.4 Query View	12
4.4.1 Changing query range	13
	13
4.4 Downloading results	13
4.4.1 Downloading Images	13
4.4.2 Downloading Tables	14
4.4.3 Downloading Graphs	14
4.4 Changing MapOptics Aesthetics	14
4.5.1 Changing Graph Styles	14
4.5.2 Changing Colours	14
5 File Formats	16
5.1 fasta	16
5.2 Cmap	16
5.3 Xmap	16
5.4 Annotation	16
5.5 Karyotype	16
6 Server set-up	17
6.1 Hardware requirements	17
6.2 Software requirements	17
6.3 Working directory	17

1 Installation instructions

Download source code from <https://github.com/franpeters/MapOptics>

1.1 System requirements

1. Requires the latest version of Java to be installed.
2. Access to a Linux server with run_job.sh pre-installed (see Server Handling for more information) with all of run_job.sh dependencies (See run_job.sh user guide for further information). There must be enough memory on the server to store the uploaded files and results.

This program has been tested on Windows and MacOS.

1.2 Launching the program

Click on the “MapOptics.jar” file to execute the program.

2 Start Up User interface

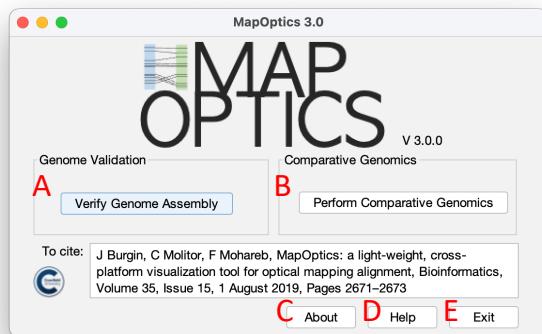


Figure 1 Start Up Screen of MapOptics

On start-up of MapOptics (Figure 1), a window will launch whereby you can choose to launch the ‘Verify Genome Assembly’ pipeline by clicking button **A**. To launch the ‘Perform Comparative Genomics’ pipeline click button **B**. Button **C** will show information about program and button **D** will launch a window showing this help document. Pressing button **E** will exit the program.

3 Comparative Genomics Start Screen

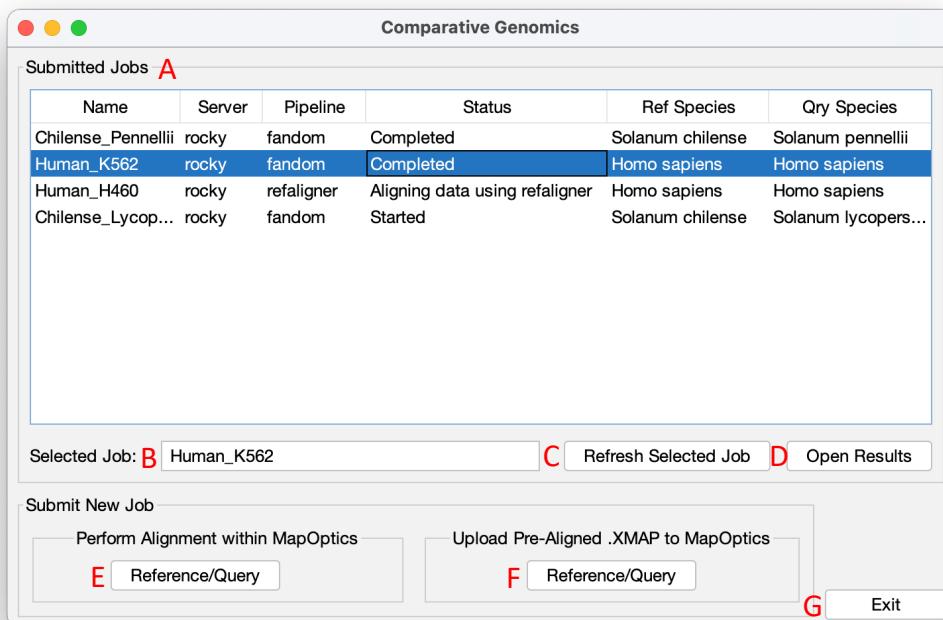


Figure 2 Comparative Genomics Start Window

Upon launching the comparative genomics pipeline a window showing will launch (Figure 2), with all previously submitted jobs to MapOptics shown in table **A**. By clicking on a row in the table the job name will appear in text field **B**.

