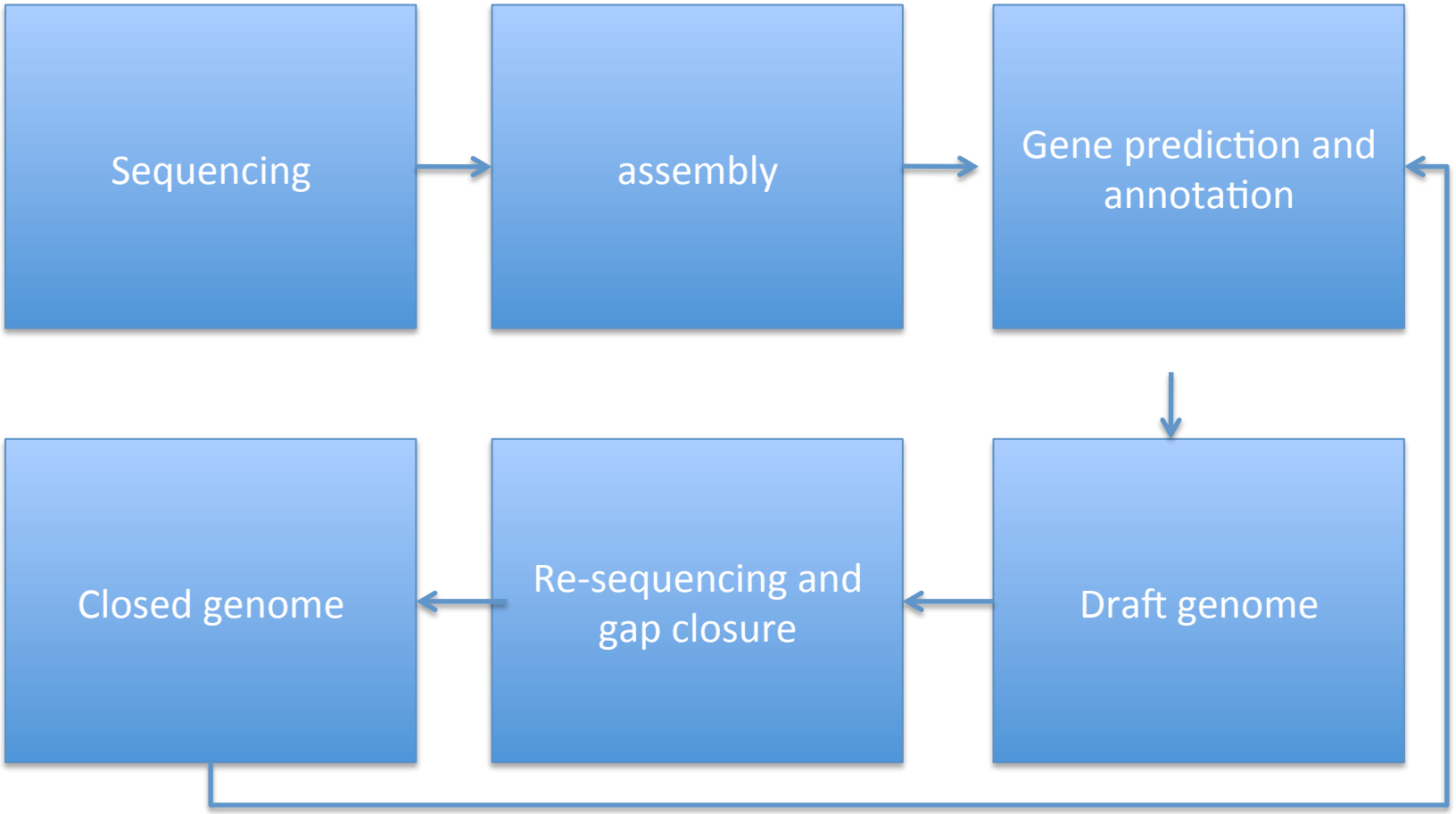
Lecture 8

Genome data analysis 6: Post assembly genome improvement

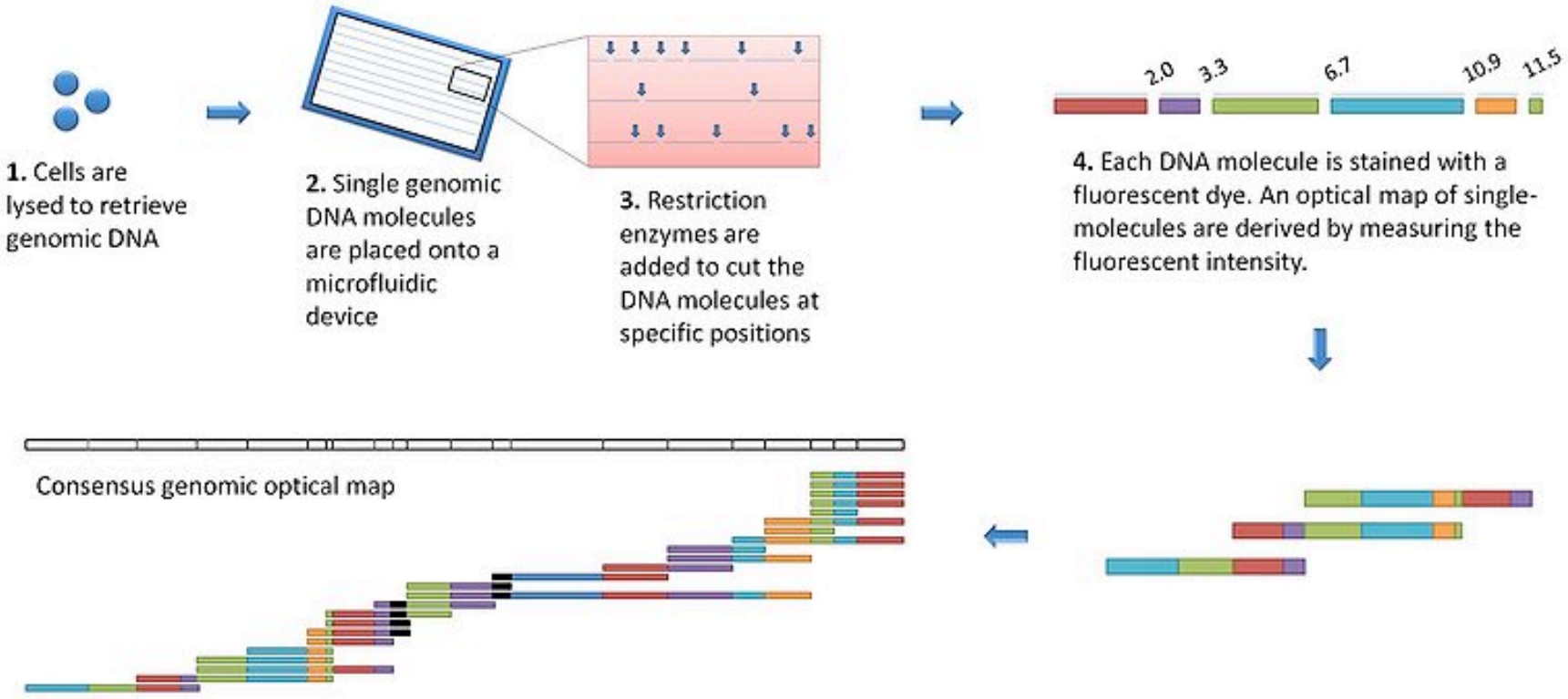
The whole process



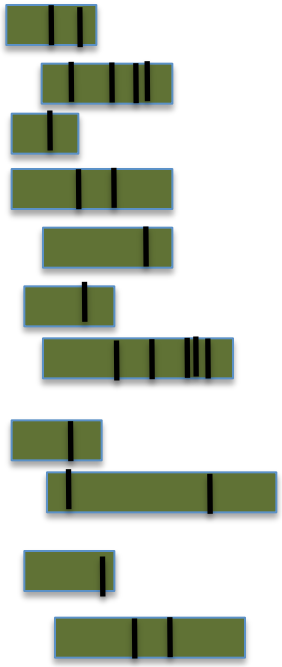
Post assembly genome improvement

- Scaffolding
- Gap closure
- Re-sequencing (Mapping and fixing errors)

