LTR Parser Results

Kevin McCormick

22 June, 2020

Overview

This report contains the results of attempted primer & LTR identification in the following FASTQ file:

/home/kevin/dev/ltrparser/testdata/demo/fastq.gz/hivTest1_u3.R2.fastq.gz

Siliciano project, sample: GTSP3034-4 (U3)

Analysis of common primer-LTR pairs

A total of 4 potential primer-LTR pairs were identified.
This table shows how the reads captured by each primer-LTR pair compare to each other and to the raw FASTQ file.
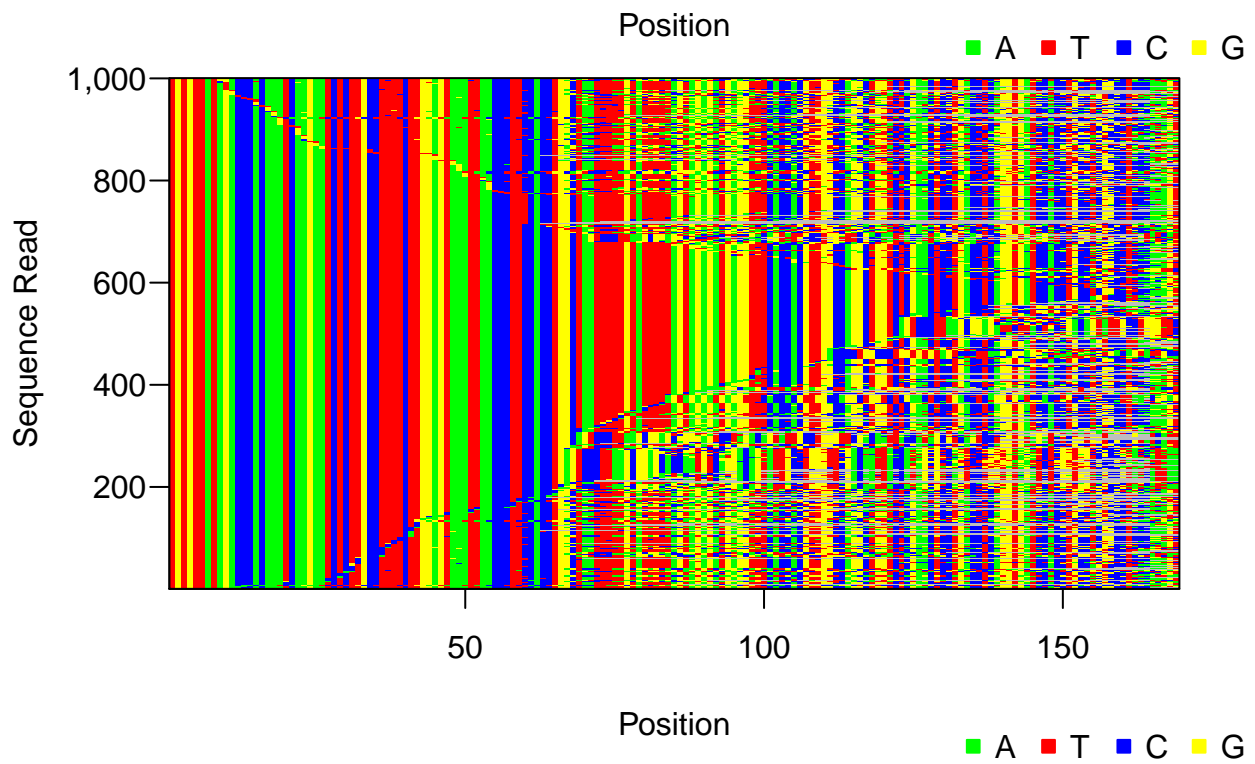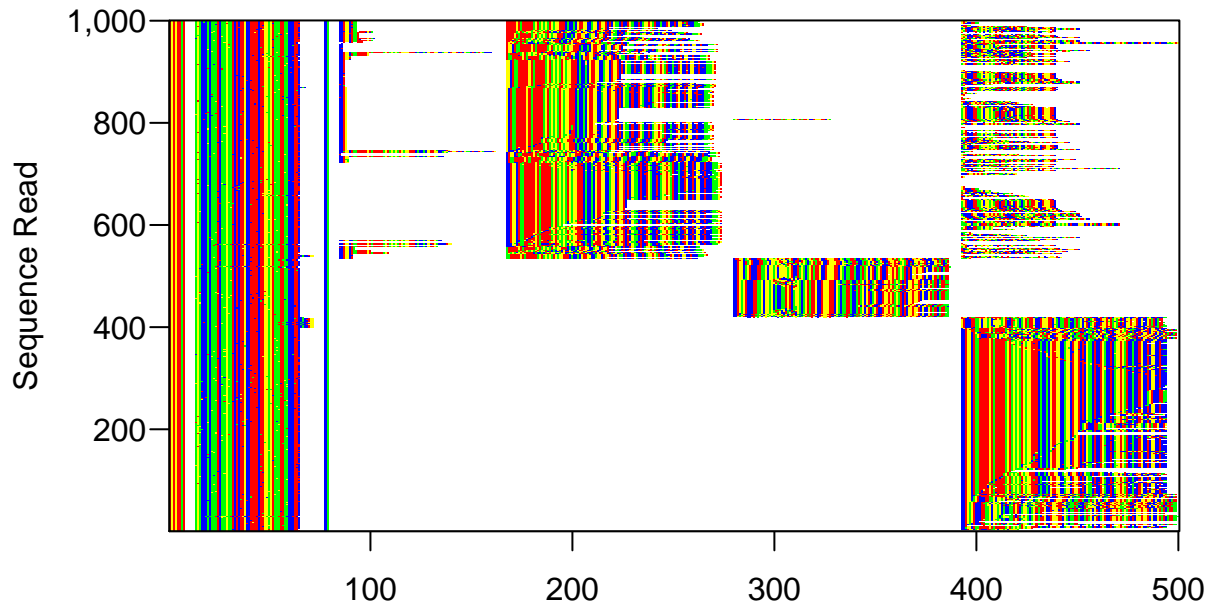
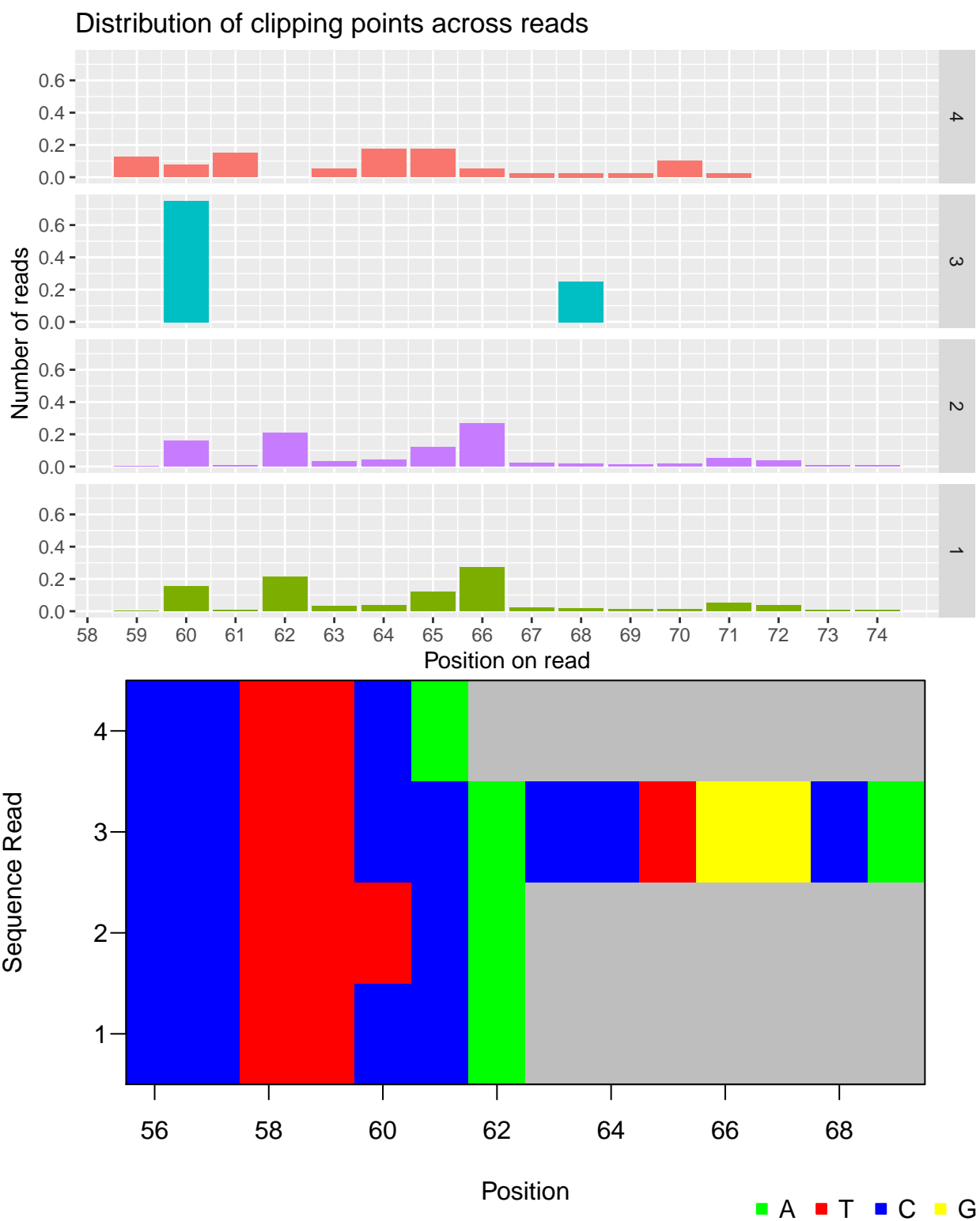| id	primer	LTR	expectedLTR	LANL
1	TGTGTTAT	AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCCA	FALSE	TRUE
2	TGTGTTAT	AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTTCA	FALSE	TRUE
3	TGTGTTAT	AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCCACCTGGCA	FALSE	TRUE
4	TGTGTTAT	AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCA	FALSE	TRUE

