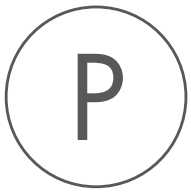


Whole Genome Alignment

Plugin

USER MANUAL

User manual for

Whole Genome Alignment 25.0.1

Windows, macOS and Linux

February 13, 2025

This software is for research purposes only.

QIAGEN Aarhus AS
Kalkværksvej 5, 11.
DK - 8000 Aarhus C
Denmark

Contents

Chapter 1

Introduction

The Whole Genome Alignment plugin enables users to compare small and medium-sized genomes (up to 100 Mb) and to explore and visualize their evolutionary relationships (figure 1.1).

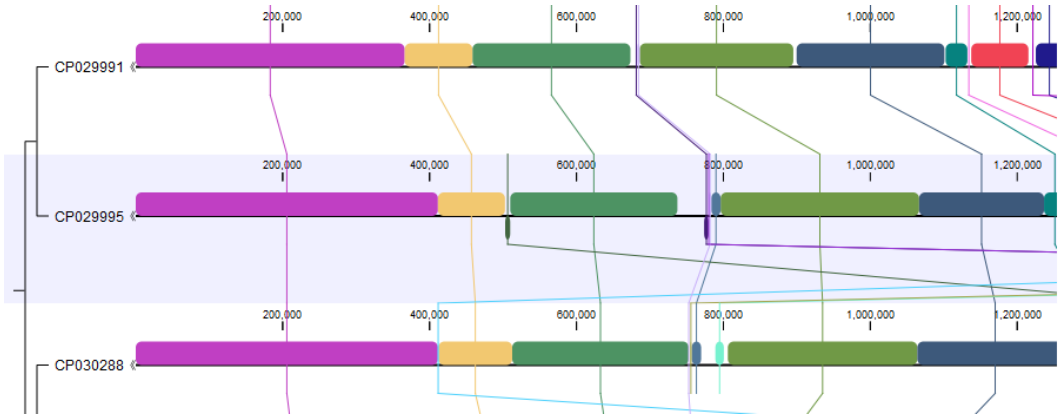


Figure 1.1: *A Whole Genome Alignment view.*

The plugin delivers tools that are put into a Whole Genome Alignment folder under the Tools menu (figure 1.2), and adds relevant file formats to the Export and Import functionality of the Workbench.

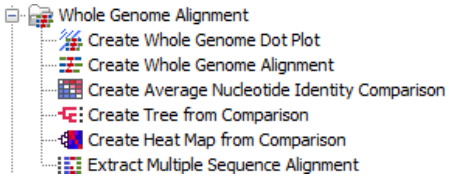


Figure 1.2: *The tools delivered by this plugin are installed in a folder called Whole Genome Alignment, under the Tools menu.*

The plugin contains functionality for:

- Fast whole genome dot plots
- Aligning multiple genomes

- Visualizing alignments of multiple genomes, which can contain large-scale events such as inversions and translocations
- Calculating the Average Nucleotide Identity
- Transferring annotations from a reference genome to other genomes
- Creating evolutionary trees and heat maps based on the Average Nucleotide Identity
- Importing and exporting standard Whole Genome Alignments formats (MAF and XMFA)
- Extracting multiple sequence alignments, optionally restricting to annotations such as coding regions

Chapter 2

Create Whole Genome Dot Plot

The Create Whole Genome Dot Plot tool makes it possible to compare two genomes and get a quick, initial overview of the similarities between them.

Note that the dot plots created by the Create Whole Genome Dot Plot are fundamentally different than the dot plots created by the Create Dot Plot tool that comes as part of the Workbench's tool set (see http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Create_dot_plots.html). Whereas the classic Create Dot Plot tool compares all positions between two sequences, the Create Whole Genome Dot Plot tool works by identifying seeds, i.e., short stretches of nucleotide sequences that are shared between two genomes but not present multiple times on the same genome - which is the same approach as the one used in the initial stages of the Create Whole Genome Alignment tool. So when compared to the classic Create Dot Plot tool, the Create Whole Genome Dot Plot tool will produce less noise, but might also miss homologous regions if better matches are found in a different region on a genome.

To run the **Create Whole Genome Dot Plot** tool, go to:

Tools | **Whole Genome Alignment** ( ) | **Create Whole Genome Dot Plot ( )**

Once the tool wizard has opened (figure 2.1), choose two or more nucleotide sequences or nucleotide sequence lists. If the input elements are nucleotide sequence lists, each sequence in the list (chromosomes or contigs) is considered to be part of the same genome.

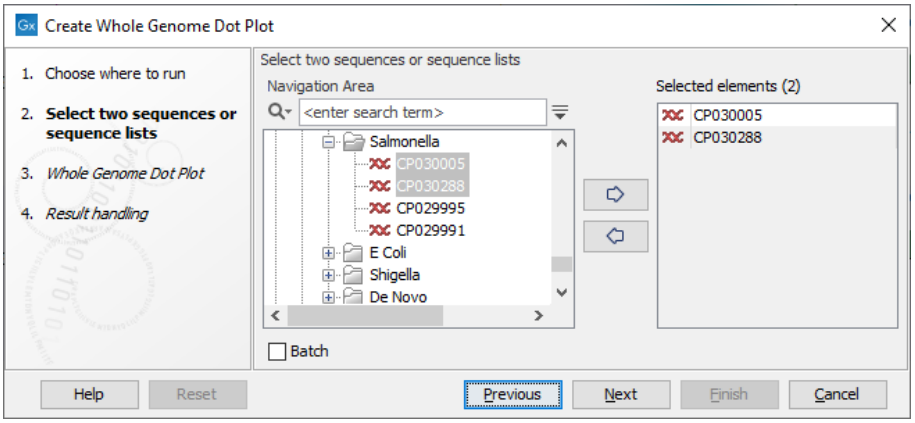


Figure 2.1: *Select input for the Create Whole Genome Dot Plot tool.*

You can set the following parameters (figure 2.2):

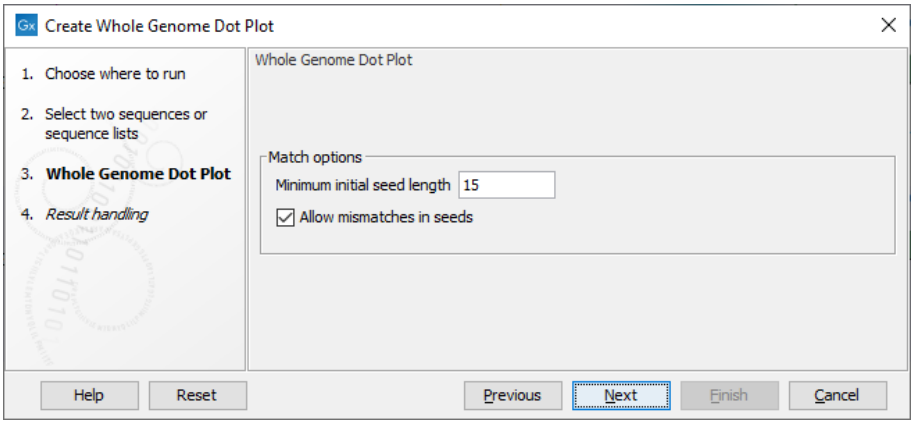


Figure 2.2: *Configurable parameters for the Create Whole Genome Dot Plot tool.*

- **Minimum initial seed length** The tool works by identifying seeds - short stretches that are similar - in the two genomes. This option determines the smallest number of nucleotides required before a seed is shown in the dot plot.
- **Allow mismatches in seeds** When this option is enabled, the search for initial seeds will allow for mismatches in the seeds. This makes it possible to visualize more divergent genome pairs, but may also introduce more noise.

