

LTR Parser Results

Kevin McCormick

26 May, 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/hivIntSite2.fastq.gz

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.

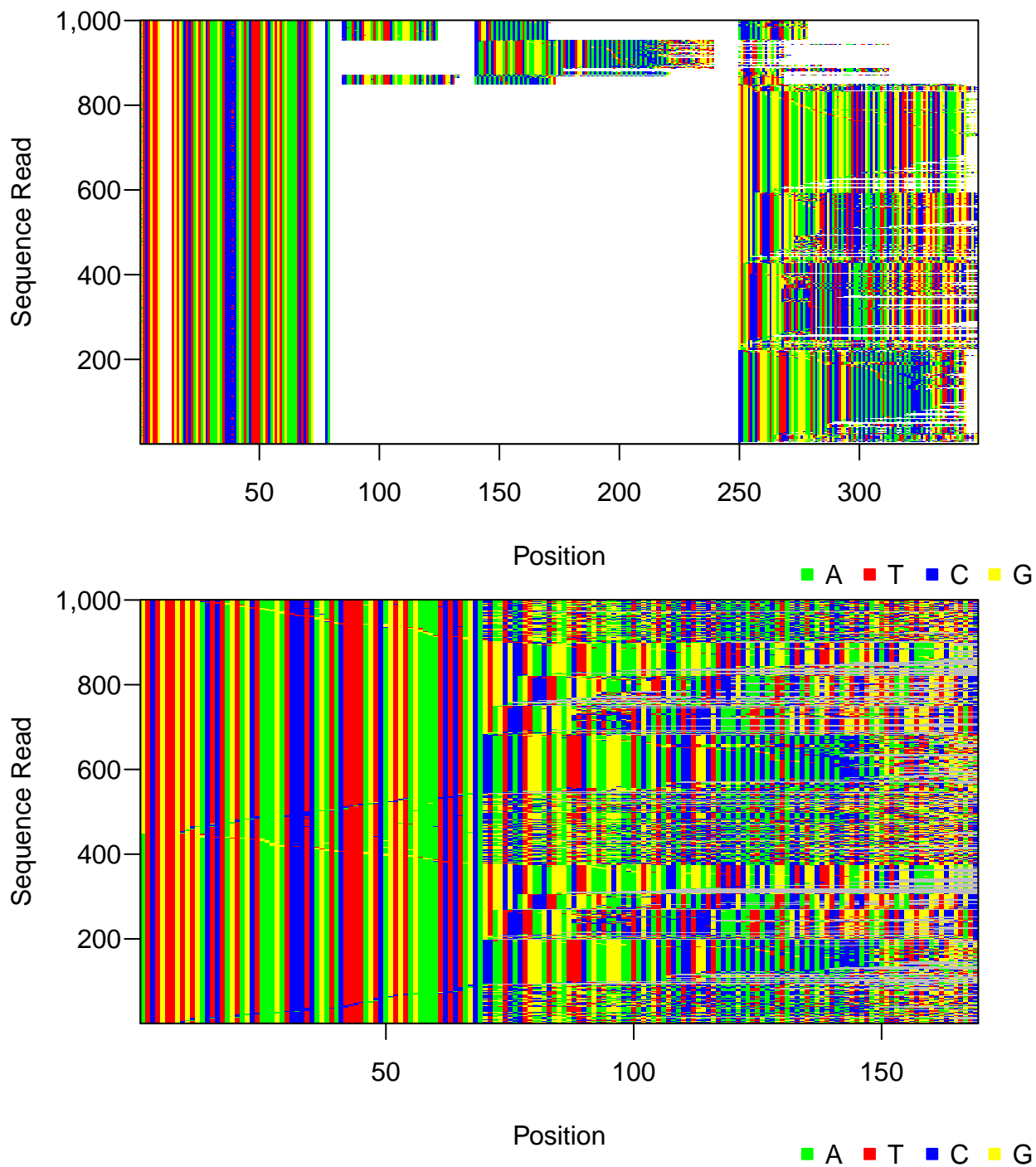
id	primer	LTR	expectedLTR	LANL
1	ATCTGTTG	TGTGACTCTAGTAGCTAAAGATCCCCCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA	FALSE	TRUE
2	ATCTGTTG	TGTGACTCTAGTAGCTAAAGATCCCTCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA	FALSE	TRUE
3	GTCTGTTG	TGTGACTCTAGTAGCTAAAGATCCCCCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA	FALSE	TRUE
4	GTCTGTTG	TGTGACTCTAGTAGCTAAAGATCCCTCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA	FALSE	TRUE

