Gene xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

V-GAP: Viral genome assembly pipeline

Yoji Nakamura [a,*], Motoshige Yasuike [a], Issei Nishiki [a], Yuki Iwasaki [a], Atushi Fujiwara [a,*], Yasuhiko Kawato [b], Toshihiro Nakai [c], Satoshi Nagai [a], Takanori Kobayashi [a], Takashi Gojobori [d,e], Mitsuru Ototake [a]

[a] Research Center for Aquatic Genomics, National Research Institute of Fisheries Science, Fisheries Research Agency, 2-12-4 Fukuura, Kanazawa, Yokohama 236-8648, Kanagawa, Japan
[b] National Research Institute of Aquaculture, Fisheries Research Agency, 422-1 Nakatsuhamaura, Minami-ise, Mie 516-0193, Japan
[c] Graduate School of Biosphere Science, Hiroshima University, Higashi-Hiroshima, Hiroshima 739-8528, Japan
[d] Computational Bioscience Research Center, Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia
[e] Center for Information Biology, National Institute of Genetics, 1111 Yata, Mishima 411-8540, Japan

ARTICLE INFO

Available online xxxx

Keywords:
Next-generation sequencing
De novo assembly
Shotgun sequences
Virus genomes

ABSTRACT

Next-generation sequencing technologies have allowed the rapid determination of the complete genomes of many organisms. Although shotgun sequences from large genome organisms are still difficult to reconstruct perfect contigs each of which represents a full chromosome, those from small genomes have been assembled successfully into a very small number of contigs. In this study, we show that shotgun reads from phage genomes can be reconstructed into a single contig by controlling the number of read sequences used in *de novo* assembly. We have developed a pipeline to assemble small viral genomes with good reliability using a resampling method from shotgun data. This pipeline, named V-GAP (Viral Genome Assembly Pipeline), will contribute to the rapid genome typing of viruses, which are highly divergent, and thus will meet the increasing need for viral genome comparisons in metagenomic studies.

1. Introduction

Next-generation sequencing (NGS) technologies have caused a paradigm shift in the field of molecular biology because genomic sequences can now be determined rapidly and at low cost (Metzker, 2010). In general, however, genome sequences have not been completely reconstructed even when massive numbers of shotgun reads were obtained. One reason for this is that NGS platforms often produce low-quality bases as noise compared with the traditional Sanger-based method. Another and more serious reason is that large genomes (particularly the eukaryotic genomes) have many repetitive elements such as transposons and microsatellite repeats, which disturb contig extension in the assembly process. On the other hand, for the smaller and simpler genomes of organisms such as bacteria and viruses, single chromosomes (or segments) are often rebuilt in the assembly process. Previously, we successfully determined the complete genomes of bacteriophages (phages) using a pyrosequencing method (Yasuike et al. 2013a,b, 2014). Based on this experience, we developed an automated pipeline to reconstruct complete genome sequences of viruses. The pipeline, V-GAP (Viral Genome Assembly Pipeline), can be applied to any virus composed of a single genomic segment. To evaluate the pipeline, we sequenced the genomes of four marine phages that infect the Gram-positive bacteria, *Lactococcus garvieae* (formerly *Enterococcus seriolicida*) (Kusuda et al. 1991) and *Streptococcus iniae* (Pier and Madin, 1976). Both bacteria are fish pathogens that commonly cause losses of cultured fish; hence, preventing infections caused by these bacteria is an important issue in aquaculture. The four marine phages are considered good candidate reagents for phage therapy because of their bacteriolysis ability (Nakai et al. 1999; Matsuoka et al. 2007; Kawato and Nakai, 2012).

From the environmental viewpoint, viruses have a variety of roles in ecology and biodiversity, just as other organisms do (Danovaro et al. 2008, 2011). Viruses (Phages) have been used to detect a small amount of bacteria in sea water (Matsuoka and Nakai, 2004), where viral particles amplified in the infected cells are released by bacteriolysis. Recent studies have suggested that many viruses (known and unknown) exist in the sea (Ortmann and Suttle, 2005; Suttle, 2005), raising concern in the field of metagenomics about the detection of viruses using NGS technologies (Labonté and Suttle, 2013). In parallel with the high-throughput sequencing used in metagenomic studies, determination of the genomes of isolated viral strains is also necessary for the identification and classification of each species. In this study, we show that V-GAP can meet this need, and can contribute to the cataloging of viral genome data produced by NGS platforms.

