

Supplementary Table 2. Assembly statistics using CLC Genomics Workbench with different datasets.

	Untrimmed dataset	Trimmed dataset 1	Trimmed dataset 2
Total reads	106,660,317	106,660,317	106,660,317
Matched reads	88,140,071	88,879,703	91,038,792
Not matched reads	18,520,246	17,780,614	15,621,525
Contigs total	128,281	123,737	114,708
Contigs ≥ 100 bp	127,272	122,790	113,893
N50 (bp)	978	1,029	1,151
Largest contig length (bp)	15,709	15,684	15,684
Smallest contig length (bp)	100	100	100
Average contig length (bp)	398.32	406.40	428.47

The N50, largest, smallest and average contige length are given for only the contigs ≥ 100 bp.