

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/hivTest2_u5.R2.fastq.gz

Jones project, sample: GTSP3588-1 (U5)

Analysis of common primer-LTR pairs

A total of 8 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.

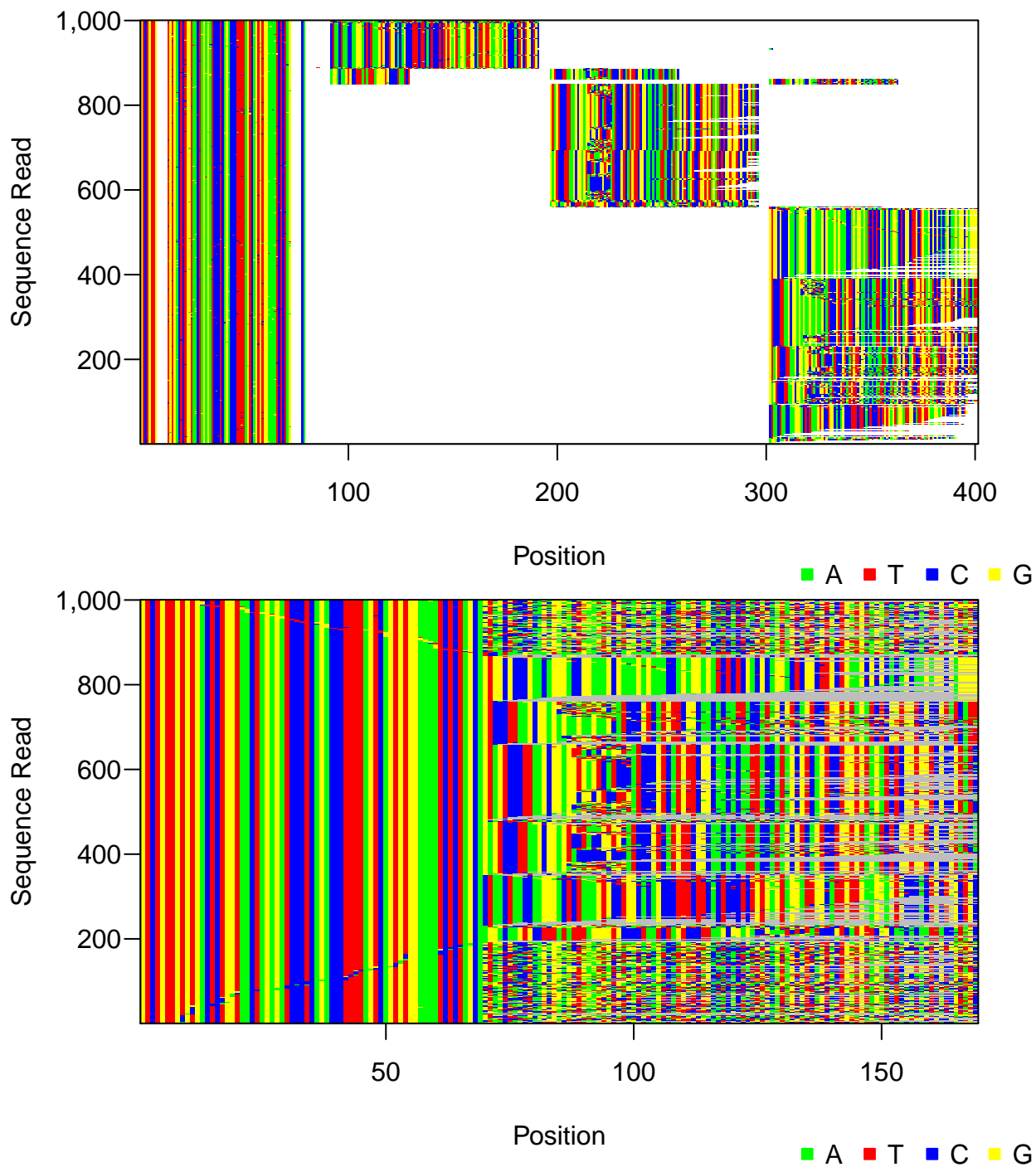
id	primer	LTR	expectedLTR	LANL
1	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	TRUE	TRUE
2	GTCTGTTG	TGCGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
3	GTCTGTTG	CGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
4	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
5	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
6	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
7	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE
8	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA	FALSE	TRUE

