

METHODOLOGY ARTICLE

Open Access

A novel termini analysis theory using HTS data alone for the identification of *Enterococcus* phage EF4-like genome termini

Xianglilan Zhang^{1†}, Yahui Wang^{1,2†}, Shasha Li¹, Xiaoping An¹, Guangqian Pei¹, Yong Huang¹, Hang Fan¹, Zhiqiang Mi¹, Zhiyi Zhang¹, Wei Wang¹, Yubao Chen^{3*} and Yigang Tong^{1*}

Abstract

Background: *Enterococcus faecalis* and *Enterococcus faecium* are typical enterococcal bacterial pathogens. Antibiotic resistance means that the identification of novel *E. faecalis* and *E. faecium* phages against antibiotic-resistant *Enterococcus* have an important impact on public health. In this study, the *E. faecalis* phage IME-EF4, *E. faecium* phage IME-EFm1, and both their hosts were antibiotic resistant. To characterize the genome termini of these two phages, a termini analysis theory was developed to provide a wealth of terminal sequence information directly, using only high-throughput sequencing (HTS) read frequency statistics.

Results: The complete genome sequences of phages IME-EF4 and IME-EFm1 were determined, and our termini analysis theory was used to determine the genome termini of these two phages. Results showed 9 bp 3' protruding cohesive ends in both IME-EF4 and IME-EFm1 genomes by analyzing frequencies of HTS reads. For the positive strands of their genomes, the 9 nt 3' protruding cohesive ends are 5'-TCATCACCG-3' (IME-EF4) and 5'-GGGTCAGCG-3' (IME-EFm1). Further experiments confirmed these results. These experiments included mega-primer polymerase chain reaction sequencing, terminal run-off sequencing, and adaptor ligation followed by run-off sequencing.

Conclusion: Using this termini analysis theory, the termini of two newly isolated antibiotic-resistant *Enterococcus* phages, IME-EF4 and IME-EFm1, were identified as the byproduct of HTS. Molecular biology experiments confirmed the identification. Because it does not require time-consuming wet lab termini analysis experiments, the termini analysis theory is a fast and easy means of identifying phage DNA genome termini using HTS read frequency statistics alone. It may aid understanding of phage DNA packaging.

Keywords: Termini analysis theory, Antibiotic-resistant *Enterococcus*, Phage, Genome termini

Background

Enterococcus faecalis and *Enterococcus faecium* are the most common enterococcal bacteria cultured from humans, making up more than 90 % of clinical isolates [1]. Since the 1990s, *Enterococcus* strains have been found to be opportunistic Gram-positive pathogens, responsible for various diseases [2–8]. Recently, strains of enterococci resistant to vancomycin have emerged, with an increased incidence reported worldwide [9–11].

Vancomycin was one of the last antibiotics to remain reliably effective against enterococci [12]. The continuous overuse and misuse of antibiotics has produced antibiotic-resistant *Enterococcus*. These, most notably vancomycin-resistant *Enterococcus*, have become a threat in nosocomial settings. However, phage therapy has great promise in antibiotic-resistant *Enterococcus* treatment. For this reason, the identification and study of phages for antibiotic-resistant *Enterococcus* may have a significant medical impact in the near future.

The most important information concerning a phage is its genome packaging, which affects its entire life cycle from initiation [13] to viral DNA replication [14], termination, and regulation of transcription [15]. Identifying

* Correspondence: allexchen@126.com; tong.yigang@gmail.com

†Equal contributors

³Beijing Computing Center, Beijing 100094, P.R. China

¹State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, P.R. China

Full list of author information is available at the end of the article

genome termini identification is crucial to the entire DNA packaging process. High-throughput sequencing (HTS) is an effective means of performing phage genome sequence analysis [16–19], including genome termini analysis. HTS generates a large number of reads. Mining useful sequence information from these large datasets is a key problem in bioinformatics. Conventional methods first use these data to assemble the full sequence of a phage genome and then carry out molecular biology experiments to identify its termini. Unlike these conventional methods, our termini analysis theory can be used to study the phage's genome termini and genome packaging directly using only the data from the HTS reads. This approach can prove a phage's termini information without the secondary molecular biological experiments required by conventional methods. It also reduces the time and expense of DNA packaging analysis.

In this study, two newly introduced antibiotic-resistant *Enterococcus* phages, IME-EF4 and IME-EFm1, were analyzed using the current method. By using the termini analysis theory, we easily identified the IME-EF4/IME-EFm1's termini and hypothesized that the IME-EF4 and IME-EFm1 both have special 9 bp 3' protruding cohesive ends. Further molecular biological experiments confirmed this hypothesis. This paper is the first to propose the termini analysis theory, which is a means of identifying a phage's termini without any wet-lab termini conformation experiment.

Methods

IME-EF4, IME-EFm1, and their bacterial strains

The lytic *Enterococcus* phages IME-EF4 and IME-EFm1 were isolated from sewage from PLA Hospital 307 (Beijing, China), and the host bacteria for the two phages were isolated from clinical samples in the same hospital. The collaboration between the PLA Hospital 307 and the present laboratory meant that no specific permits were required for these studies. The collected samples were neither privately owned nor protected and did not involve any endangered or protected species.

IME-EF4/IME-EFm1 isolation

Enrichment cultures were used to isolate IME-EF4 and IME-EFm1 from sewage [20]. Specifically, approximately

2 mL of filtered (Millipore membranes, pore diameter 0.45 μ M) sewage was mixed with 2 mL of 3*liquid LB medium and 100 μ L *E. faecalis* or *E. faecium* that had been cultured overnight. To amplify the IME-EF4/IME-EFm1, the enrichment culture was incubated at 37 °C for at least 14 h with agitation and then centrifuged (10 min, 10,000 \times g, 4 °C). The supernatant was filtered (Millipore membranes, pore diameter 0.45 μ M) to remove the residual bacterial cells. Then 100 μ L of phage stock solution was mixed with 500 μ L EF4/EFm1 cells in the exponential growth phase ($OD_{600} = 0.2$ to 0.5). The mixture was incubated at 37 °C for 5 min. Then 5 mL top agar (LB with 0.75 % agar) was added at 50 °C, and the mixture was poured into an LB plate that had been pre-warmed to 37 °C (double-layer method). This plate was then incubated overnight at 37 °C to produce phage plaques [21].

IME-EF4/IME-EFm1 genome DNA extraction

IME-EF4/IME-EFm1 DNA was extracted based on a method described in a previous work [22]. In brief, DNase I and RNase A (Thermo Scientific, America) were added to the phage IME-EF4 stock solution to a final concentration of 1 μ g/ml. The mixture was incubated overnight at 37 °C. Then samples were incubated at 80 °C for 15 min to deactivate the DNase I. Lysis buffer (final concentration, 0.5 % sodium dodecyl sulfate, 20 mM EDTA, and 50 μ g/ml proteinase K) was added to samples, which were then incubated at 56 °C for 1 h. An equal volume of phenol was added to extract the DNA. After centrifugation at 7000 \times g for 5 min, the aqueous layer was removed to a fresh tube containing an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) and centrifuged at 7000 \times g for 5 min. The aqueous layer was collected, mixed with an equal volume of isopropanol, and stored overnight at -20 °C. The mixture was centrifuged at 4 °C for 20 min at 10,000 \times g, and the DNA pellet was washed with 75 % ethanol. The DNA was then air dried at room temperature, resuspended in deionized water, and stored at -20 °C.

IME-EF4/IME-EFm1 high-throughput sequencing

After extraction of the phage IME-EF4 and IME-EFm1 genome DNA, the genomes were using the semiconductor

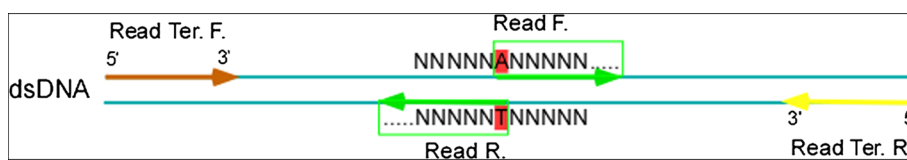
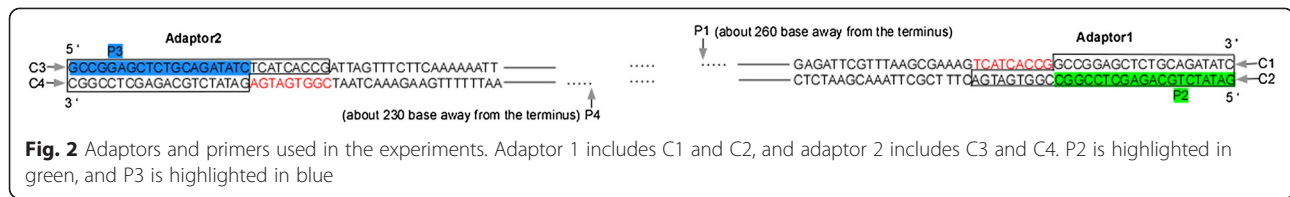


Fig. 1 Generation of reads in dsDNA using high-throughput sequencing (HTS). F represents forward sequences and R represents reverse sequences. Ter. refers to termini. The sequence of Read F starts with base A. The sequence of Read R starts with base T. The starting bases are shown with red backgrounds. Reads Ter. F and Ter. R begin with natural termini



sequencer in the Life Technologies Ion Torrent Personal Genome Machine (PGM) IonTorrent sequencer (IonTorrent). This technology uses emulsion polymerase chain reaction (PCR) and a sequencing-by-synthesis approach [23]. The library preparation, amplification, and sequencing were performed according to the Ion Torrent™ sequencing protocols. The IME-EF4 and IME-EFm1 genome DNA samples were sheared using Ion Shear™ Plus Reagents. These DNA fragments were then ligated to Ion Torrent adapters for subsequent nick repair and purification. Purified DNA fragments of about 300 bp were selected by using E-Gel® SizeSelect™ agarose gel. After amplification and purification of the selected library, emulsion PCR was used to process it. PCR was performed in a water-in-oil microreactor containing a single DNA molecule on a bead [24]. The H⁺ Ion Torrent signal was detected during the sequencing-by-synthesis. In this process, four fluorescently labeled nucleotides were added to the flowcell channels during DNA synthesis. The florescent light signals were detected using a genome analyzer, which was used for base calling [25].

Termini analysis theory

A phage with a linear double-stranded DNA (dsDNA) has terminal repetitions. These repetitions are used for homologous recombination during the phage’s DNA replication process. The phage’s dsDNA can be circularized through the genome terminal repetitions. This makes the phage’s natural genome termini, which are cleaved by a terminase, difficult to identify. In this study, a termini analysis theory, which can be used to find natural termini using the read frequency, is described.

Suppose that there are *m* identical genomes and the length of each genome is *L*. All the genomes are divided into *N_r* short sequences. Each short sequence is called a read. The average length of the reads is *L_{reads}*.

Table 1 Prepared adaptor list

Adaptor		Sequence (From 5' to 3')
Adaptor 1	C1-P	P-GCCGGAGCTCTGCAGATATC
1	C2	GATATCTGCAGAGCTCCGGC-CGGTGATGA
Adaptor 2	C3	GCCGGAGCTCTGCAGATATC-TCATCACCG
2	C4-P	P-GATATCTGCAGAGCTCCGGC

Theorem 1 $R = \frac{Freq_{ter}}{Freq_{ave}} = 2 * L_{reads}$ (1)

Proof 1. There are *m* identical genomes. As illustrated in Fig. 1, the high-throughput sequencing (HTS) machine reads each read from 5' to 3'. In this way, each genome with dsDNA has two termini. The frequency of the reads starting with a natural terminus is as follows:

$Freq_{ter} = m$ (2)

There are a total of *N_r* reads in the *m* genomes. As shown in Fig. 1, two different reads, F and R, start with base A and base T. In this way, the average frequency of all reads is as follows:

$Freq_{ave} = \frac{Nr}{2 * L}$ (3)

The ratio of *Freq_{ter}* to *Freq_{ave}* is as follows:

$R = \frac{Freq_{ter}}{Freq_{ave}} = \frac{m}{\frac{Nr}{2 * L}} = \frac{2 * m * L}{Nr} = 2 * L_{reads}$ (4)

Our termini analysis theory provides a theoretical frequency ratio of terminal reads to general reads. In the experiment, we choose only 300 bp reads rather than the whole amount of reads for analysis. Such a library selection will decrease the practical frequency ratio. Fortunately, we can still practically distinguish the termini from the other reads (see results).

Mega-primer PCR sequencing

PCR amplification was performed on the IME-EF4 DNA genome to determine more about the IME-EF4 complete genome sequence acquired using the HTS data. The PCR involved mega-primer-guided polymerization through the protruding cohesive end.

Table 2 Prepared primer list

Primer	Sequence (From 5' to 3')
P1	CTCTAGTTTGTTCGCTGCGTAAATC
P2	GATATCTGCAGAGCTCCGGC
P3	GCCGGAGCTCTGCAGATATC
P4	AGGTACGGACCGCAATGGGTTGGGA

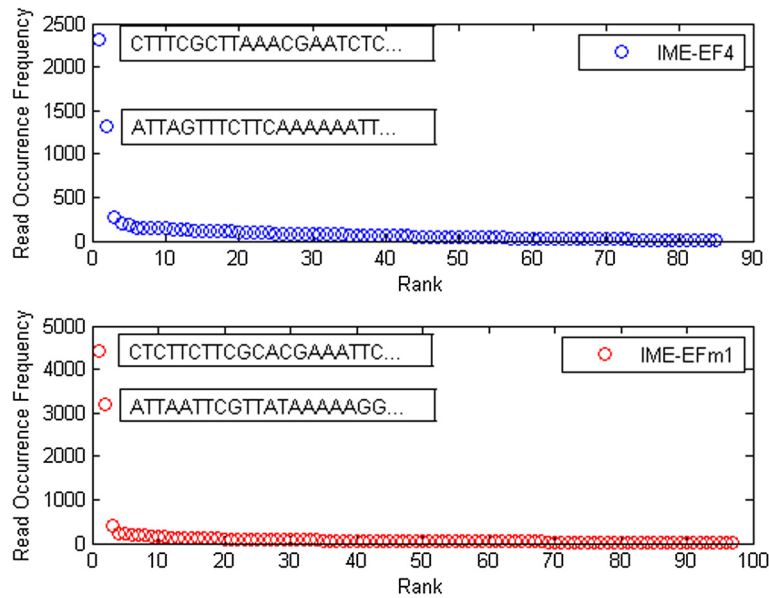


Fig. 3 Read occurrence distribution. Blue and red circles indicate the HTS reads of IME-EF4 and IME-EFm1 samples, respectively. The rank in the x-axis refers to the relative frequency of each read

Terminal run-off sequencing

The IME-EF4 complete genome was used as the template for terminal run-off sequencing, the process of which is described in [26]. The 3' end of the IME-EF4 genome was marked using the primer P1 (5'-CTCTAGTTTGTTCGTGCGTAAATC-3'). The 5' end of the IME-EF4 genome was marked using the primer P4 (5'-AGGTACG

GACCGCAATGGGTTGGGA-3'). The Sangon Company synthesized the primers (Beijing, China).

Ligation of the adaptors to the termini

To prove that IME-EF4 is linear dsDNA ending with a 3' 9 nt protruding cohesive end, two pairs of adaptors ligated with the ends of IME-EF4 we created. They are

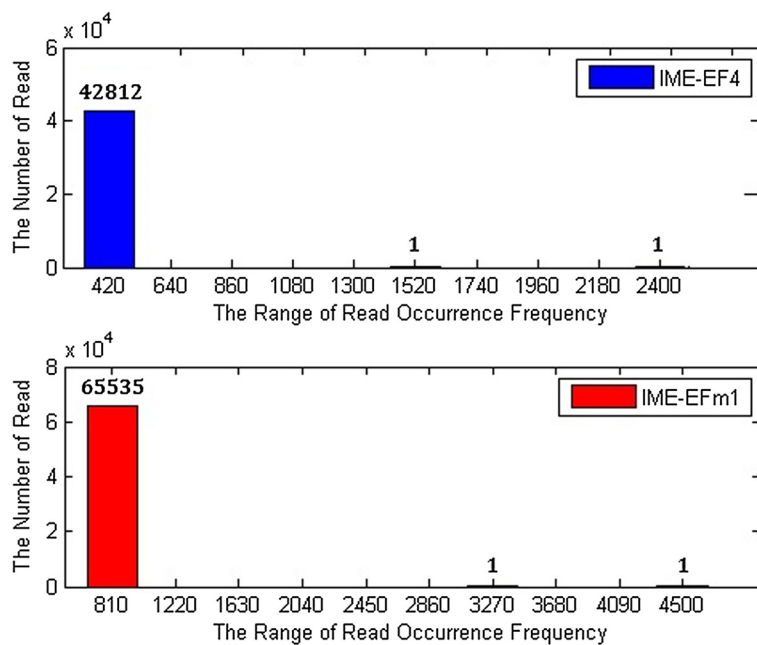


Fig. 4 Occurrence rates of numbers of reads. Blue and red bars indicate the HTS reads of IME-EF4 and IME-EFm1 samples, respectively. The number on the x-axis represents the range from the last number to the next number

Table 3 IME-EF4 and IME-EFm1 terminal sequence frequency statistics

Phage	Strand	Ave. Freq.	Ter. Freq.	Ter.Freq. Ave.Freq.	Terminal Sequence
IME-EF4	Positive	6.73	1,322	196	ATTAGTTTCTTCAAAAATT
	Negative	6.73	2,318	344	CTTTCGCTTAAACGAATCTC
IME-EFm1	Positive	12.95	3,194	246	ATTAATTCGTTATAAAAGG
	Negative	12.95	4,412	341	CTCTTCTTCGCACGAAATTC

shown in Fig. 2. Adaptor 1 was created by combining C1 and C2 and adaptor 2 by combining C3 and C4. To ligate adaptors with IME-EF4 terminal sequences, phosphoric acids were added to C1 and C4. The prepared adaptors are shown in Table 1. At the same time, primers were prepared for use in the next step of PCR amplification, as illustrated in Fig. 2. The primer sequences are shown in Table 2. The Sangon Company prepared the oligonucleotides for these adaptors and primers. The adaptor oligonucleotide mixtures were hybridized by running the ligation program on a PCR machine. To ligate the IME-EF4 genome with the prepared adaptor, phosphoric acid was added for the IME-EF4 genome end repair. Specifically, 800 ng of the IME-EF4 genome was diluted to 16 μ L using nuclease-free water, and prepared the end repair mix. Then the phosphate groups were added, also through PCR. The adaptor-ligated DNA was purified as listed in the NEBNext[®] Fast DNA Library Prep Set for Ion Torrent[™] instruction manual (version 3.1). The purified DNA samples were used as templates for PCR amplification and sequencing with the primer in two sets, P1 and P2, and P3 and P4.

Results and discussion

IME-EF4/IME-EFm1 high-throughput sequencing (HTS) analysis

IME-EF4/IME-EFm1 high-frequency read statistics

All HTS reads from the IME-EF4 and IME-EFm1 samples were statistically analyzed and then their frequencies were ranked in descending order. As shown in Fig. 3, IME-EF4 and IME-EFm1 HTS read data have two significant high-frequency reads, beginning with CTTTCGCTTAAACGAATCTC and ATTAGTTTCTTCAAAAATT, respectively. The HTS data of IME-EF4 and IME-EFm1 share similar sequence occurrence frequency curves. Fig. 4 shows that more than 99 % of reads have frequencies below 420 for IME-EF4 and below 810 for IME-EFm1. Using the termini analysis theory, it was concluded that the two reads with the highest frequencies were the

termini of phage IME-EF4 and IME-EFm1. As shown in Table 3, the occurrence rates of terminal reads to general reads were 196 for the positive strand of IME-EF4, 344 for the negative strand of IME-EF4, 246 for the positive strand of IME-EFm1, and 341 for the negative strand of IME-EFm1. Biological experiments were used to verify our conclusion (see behind).

IME-EF4/IME-EFm1 sequence assembly and identification of termini

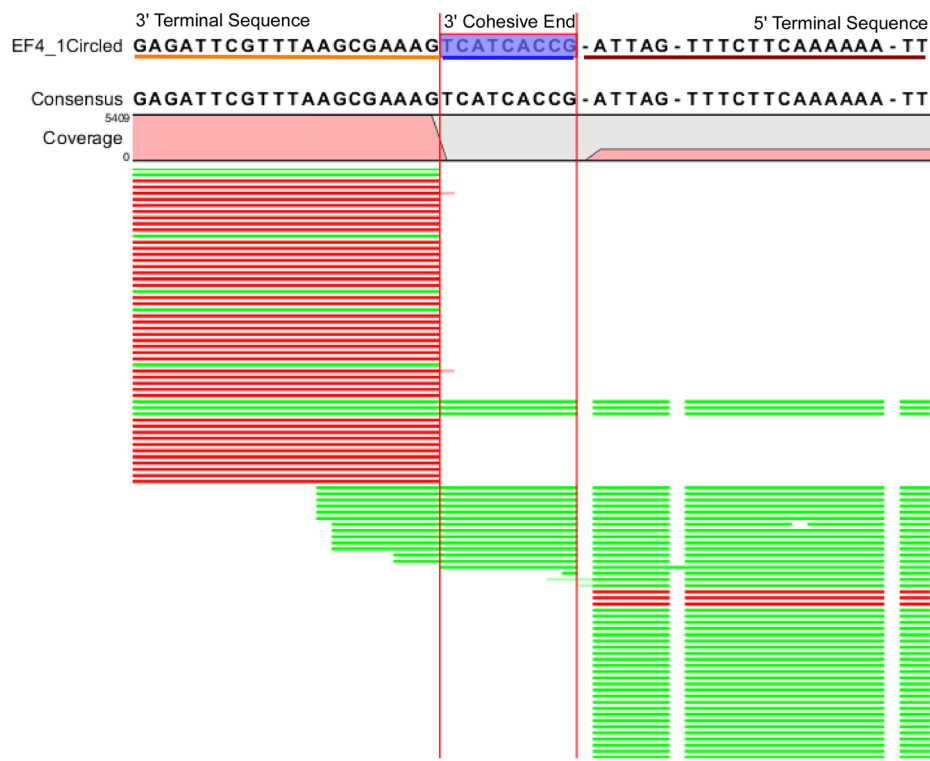
The IME-EF4 and IME-EFm1 sequence assembly results were acquired using Newbler (version 2.9, Roche). As shown in Table 4, the genome size of IME-EF4 is 40,713 bp and the read matches percentage was more than 97 %, while the genome size of IME-EFm1 is 42,599 bp and the read matches percentage was more than 95 %.

The reads were then aligned to the reference assembly using a CLC Genomics Workbench (version 3.6.1). The CLC module was run with default parameters (mismatch cost 2, insertion cost 3, deletion cost 3, length fraction 0.5, and similarity 0.8) using FASTQ files as read input and using FNA files as reference input. The results are shown in Fig. 5. The reference genomes are indexed as IME-EF4_Circled and IME-EFm1_Circled, respectively. For IME-EF4, the sequence TCATCACCG is the 9 nt tail that connects the 5' and 3' ends. For IME-EFm1, the sequence GGGTCAGCG is the 9 nt tail that connects 5' and 3' ends. Two 5' ends start with ATTAGTTTCTTCAAAAATT (IME-EF4) and ATTAATTCGTTATAAAAGG (IME-EFm1). Two 3' ends start with GAGATTCGTTTAAAGCGAAAG (IME-EF4) and CTCTTCTTCGCACGAAATTC (IME-EFm1). Fig. 6 shows the positions of terminal sequences in IME-EF4 and IME-EFm1, each with a 9 nt protruding cohesive end. The last bases of the 3' termini were located at position 40,704 (IME-EF4, position = genome size (40,713) – protruding cohesive end size (9)) and 42,590 (IME-EFm1, position = genome size (42,599) – protruding cohesive end size (9)).

As shown in Fig. 5, there were fewer than 20 cases of short reads with the 9 nt protruding cohesive end. As

Table 4 IME-EF4 and IME-EFm1 sample sequence assembly results

Phage	Length	# Matched Reads	# Total Reads	Relative Matches	Min Coverage	Max Coverage
IME-EF4	40,713	548,309	563,429	97.32%	11	7,215
IME-EFm1	42,599	1,103,439	1,154,818	95.55%	51	9,657



(a) IME-EF4 Mapping Result



(b) IME-EFm1 Mapping Result

Fig. 5 (a) IME-EF4 and (b) IME-EFm1 mapping results. The mapped reads were acquired from the original HTS data. The 3' terminal sequences are underlined in orange, the 5' terminal sequences are underlined in dark red, and the 3' protruding cohesive ends are underlined in blue

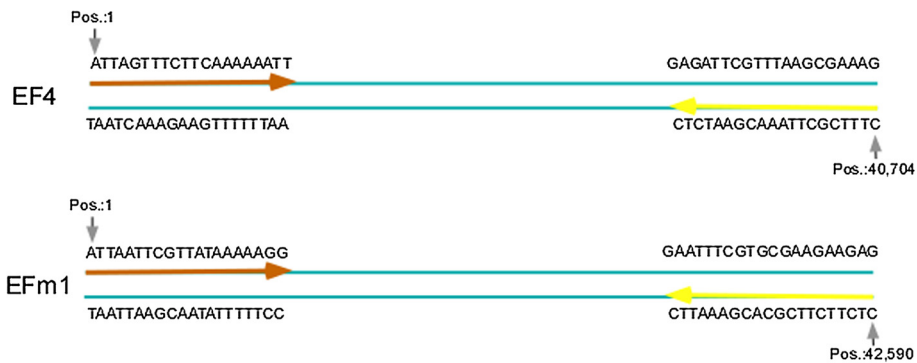
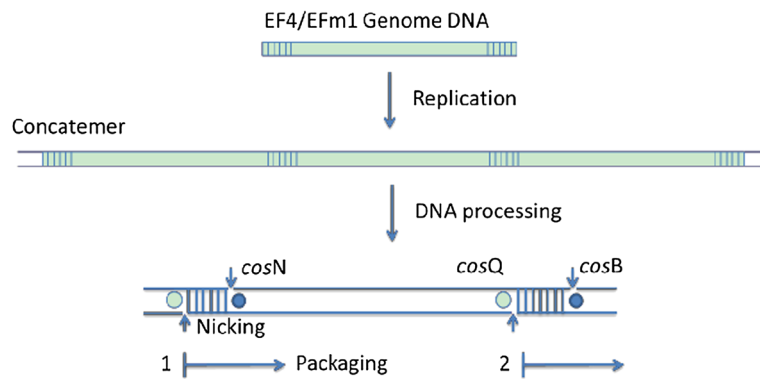


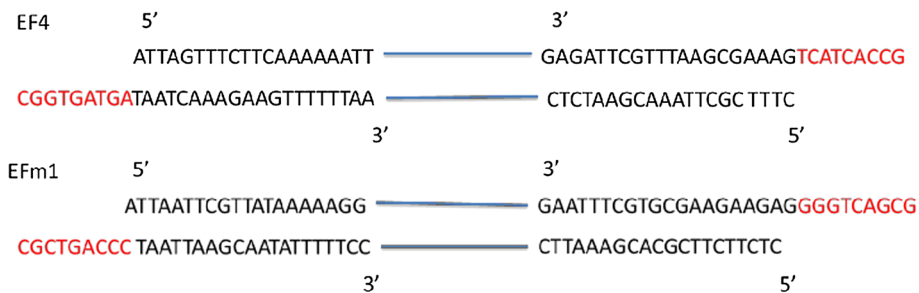
Fig. 6 Terminal sequences and their positions. Pos. refers to base positions. The dark orange arrows indicate the 5' terminal sequences in HTS read data, and the light yellow arrows indicate the 3' terminal sequences in HTS read data

shown in Table 3 and Fig. 3, the 9 nt protruding cohesive end was not detected in the most frequently occurring reads. Considering that HTS generates the dsDNA reads, it is possible that mature phage IME-EF4 and IME-EFm1 DNA molecules are the ones that have the 9 nt single-stranded 3' ends, which are called protruding

cohesive ends. The reason is explained as follows. If a terminase introduces a double-stranded break into the cos site of the IME-EF4/IME-EFm1, the 5' and 3' terminal sequences should ligate to each other without any internal sequence between them. However, this does not match the results of the present assembly experiments,

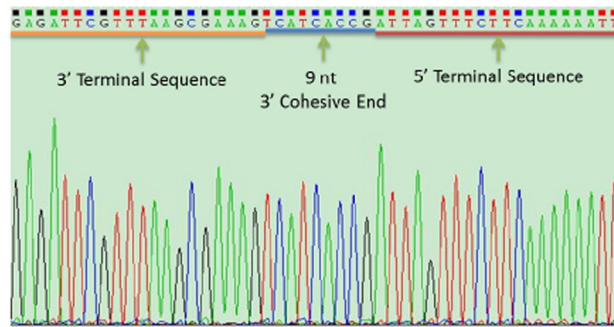


(a) EF4/EFm1 Concatemer Processing Pathway

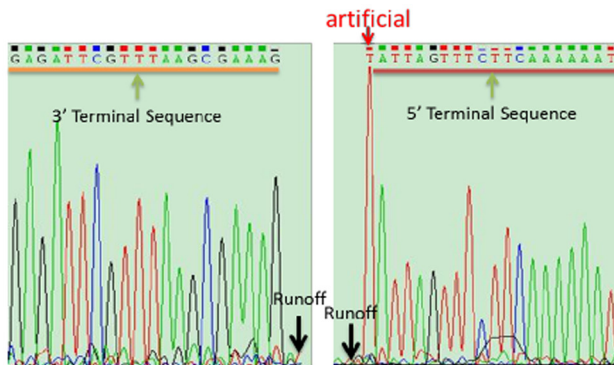


(b) Terminal Sequence Details

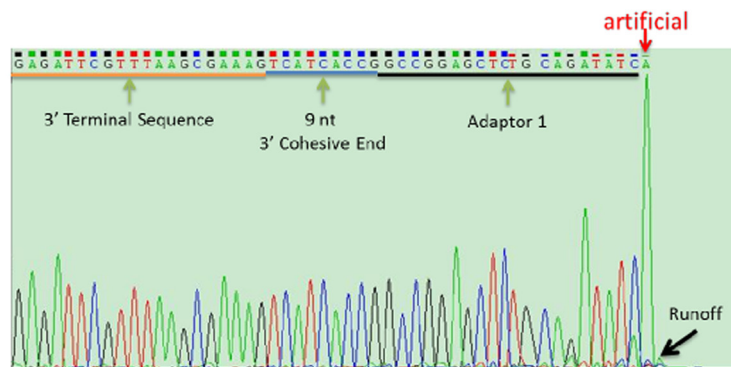
Fig. 7 (a) IME-EF4 and IME-EFm1 concatemer processing pathways and (b) terminal sequence details. The packaging of concatemeric dsDNA is initiated from the end generated by the initiation cleavage, which then proceeds unidirectionally before being terminated by the termination cleavage. IME-EF4/IME-EFm1 terminases introduce nicks staggered by 9 bp concatemer (highlighted in red) at cosN and melts the strands by its helicase activity after recognition of cosB (●) and cosQ (○) for initiation and termination cleavage, respectively



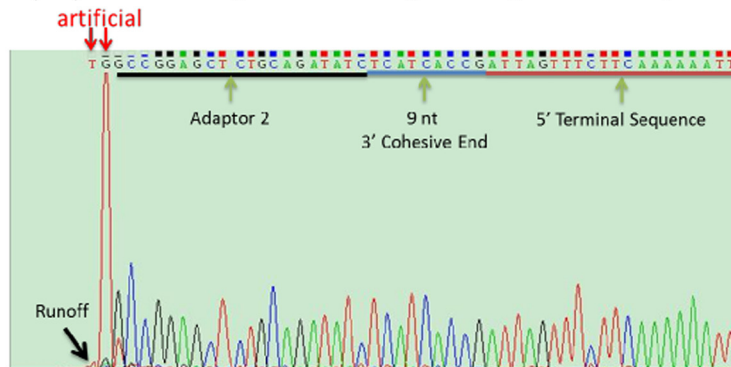
(a) The chromatogram of IME-EF4 terminal sequence



(b) The chromatograms (the right one is reversed) of IME-EF4 terminal run-off sequencing

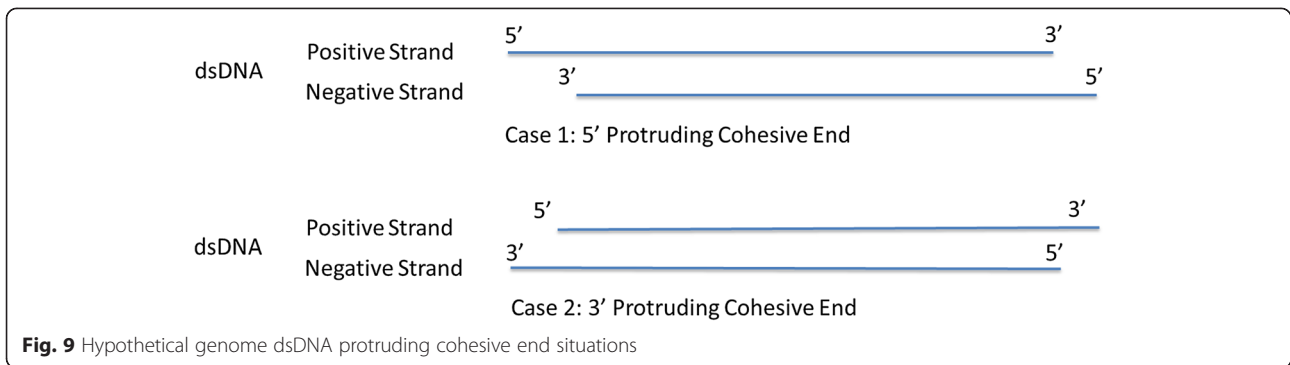


(c.1) The chromatogram of IME-EF4 genome ligated with adaptor 1



(c.2) The chromatogram (reversed) of IME-EF4 genome ligated with adaptor 2

Fig. 8 Chromatograms of the three molecular biology experiments, including (a) IME-EF4 complete genome sequencing, (b) IME-EF4 terminal run-off sequencing, and (c) IME-EF4 adaptor ligation to the terminal sequences



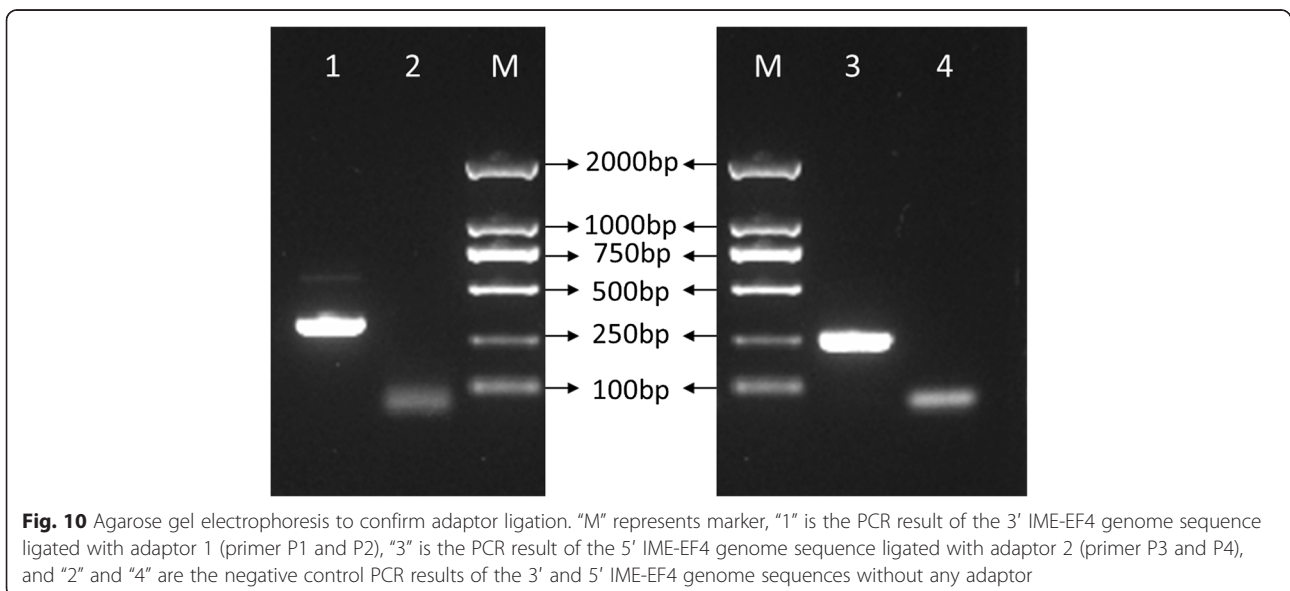
as shown in Fig. 5. In another assumption, the IME-EF4/IME-EFm1 had a 5' protruding cohesive end. Because the HTS experiment uses the T4 DNA polymerase and because T4 DNA polymerase catalyzes the synthesis of DNA in the 5' → 3' direction with 3' → 5' exonuclease activity rather than with 5' → 3' exonuclease function, the 5' protruding cohesive end must be repaired. This means that a repeated sequence between the 5' and 3' terminal sequences should be detectable. However, there were few short reads including both 5' and 3' terminal sequences in the HTS data. Since the linear phage genome was circularized through the genome terminal repetitions during the replication of phage genome, the short reads including 5' and 3' terminal sequences found in the HTS data might be the intermediate form of the circularized linear phage genome. Therefore, IME-EF4 and IME-EFm1 may both have 9 bp 3' protruding cohesive ends. As shown in Fig. 7, the terminase cleaves the IME-EF4/IME-EFm1 dsDNA from cosB and cosQ, which has a 9 nt protruding cohesive end between them.

Experimental verification

To confirm the conclusions drawn from the HTS data analysis, three molecular biology experiments were performed on IME-EF4. These experiments showed that the high-frequency reads were indeed the termini, which means that this termini analysis theory can be used to locate the phage's termini and their positions using only HTS data even if the procedure is performed quickly.

Mega-primer PCR sequencing

To identify the IME-EF4 complete genome sequence acquired by the HTS data, PCR amplification was performed on the IME-EF4 DNA genome. The PCR involved mega-primer-guided polymerization through the protruding cohesive end. The genome sequence snapshot including the 9base is shown in Fig. 8(a), where the upstream sequence was ...GAGATTCGTTT AAGCGAAAG and the downstream sequence was AT TAGTTTCTTCAAAAAATT... The results of PCR were consistent with those of HTS data statistical analysis (Fig. 5(a)). It proves that both the IME-EF4



complete genome and the protruding cohesive ends acquired from the HTS data statistics are correct.

Terminal Run-off sequencing

As shown in Fig. 9, two different cases of dsDNA protruding cohesive ends hypothetically exist. Case 1 shows the 5' protruding cohesive end situation, and case 2 describes the putative IME-EF4 situation of the 3' protruding cohesive end. If the IME-EF4 has the protruding cohesive end situation according to cases 1, then there would be signals after either the terminal sequence GAGATTCGTTTAAAGCGAAAG, ATTTTTTGAAGAACTAATA, or both of them in the terminal run-off sequencing result. However, as shown in Fig. 8(b), no signal was detected after the termini GAGATTCGT TTAAGCGAAAG in the positive strand or after the termini ATTTTTTGAAGAACTAATA in the negative strand. These results confirmed the conclusion that IME-EF4 has a linear, double-stranded DNA genome with a 9 nt 3' protruding cohesive end, represented by case 2—the 3' protruding cohesive end situation.

Adaptor ligation to the termini

As shown in Fig. 10, the 3' IME-EF4 genome sequence ligated to adaptor 1 was about 280 bp in size, and 5' IME-EF4 genome sequence ligated to adaptor 2 was about 250 bp in size. These findings were consistent with experimental design (P1 and P2 illustrations in Fig. 2). The sequencing results are shown in Fig. 8(c). This further proves that the two adaptors (adaptor 1 and adaptor 2) had successfully ligated to the IME-EF4 genome, which means the IME-EF4 genome has a 9 bp 3' protruding cohesive end.

Conclusions

In this study, using a novel termini analysis theory, the complete genome sequences and termini of newly isolated antibiotic-resistant *E. faecalis* phage IME-EF4 and antibiotic-resistant *E. faecium* phage IME-EFm1 were extracted and identified. Analysis of the IME-EF4/IME-EFm1's HTS data indicated that both IME-EF4 and IME-EFm1 have 9 bp 3' protruding cohesive ends (TCATCACCG in IME-EF4 positive strand, and GGGTCAGCG in IME-EFm1 positive strand). Further molecular biological experiments, including megaprimer PCR sequencing, terminal run-off sequencing, and adaptor ligation to the termini, fully supported these conclusions. Results indicated that this termini analysis theory facilitated acquisition of important phage genome termini information through analysis of HTS data alone, with no further molecular biology experiments required. In addition, the current analysis of the newly isolated antibiotic-resistant *E. faecalis* and antibiotic-resistant *E. faecium* phages, IME-EF4 and

IME-EFm1, have enriched the knowledge of antibiotic-resistant *Enterococcus* phages, which is essential to future antibiotic-resistant *Enterococcus* phage therapy.