Variable	Total	1	2	3	4
Distinct	5989	5243 (87.5%)	4928 (82.3%)	218 (3.6%)	97 (1.6%)
DistinctHuman	2670	2417 (90.5%)	2263 (84.8%)	5 (0.2%)	42 (1.6%)
DistinctLinker	796	668 (83.9%)	626 (78.6%)	1 (0.1%)	35 (4.4%)
Max	3854	3854 (100%)	3854 (100%)	222 (5.8%)	18 (0.5%)
MaxHuman	3854	3854 (100%)	3854 (100%)	2 (0.1%)	9 (0.2%)
MaxLinker	1416	1416 (100%)	1416 (100%)	1 (0.1%)	18 (1.3%)
Total	24100	21797 (90.4%)	21479 (89.1%)	548 (2.3%)	163 (0.7%)
TotalHuman	11707	11344 (96.9%)	11188 (95.6%)	6 (0.1%)	78 (0.7%)
TotalLinker	3113	2930 (94.1%)	2888 (92.8%)	1 (0%)	72 (2.3%)

Visualization of reads

Reads are sectioned off into sections separated by white (nucleotide color legend is below the plot).

Primer /// LTR /// CA /// (unmatched) /// human-matching /// linker-matching /// (unmatched)

Distribution of clipping points across reads

```
logo      AGAcCCACAaATCAAGAATcTCTTGcCTTtTcTtGGAGTaAATTAaCCCTTccA CCTGGCA

4         AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCA........    53
3         AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCCACCTGGCA    61
2         AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTTCA.......    54
1         AGACCCACAAATCAAGAATCTCTTGCCTTTTCTTGGAGTAAATTAACCCTTCCA.......    54
HXB2      AGATCCACAGATCAAGGATATCTTGTCTTCTTTGGGAGTGAATTAGCCCTTCCA.......    54
consensus !!!*!!!!!*!!!!!!!*!!*!!!!!*!!!*!*!*!!!!!*!!!!!*!!!!!***
```

X   non conserved

X   ≥ 50% conserved