Optical Mapping



Optical Mapping

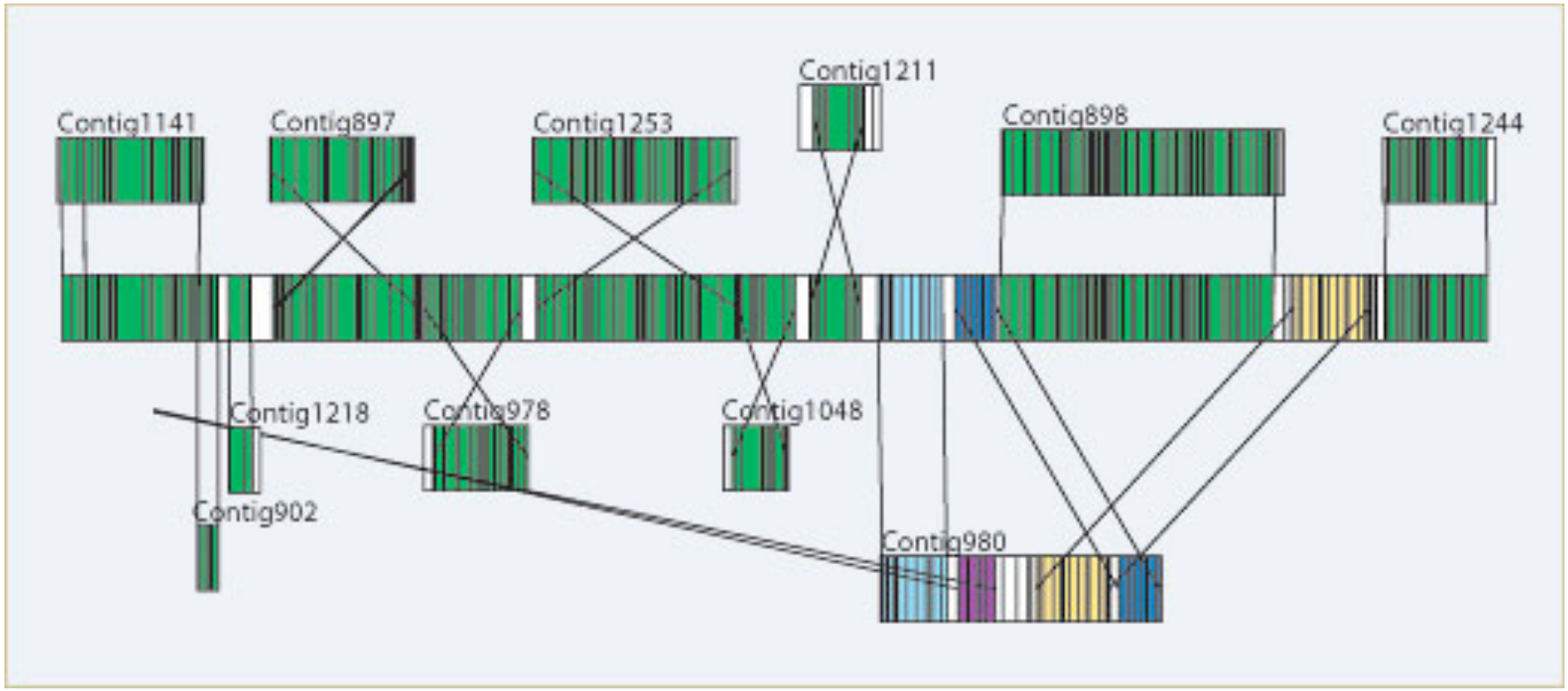


Contigs

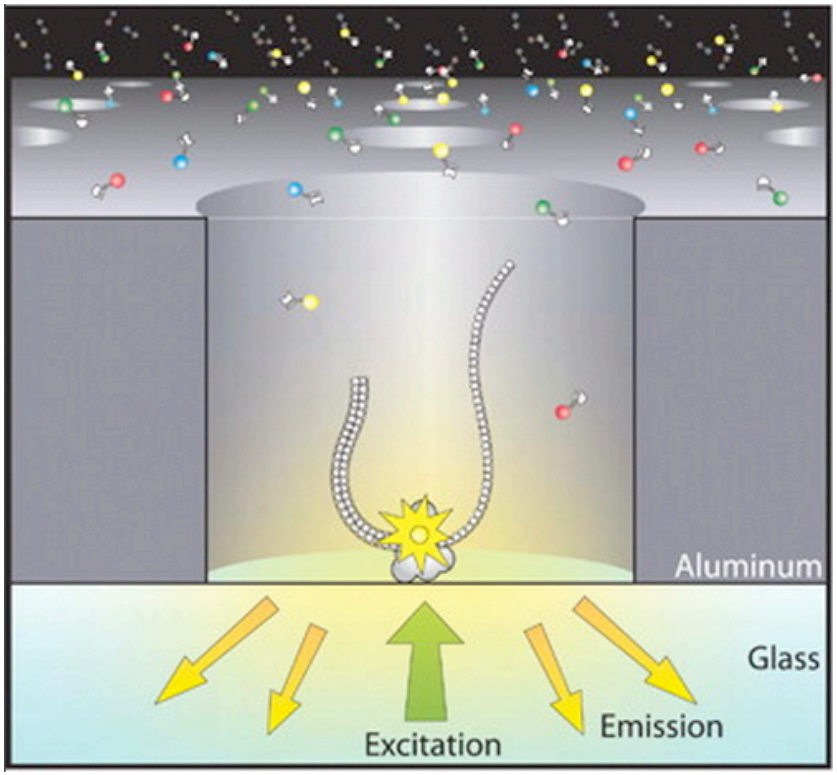
Map the contigs into the optical map

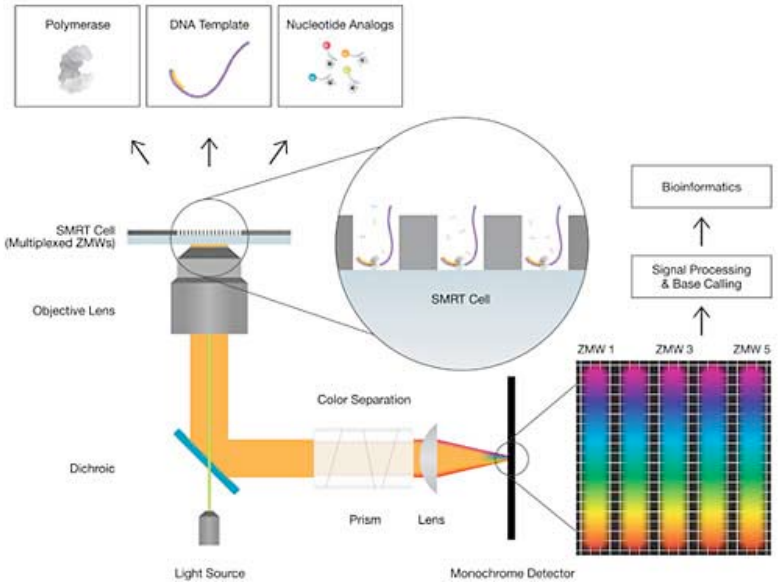
Optical map

Optical Mapping



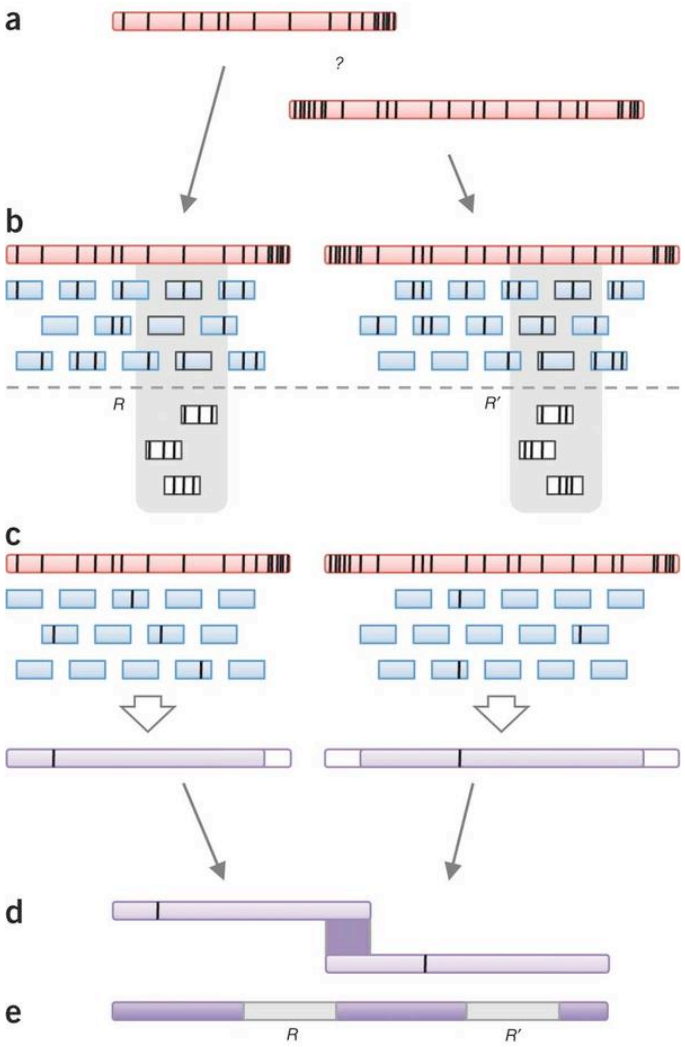
PacBio Sequencing



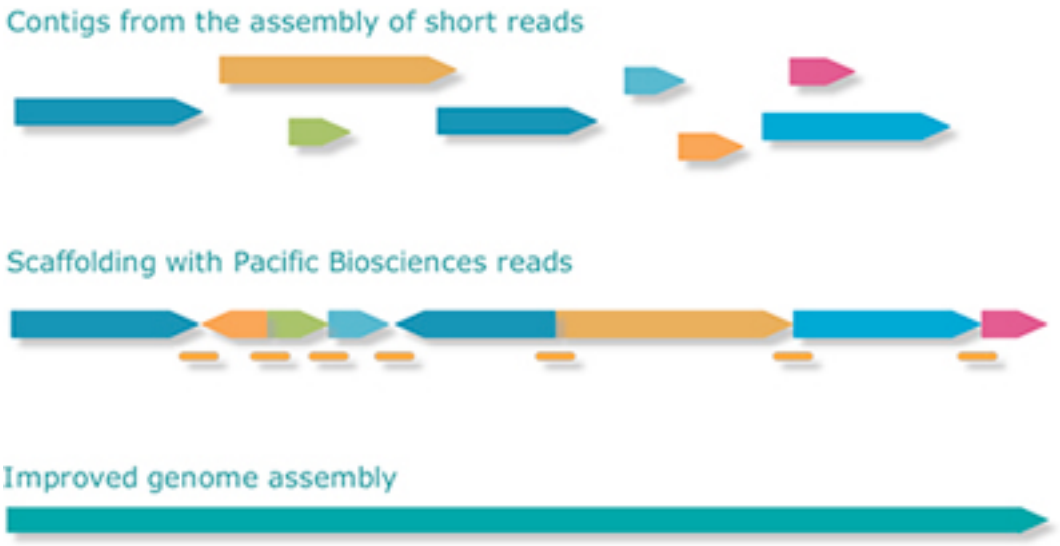


PacBio Sequencing

Error correction with short reads



PacBio Sequencing



Alignment with referential genome

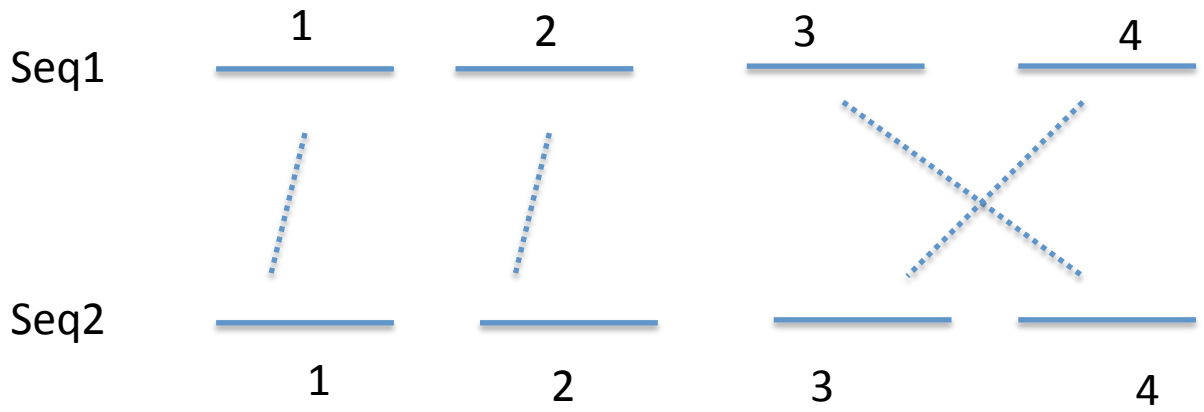
MUMMER ANI and ABACAS

Mummer

- Obtaining MUMs: Maximal Unique Matches

MUMs occur exactly one in each sequence (repeats sequences are ignored) 20 bp minimal