Abbreviation: NGS, next-generation sequencing.

* Corresponding authors.
E-mail addresses: yojnakam@affrc.go.jp (Y. Nakamura), jiwara@affrc.go.jp (A. Fujiwara).

http://dx.doi.org/10.1016/j.gene.2015.10.029

2. Materials and methods

2.1. Bacteriophage strains

The bacteriophage strains sequenced in this study are shown in Table 1. We sequenced four strains of double-stranded DNA viruses of the family Siphoviridae (Hendrix and Casjens, 2005); two of them (PLgW-1 and PLgY-16) infect *L. garvieae* (Park et al. 1997, 1998) and two (PSiJ31 and PSiJ32) infect *S. iniae* (Matsuoka et al. 2007). The phage purification procedure was according to a previously reported method (Kawato and Nakai, 2012).

2.2. DNA preparation and sequencing

One hundred microliter of purified phages (10^{10} PFU/mL), 400 μL of TNES-UREA buffer (6 M urea, 10 mM Tris–HCl (pH 7.5), 125 mM NaCl, 10 mM EDTA, 1% sodium dodecyl sulfate) and 5 μL of proteinase K (20 mg/mL) were mixed and incubated for 6 h at 37 °C. Then, the mixture was incubated with 1 μL of RNase A (20 mg/mL) for 2 h at 37 °C. Next, 50 μL of 5 M NaCl was added and the solution was deproteinated by extraction with an equal volume of phenol-chloroform (1:1). The DNA in the aqueous phase was mixed with 800 μL of ethanol and precipitated by centrifugation (12,000 × *g*, 10 min, 4 °C). The precipitated DNA was washed with 70% ethanol, and finally the DNA pellet was dissolved in 50 μL of TE buffer (10 mM Tris–HCl (pH 8.0), 1 mM EDTA). The concentration and purity of the extracted phage genomic DNA were calculated by optical density using the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE, USA).

Using a Covaris instrument (Covaris Inc., Woburn, MA, USA), 1 μg of the genomic DNA was sheared into 1500-bp fragments. Then, 454-pyrosequencing phage libraries were constructed from the sheared DNA using a GS Titanium Rapid Library Preparation Kit (Roche Diagnostics, Branford, CT, USA). Unique sequence tags were added to the DNA from each phage using a GS FLX Titanium Rapid Library MID Adaptor Kit (Roche Diagnostics). Subsequently, the four libraries were mixed in equal amounts (10^6 molecules/μL) and subjected to emulsion PCR (emPCR) using a GS Titanium SV emPCR (Lib-L) Kit (Roche Diagnostics). The captured beads with bound DNA were enriched and the number of enriched beads was estimated using a Z1 coulter particle counter (Beckman Counter, Inc., Hialeah, FL, USA). A total of 600,000 enriched beads were loaded onto one-quarter of the area of a 70 mm × 75 mm Titanium PicoTiter plate. Pyrosequencing was performed using a Roche 454 GS FLX + sequencer according to the manufacturer's protocols (Roche Diagnostics).

2.3. Genome assembly pipeline

The V-GAP flowchart is illustrated in Fig. 1A. In principle, V-GAP repeats *de novo* assembly using resampled reads from the shotgun sequencing data. In the flowchart, N is the number of reads resampled ($N = N_0, \ldots, N_{max}$ with step size $= \Delta N$). The N reads are chosen randomly from an input sff or fastq file, and assembled using Newbler ver. 2.9 (Roche Diagnostics). This step (i.e. read resampling and assembly) is repeated T times. If a single contig is produced successfully in the assembly, the result is stored, otherwise it is discarded. S_N ($S_N \le T$) is the number of successes out of T assembly trials using N reads. Finally, the best N (denoted by N') is selected when S_N is the largest. When there are several maximum S_Ns, V-GAP selects the one in which the average number of contigs produced was the smallest. The N' can be also manually selected after the resampling simulation and then the subsequent steps (i.e. multiple alignment and consensus sequence construction) can be resumed. The $S_{N'}$ single contigs are then aligned using MAFFT (Katoh and Toh, 2008), and a consensus sequence is constructed from the multiple alignment. In this study, we conducted 200 assembly simulations for each N ($T = 200$). For the multiple alignments, the contig closest in size to the mean of the $S_{N'}$ contigs was chosen as the seed contig, and the 5′ end of the other contigs were adjusted to that of the seed contig by trim and shift (Fig. 1B) based on BLASTN (Altschul et al. 1990) matches with ≥98% identity. Then, uniform distribution of the trim sizes of $S_{N'}$-1 contigs to the seed contig was evaluated using the χ^2 test with a significance level of 0.01. The trim sizes were split into M intervals and the uniform distribution across the M categories was tested. Here, M was determined by power analysis using the R *pwr* package, with the sample size, significance level, power, and effect size set to $S_{N'}$ - 1, 0.01, 0.8, and 0.3, respectively. At each site in a multiple alignment, the consensus nucleotide was determined by majority voting, that is, the most frequent nucleotide at each site was selected. In particular, the nucleotide conserved in more than 95% of the $S_{N'}$ contigs was considered a valid nucleotide.

2.4. Gene prediction and comparison

Open reading frames (ORFs) in each assembled genome sequence were predicted using two gene-finding programs, Glimmer3 (Delcher et al. 2007) and GeneMarkS (Besemer et al. 2001). ORFs predicted by either of these programs were considered as potential protein-coding genes. Orthologous gene pairs between phage strains were defined as reciprocal best hit pairs in BLASTP (Altschul et al. 1997) matches with an E-value of $<10^{-10}$. In gene annotation, BLASTP searches with an E-value $<10^{-5}$ were conducted against the virus protein sequences in the NCBI non-redundant database as of 20 February 2014. In genome sequence comparison among phage strains, BLASTN (Altschul et al. 1990) was used (E-value $<10^{-10}$).

3. Results and discussion

3.1. Assembly of the four phage genomes

The statistics of the sequenced phage genome assemblies are summarized in Table 1. For all four phage strains, single contigs were obtained successfully using from two to ten thousand resampled reads (Fig. 2A). For each strain, the number of reads that produced single contigs the most number of times (200 times) was chosen, and the 200 contigs produced were used for the multiple alignments (Table 1). Consensus sequences for the four genomes were constructed from the multiple alignments using MAFFT, and the genome sizes were 30,189 bp for PLgW-1, 29,342 bp for PLgY-16, and 40,426 bp for both PSiJ31 and PSiJ32. The proportions of valid nucleotides (i.e. the nucleotides conserved in more than 95% of the contigs) were estimated to be >99.99% in the four whole genome sequences, respectively.

Table 1
Genome assembly statistics for the bacteriophage strains sequenced in this study.

Strain	Host bacterium	Total reads	Resampled reads[a]	Number of successes[b]	Genome size (bp)	Coverage	Valid bases (bp)	Valid ratio (%)
PLgW-1	*Lactococcus garvieae*	34,633	6000	200	30,189	81	30,187	99.993
PLgY-16	*Lactococcus garvieae*	34,826	4000	200	29,342	55	29,340	99.993
PSiJ31	*Streptococcus iniae*	11,875	3000	200	40,426	29	40,424	99.995
PSiJ32	*Streptococcus iniae*	26,019	4000	200	40,426	38	40,424	99.995

[a] Number of reads assembled for the construction of consensus sequence.
[b] Number of times single contigs were obtained out of 200 simulations ($T = 200$).