The tool outputs a dot plot (figure 2.3):

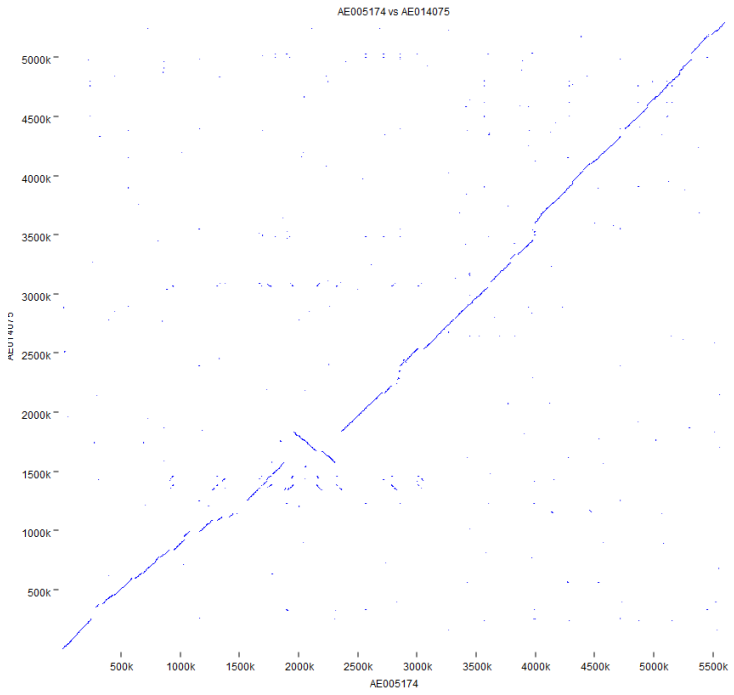


Figure 2.3: *Dot plot output from the Create Whole Genome Dot Plot tool.*

Chapter 3

Create Whole Genome Alignment

The Create Whole Genome Alignment tool works by identifying seeds, i.e., short stretches of nucleotide sequence that are shared between multiple genomes but not present multiple times on the same genome. These seeds are then extended using a HOXD scoring matrix [Chiaromonte et al., 2002] until the local alignment score drops below a fixed threshold. From the initial extended seed matches, a distance matrix between the input genomes is calculated. This distance matrix is used for the subsequent pairwise processing, where the most similar genomes are processed first. Proceeding iteratively on the most similar pair of genomes, the tool will then extend and merge seed matches to create longer alignment blocks. These blocks may be present on two or more genomes, and may align to both strands of the genomes (allowing for the identification of inversions). Similar to progressiveMauve [Darling AE, 2010], we combine the HOXD substitution score with an adjustment term based on kmer frequency. This is done to avoid spurious matches to repetitive regions in the genome.

To run the **Create Whole Genome Alignment** tool, go to:

Tools | **Whole Genome Alignment** ( ) | **Create Whole Genome Alignment** ( )

Once the tool wizard has opened (figure 3.1), choose two nucleotide sequences or nucleotide sequence lists. If the input objects are nucleotide sequence lists (chromosomes or contigs) , each sequence in the list is considered to be part of the same genome.

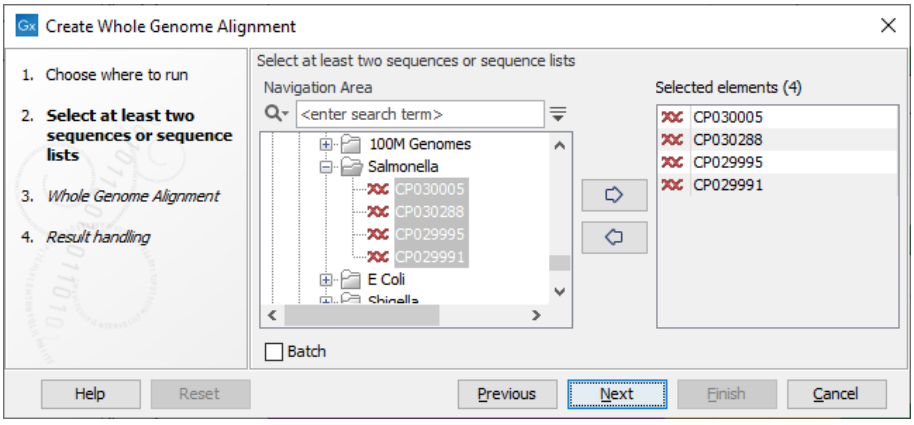


Figure 3.1: *Select input for the Create Whole Genome Alignment tool.*

The tool has the following alignment options (figure 3.2):

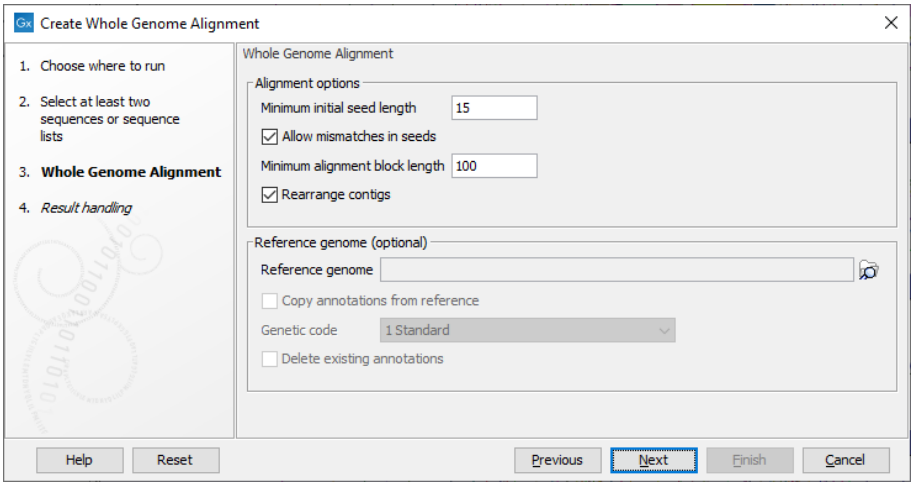


Figure 3.2: *Configurable parameters for the Create Whole Genome Alignment tool.*

- **Minimum initial seed length** The tool works by identifying seeds - short stretches that are similar - in the genomes. This option determines the number of nucleotides required for a seed to be included in the later stages of the alignment. **Note about memory requirements**: as a rule of thumb, the tool needs 10 bytes of memory for each nucleotide in the input genomes, plus additional memory for storing the found seeds. This means that smaller values of the "Minimum initial seed length" option will require more memory. As a consequence, the tool is not designed for large eukaryotic genomes, but for aligning bacterial size genomes.

- **Allow mismatches in seeds** When this option is enabled, the search for initial seeds will allow for mismatches in the seeds. This makes it possible to visualize more divergent genome pairs, but may also introduce more noise.

- **Minimum alignment block length** Only alignment blocks larger than this length will be output. This parameter may be used to filter away smaller and potentially erroneous alignments.

- **Rearrange contigs** If the 'Rearrange contigs' option is enabled, the input genomes will be rearranged to more closely match each other. The possible rearrangements are: reordering the contigs in each genome, reverse complementing contigs, and shifting of circular contigs. Rearrangements are performed in order to minimize the number of crossing lines between alignments blocks in the final whole genome alignments. Notice, that if a reference genome is specified, the rearrangements will be performed to most closely match the reference genome. The reference genome itself is not modified. If no reference genome is specified, preference will be given to the genome with the highest average contig length, though it may still undergo rearrangements.

The tool also has options for working with a reference genome (figure 3.2):

- **Reference genome** This option can be used to specify a reference genome. A reference genome can be used when copying annotations (see below), or to guide the rearrangements (see the Rearrange contigs option above). It will not make a difference whether the specified reference genome is included as part of the input genomes or not - in either case, only a single copy of the reference genome will appear in the whole genome alignment.