Availability of supporting data

The data sets supporting the results of this article are available in the NCBI GenBank repository, [IME-EF4 complete genome <http://www.ncbi.nlm.nih.gov/nucleotide/589892984?report=genbank> and <http://www.ncbi.nlm.nih.gov/nucleotide/641468964>].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XZ carried out the data analysis and drafted and revised the manuscript. YW performed the molecular biology experiments. SL, XA, and GP participated in the HTS implementation. YH and HF contributed to the proof of the termini analysis theory. ZM and ZZ adjusted the experimental design. VVV contributed to the preparation of the figures. YC and YT designed and managed the experiments and revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgments

This research was supported by a grant from the National Hi-Tech Research and Development (863) Program of China (No. 2012AA022003 and No. 2014AA021402), the China Mega-Project on Infectious Disease Prevention (No. 2013ZX10004605, No. 2011ZX10004001, No. 2013ZX10004607-004 and No. 2013ZX10004217-002-003), and the State Key Laboratory of Pathogen and BioSecurity Program (No. SKLPBS1113).

Author details

¹State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, P.R. China. ²School of Life Science & Technology, China Pharmaceutical University, 24 Tong Jia Xiang, Nanjing 210009, P.R. China. ³Beijing Computing Center, Beijing 100094, P.R. China.

Received: 4 July 2014 Accepted: 4 May 2015

Published online: 28 May 2015

References

- Fraser SL, Lim J, Donskey J, Salata R. Enterococcal Infections. Retrieved June. 2008;12:2009.
- Moellering Jr RC. Emergence of Enterococcus as a significant pathogen. *Clin Infect Dis*. 1992;1173–1176.
- Edgeworth JD, Treacher DF, Eykyn SJ. A 25-year study of nosocomial bacteremia in an adult intensive care unit. *Crit Care Med*. 1999;27(8):1421–8.
- Richards MJ, Edwards JR, Culver DH, Gaynes RP. Nosocomial infections in combined medical-surgical intensive care units in the United States. *Infect Control Hosp Epidemiol*. 2000;21(8):510–5.
- Megran DW. Enterococcal endocarditis. *Clin Infect Dis*. 1992;15(1):63–71.
- Gerberding J, Gaynes R, Horan T, Abshire J, Alonso-Echanove J, Edwards J, et al. National nosocomial infections surveillance (NNIS) system report, data summary from January 1990-May 1999, issued June 1999. *Am J Infect Control*. 1999;27(6):520–32.
- Stuart CH, Schwartz SA, Beeson TJ, Owatz CB. *Enterococcus faecalis*: Its Role in Root Canal Treatment Failure and Current Concepts in Retreatment. *J Endod*. 2006;32(2):93–8.
- Rams T, Feik D, Young V, Hammond B, Slots J. Enterococci in human periodontitis. *Oral Microbiol Immunol*. 1992;7(4):249–52.
- Bonten MJ, Willems R, Weinstein RA. Vancomycin-resistant enterococci: why are they here, and where do they come from? *Lancet Infect Dis*. 2001;1(5):314–25.
- Deshpande LM, Fritsche TR, Moet GJ, Biedenbach DJ, Jones RN. Antimicrobial resistance and molecular epidemiology of vancomycin-resistant enterococci from North America and Europe: a report from the SENTRY antimicrobial surveillance program. *Diagn Microbiol Infect Dis*. 2007;58(2):163–70.

11. Hoshuyama T, Moriguchi H, Muratani T, Matsumoto T. Vancomycin-resistant enterococci (VRE) outbreak at a university hospital in Kitakyushu, Japan: case-control studies. *J Infect Chemother.* 2008;14(5):354–60.
12. Stevens RH, Ektefaie MR, Fouts DE. The annotated complete DNA sequence of *Enterococcus faecalis* phage ϕ Ef11 and its comparison with all available phage and predicted prophage genomes. *FEMS Microbiol Lett.* 2011;317(1):9–26.
13. Fujisawa H, Morita M. Phage DNA packaging. *Genes Cells.* 1997;2(9):537–45.
14. Zaballos A, Mellado RP, Salas M. Initiation of phage ϕ 29 DNA replication by mutants with deletions at the amino end of the terminal protein. *Gene.* 1988;63(1):113–21.
15. Stewart CR, Casjens SR, Cresawn SG, Houtz JM, Smith AL, Ford ME, et al. The Genome of *Bacillus subtilis* Phage SPO1. *J Mol Biol.* 2009;388(1):48–70.
16. Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, et al. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. *BMC Genomics.* 2010;11(1):120.
17. Bannwarth L, Goldberg AB, Chen C, Turk BE. Identification of exosite-targeting inhibitors of anthrax lethal factor by high-throughput screening. *Chem Biol.* 2012;19(7):875–82.
18. Chen CG, Fabri LJ, Wilson MJ, Panousis C. One-step zero-background IgG reformatting of phage-displayed antibody fragments enabling rapid and high-throughput lead identification. *Nucleic Acids Res.* 2014;42(4):e26. doi:10.1093/nar/gkt1142.
19. Li S, Fan H, An X, Fan H, Jiang H, Chen Y, et al. Scrutinizing Virus Genome Termini by High-Throughput Sequencing. *Plos One.* 2014;9(1), e85806.
20. Adams MH. *Phages.* New York: Interscience; 1959.
21. Sheng W, Huanhuan J, Jiankui C, Dabin L, Cun L, Bo P, et al. Isolation and rapid genetic characterization of a novel T4-like phage. *J Med Coll PLA.* 2010;25(6):331–40.
22. Joseph S, David WR. *Molecular cloning: a laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York 2001, 2.
23. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol.* 2012;30(5):434–9.
24. Pennisi E. Semiconductors inspire new sequencing technologies. *Science.* 2010;327(5970):1190–0.
25. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet Genomics.* 2011;38(3):95–109.
26. Lu S, Le S, Tan Y, Zhu J, Li M, Rao X, et al. Genomic and proteomic analyses of the terminally redundant genome of the *Pseudomonas aeruginosa* phage PaP1: establishment of genus PaP1-like phages. *PLoS One.* 2013;8(5):e62933.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

