Telogator: a method for reporting chromosome-specific TL in humans from long read sequencing data – Supplementary Data

December 10, 2021
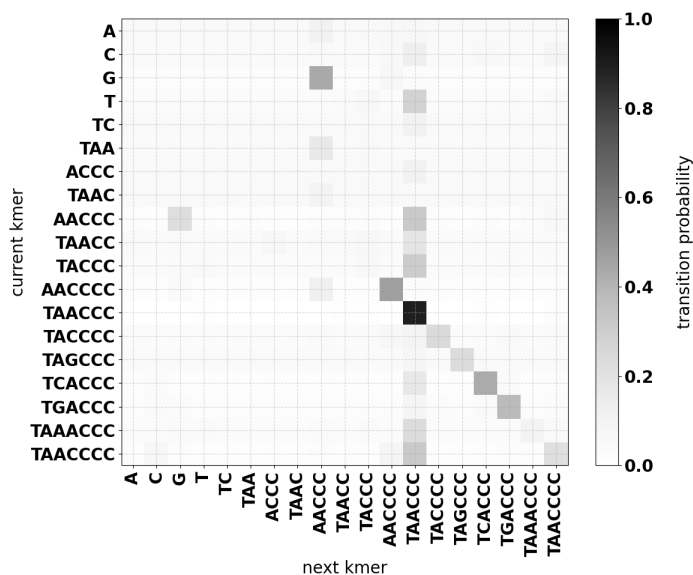

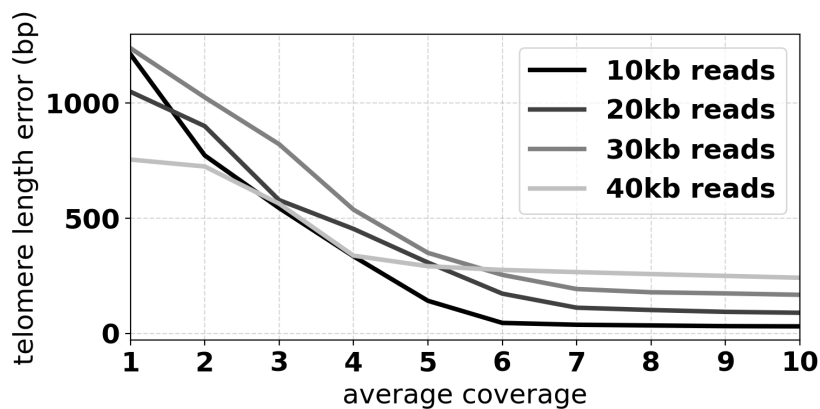Figure 1: Telomere kmer transition matrix computed from T2T telomere sequences.

Figure 2: Median TL error vs. coverage.

kmer sequence	type
CCCCAA	j-type
CCCTAA	t-type
CCCTCA	g-type
CCCTGA	c-type
CCCAA	other variations
CCCCTA	
CCCCTAA	
CCCGA	
CCCTA	
CCCTAAA	
CCCTAG	
CCTA	
CCTAA	

Table 1: Frequent kmers extracted from T2T telomere sequences (forward strand).

SRA Run ID (short reads)	SRA Run ID (long reads)	Sample ID
SRR14724532	SRR11008518	HG002
ERR3241754	ERR3861414	HG00731
ERR3241755	ERR3861405	HG00732
ERR3988973	ERR3861412	HG02011
ERR3989019	ERR3861400	HG02492
ERR3989118	ERR3861397	HG03065
ERR3989162	ERR3861407	HG03371
ERR3989199	ERR3861403	HG03683
ERR3989317	ERR3861409	NA12329
ERR3989454	ERR3861402	NA19983

Table 2: SRA run IDs for short and long reads for 10 samples processed by Telogator.
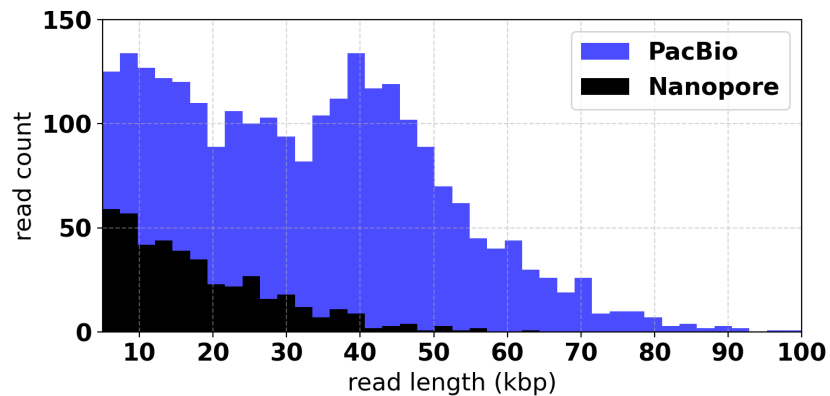
Figure 3: Read count and length for reads spanning telomere/subtelomere boundaries in HG002 PacBio vs. Nanopore.
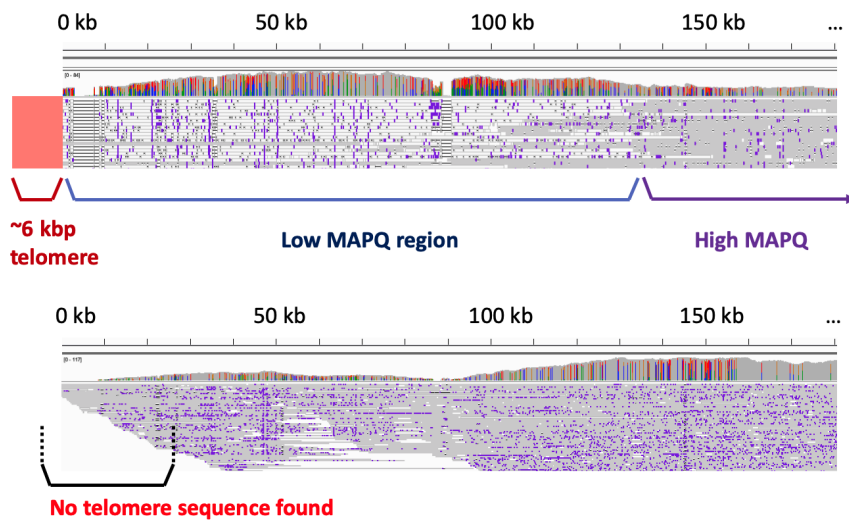

Figure 4: Annotated screenshot showing very long Nanopore reads (top) and PacBio CLR reads (bottom) aligned to alternate subtelomere 8p assembly. The large region of low mapping quality is due to sequence similarity with subtelomere 19p.