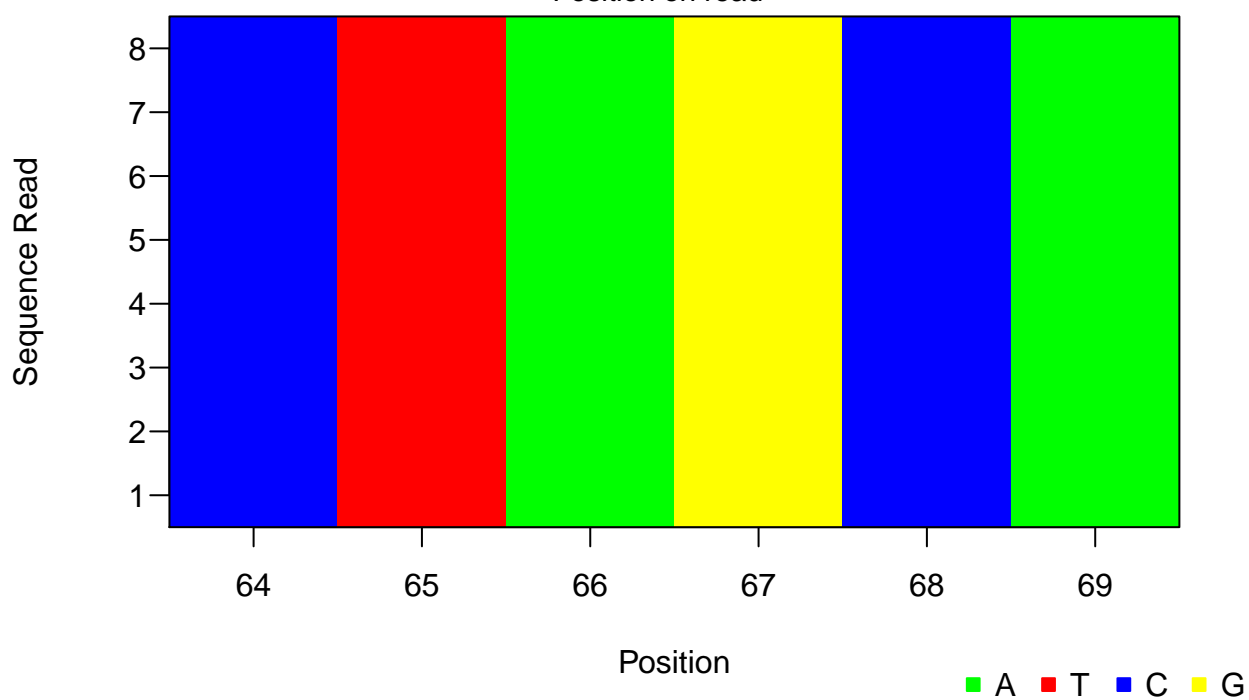
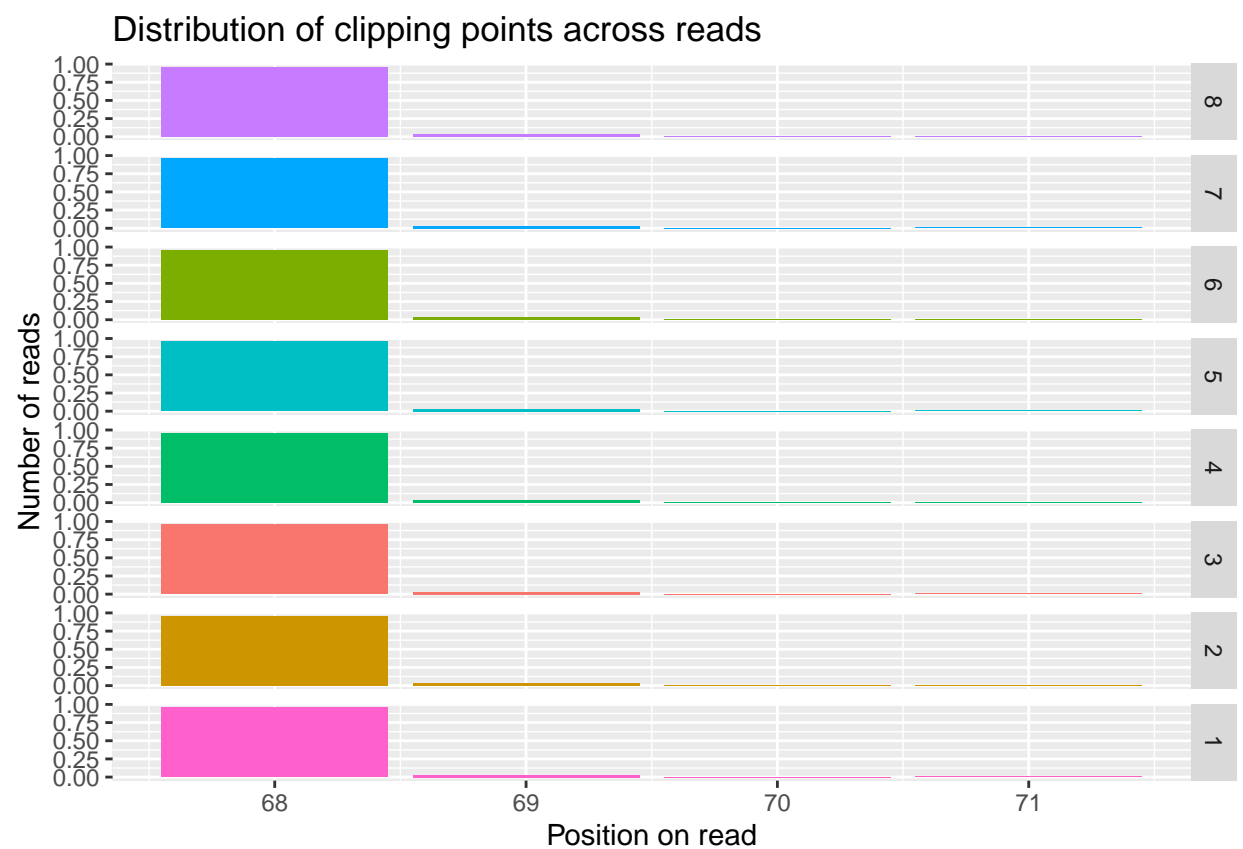
Variable	Total	1	2	3	4	5	6
Distinct	5751	3134 (54.5%)	3068 (53.3%)	3065 (53.3%)	3060 (53.2%)	3060 (53.2%)	3059 (53.2%)
DistinctHuman	547	495 (90.5%)	477 (87.2%)	476 (87%)	474 (86.7%)	473 (86.5%)	474 (86.5%)
DistinctLinker	2091	988 (47.3%)	978 (46.8%)	971 (46.4%)	971 (46.4%)	973 (46.5%)	972 (46.5%)
Max	1708	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)
MaxHuman	1708	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)	1708 (100%)
MaxLinker	848	848 (100%)	848 (100%)	848 (100%)	848 (100%)	848 (100%)	848 (100%)
Total	22994	12690 (55.2%)	12624 (54.9%)	12621 (54.9%)	12616 (54.9%)	12616 (54.9%)	12615 (54.9%)
TotalHuman	3081	2999 (97.3%)	2981 (96.8%)	2980 (96.7%)	2978 (96.7%)	2977 (96.6%)	2978 (96.6%)
TotalLinker	7909	3737 (47.2%)	3727 (47.1%)	3720 (47%)	3720 (47%)	3722 (47.1%)	3721 (47.1%)

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)





TGTGACTTGGTAACTAGAGATCCCTCAGACCCCTTCTAGTCAGTGCGAAAATCTCTAGCA

TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGCCTGGAAAAATCTCTAGCA
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTCTAGTCAGTGTGGAAAAATCTCTAGCA
TGTGACTCTCGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA
CGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA
TGCCTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA
TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTGTAGTCAGTGTGGAAAAATCTCTAGCA

☐ non conserved
☒ $\geq 50\%$ conserved