To query the progress of the selected job, click on button **C** and the log.txt file corresponding to that job will be queried and the latest update will appear in table **A**.

To open the results into the Comparative Genomics View window (see section 4) click on button **D**.

To launch the window to begin performing alignment of two genomes within MapOptics click on button **E**, otherwise choose option **F** to upload files relating to a completed alignment of two genomes, see section 5 for information on required files and accepted file formats.

3.1 Performing Alignment within MapOptics

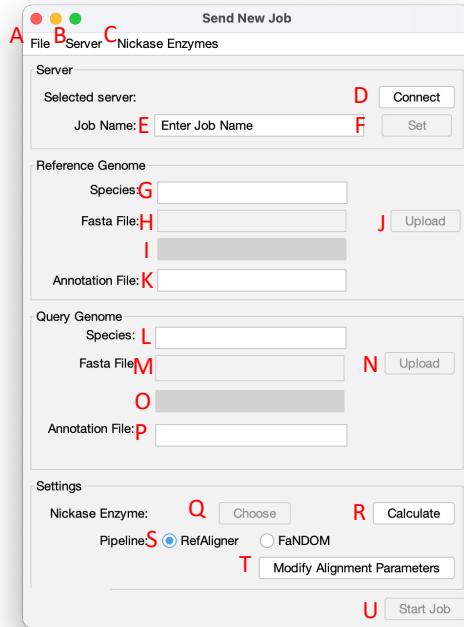


Figure 3 Window to submit a new job for alignment within MapOptics

To send a new alignment job six files are required alongside a job name, chosen digestion enzyme and pipeline. The send new job window (Figure 3) will be used to upload these files and submit the new job to the server. To upload files to a server, MapOptics must be connected to a sever. This is achieved using button **D** (See 3.1.1 for more detail). The button will change to green if a connection is established successfully (Figure 3).



Figure 4 Screenshot showing successful connection to an external server

Next a Job name must be selected. Note: a job name must not contain any numbers as the first character nor any spaces. The buttons **J**, **N**, **Q** and **U** (Figure 4) will now become enabled and uploading the required files to the server can begin. See 3.1.2 and 3.1.3 for information of uploading reference and query files.

Next selection of restriction enzyme can occur, either using an enzyme chosen from a list of supported enzymes (button **C**, see 3.1.4) or an enzyme that has been analysed by MapOptics (button **R**, see 3.1.5). Note: if using experimental cmap data for your query genome the chosen enzyme must match the enzyme used to generate this data else the alignment will not run.

Once the digestion enzyme has been selected (See section 3.1.4 for more information) the alignment algorithm must be selected. MapOptics currently supports two algorithms, RefAligner and FaNDOM. Using the radio buttons (**S**) either one can be selected. The send new job will now resemble what is shown in Figure 5.

Once the form has been completed, the alignment job can be executed on the server by clicking button **U**. This window will close, and you will return to the previous window (Figure 2).

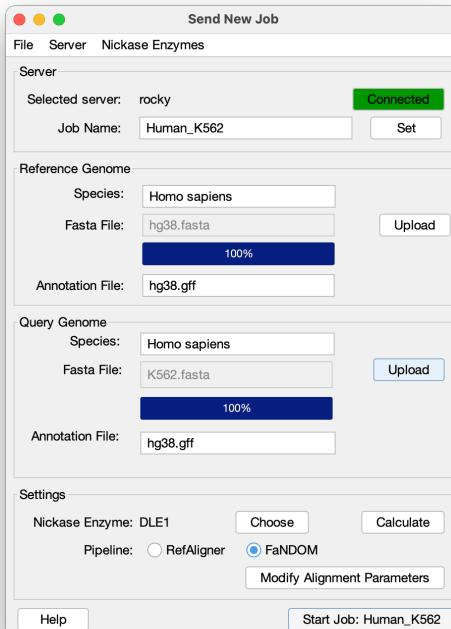


Figure 5 Send new job window once all files have been uploaded and pipeline choices made

3.1.1 Connecting to a server

Before any files are submitted MapOptics must connect onto a Linux server. Button **B** on the menu bar enables the user to launch further menu bar items show in Figure 6.

Server info provides information about the currently connected server if there is one. If not, the window in Figure 7 will display:

For further information on the set-up requirements of the server see section 6.

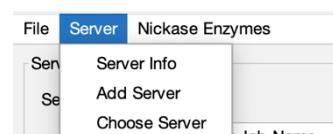


Figure 6 Menu bar buttons relating to external server connection to MapOptics



Figure 7 Error message if server info is requested when there is no server selected

3.1.2 Uploading Reference Genome Files

MapOptics supports the uploading of reference fasta files that are both local and via a URL and can be performed using the buttons shown in Figure 9. Further information on the required format of the fasta file can be found in section 5.1.

Currently only a local version of the annotation file is accepted, with further information on the accepted formats of the annotation file found in section 5.4.

3.1.3 Uploading Query Genome Files

MapOptics supports the uploading of query cmap or fasta files that are both local and via a URL. Further information on the required format of the fasta file can be found in section 5.1 and 5.2.

Currently only a local version of the annotation file is accepted, with further information on the accepted formats of the annotation file found in section 5.4.

3.1.4 Choosing Digestion Enzyme

MapOptics provides the means to query which enzymes are supported and available to the user by clicking on ‘Browse Enzymes’ (Figure 10). Please note MapOptics must be connected to a server for supported enzymes to be accessed.

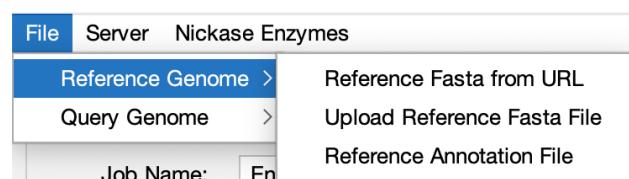


Figure 8 Options available to the user when uploading the reference genome files

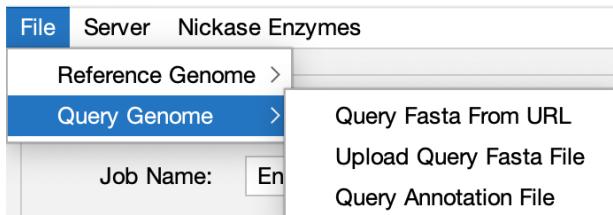


Figure 9 Options available to the user when uploading the query genome files

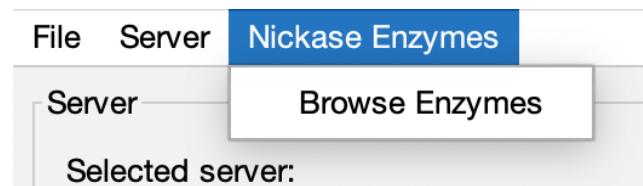


Figure 10 Browsing MapOptics supported enzymes

Enzyme name	Restriction Site
BspQ1	^GCTCTTC
BssS1	^CACGAG
DLE1	^CTTAAG
BseC1	^ATCGAT
test	^GCAT
test	^GCAT
new	^GACT

The supported enzymes are then viewed in a separate window for the user to select (Figure 11).