- **Copy annotations from reference** When enabled, annotations from alignments blocks are copied from the reference to all aligned genomes. Copied annotations will have an added 'copied_from=GENOME_NAME' qualifier. Annotations on the reference sequence may extend beyond alignments blocks, in which case the transferred annotations become partial copies: These annotations will have an additional 'partial_copy=true' qualifier. For CDS annotations, the reading frame may change. To avoid this, all 3 possible reading frames are checked and the one with the fewest stop codons is chosen. The best reading frame is annotated using a 'codon_start' tag. The 'genetic code' parameter is used when identifying stop codons. If the best reading frame still contains (non-trailing) stop codons, a qualifier 'copy_contains_stopcodons=true' is added to the annotation. Any pre-existing translation, codon_start, or frame qualifiers present on the original reference annotation, are removed on the copied annotations.
- **Genetic code** Used when transferring CDS annotations (see above).
- **Delete existing annotations** When enabled, all annotations on the input genomes will be removed.

The tool outputs a Whole Genome Alignment showing the aligned regions between the genomes (figure 3.3).

The output option **Output genomes after alignment** makes it possible to output the genomes, including any modifications, such as contig rearrangements and added annotations. This can also be used in workflows for automated annotations of genomes against a reference genome.

Note that the tree that is output per default in the Whole Genome Alignment view is a Neighbor-Joining tree based on the distance matrix (see the beginning of this section).

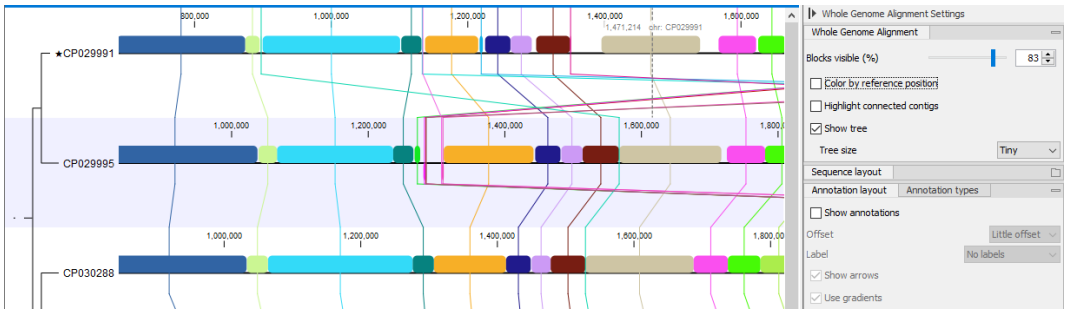


Figure 3.3: *Whole Genome Alignment view. The star next to the top genome name indicates that this genome was chosen as a reference.*

An alignment block (shown as a colored box) corresponds to a region of the genome that is aligned to a region on at least one other genome. The position of the box relative to the sequence indicates the strand on which the alignment was identified: above the sequence for the plus strand, below the sequence for the minus strand. When hovering the mouse over a block, the corresponding alignment blocks on the aligned genomes will be highlighted. The connected alignment blocks (which will share the same color) can be thought of as an ordinary linear multiple sequence alignment: they will not contain any internal rearrangements.

When clicking on a position on a genome, the view will automatically modify so that the aligned positions are centered on top of each other. When double clicking an alignment block, the regions covered by the connected alignment blocks will be selected.

The Whole Genome Alignment view shares most of the functionality of the ordinary sequence viewer: this includes the ability to show any annotations on the genomes (such as CDS or Gene annotations), searching for gene names (using the "Find" panel), and zooming down to the nucleotide level.

The Whole Genome Alignment view has a few special options:

- **Blocks visible (%)** This option makes it possible to hide the smallest alignment blocks. A value of 20 means that only the longest 20% of the alignment blocks will be shown. The default value is chosen such that 100 alignment blocks are visible.
- **Color by reference position** Alignment blocks are colored according to their position on the chromosomes/contigs of the reference genome. This option is only available if a reference genome is chosen. Reference genomes are indicated by star prepending the genome name (see figure 3.3).
- **Highlight connected contigs** When hovering an alignment block, chromosomes/contigs not part of the alignment will be dimmed.
- **Show tree** When the editor is opened, a tree is shown to the left of the genomes. This is the tree created by the Create Whole Genome Alignment tool using Neighbor-Joining on the distance matrix calculated from the extended seeds. This tree will provide a rough overview of the relations between the genomes: closely related genomes should be located close to each other. It is possible to change the ordering of the genomes by dragging them in the view, but this requires that the "Show tree" option in the side panel has been disabled.

Extracting multiple sequence alignments: When selecting part of a sequence in an alignment block, it is possible to use the context menu to extract the selection into an ordinary multiple sequence alignment (figure 3.4):

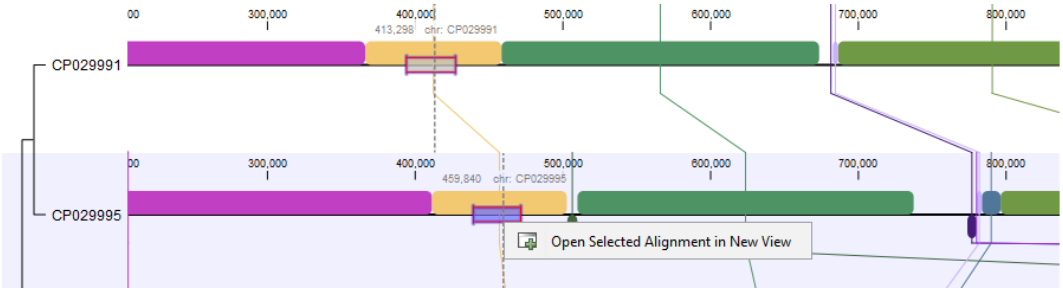


Figure 3.4: *Whole genome alignment viewer.*

Open as Sequence List in New View: When using the context menu on a genome (figure 3.5), it is possible to open the genome as a new sequence list, including any re-ordering, shifting, and reverse complementing done as part of the alignment.

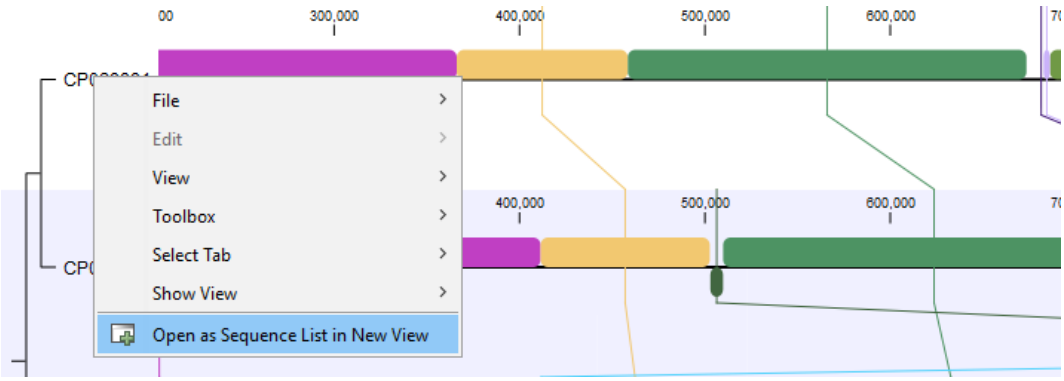


Figure 3.5: *Whole genome alignment viewer.*

Chapter 4

Create Average Nucleotide Identity Comparison

The Create Average Nucleotide Identity Comparison tool is useful for getting a quantitative measure of the similarity between genomes. This tool takes a Whole Genome Alignment as input, and for each pair of genomes, the regions aligned between the two genomes are identified, and the following two measures are calculated:

- The **Alignment Percentage** is the average percentage of the two genomes which is aligned.
- The **Average Nucleotide Identity** is the percentage of exactly matching nucleotides for these aligned regions.

