QUAST assembly evaluation

Karin Lagesen

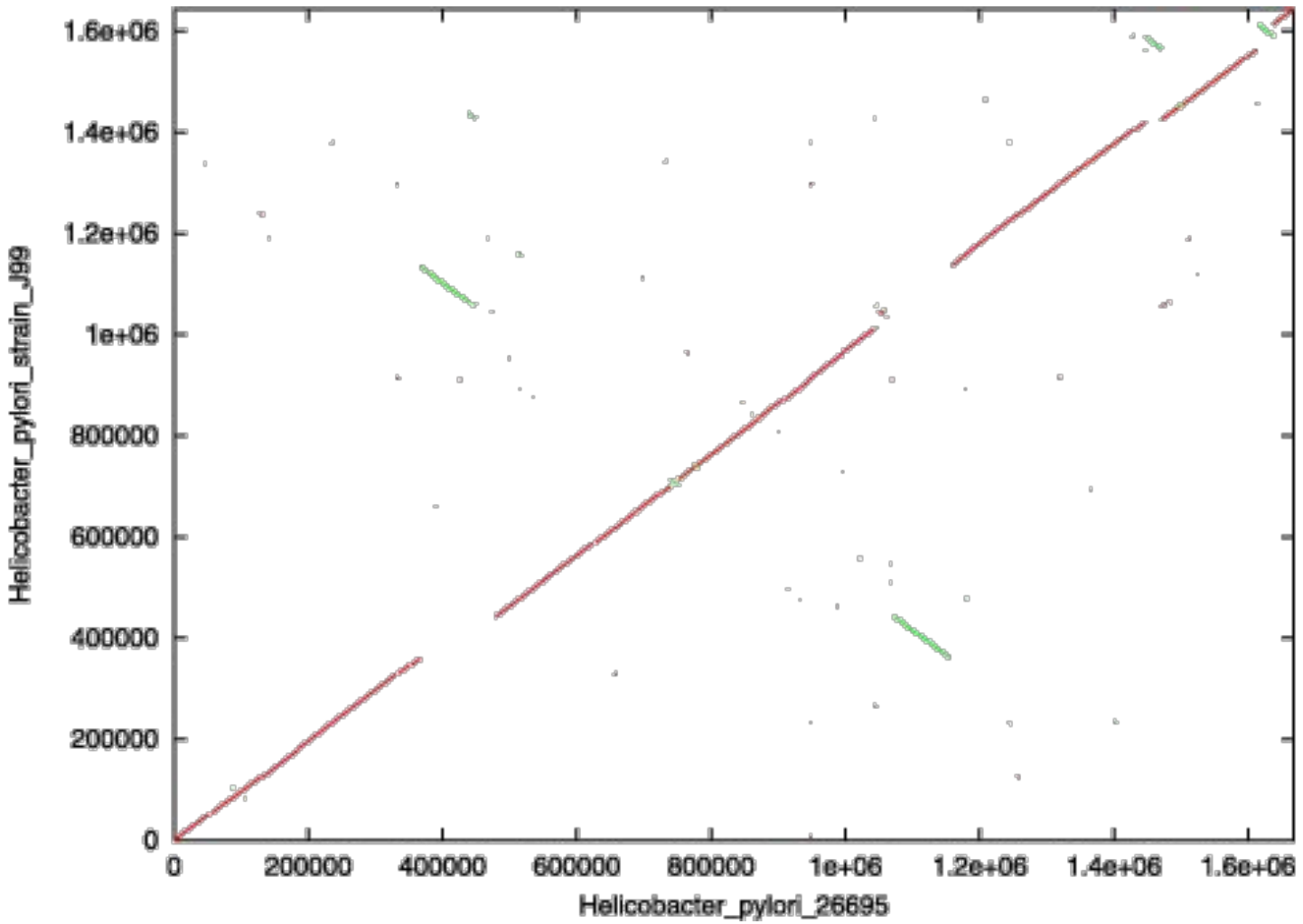
karin.lagesen@vetinst.no

QUAST statistics

- Contig size information
- Misassemblies and structural variations
- Genome features found in the assembly
- Variations on N50 statistics