Figure 11 Results of querying which enzymes are supported by MapOptics

3.2 Uploading completed alignment to MapOptics

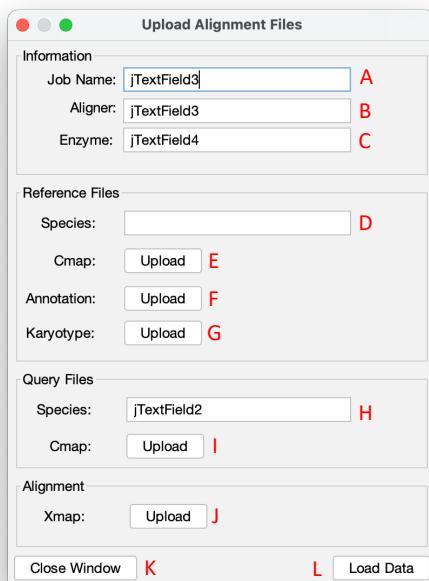


Figure 12 Window to upload files for a completed genomic alignment

This pipeline requires no external server to use MapOptics as a tool to compare two genomes where the alignment xmap file has already been produced. A window to upload the required files and information is launched (Figure 12) when this pipeline is chosen.

Some information about the job uploaded is required, including a name (**A**) (Here no rules are applied as they were for job name in 3.1), alignment algorithm (**B**) and enzyme used (**C**).

Next the files and information for the reference genome are required, including the species (**D**), cmap (**E**), annotation (See 5.4 for supported formats) (**F**) and the karyotype file (See 5.5 for required format) (**G**).

The query genome only requires a species name (**H**) and cmap (**I**) file. Finally, the alignment xmap file (**J**) is to be uploaded before the data can be loaded into Comparative Genomics View (**L**).

4 Viewing results in Comparative Genomics View

Whilst the data is being downloaded and parsed into MapOptics, a dialog will display (Figure 13). Once the downloading and parsing of data is completed, this dialog will disappear. Typical parsing time for a human v human alignment is ~1 min on an Intel i5 core 8GB laptop. Information on downloading the views and tables generated can be found in section 4.4.



Figure 13 Dialog displayed whilst data is parsed

4.1 Navigation

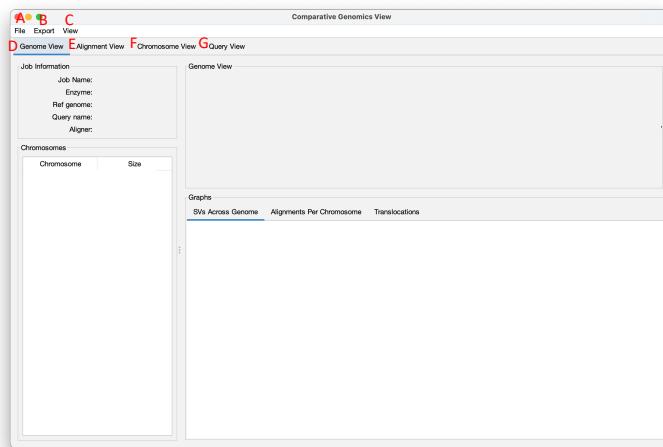


Figure 14 Screenshot of the comparative genomics view window

Figure 14 shows how navigation between the different views is achieved by clicking on the different tabs, **D** (genome view), **E** (Alignment view), **F** (Chromosome view) and **G** (Query View).

4.2 Genome View

The first tab of comparative genomics view shows genome view, alongside information about the job, a table containing all the chromosomes and graphs. There is also a tab for showing any translocation events that have occurred across the genome.

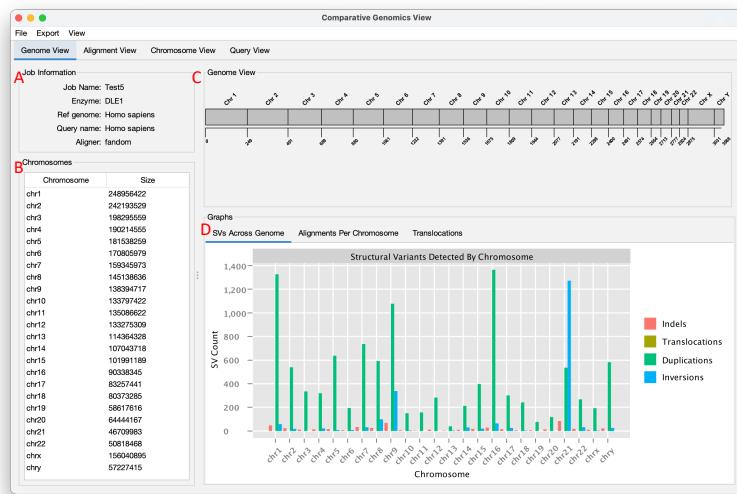


Figure 15 Genome view of comparative genomics view

Genome view (Figure 15) is divided into four main sections, **A** which lists information about the job the data was generated from, **B** contains a table with all the chromosomes contained within the reference genome, **C** which is a graphical representation of all the chromosomes within the reference genome and **D** which is a tabbed pane containing graphs of structural variants across the genome, number of alignments per chromosome and also a graphic of all the translocation events detected.

The chromosomes in the chromosome table (**C**) can be clicked on, which will change the selected tab to ‘chromosome view’. The chromosome being viewed in query view will also be set to this selected chromosome. Further information about chromosome view and query view are in sections 4.3 and 4.4 respectively.

4.3 Chromosome View

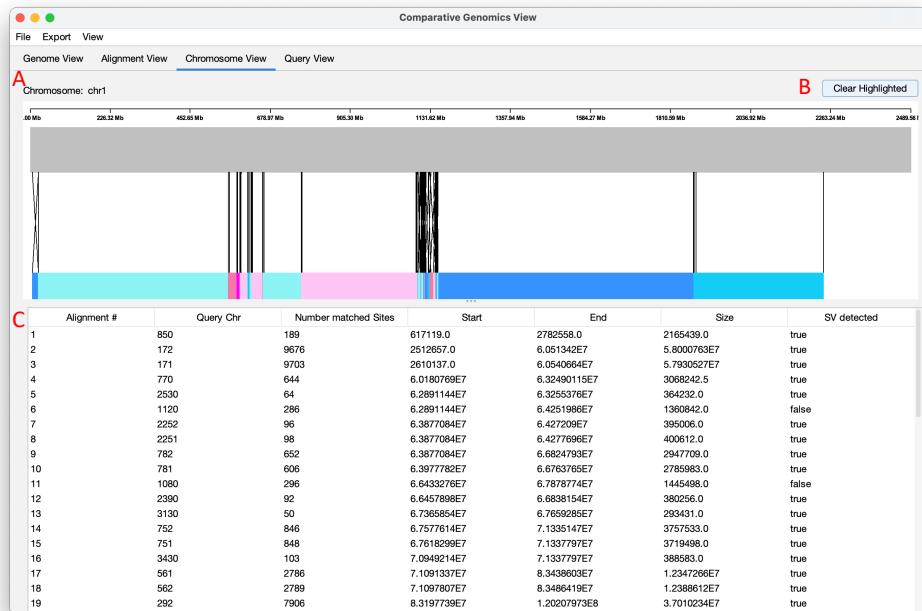


Figure 16 Chromosome view of comparative genomics view

Chromosome view (Figure 16) shows all the different alignments across the whole chromosome in panel **A**. The chromosome is represented by the grey rectangle and a scale bar is drawn at the top of the panel for reference. Each alternating alignment is assigned a different colour to distinguish between different adjacent alignments.

An alignment which has a '-' alignment will be indicated by the two alignment lines forming a 'X' shape. Those which have a '+' alignment will be drawn perpendicularly to the chromosome rectangle. Using the table **C**, a user can click on an alignment to highlight its position. Finally, button **B** can be used to clear any previous highlighting.

4.4 Query View

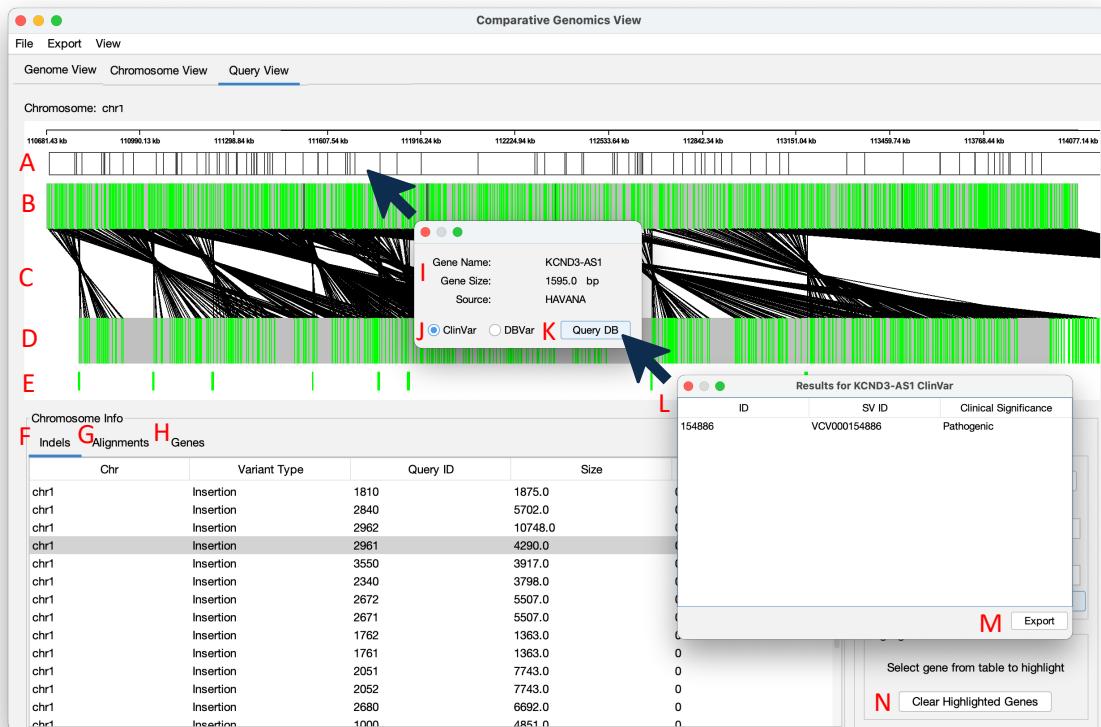


Figure 17 The query view is interactive, allowing the user to click on genes and query databases

Query view (Figure 17) allows the user to zoom in on alignments within a query range given by the user. Information on how to change the range can be found in section 4.4.1.

The user is provided with an annotated view of the alignment to the chromosome of choice, which must be selected in genome view (Figure 15) prior to using this view. Track **A** shows locations of the annotations provided in the annotation file.

The reference chromosome is denoted by a grey rectangle (**B**), with sites that have been aligned to the reference drawn as a green line and those which have not been aligned a black line. Black lines are drawn to match query sites to reference sites (**C**).

The query (**D**) is drawn as another grey rectangle, again with matching sites in green and non-matched sites in black. Finally, any identified insertions or deletions (**E**) are

drawn as green or red rectangles for the region they span for insertion or deletion events of more than 500bp.

Additional information is provided in the tabs **F**, **G** and **H**, where tables containing the identified indels, which can be clicked on to populate the query range with where the indel is located.