It is possible to calculate these measures either for the full set of aligned regions or for a set of regions defined by annotations on the genomes (such as the CDS annotations of protein coding regions).

To run the **Create Average Nucleotide Identity Comparison** tool, go to:

> **Tools** | **Whole Genome Alignment** ( ) | **Create Average Nucleotide Identity Comparison ( )**

Once the tool wizard has opened (figure 4.1), choose the alignment you would like to analyze.

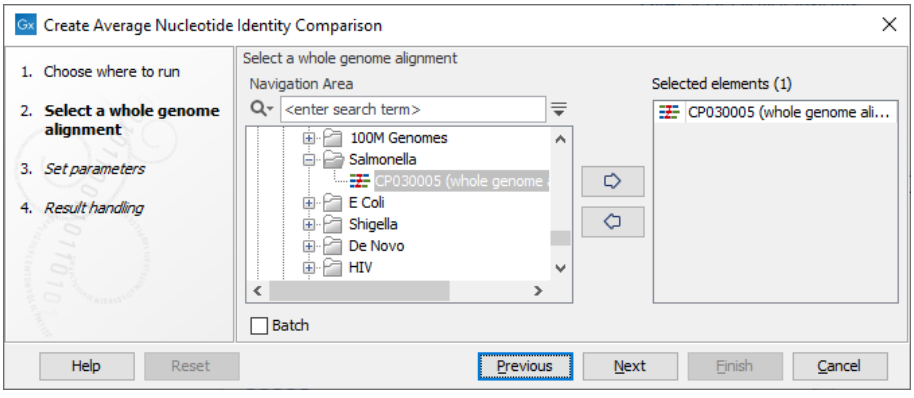


Figure 4.1: *Select a Whole Genome Alignment.*

In the next dialog, set the parameters for the tool (figure 4.2).

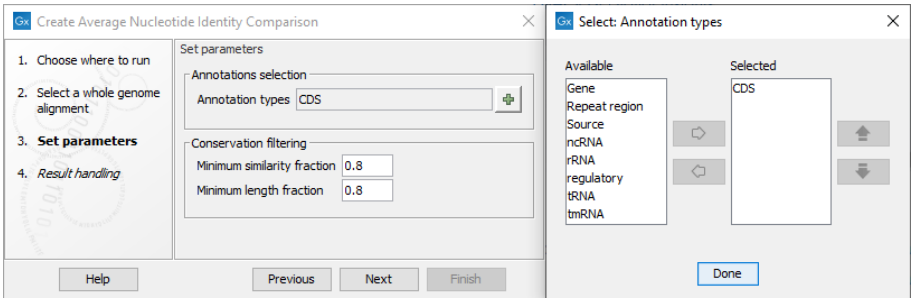
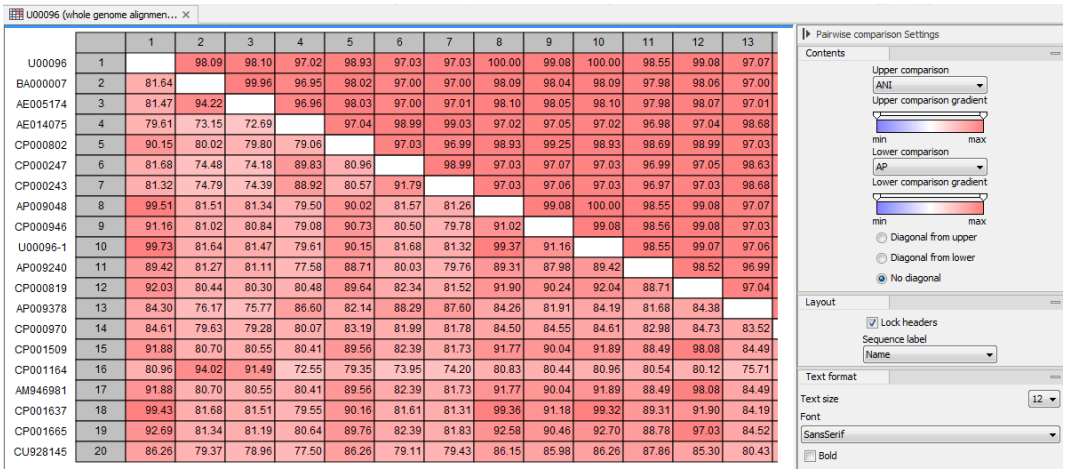


Figure 4.2: *Parameters for the tool, and a view of the annotations that can be specified to restrict the Pairwise Comparison.*

Click on the green plus sign to select the **Annotations types** you wish to restrict your analysis to, as can be seen to the right of figure 4.2. If the tool is restricted to a set of annotations, each annotation must satisfy the Conservation filtering criteria: there must be a minimum required number of identical matches, the **Minimum similarity fraction**, over at least a specified fraction of the annotation length, the **Minimum length fraction**. In cases where no annotations are specified, the tool will divide all the aligned regions into smaller regions of 1024 nucleotides. Each of these smaller regions must then satisfy the conservation filtering criteria.

The tool outputs a Pairwise Comparison table as the one seen in figure 4.3.



		1	2	3	4	5	6	7	8	9	10	11	12	13
U00096	1		98.09	98.10	97.02	98.93	97.03	97.03	100.00	99.08	100.00	98.55	99.08	97.07
BA000007	2	81.64		99.96	96.95	98.02	97.00	97.00	98.09	98.04	98.09	97.98	98.06	97.00
AE005174	3	81.47	94.22		96.96	98.03	97.00	97.01	98.10	98.05	98.10	97.98	98.07	97.01
AE014075	4	79.61	73.15	72.69		97.04	98.99	99.03	97.02	97.05	97.02	96.98	97.04	98.68
CP000802	5	90.15	80.02	79.80	79.06		97.03	96.99	98.93	99.25	98.93	98.69	98.99	97.03
CP000247	6	81.68	74.48	74.18	89.83	80.96		98.99	97.03	97.07	97.03	96.99	97.05	98.63
CP000243	7	81.32	74.79	74.39	88.92	80.57	91.79		97.03	97.06	97.03	96.97	97.03	98.68
AP009048	8	99.51	81.51	81.34	79.50	90.02	81.57	81.26		99.08	100.00	98.55	99.08	97.07
CP000946	9	91.16	81.02	80.84	79.08	90.73	80.50	79.78	91.02		99.08	98.56	99.08	97.03
U00096-1	10	99.73	81.64	81.47	79.61	90.15	81.68	81.32	99.37	91.16		98.55	99.07	97.06
AP009240	11	89.42	81.27	81.11	77.58	88.71	80.03	79.76	89.31	87.98	89.42		98.52	96.99
CP000819	12	92.03	80.44	80.30	80.48	89.64	82.34	81.52	91.90	90.24	92.04	88.71		97.04
AP009378	13	84.30	76.17	75.77	86.60	82.14	88.29	87.60	84.26	81.91	84.19	81.68	84.38	
CP000970	14	84.61	79.63	79.28	80.07	83.19	81.99	81.78	84.50	84.55	84.61	82.98	84.73	83.52
CP001509	15	91.88	80.70	80.55	80.41	89.56	82.39	81.73	91.77	90.04	91.89	88.49	98.08	84.49
CP001164	16	80.96	94.02	91.49	72.55	79.35	73.95	74.20	80.83	80.44	80.96	80.54	80.12	75.71
AM946981	17	91.88	80.70	80.55	80.41	89.56	82.39	81.73	91.77	90.04	91.89	88.49	98.08	84.49
CP001637	18	99.43	81.68	81.51	79.55	90.16	81.61	81.31	99.36	91.18	99.32	89.31	91.90	84.19
CP001665	19	92.69	81.34	81.19	80.64	89.76	82.39	81.83	92.58	90.46	92.70	88.78	97.03	84.52
CU928145	20	86.26	79.37	78.96	77.50	86.26	79.11	79.43	86.15	85.98	86.26	87.86	85.30	80.43

Figure 4.3: *Average Nucleotide Identity Comparison table.*

It is possible to export the table using the Pairwise Comparison CSV exporter (see section 8.2). It is also possible to input this file in the Create Tree from Comparison and Create Heat Map from Comparison tools for better visualization of the results.