Please cite this article as: Nakamura, Y., et al., V-GAP: Viral genome assembly pipeline, Gene (2015), http://dx.doi.org/10.1016/j.gene.2015.10.029

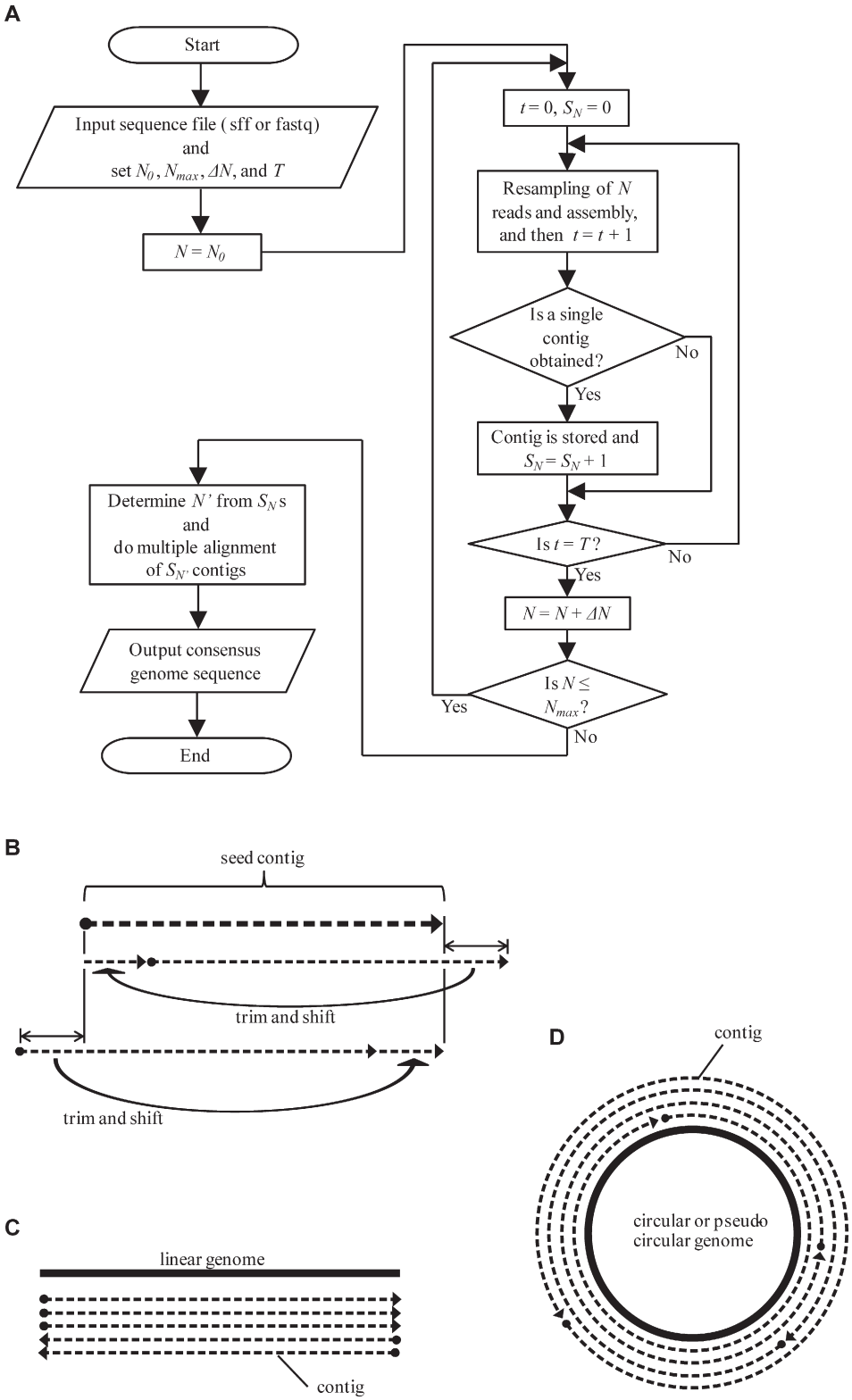

Fig. 1. Flowchart and description of the processes used in the viral genome assembly pipeline (V-GAP). (A) The V-GAP flowchart. (B) Processing of contigs by trim and shift for multiple alignment. Filled circle and arrowhead on each contig (dashed lines) indicate the start and end positions, respectively. (C) Patterns of aligned contig produced when a linear-genome virus is sequenced. (D) Patterns of aligned contig produced when a circular-genome virus is sequenced.

3.2. Estimating circular and linear genomes from the simulation data

Here, we focused on the start positions of the contigs produced in the assembly simulations. Viruses can have circular or linear genomes depending on their life stage or evolutionary origin. If the virus genome is linear, most of the contigs produced in the simulations should start from the same or close nucleotide position (Fig. 1C). An exception to this may occur if both ends of a linear genome sequence are very similar to each other; for example, the direct repeat found in the T7 phage genome (Dunn and Studier, 1983). In such a case, even if the genome is

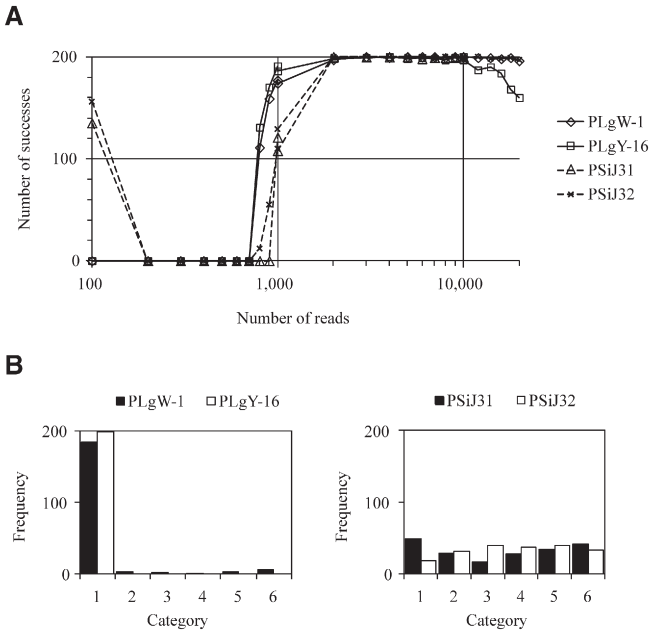

Fig. 2. Assembly simulations and distribution of trim sizes of contigs produced by V-GAP. (A) Relationship between the number of resampled reads and assembly success. The x-axis indicates the number of reads resampled from the original shotgun data, and the y-axis indicates the number of successes (i.e. the number of times a single contig was obtained) out of 200 assembly simulations. The simulations were performed under three read number N conditions (N = 100, …, 1000 with step size = 100; N = 1000, …, 10,000 with step size = 1000; and N = 10,000, …, 20,000 with step size = 2000). The simulations with N = 10,000, …, 20,000 were not performed for PSiJ31 because the total number of original reads were less than 12,000 (Table 1). (B) Distribution of the contig trim sizes prior to the multiple alignments, for *L. garvieae* phages (left panel) and *S. iniae* phages (right panel). The trim sizes were divided into six intervals, where a smaller category number indicates smaller trim size. The y-axis indicates the number of contigs produced in the simulations.

linear, the assembly algorithm might close the ends (we named this "pseudo-circular"). If the virus genome is circular (i.e. the genome sequence has no start or end position), when random shotgun reads are assembled, the contig may start from any position in the sequence (Fig. 1D). These possibilities can be tested roughly by examining the trim sizes of the contigs produced in the assembly simulations (i.e. start position shifts). The null hypothesis is that trim sizes should be uniformly distributed to the contig sizes. Almost all the contigs of the two *L. garvieae* phages had similar small trim sizes (Fig. 2B, left panel), and the uniform distributions of the contig trim sizes were rejected strongly at the significance level of $\alpha = 0.01$ (both P values $<10^{-10}$). In details, we confirmed that the start positions of almost all the contigs were identical, indicating that PLgW-1 and PLgY-16 both have linear genomes that were never closed in the simulations. Conversely, the trim sizes of the *S. iniae* phage contigs varied widely (Fig. 2B, right panel). Although the null hypothesis of uniform distribution of trim size for PSiJ31 was rejected ($P = 0.002$), it was not rejected for PSiJ32 ($P = 0.11$). This result implies that the *S. iniae* phages may have circular or pseudo-circular genomes. In V-GAP, there are two important beneficial aspects of using read resampling in assembly simulations: i) because it is based on majority voting, which minimizes assembly errors, the most accurate genome sequence possible (i.e. consensus sequence) can be obtained from shotgun data; and ii) the distribution of the start positions of the produced contigs can be tested statistically, which allows the genomic form of the virus to be predicted.