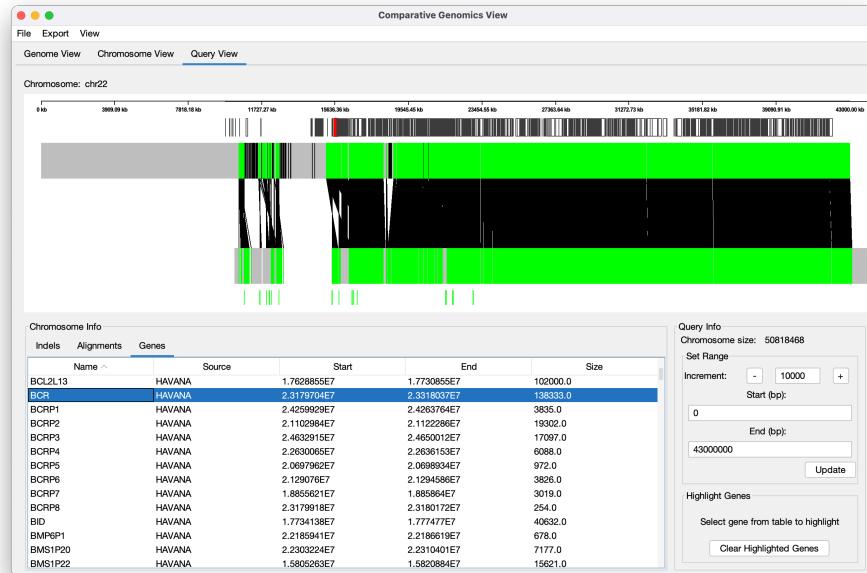


Figure 18 Illustrating highlighting gene

There is also a graph showing the number of alignments per query cmap id and genes within the query, which can be clicked on to highlight their position, as shown in Figure 18.

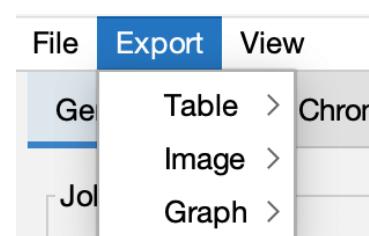
4.4.1 Changing query range

The user has full control over the region of the chromosome being viewed and can change it using the panel shown in Figure 19. The start and end range can be set by changing the values in the text fields and hitting 'Update' which will redraw the query.

4.4 Downloading results

All graphics, tables and graphs can be downloaded easily to the users local hard drive. This can be achieved using the Comparative Genomics View Export button in the menu bar, which opens up a dropdown menu shown in

Figure 19 Panel to change the range of the query of the selected chromosome



4.4.1 Downloading Images

To download any of the views, simply click on the Image button from the dropdown menu in Figure 20 and choose which view you wish to download (Figure 21).

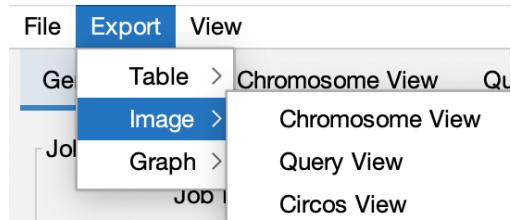


Figure 21 The different views available to the user to download

Figure 20 Dropdown menu to export data from MapOptics

4.4.2 Downloading Tables

To download any of the views, simply click on the Table button from the dropdown menu in Figure 20 and choose which view you wish to download (Figure 22).

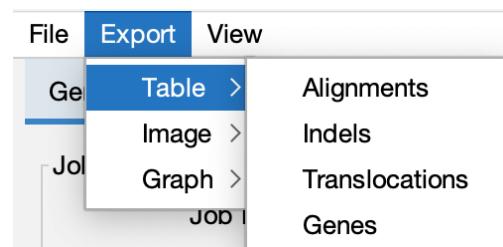


Figure 22 The different tables that can be exported

4.4.3 Downloading Graphs

To download any of the graphs simply click on the Graph button from the dropdown menu in Figure 20 and choose which view you wish to download (Figure 23).