Chapter 5

Create Tree from Comparison

The Create Tree from Comparison tool builds a tree from a Pairwise Comparison such as those generated by Create Average Nucleotide Identity Comparison tool.

To run the **Create Tree from Comparison** tool, go to:

Tools | **Whole Genome Alignment** ( ) | **Create Tree from Comparison ( )**

Once the tool wizard has opened (figure 5.1), choose the Pairwise Comparison table you would like to use.

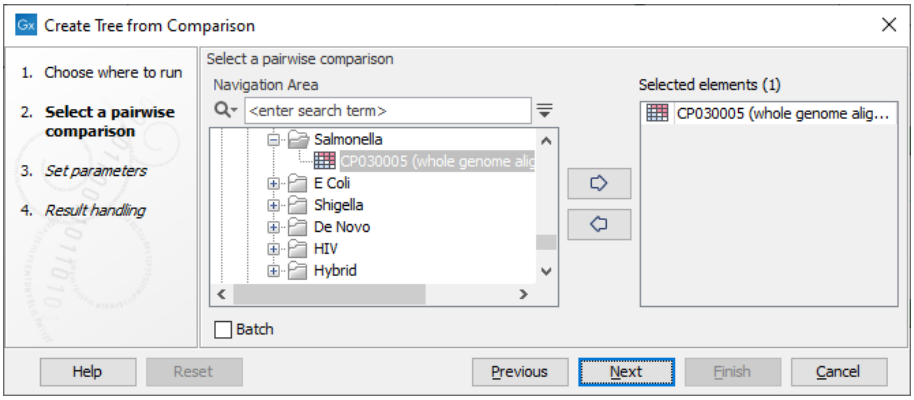


Figure 5.1: *Select a Pairwise Comparison table.*

In the next dialog (figure 5.2), you can set the following parameters:

- **Table types** The possible table types are extracted from the Pairwise Comparison table input. In the case of a Pairwise Comparison table obtained from Create Average Nucleotide Identity Comparison, these are: ANI (Average Nucleotide Identity) or AP (Alignment Percentage). If left empty, as it is by default, both types will be used.
- **Tree construction method** Neighbor Joining, well suited for trees with varying rates of evolution, and / or UPGMA, method that assumes a constant rate of evolution. Learn more about distance based reconstruction methods here: `http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Distance_based_reconstruction_methods.html`.

Note that the tool outputs a tree for each combination of table types and tree construction

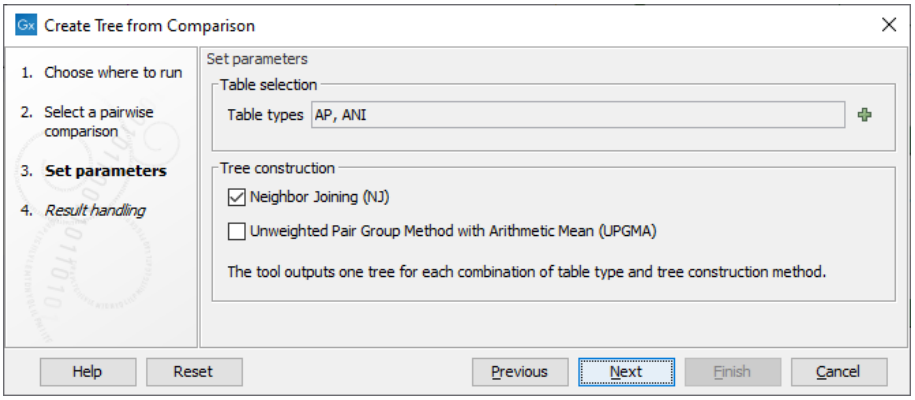


Figure 5.2: *Select the table types and tree construction methods you would like to use for building trees.*

method. In addition, metadata from the Pairwise Comparison is automatically transferred to the tree. Sequence metadata containing taxonomy information is also added if this information was present in the inputs.

Learn more about visualizing trees here: http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Tree_Settings.html.

The Pairwise Comparison table input is either a distance or similarity matrix. The tool automatically detects the type of each table by checking the values on the diagonal: if the diagonal contains only zeros, then the table represents a distance matrix, otherwise it is a similarity matrix. From the table, a symmetric distance matrix d is calculated as follows:

`d[i][j] = (t[i][j] + t[j][i]) / 2` if the table is a distance matrix,

`d[i][j] = (1 - t[i][j] + 1 - t[j][i]) / 2` if the table is a similarity matrix,

where `t[i][j]` is the relative value (between 0 and 1) found in the table in row i and column j.

A tree is then created from the distance matrix d using the specified tree construction method. The tree is then generated such that the distance between two leaves (calculated as the sum of lengths of the branches connecting the leaves) is below 1, as the relative distance was used above. The branch lengths are then scaled so that the distance between two leaves reflects the absolute distance between them to match the entries in the table.

Chapter 6

Create Heat Map from Comparison

The Create Heat Map from Comparison tool builds a heat map from a Pairwise Comparison such as those generated by Create Average Nucleotide Identity Comparison tool.

To run the **Create Heat Map from Comparison** tool, go to:

Tools | **Whole Genome Alignment** ( ) | **Create Heat Map from Comparison** ( )

Once the tool wizard has opened (figure 6.1), choose the Pairwise Comparison table you would like to use.

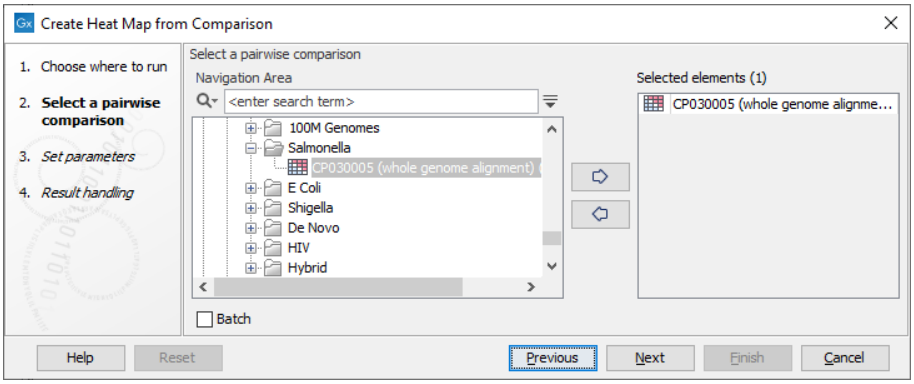


Figure 6.1: *Select a Pairwise Comparison table.*

In the next dialog (figure 6.2), you can set the following parameters:

- **Table types** The possible table types are extracted from the Pairwise Comparison table input. In the case of a Pairwise Comparison table obtained from Create Average Nucleotide Identity Comparison, these are: ANI (Average Nucleotide Identity) or AP (Alignment Percentage). If left empty, as it is by default, both types will be used.
- **Clusters construction methods**

  There are three kinds of **distance measures**:

  - **Euclidean distance**. The length of the segment connecting two points. If $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$, then the Euclidean distance between u and v

