Complete Genome Sequence of VpKK5, a Novel *Vibrio parahaemolyticus* Lytic Siphophage

Tamrin M. Lal, Julian Ransangan

Borneo Marine Research Institute, Universiti Malaysia Sabah, Jalan UMS, Kota Kinabalu, Sabah, Malaysia

This paper describes the complete sequence of a novel lytic marine siphophage, VpKK5, that is specific to *Vibrio parahemolyticus*.

Address correspondence to Julian Ransangan, liandra@ums.edu.my.

*V*ibrio parahaemolyticus is an emergence of bacterial pathogens implicated in fish vibriosis (1, 2). Here, we describe the complete sequence of the *V. parahaemolyticus*-specific phage designated VpKK5.

The VpKK5 siphophage was isolated from coastal sand sediment of Sabah, Malaysia. The genome was extracted and purified using the DNeasy blood and tissue kit (Qiagen) according to the manufacturer's instructions. A DNA library was prepared using Nextera XT (Illumina) and sequenced using NGS Illumina Miseq PE sequencing (AITBiotech, Singapore). A set of reads (2 × 250,000 samples) with an average read size of 250 bp were *de novo* assembled using Velvet 1.1 (Zerbino, European Bioinformatics Institute, United Kingdom) into a single contig. The genome terminal was predicted using a tandem repeat finder (3). The complete genome sequence was then subjected to BLASTn. The open reading frames (ORFs) of the genome were predicted using three bioinformatics programs, GeneMarkS (4), GLIMMER v3.02 (5), and ORF Finder (6). The function of each ORF was predicted using the PSI-BLAST (6), ScanProsite (7), Pfam (8), InterPro (9), and NCBI Conserved Domain databases (6). The sequences of tRNAs were predicted using the tRNAscan-SE program (10). The virulence factor was analyzed against VFDB (11) and MvirDB (12) databases.

The sequencing analysis revealed that the complete genome of VpKK5 is 56,637 bp in length and has a 51.32% G+C content. It consists of 80 predicted coding sequences (CDSs) with no tRNA detected. The 80 CDSs represent 90.66% of the total genome. The genes varied from 138 bp (orf47) to 3,171 bp (orf39). Thirty-seven CDSs were hypothetically novel while the others 43 CDSs showed homology but at low identity (<62%). The protein function analyses showed some CDSs are related to the DNA replication and packaging (orf15, orf19, orf24, orf34, orf35, orf60, orf63), head structure (orf45, orf56 and orf58), tail structure (orf39, orf40, orf41, orf42, orf43), and phage-bacteria interaction property (orf62). Interestingly, the genome sequence of the VpKK5 did not exhibit homology to any virulence factors. Unfortunately, the genome end cannot be determined in this study, but the deposited VpKK5 genome was arranged from replication to structural genes.

The study concluded that the genome of the *Vibrio* phage VpKK5 is novel. Lack of virulence factors would allow the phage to be used in phage therapy. The future applications of this novel phage are promising, particularly in therapy against *V. parahemolyticus* infection.

Nucleotide sequence accession number. The complete sequence of the VpKK5 genome was deposited in GenBank under the accession no. KM378617.

ACKNOWLEDGMENTS

REFERENCES

1. **Khouadja S, Lamari F, Backhrouf A.** 2013. Characterization of *Vibrio parahaemolyticus* isolated from farm sea bass (*Dicentrarchus labrax*) during disease outbreaks. Int Aquat Res **5:**13. http://dx.doi.org/10.1186/2008-6970-5-13.
2. **Ransangan J, Lal MMT.** 2013. Simultaneous detection of *Photobacterium damselae*, *Vibrio alginolyticus*, *Vibrio harveyi* and *Vibrio parahaemolyticus* using multiplex PCR amplification method. Int Res J Biol Sci **2:**79–84.
3. **Benson G.** 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res **27:**573–580. http://dx.doi.org/10.1093/nar/27.2.573.
4. **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. http://dx.doi.org/10.1093/nar/29.12.2607.
5. **Delcher AL, Harmon D, