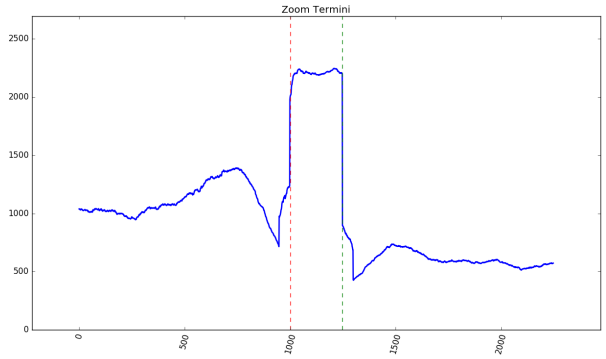
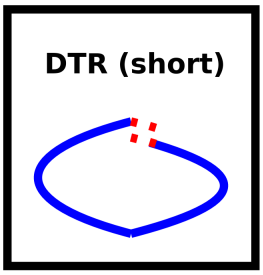
my_phage PhageTerm Analysis





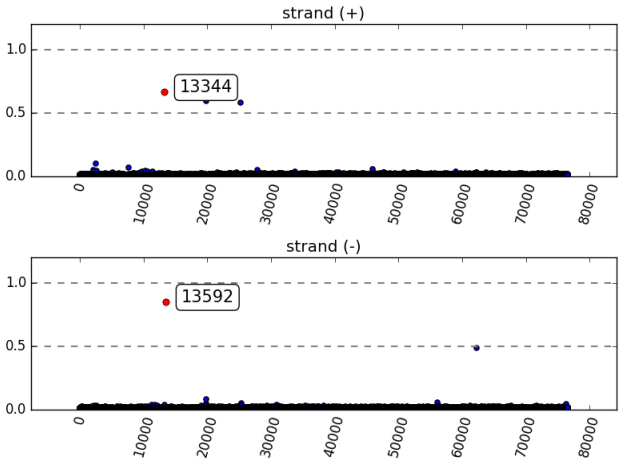
PhageTerm Method

Ends	Left (red)	Right (green)	Permuted	Orientation	Class	Type
Redundant	13344	13592	No	NA	DTR (short)	T7

**Direct Terminal Repeats: 249 bp*

Strand	Location	T	pvalue
+	13344	0.66	0.00e+00
	19862	0.59	9.85e-112
	25285	0.58	8.69e-84
	2483	0.10	8.49e-10
	7607	0.07	2.73e-07
-	13592	0.85	1.51e-318
	62298	0.48	1.02e-67
	19878	0.08	5.25e-05
	56061	0.06	1.27e-05
	25293	0.05	3.26e-03

T (Start. Pos. Cov. / Whole Cov.)



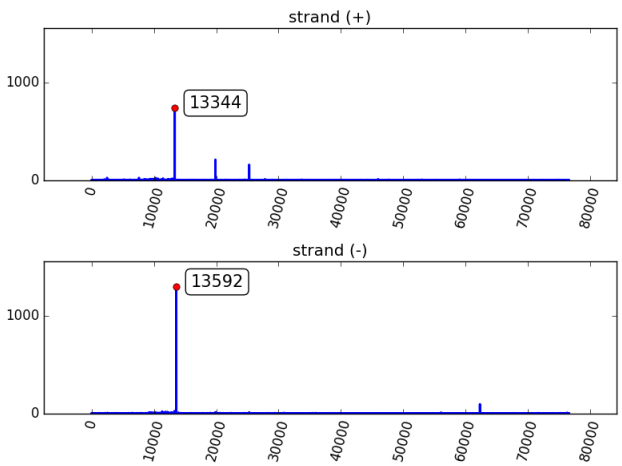
Li's Method

Packaging	Termini	Forward	Reverse	Orientation
COS	Fixed	Obvious Termini	Obvious Termini	Reverse

**Direct Terminal Repeats: 249 bp*

Strand	Location	SPC	R
+	13344	740	3.0
	19862	212	-
	25285	160	-
	7607	26	-
	2483	26	-
-	13592	1297	13.0
	62298	97	-
	13320	28	-
	11336	21	-
	11845	19	-

SPC

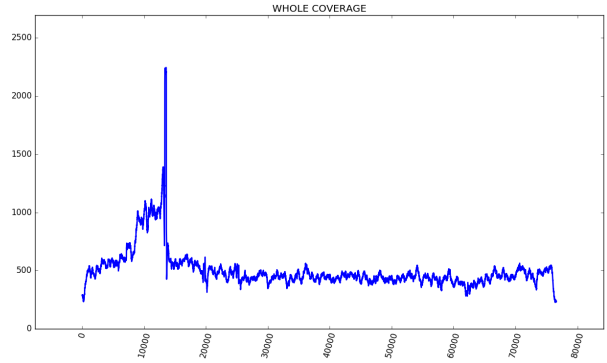


Analysis Methodology

PhageTerm software uses raw reads of a phage sequenced with a sequencing technology using random fragmentation and its genomic reference sequence to determine the termini position. The process starts with the alignment of NGS reads to the phage genome in order to calculate the starting position coverage (SPC), where a hit is given only to the position of the first base in a successfully aligned read (the alignment algorithm uses the lenght of the seed (default: 20) for mapping and does not accept gap or missmatch to speed up the process). Then the program apply 2 distinct scoring methods: i) a statistical approach based on the Gamma law; and ii) a method derived from LI and al. 2014 paper.

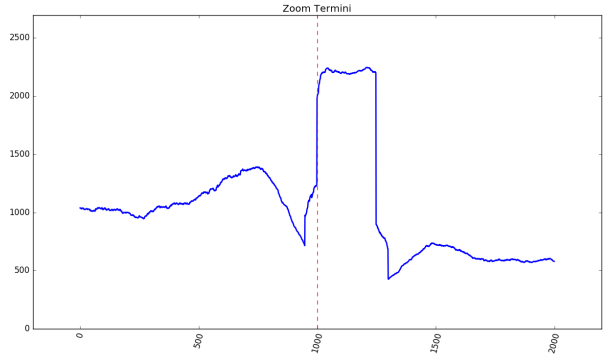
General set-up and mapping informations

Phage Genome	76572 bp
Sequencing Reads	718152
Mapping Reads	89 %
OPTIONS	
Mapping Seed	*20*
Surrounding	*20*
Host Analysis	*No*

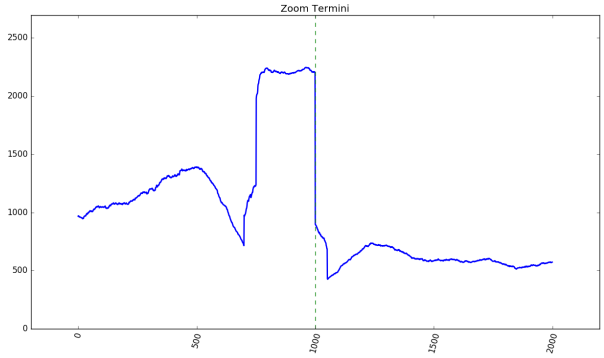


Highest peak of each side coverage graphics

Whole Coverage Zoom (Left)



Whole Coverage Zoom (Right)



General controls information

Whole genome coverage	251	OK
Weak genome coverage	0.0 %	OK
Insert mean size	483	Mean insert estimated from paired-end reads
Reads lost during alignment	10.2 %	OK

i) PhageTerm method

Reads are mapped on the reference to determine the starting position coverage (SPC) as well as the coverage (COV) in each orientation. These values are then used to compute the variable T = SPC / COV. The average value of T at positions along the genome that are not termini is expected to be 1/F, where F is the average fragment size. For the termini that depends of the packaging mode. Cos Phages: no reads should start before the terminus and therefore X=1. DTR phages: for N phages present in the sample, there should be N fragments that start at the terminus and N fragments that cover the edge of the repeat on the other side of the genome as a results T is expected to be 0.5. Pac phages: for N phages in the sample, there should be N/C fragments starting at the pac site, where C is the number of phage genome copies per concatemer. In the same sample N fragments should cover the pac site position, T is expected to be (N/C)/(N+N/C) = 1/(1+C). To assess whether the number of reads starting at a given position along the genome can be considered a significant outlier, PhageTerm first segments the genome according to coverage using a regression tree. A gamma distribution is then fitted to SPC for each segment and an adjusted p-value is computed for each position. Finally if several significant peaks are detected within a small sequence window (default: 20bp), their T values are merged.

Nearby Termini (Forward / Reverse)	0 / 0	Peaks localized 20 bases around the maximum

ii) Li's method

The second approach is based on the calculation and interpretation of three specific ratios R1, R2 and R3 as suggested in a previous publication from Li et al. 2014. The first ratio, is calculated as follow: the highest starting frequency found on either the forward or reverse strands is divided by the average starting frequency, R1 = (highest frequency/average frequency). Li's et al. have proposed three possible interpretation of the R1 ratio. First, if R1 < 30, the phage genome does not have any termini, and is either circular or completely permuted and terminally redundant. The second interpretation for R1 is when 30 ≤ R1 ≥ 100, suggesting the presence of preferred termini with terminal redundancy and apparition of partially circular permutations. At last if R1 > 100 that is an indication that at least one fixed termini is present with terminase recognizing a specific site. The two other ratios are R2 and R3 and the calculation is done in a similar manner. R2 is calculated using the highest two frequencies (T1-F and T2-F) found on the forward strand and R3 is calculated using the highest two frequencies (T1-R and T2-R) found on the reverse strand. To calculate these two ratios, we divide the highest frequency by the second highest frequency T2. So R2 = (T1-F / T2-F) and R3 = (T1-R / T2-R). These two ratios are used to analyze termini characteristics on each strand taken individually. Li et al. suggested two possible interpretations for R2 and R3 ratios combine to R1. When R1 < 30 and R2 < 3, we either have no obvious termini on the forward strand, or we have multiple preferred termini on the forward strand, if 30 ≤ R1 ≤ 100. If R2 > 3, it is suggested that there is an obvious unique termini on the forward strand. The same reasoning is applicable for the result of R3. Combining the results for ratios found with this approach, it is possible to make the first prediction for the viral packaging mode of the analyzed phage. A unique obvious termini present at both ends (both R2 and R3 > 3) reveals the presence of a COS mode of packaging. The headful mode of packaging PAC is concluded when we have a single obvious termini only on one strand. A whole coverage around 500X is needed for this method to be reliable.

Nearby Termini (Forward / Reverse)	0 / 0	Peaks localized 20 bases around the maximum
R1 - highest freq./average freq.	1361	At least one fixed termini is present with terminase recognizing a specific site.
R2 Forw - highest freq./second freq.	3	Unique termini on the forward strand.
R3 Rev - highest freq./second freq.	13	Unique termini on the reverse strand.

Please cite: Sci. Rep. DOI 10.1038/s41598-017-07910-5

Garneau, Depardieu, Fortier, Bikard and Monot. PhageTerm: Determining Bacteriophage Termini and Packaging using NGS data.

Report generated : Tue Dec 25 01:54:07 2018