LTR Parser Results

Kevin McCormick

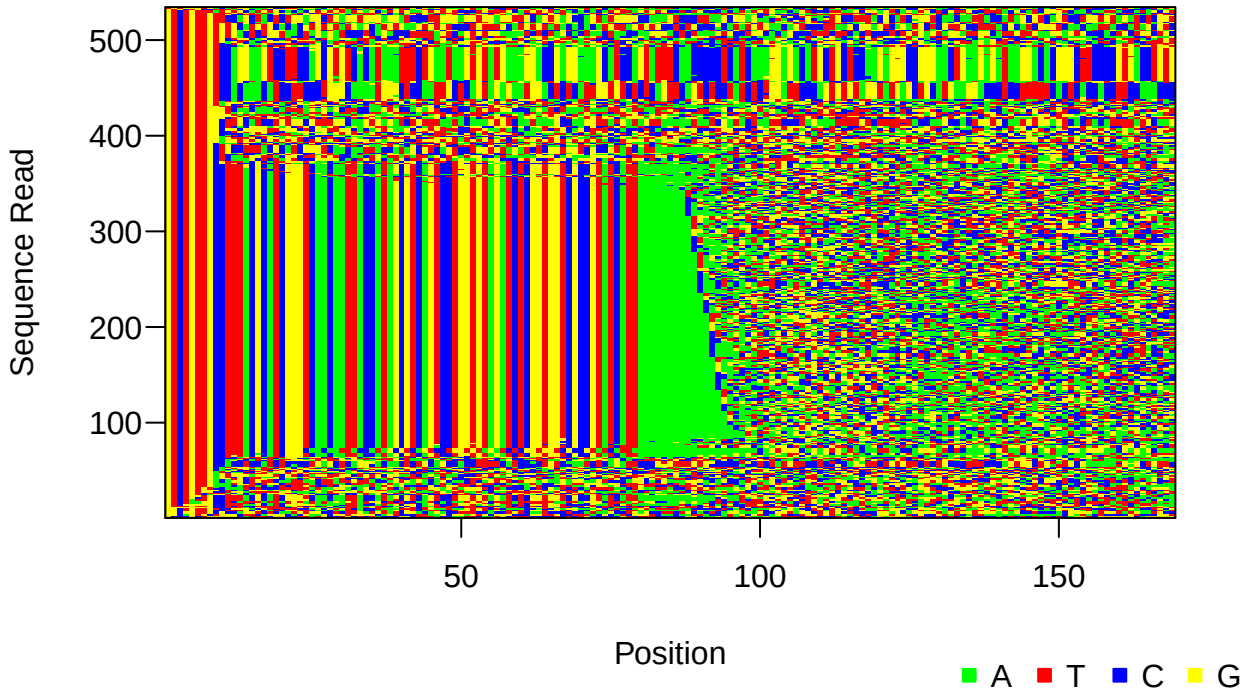
20 April, 2020

Overview

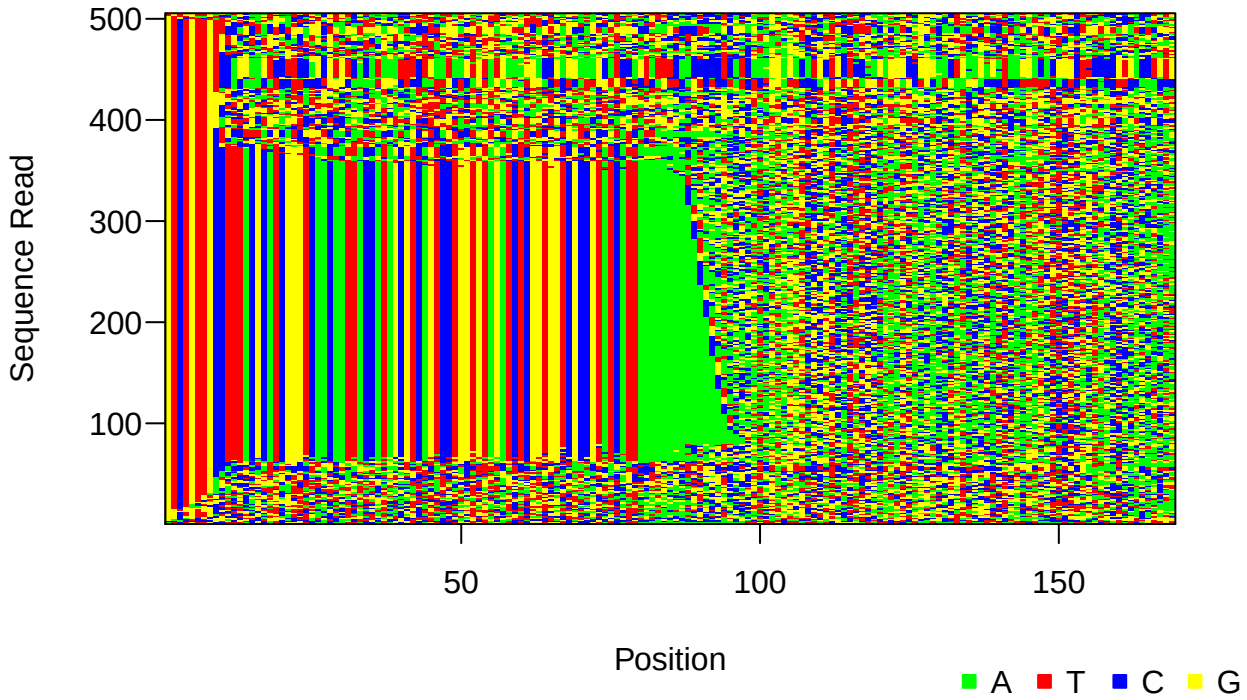
This report contains the results of attempted primer & LTR identification in the following FASTQ file:
/home/kevin/dev/ltrparser/testdata/20200303293TControl-1_u5.R2.fastq.gz

Total number of reads in sample	534
Number of distinct reads in sample	505
Number of copies of most common read	16

Visualization of 534 sequencing reads, selected randomly and arranged by nucleotide sequence:



Visualization of 505 distinct sequencing reads (after removing duplicates), selected randomly and arrangedy by nucleotide sequence:



We found 4 potential primer-LTR pairs, 0 of which has an LTR sequence that mapped to an LTR region of at least one genome in the Los Alamos National Laboratories HIV database. The "count" column here is the number of reads with each LTR-primer pair that also mapped to the human genome.

primer	count	LTR	LANL
GTCTGTTG	18	CCTTTACGCATCGGGTCAACAATTACCATAGCGTCA	FALSE
GTCTGTTG	18	CCTTTACGCATCGGGTCAACAATTACCATAGCGTCAGTCCTGGTGTAGATCTCGGTGGTCGCA	FALSE
GTCTGTTG	18	CCTTTACGCATCGGGTCAACAATTACCATAGCGTCA	FALSE
GTCTGTTG	18	CCTTTACGCATCGGGTCAACAATTACCATAGCGTCAGTCCTGGTGTAGATCTCGGTGGTCGCA	FALSE

Visualization of 1000 sequencing reads filtered for common primer & LTR, selected randomly and arranged by nucleotide sequence (these reads may or may not map to the human genome):