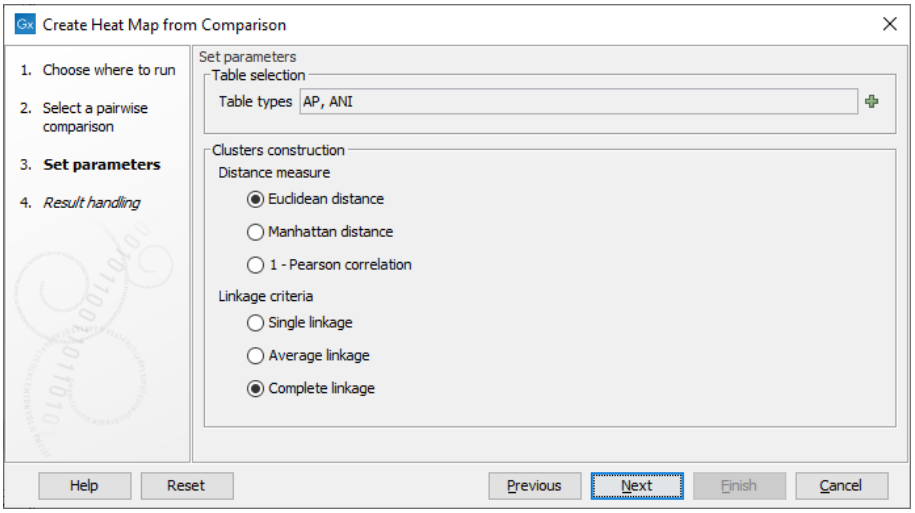


Figure 6.2: *Select the table types and clusters construction methods you would like to use for building the heat maps.*

is

$$|u - v| = \sqrt{\sum_{i=1}^{n}(u_i - v_i)^2}.$$

– **Manhattan distance**. The distance between two points measured along axes at right angles. If $u = (u_1, u_2, \ldots, u_n)$ and $v = (v_1, v_2, \ldots, v_n)$, then the Manhattan distance between u and v is

$$|u - v| = \sum_{i=1}^{n} |u_i - v_i|.$$

– **1 - Pearson correlation**. The Pearson correlation coefficient between $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ is defined as

$$r = \frac{1}{n-1} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s_x} \right) \cdot \left(\frac{y_i - \overline{y}}{s_y} \right)$$

where $\overline{x}/\overline{y}$ and s_x/s_y are the average and sample standard deviation, respectively, of the values in x/y values.

The Pearson correlation coefficient ranges from -1 to 1, with high absolute values indicating strong correlation, and values near 0 suggesting little to no relationship between the elements.

Using 1 - | Pearson correlation | as the distance measure ensures that highly correlated elements have a shorter distance, while elements with low correlation are farther apart.

The distance between two clusters is determined using one of the following linkage types:

– **Single linkage**. The distance between the two closest elements in the two clusters.

– **Average linkage**. The average distance between elements in the first cluster and elements in the second cluster.

– **Complete linkage**. The distance between the two farthest elements in the two clusters.

The Pairwise Comparison table input is either a distance or similarity matrix. The tool automatically detects the type of each table by checking the values on the diagonal: if the diagonal contains only zeros, then the table represents a distance matrix, otherwise a similarity matrix. If the table is distance matrix, a similarity matrix s is calculated as follows:

`s[i][j] = min + (1 - t[i][j]) * (max - min)` if the table is a distance matrix,

where `t[i][j]` is the relative value (between 0 and 1) found in the table in row i and column j,

and min and max are the minimum and maximum magnitude of the table.

A heat map (figure 6.3) is then created from the similarity matrix s according to the specified clustering options and using a hierarchical clustering algorithm. Note that the tool outputs a heat map for each chosen table type and its name contains the table type used.

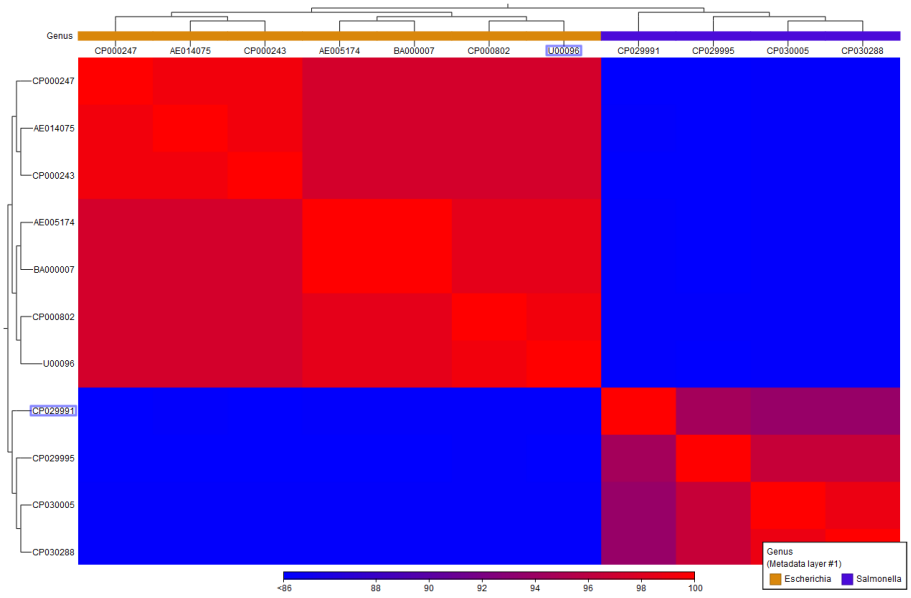


Figure 6.3: *A Comparison Heat Map.*

Metadata from the Pairwise Comparison is transferred to the map. Additionally, sequence metadata containing taxonomy information is added if this information was present in the inputs. You can learn more about heat map views here: `http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=_heat_map_view.html`.

Chapter 7

Extract Multiple Sequence Alignment

The Extract Multiple Sequence Alignment tool traverses all aligned blocks in a Whole Genome Alignment and creates a linear concatenated Multiple Sequence Alignment. In contrast to the Whole Genome Alignment, a Multiple Sequence Alignment has all aligned blocks occurring in the same order and orientation.

Note: if the extracted Multiple Sequence Alignment contains long sequences, viewing the Multiple Sequence Alignment may be slow. The purpose of this tool is to make it possible to export the extracted alignment in Nexus format for example, so it can be used in third-party software that cannot process whole genome alignments formats (MAF and XMFA).

To run the **Extract Multiple Sequence Alignment** tool, go to:

Tools | **Whole Genome Alignment** ( ) | **Extract Multiple Sequence Alignment ( )**

Once the tool wizard has opened (figure 7.1), choose the Whole Genome Alignment you would like to use.

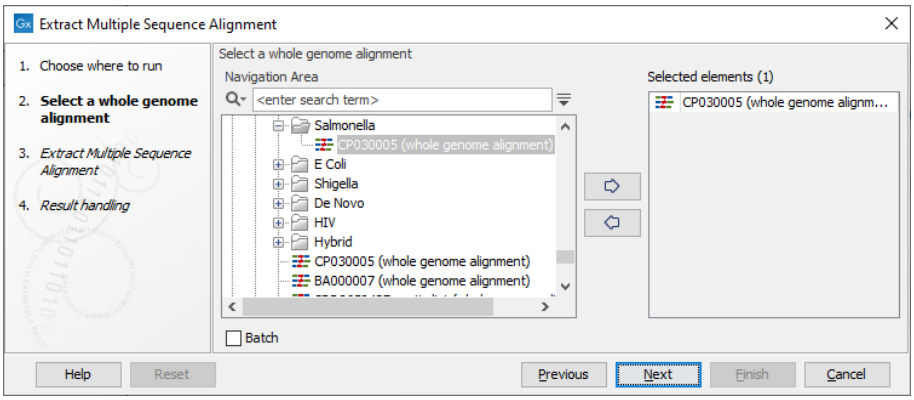


Figure 7.1: *Select a Whole Genome Alignment.*

In the next dialog (figure 7.2), you can set the following parameters:

- **Minimum percentage of genomes** An alignment block is only included if it is present on at least this fraction of the genomes. Setting this value to 100 means that only alignment blocks covering all genomes are included.
- **Extract only annotated blocks** It is possible to restrict the extraction only to regions of

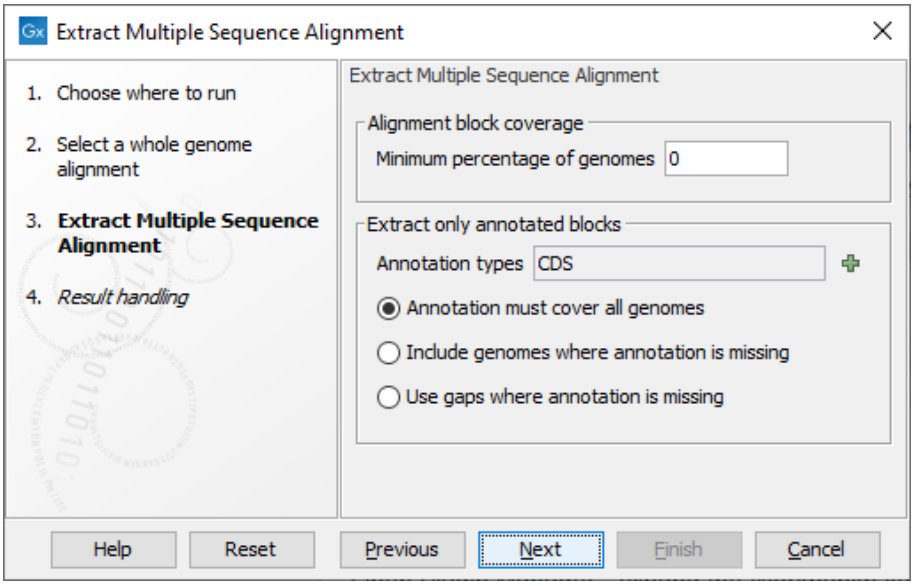


Figure 7.2: *Select the table types and clusters construction methods you would like to use for building the heat maps.*

the genome covered by an annotation (such as the CDS annotations for protein coding regions). If an alignment block is partly covered by an annotation, only the intersection of the annotation region and alignment region is included. Choose among the following options:

- **Annotation types** Which annotations types to restrict the extraction to, such as CDS or Gene types.
- **Annotation must cover all genomes** The intersection of the alignment block and the annotations is extracted only if the block intersects an annotation on all genomes the alignment block covers.
- **Include genomes where annotation is missing** An alignment block only needs to intersect an annotation for some of the genomes it covers in order for the intersection of the block and the annotations to be extracted.
- **Use gaps where annotation is missing** If an alignment block covers a region of a genome without an annotation, that region will be represented as gaps in the extracted alignment.

Chapter 8

Import and export

8.1 Whole Genome Alignment import and export

Whole Genome Alignments can be exported and imported as MAF and XMFA files using the standard Import and Export functionality of the Workbench (figure 8.1).

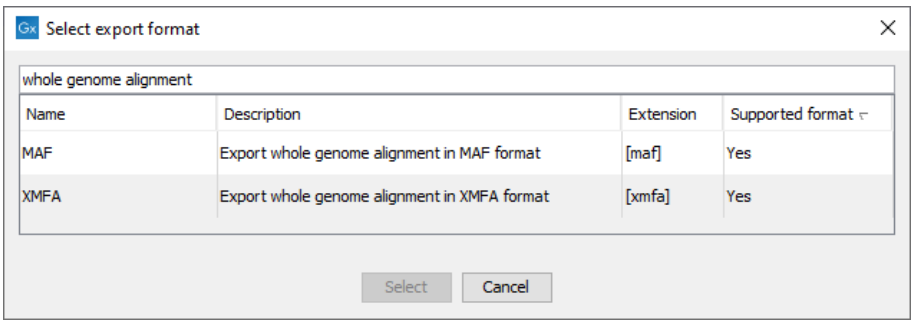


Figure 8.1: *Export and import of Whole Genome Alignments can be done using the MAF and XMFA formats.*

For MAF and XMFA import, an alignment block must contain each genome only once. We do not support alignment blocks that map multiple times on the same genome.

The MAF importer does not support the optional line types (such as i and e) and these are ignored if found in the file.

The XMFA importer tries to extract the sequence names found as comments in the sequence headers. If the extraction fails, the sequences are named using the numbers found in the file and the comments are ignored.

8.2 Pairwise Comparison export

Pairwise Comparison tables can be exported as CSV file (figure 8.2).

In that case, the following options are available (figure 8.3).

The exporter creates a CSV file from a Pairwise Comparison element. It can export the element in two different formats: a long format (which is the default option), or a shorter one called matrix format (when the option is checked).

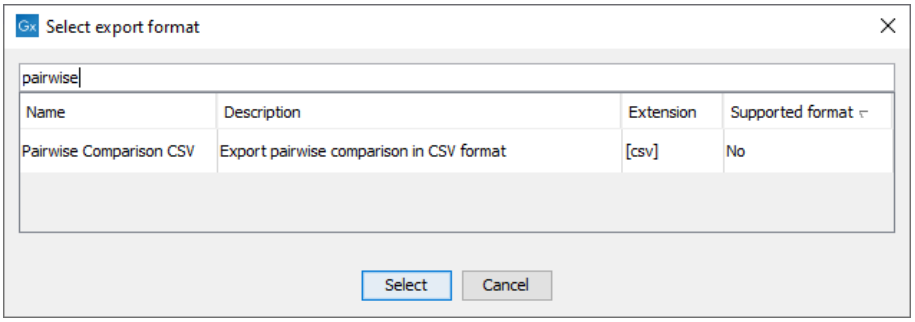


Figure 8.2: *Export of Pairwise Comparisons can be done using the CSV format.*

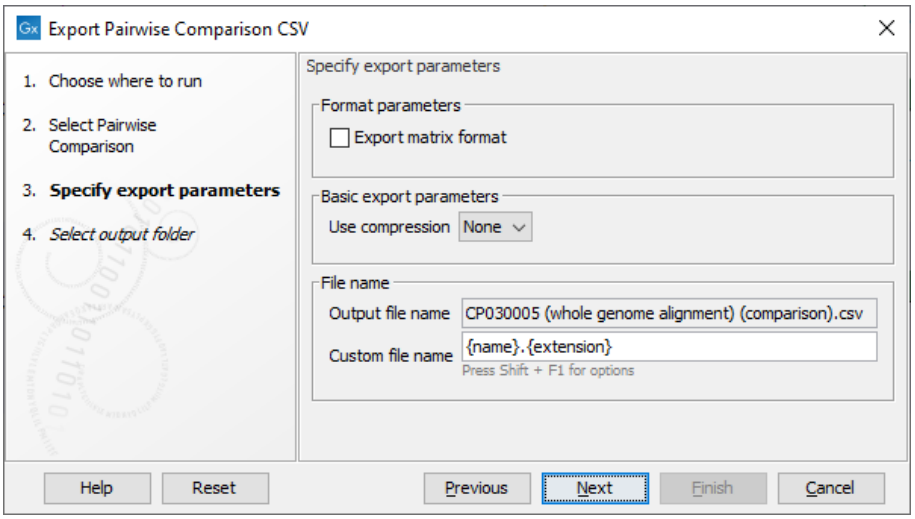


Figure 8.3: *Options for CSV export.*

The long format will export the data with a pair of samples for each line (figure 8.4), with the different values found in the Pairwise Comparison as columns:

```
"Sample 1","Sample 2","ANI","AP"
"CP000802","CP000802","100","100"
"CP000802","CP030005","84.3154618995","35.8810003247"
"CP000802","CP029991","84.1991180688","33.5905271219"
"CP000802","U00096","98.9645052659","90.2435851685"
"CP000802","CP030288","84.2625397122","35.3338287794"
"CP000802","AE014075","97.0278468967","79.5646373385"
"CP000802","CP000243","97.0026908053","80.8368681865"
```

