Introduction to CLC Main workbench

Dedan Githae, Bioinformatics
d.githae@cgiar.org

15th Anniversary biosciences eastern and central africa

NEPAD TRANSFORMING AFRICA

NC STATE UNIVERSITY

ILRI INTERNATIONAL LIVESTOCK RESEARCH INSTITUTE

Prelude

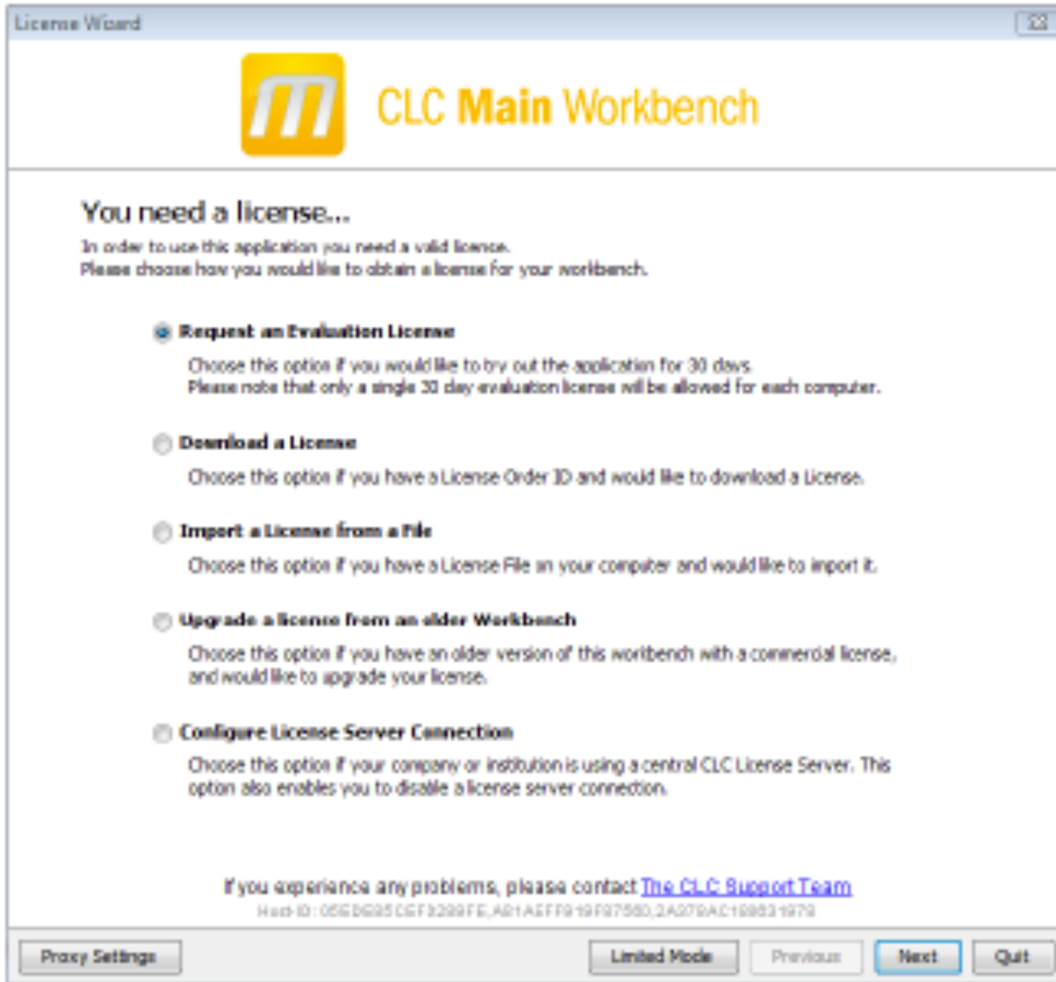
- Download and Installation

We have pre-downloaded different CLC main workbench softwares. Select the appropriate CLC program for your computer from the course website, according to your OS and install.

- Licences

Licences have been acquired for every participant. When you run CLC for the first time, the license assistant wizard pops up.

*Please note you need to have administrative privileges to install programs



Select the:

Download a license.
When you purchase a license, you will get a license ID from CLC bio. Using this option, you will get a license based on this ID.

Importing sequences

Importing sequences means reading the raw data generated, into the software workspace for analysis.

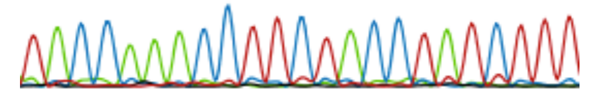
Import | Browse to where the sequences are | Select "filetype: Trace files (.abi/.ab1/.scf/.phd**)" Next | Select / Create folder where to save the sequences | finish**

Viewing the sequences

To view the sequences / output from the sequencer, double click on any of the sequences.