- MUM choosing for alignment
- Gap closing



NUCMER and PROMER

NUCMER
-Finds maximal exact matches of a given length
-Clusters matches to form larger inexact alignment regions
-Extends alignments outward from each of the matches to join the clusters into a single high scoring pair-wise alignment.
NUCMER for comparison of close related organisms

PROMER
Same steps used in nucmer but before any of the exact matching takes place, the input sequences are translated in all six amino acid reading frames.
PROMER for comparison more “distant related” organisms

WORKING with Mummer

- Compare *Candidatus liberibacter strains*
- Compare *Candidatus liberibacter* solanacearum
- With *Candidatus liberibacter americanus S Paulo St*
- Compare your assembly with file CLsolanacearum.fasta (*Candidatus liberibacter* solanacearum
- The basic command
- nucmer options Reference query
- This will generate a out.delta file
- The out.delta can be used for the downstream analysis:

WORKING with Mummer

- Basic commands you could use additional parameters to tune the out put
- Plot the alignment:
- mummerplot -p results --postscript out.delta
- show-snps out.delta > out.snps
- show-coords out.delta > out.coor

Use the Average nucleotide index ANI for your assembly

ANI calculators

http://enve-omics.ce.gatech.edu/ani/

ANI.pl : Scripts that uses blast to compare two seuqneces and avarege the nucleotide indetity. This scripts uses blast , Instead of blast mummer could be used

ANI.pl --fd formatdb --bl blastall --qr one strain genome --sb the other strain genome --od output directory

Run your assembly and obtain the ANI in comparison to CLsolanacearum.fasta

Alignment of microbial genomes with parsnp

-Microbial core genome alignment and SNP detection
-Generation of phylogenetic tree based on core SNPs
-SNPs between genomes

Basic command:
parsnp –r reference_genome –d folder_with your_sequences –o results

Open the tree file in seawiew or other tree editor program