Figure 8.4: *Options for CSV export.*

The short (matrix) format will export the data with a pair of samples for each line(figure 8.5), with all the tables found in the Pairwise Comparison after each other:

```
"Table","Sample","CP000802","CP030005","CP029991","U00096","CP030288","AE014075","CP000243","CP000247","AE005174","BA000007","CP029995"
"ANI","CP000802","100","84.3154618995","84.1991180688","98.9645052659","84.2625397122","97.0278468967","97.0026908053","97.0313818959","98.0325803583","98.0281368851","84.2203412695"
"ANI","CP030005","84.3154618995","100","93.4101171015","84.2287526226","98.4517465391","84.2310544659","84.2434695755","84.263733624","84.254010969","84.2401228065","96.1582469548"
"ANI","CP029991","84.1991180688","93.4101171015","100","84.2219701754","93.4568137042","84.2054043512","84.1491290348","84.1424635541","84.1818373105","84.180798223","93.959804916"
"ANI","U00096","98.9645052659","84.2287526226","84.2219701754","100","84.2422754763","97.0124295541","97.0202359769","97.0276473463","98.1153176503","98.1186826118","84.1783415996"
```

Figure 8.5: *Options for CSV export.*

Learn about the other export options here:

http://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/current/index.php?manual=Export_parameters.html.

Chapter 9

Install and uninstall plugins

Whole Genome Alignment is installed as a plugin.

9.1 Installation of plugins

Note: In order to install plugins and modules, the *CLC Workbench* must be run in administrator mode. On Windows, you can do this by right-clicking the program shortcut and choosing "Run as Administrator". On Linux and Mac, it means you must launch the program such that it is run by an administrative user.

Plugins and modules are installed and uninstalled using the Workbench Plugin Manager. To open the Plugin Manager, click on the **Plugins ( ) button** in the top Toolbar, or go to the menu option:

Utilities | **Manage Plugins... ( )**

The Plugin Manager has two tabs at the top:

- **Manage Plugins** An overview of your installed plugins and modules is provided under this tab.
- **Download Plugins** Plugins and modules available to download and install are listed in this tab.

To install a plugin, click on the **Download Plugins** tab (figure 9.1). Select a plugin. Information about it will be shown in the right hand panel. Click on the **Download and Install** button to install the plugin.

Accepting the license agreement

The End User License Agreement (EULA) must be read and accepted as part of the installation process. Please read the EULA text carefully, and if you agree to it, check the box next to the text **I accept these terms**. If further information is requested from you, please fill this in before clicking on the **Finish** button.

Installing a cpa file

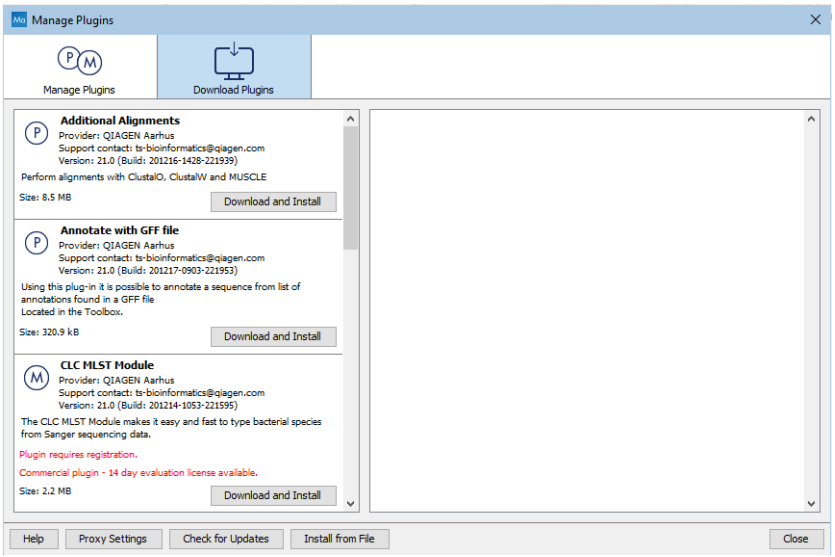


Figure 9.1: *Plugins and modules available for installation are listed in the Plugin Manager under the Download Plugins tab.*

If you have a .cpa installer file for Whole Genome Alignment, you can install it by clicking on the **Install from File** button at the bottom of the Plugin Manager.

If you are working on a system not connected to the internet, plugin and module .cpa files can be downloaded from `https://digitalinsights.qiagen.com/products-overview/plugins/`using a networked machine, and then transferred to the non-networked machine for installation.

Restart to complete the installation

Newly installed plugins and modules will be available for use after restarting the software. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the *CLC Workbench*.

9.2 Uninstalling plugins

Plugins and modules are uninstalled using the Workbench Plugin Manager.To open the Plugin Manager, click on the **Plugins ( ) button** in the top Toolbar, or go to the menu option:

Utilities | **Manage Plugins... ( )**

This will open the Plugin Manager (figure 9.2). Installed plugins and modules are shown under the Manage Plugins tab of the Plugins Manager.

To uninstall a plugin or module, click on its entry in the list, and click on the **Uninstall** button.

Plugins and modules are not uninstalled until the Workbench is restarted. When you close the Plugin Manager, a dialog appears offering the opportunity to restart the *CLC Workbench*.

Disabling a plugin without uninstalling it

If you do not want a plugin to be loaded the next time you start the Workbench, select it in the

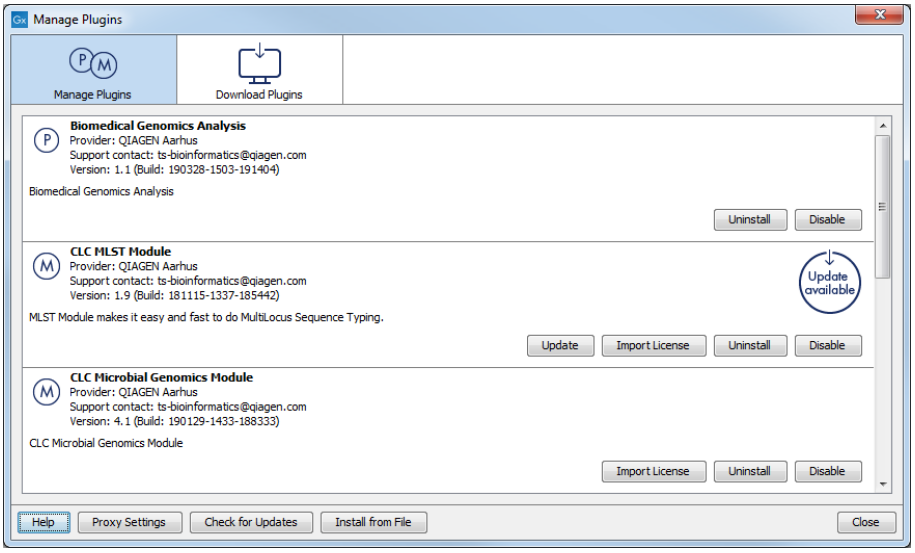


Figure 9.2: *Installed plugins and modules are listed in the Plugins Manager under the Manage Plugins tab.*

list under the Manage Plugins tab and click on the **Disable** button.

Bibliography

[Chiaromonte et al., 2002] Chiaromonte, F., Yap, V., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. *Pacific Symposium on Biocomputing*, pages 115–26.

[Darling AE, 2010] Darling AE, Mau B, P. N. (2010). progressivemauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*, 5(6).