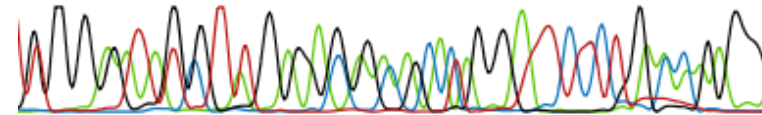
• Good quality: smooth, distinct, evenly spaced peaks, Little baseline noise.

TACCAAACCTTCTACCTATCTTT



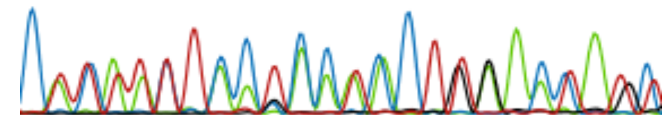
• Non-usable sequenced data: can be due to low concentration of DNA template, none or wrong primer added.

TGGAAAGGTTGAAGAAACCGGATCTTGACGGG



• Double peaks: multiple peaks of same or different height at same position: this is due to clone contamination, heterozygous positions: two bases at same position (SNP), contaminated PCR reaction.

CTTATTTCCTCCTACTTGACTATC



Ambiguous codes can be used

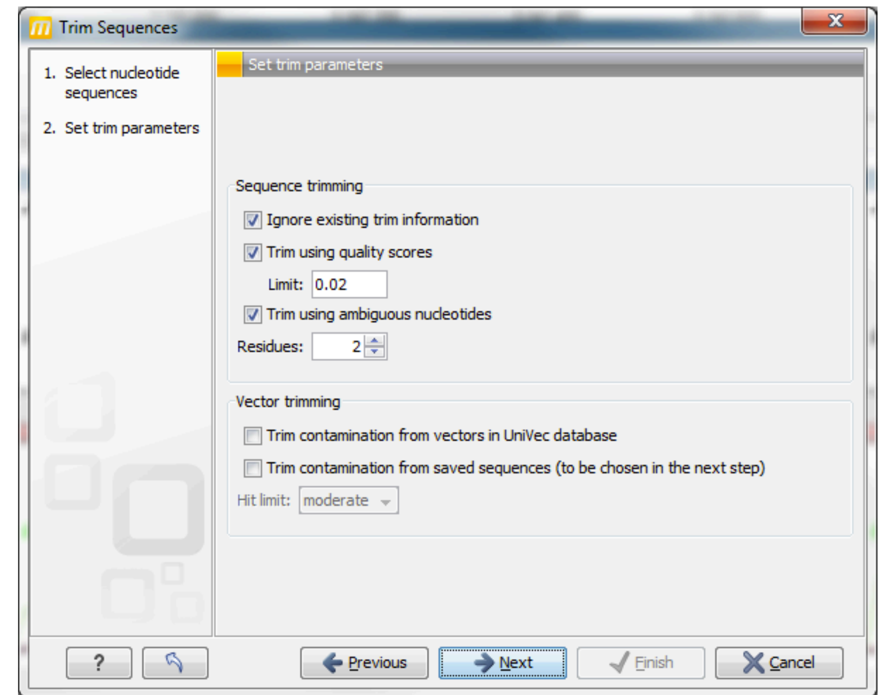
R	A or G	M	A or C
Y	C or T	B	C or G or T
S	G or C	D	A or G or T
W	A or T	H	A or C or T
K	G or T	V	A or C or G
		N	any base

Trimming sequences

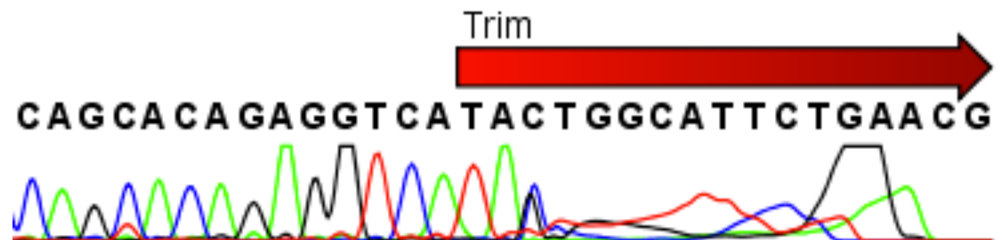
TRIMMING SEQUENCES:

The first thing to do when analyzing sequencing data is to trim the sequences.