3.3. Prediction of protein-coding genes in the assembled phage genomes

In the PLgW-1 and PLgY-16 genomes, 58 and 56 protein-coding genes were predicted, respectively, while in both the PSiJ31 and PSiJ32 genomes, 70 protein-coding genes were predicted (Table 2). Comparisons of orthologous gene pairs between the two *L. garvieae* phages and between the two *S. iniae* phages revealed that the gene orders were almost conserved in the respective genome pairs (Fig. 3). The 56 predicted genes in PLgY-16 all had orthologs in PLgW-1, and only two genes of 22 kb, *orf45* and *orf46*, were specific to PLgW-1 (Fig. 3A); although *orf46* was almost identical to a partial region of *orf44* in PLgY-16, which was estimated to be orthologous to *orf44* in PLgW-1. Therefore, an *orf45* insertion may have split *orf44* into two genes in the PLgW-1 genome. In PSiJ31 and PSiJ32, a one-to-one correspondence between all 70 of the orthologous pairs was obtained, and a perfect synteny of gene order was obtained when a 473-bp fragment at the 3′ end of the PSiJ32 genome was shifted to the beginning of the sequence (Fig. 3B). Within the shifted 473-bp region, the C-terminal end of the first protein-coding gene in the genome sequence was included, suggesting that the PSiJ32 contig sequence is circular.

3.4. Comparisons between and among the four phage genomes

The two *L. garvieae* phage genomes were more than 99% identical at the nucleotide sequence level, and hence may eventually be considered the same species. However, in addition to the PLgW-1-specific *orf45* region mentioned in section 3.3, the first 500–600 bp sequences of the PLgW-1 and PLgY-16 genomes were a bit different. In this region, a predicted protein-coding gene (*orf1*) was found to be conserved in the both genomes, but the nucleotide sequences of this ORF shared only 73% identity. The genomes of the two *S. iniae* strains were 99.99% identical to each other at the nucleotide sequence level with only three mismatches across the two genomes, indicating that these stains are also likely to be the same species. Considering the contig structure of PSiJ32 (sections 3.2 and 3.3), this species may have a circular or pseudo-circular genome. Sequence similarity searches against the NCBI non-redundant database showed that most of the predicted protein-coding genes in the *L. garvieae* and *S. iniae* genomes were similar to sequences from members of the family Siphoviridae (Table 2), consistent with the morphological observations by electron microscope of the four phage strains (Park et al. 1997; Matsuoka et al. 2007; Kawato and Nakai, 2012). However, the average sequence similarities to the best BLAST hits were quite low; namely, 54.4% for PLgW-1, 52.7% for PLgY-16, and 58.6% for both PSiJ31 and PSiJ32. Moreover, between the *L. garvieae* and *S. iniae* phages, the predicted gene sequences shared very little similarity at both the nucleotide and amino acid levels. This finding suggests that, the *L. garvieae* and *S. iniae* phages belong to two highly diverging groups, although they are classified into the same family.

Genome sequences can be good evidence for species identification, independent of morphological observations. V-GAP is a useful tool that will contribute to rapid genome determination, strain typing, and classification of viruses. Viruses often emerge quickly as new strains carrying different genotypes, and metagenomics has shown that there are still many unknown viruses in the environment. Thus, efficient pipelines, such as V-GAP, for NGS data will become increasingly important for cataloging the large numbers of reference viral genomes that are likely to be sequenced. The current version of V-GAP is available only for species with a single chromosome (or segment). Further

Table 2
Predicted ORFs and taxon of their top BLAST hits.

Family	Phage strain		
	PLgW-1	PLgY-16	PSiJ31 and PSiJ32
Siphoviridae	34	32	21
Podoviridae	1	1	1
Myoviridae	0	0	1
Unknown	22	21	29
No BLAST hit	1	2	18
Total	58	56	70

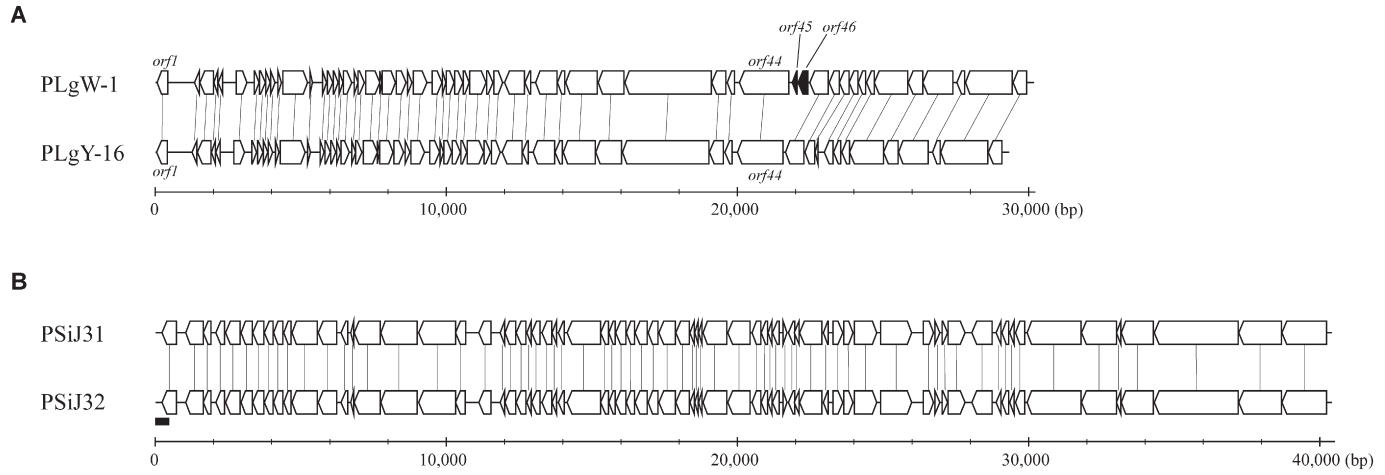

Fig. 3. Comparison of orthologous gene pairs between the phage genomes. Orthologous genes are linked by lines connecting the genome pairs. (A) The PLgW-1 and PLgY-16 genomes. The genes mentioned in sections 3.3 and 3.4 are marked. The PLgW-1-specific genes (*orf45* and *orf46*) are shown in black. (B) The PSiJ31 and PSiJ32 genomes. The 473-bp region that was shifted from the end to the beginning of the PSiJ32 genome is marked with a bold line.

improvement of V-GAP to handle viruses with multiple genomic segments, such as the influenza virus, is a challenge for the future development of the pipeline.

4. Conclusions

In viral genome sequencing using NGS, the complete genome can be reconstructed by controlling the number of reads used in *de novo* assembly. Thus, we have developed a pipeline that runs genome assembly simulations by read resampling, and applied it to the shotgun sequence data of four marine phages. Statistical tests suggested that the genomes of the two strains infecting *L. garvieae* were linear, while the genomes of the two strains infecting *S. iniae* may be circular or pseudo-circular. This result was supported by a comparison of predicted protein-coding regions in the genomes. Sequence comparisons suggested that the two *S. iniae* phages are the same species, and that the two *L. garvieae* phages may also be the same species. For the *L. garvieae* phages, however, an insertion and local diversification of the sequences were observed. Sequence similarity searches showed that all four phage strains belong to the family Siphoviridae, as previously found by morphological observations, but that they may be quite distantly related. Thus, the pipeline V-GAP allows rapid strain typing of viruses in the environment, which can enable genome-wide comparisons for further analysis.

Acknowledgments

This study was supported by a Grant-in-Aid (Marine Metagenomics for Monitoring the Coastal Microbiota) from the Ministry of Agriculture, Forestry and Fisheries of Japan.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Besemer, J., Lomsadze, A., Borodovsky, M., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29, 2607–2618.

Danovaro, R., Corinaldesi, C., Dell'anno, A., Fuhrman, J.A., Middelburg, J.J., Noble, R.T., Suttle, C.A., 2011. Marine viruses and global climate change. FEMS Microbiol. Rev. 35, 993–1034.

Danovaro, R., Dell'Anno, A., Corinaldesi, C., Magagnini, M., Noble, R., Tamburini, C., Weinbauer, M., 2008. Major viral impact on the functioning of benthic deep-sea ecosystems. Nature 454, 1084–1087.

Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics 23, 673–679.

Dunn, J.J., Studier, F.W., 1983. Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. 166, 477–535.