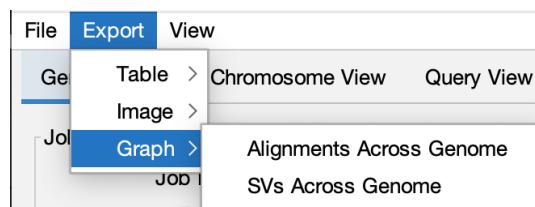


Figure 23 The different graphs that can be exported

4.4 Changing MapOptics Aesthetics

MapOptics also provides the user with some customisation options for the views, for example changing colours of the matched and unmatched sites and changing the plot style of different graphs.

4.5.1 Changing Graph Styles

To changes the plot style click on 'View' on the comparative genomics view menu bar and then 'change plot style' (Figure 24).

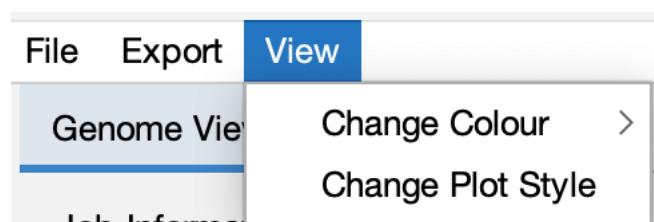


Figure 24 Different view options provided to the user

4.5.2 Changing Colours

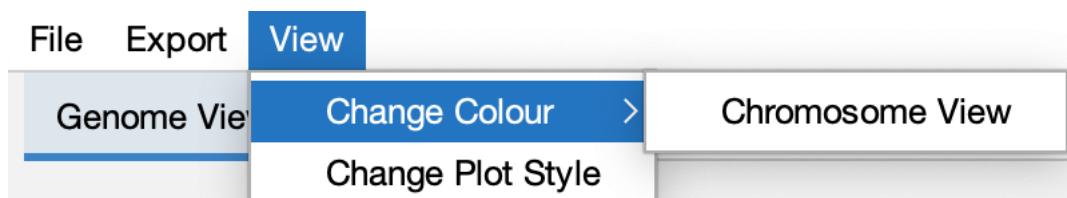


Figure 25 Changing the view colour scheme on different views

To change the colour options on different views click on ‘View’ on the comparative genomics view menu bar and then choose which view’s colour scheme you wish to edit.

5 File Formats

MapOptics Comparative Genomics requires all files uploaded to follow strict criteria to minimise chance of errors occurring.

5.1 Fasta

The fasta file containing the genomes of the reference and query genomes must follow standard fasta format. Each chromosome of the genome must be one contiguous sequence and the header contain the chromosome name.

5.2 Cmap

All cmap files must follow the format laid out by Bionano genomics found here:
<https://bionanogenomics.com/wp-content/uploads/2017/03/30039-CMAP-File-Format-Specification-Sheet.pdf>

5.3 Xmap

All xmap files must follow the format laid out by BioNano genomics found here:
<https://bionanogenomics.com/wp-content/uploads/2017/03/30040-XMAP-File-Format-Specification-Sheet.pdf>

5.4 Annotation

MapOptics Comparative Genomics currently supports annotation files in GFF3 and GTF format. Information about each can be found here:
<https://www.ensembl.org/info/website/upload/gff3.html>

5.5 Karyotype

The karyotype file used in MapOptics is a simple text file with two columns and no headers. The columns are space delimited. The first column is the length of the chromosome to one decimal place and the second column is the chromosome name, extracted from the fasta header.

6 Server set-up

In order for MapOptics to perform alignment a linux based server using either CentOS or Ubuntu is recommended. All files from the ‘Server_files’ directory must be uploaded to the server.

6.1 Hardware requirements

. At least 32GB of RAM and space to install required softwares and also upload data is required.

6.2 Software requirements

Bionano solve package, downloaded from here:

<https://bionanogenomics.com/support/software-downloads/>

6.3 Working directory

Each server must have a dedicated working directory to save the files in the ‘Server_files’ directory. It is within this directory that the data for each submitted job will be saved to.