Trimming serves a dual purpose: it both takes care of parts of the reads with poor quality, and it removes potential vector contamination. Trimming the sequencing data gives a better result in the further analysis.

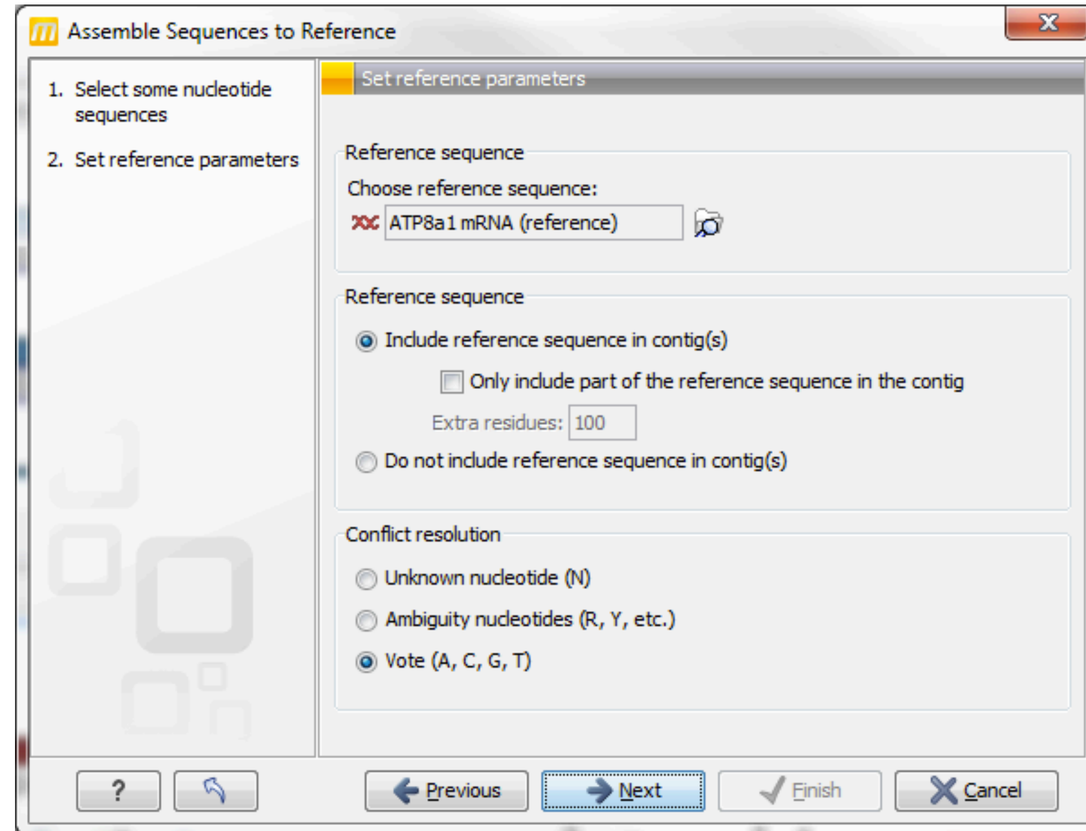


When the trimming is performed, the parts of the sequences that are trimmed are actually annotated, not removed



Sequence assembly

This is the technical term for aligning the sequences where they overlap and reverse the reverse reads to make a contiguous sequence (also called a contig)



If you have a reference sequence, use **the Assemble sequences to Reference** option;