Hendrix, R.W., Casjens, S.R., 2005. Family Siphoviridae. In: Fauquet, C., Mayo, M., Maniloff, J., Desselberger, U., Ball, L. (Eds.), Virus Taxonomy. VIIIth report of the ICTV. Elsevier Academic Press, London, United Kingdom, pp. 57–70.

Katoh, K., Toh, H., 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief. Bioinform. 9, 286–298.

Kawato, Y., Nakai, T., 2012. Infiltration of bacteriophages from intestinal tract to circulatory system in goldfish. Fish Pathol. 47, 1–6.

Kusuda, R., Kawai, K., Salati, F., Banner, C.R., Fryer, J.L., 1991. *Enterococcus seriolicida* sp. nov., a fish pathogen. Int. J. Syst. Bacteriol. 41, 406–409.

Labonté, J.M., Suttle, C.A., 2013. Metagenomic and whole-genome analysis reveals new lineages of gokushoviruses and biogeographic separation in the sea. Front. Microbiol. 4, 404.

Matsuoka, S., Nakai, T., 2004. Seasonal appearance of *Edwardsiella tarda* and its bacteriophages in the culture farms of Japanese flounder. Fish Pathol. 39, 145–152.

Matsuoka, S., Hashizume, T., Kanzaki, H., Iwamoto, E., Se Chang, P., Yoshida, T., Nakai, T., 2007. Phage therapy against β-hemolytic Streptococcicosis of Japanese flounder *Paralichthys olivaceus*. Fish Pathol. 42, 181–189.

Metzker, M.L., 2010. Sequencing technologies - the next generation. Nat. Rev. Genet. 11, 31–46.

Nakai, T., Sugimoto, R., Park, K.H., Matsuoka, S., Mori, K., Nishioka, T., Maruyama, K., 1999. Protective effects of bacteriophage on experimental *Lactococcus garvieae* infection in yellowtail. Dis. Aquat. Org. 37, 33–41.

Ortmann, A.C., Suttle, C.A., 2005. High abundances of viruses in a deep-sea hydrothermal vent system indicates viral mediated microbial mortality. Deep-Sea Res. I 52, 1515–1527.

Park, K., Matsuoka, S., Nakai, T., Muroga, K., 1997. A virulent bacteriophage of *Lactococcus garvieae* (formerly *Enterococcus seriolicida*) isolated from yellowtail *Seriola quinqueradiata*. Dis. Aquat. Org. 29, 145–149.

Park, K.H., Kato, H., Nakai, T., Muroga, K., 1998. Phage typing of *Lactococcus garvieae* (formerly *Enterococcus seriolicida*) a pathogen of cultured yellowtail. Fish. Sci. 64, 62–64.

Pier, G.B., Madin, S.H., 1976. *Streptococcus iniae* sp. nov., a beta-hemolytic *Streptococcus* isolated from an amazon freshwater dolphin, *Inia geoffrensis*. Int. J. Syst. Bacteriol. 26, 545–553.

Suttle, C.A., 2005. Viruses in the sea. Nature 437, 356–361.

Yasuike, M., Kai, W., Nakamura, Y., Fujiwara, A., Kawato, Y., Hassan, E.S., Mahmoud, M.M., Nagai, S., Kobayashi, T., Ototake, M., Nakai, T., 2014. Complete genome sequence of the *Edwardsiella ictaluri*-specific bacteriophage PEi21, isolated from river water in Japan. Genome Announc. 2.

Yasuike, M., Sugaya, E., Nakamura, Y., Shigenobu, Y., Kawato, Y., Kai, W., Fujiwara, A., Sano, M., Kobayashi, T., Nakai, T., 2013a. Complete genome sequences of *Edwardsiella tarda*-lytic bacteriophages KF-1 and IW-1. Genome Announc. 1.

Yasuike, M., Sugaya, E., Nakamura, Y., Shigenobu, Y., Kawato, Y., Kai, W., Nagai, S., Fujiwara, A., Sano, M., Kobayashi, T., Nakai, T., 2013b. Complete genome sequence of a novel myovirus which infects atypical strains of *Edwardsiella tarda*. Genome Announc. 1.