Variable	Total	1	2	3	4
Distinct	4663	1000 (21.4%)	1033 (22.2%)	1372 (29.4%)	1421 (30.5%)
DistinctHuman	407	159 (39.1%)	159 (39.1%)	213 (52.3%)	214 (52.6%)
DistinctLinker	0	0 (NaN%)	0 (NaN%)	0 (NaN%)	0 (NaN%)
Max	479	343 (71.6%)	343 (71.6%)	479 (100%)	479 (100%)
MaxHuman	64	48 (75%)	48 (75%)	64 (100%)	64 (100%)
MaxLinker	0	0 (NaN%)	0 (NaN%)	0 (NaN%)	0 (NaN%)
Total	11468	2392 (20.9%)	2425 (21.1%)	3432 (29.9%)	3481 (30.4%)
TotalHuman	751	301 (40.1%)	301 (40.1%)	414 (55.1%)	415 (55.3%)
TotalLinker	0	0 (NaN%)	0 (NaN%)	0 (NaN%)	0 (NaN%)

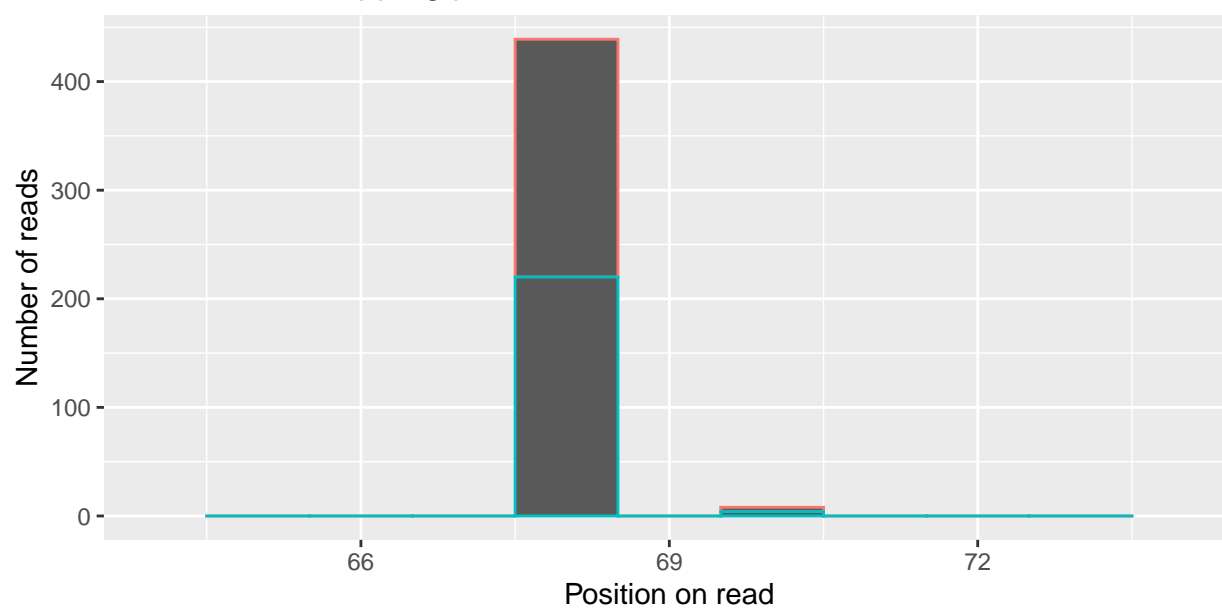
Visualization of reads

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



LTR

	TGTGACTCTAGTAGCTAAAGATCCCCCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA
	TGTGACTCTAGTAGCTAAAGATCCCTCAGATACTTTTAGTCAGTGTAGAAAATCTCTAGCA