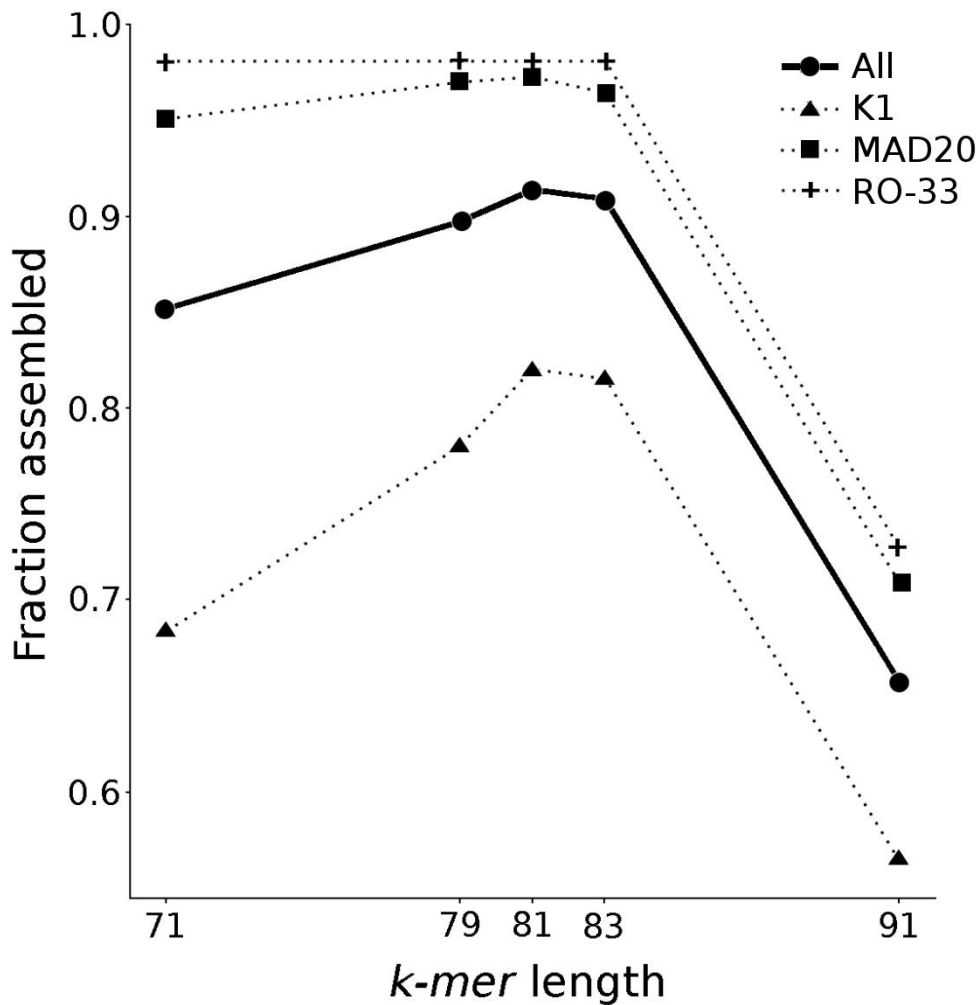


Additional file 5.



**The effect of *k-mer* length on the fraction of *msp1* block 2 sequences assembled by Velvet.** 964 *msp1* block 2 sequences from previously published long reads were used to create dummy reads of 100 bp in length with a coverage of 50-fold. These reads were assembled using Velvet (Zerbino and Birney, *Genome Res* 2008, **18**:821-829) with a range of *k-mer* lengths. The resulting contigs were then scanned for the presence of the correct *msp1* block 2 sequence. The fraction of *msp1* block 2 sequences that were fully assembled for all 964 sequences is shown (solid line). The fraction of sequences fully assembled for *K1-like* sequences (triangles, n = 392), *MAD20-like* sequences (squares, n = 354) and *RO33-like* sequences (crosses, n = 204) are also shown (dotted lines).