- Visualization

QUAST uses *nucmer* from MUMmer package



Gene finding and genes

- Will also look for reference genes in genome
- Can ask for gene finding too - can show discrepancy
- Can also show rRNA genes

- Genes found indication of genome completeness
- Lots of partial genes: broken assembly

Mapped reads

- Can input reads
- Get coverage graph

- Informative regarding high copy regions/”lost” regions

- Note: can only “see” what is in the reference

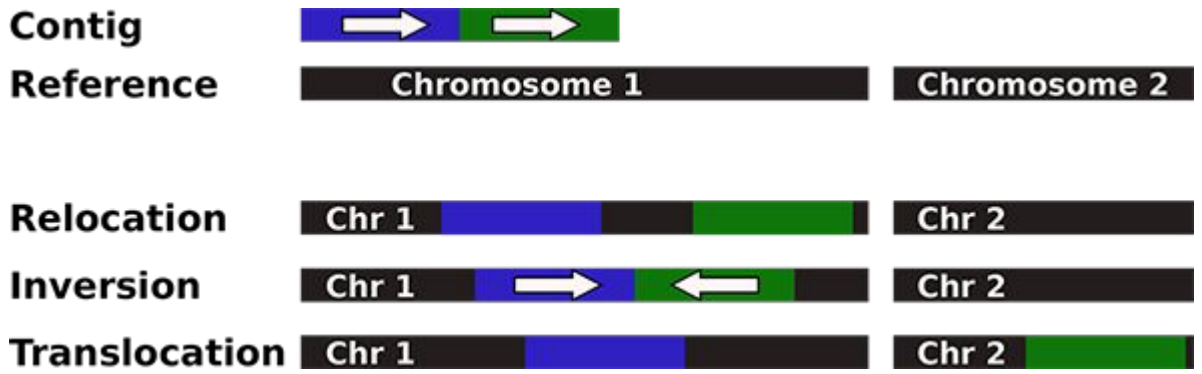
Metrics

Aligned to "BCep_ref" | 8 605 945 bp | 4 fragments | 66.61% G+C
7705 genomic features

Worst Median Best ☑ Show heatmap

Genome statistics	spades_default	spades_kmers_careful	megahit_default	megahit_min_count_3
Genome fraction (%)	98.14	98.421	98.6	98.603
Duplication ratio	1	1	1.001	1.001
# genomic features	7539 + 75 part	7563 + 62 part	7540 + 105 part	7540 + 104 part
Largest alignment	455 950	505 898	350 746	350 746
Total aligned length	8 436 553	8 470 789	8 492 473	8 493 177
NGA50	143 431	198 969	144 083	125 159
LGA50	18	13	20	21
Misassemblies				
# misassemblies	6	6	9	8
Misassembled contigs length	1 469 048	1 719 134	1 200 775	1 050 989
Mismatches				
# mismatches per 100 kbp	3.85	2.72	2.38	2.3
# indels per 100 kbp	0.67	0.45	0.32	0.28
# N's per 100 kbp	0	0	0	0
Statistics without reference				
# contigs	132	91	156	158
Largest contig	754 490	961 949	539 126	539 126
Total length	8 447 218	8 472 540	8 492 975	8 493 797
Total length (>= 1000 bp)	8 447 218	8 472 540	8 492 975	8 493 797
Total length (>= 10000 bp)	8 324 069	8 384 754	8 296 360	8 308 919
Total length (>= 50000 bp)	7 438 644	7 723 080	6 917 725	6 910 273

Misassemblies



- Relocation is a misassembly event (breakpoint) where the left flanking sequence aligns over 1 kbp away from the right flanking sequence on the reference genome, or they overlap by more than 1 kbp, and both flanking sequences align on the same chromosome.
- Translocation is a misassembly event (breakpoint) where the flanking sequences align on different chromosomes.
- Inversion is a misassembly event (breakpoint) where the flanking sequences align on opposite strands of the same chromosome.

Icarus viewer