else

Use **Assemble sequences** to pair the forward and reverse sequences

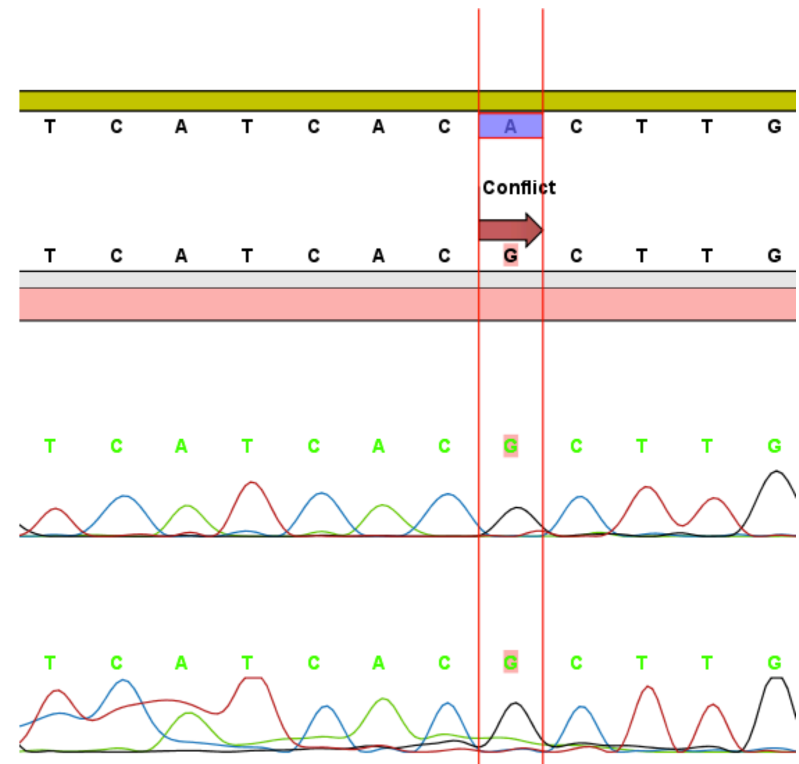
Finding and editing conflicts

- Click the **Find Conflict** button at the top of the Side Panel or press the Space key to find the first position where there is disagreement between the reads
- For example, supposing the first read has a "T", whereas the second line has a gap. In order to determine which of the reads we should trust, we assess the

Sequence assembly

You can edit the trace files whenever
 there is a conflict using your judgement.

One other way of handling conflicts is using
ambiguous codes. This is particulary important
In instances that both trace files are clear and
Good quality, thus showing heterozygosity



Consensus sequences

For further analysis, use the consensus sequences generated by resolving all conflicts.

Right-click the name "Consensus" | Open Sequence | Save

The sequence can be exported to a folder in your computer and used for further analysis.

Export | Export sequences and sequence lists in “fasta “ format

Other functionalities

Other functionalities of CLC main workbench include:

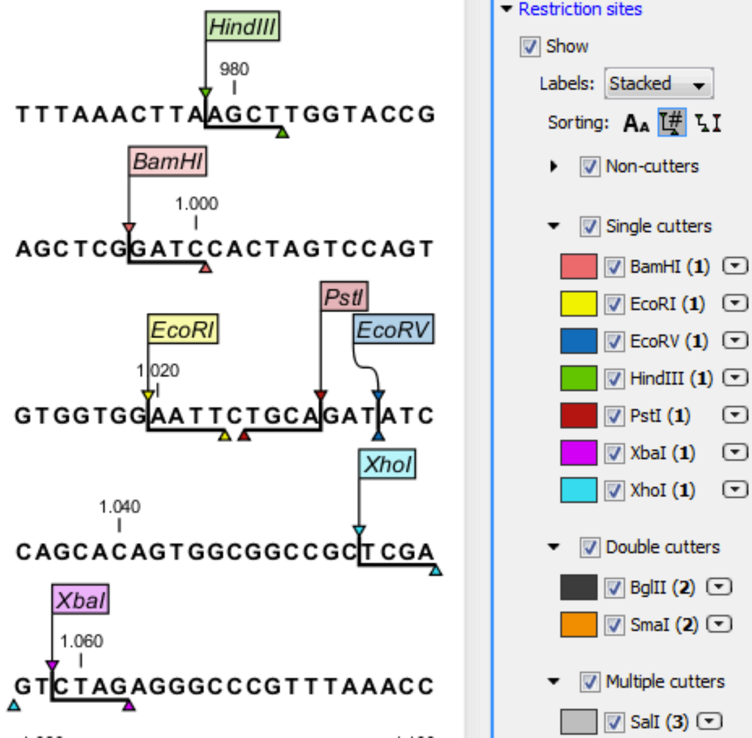
- Primer design
- Restriction enzyme analysis
- Sequence alignments
- Creation of phylogeny trees
- DNA > protein translation
- DNA <-> RNA conversion
- Reverse complementing sequences
- Searching for ORFs
- More advanced analysis such as
  - Protein charge plots
  - Predicting domains and searching using pfam
  - Finding cleavage sites
  - Online BLAST searching

Restriction analysis

There are two ways of finding and showing restriction sites.

- The dynamic restriction sites found in the Side Panel of sequence view (Fig 1)
- The toolbox way: **Toolbox | Cloning and Restriction Sites | Restriction Site Analysis**.

This way provides more control of the analysis and gives you more output options, e.g. a table of restriction sites and a list of restriction enzymes that can be saved for later use (Fig 2)



You can view a virtual gel from the Resulting fragments.

Hovering over the icons on the Restriction map, click on the "**show gel**" option. This also gives you The opton to input show a marker ladder, And the base pair sizes seperated with comas.

Enjoy CLC during the hands-on session

Sequence assembly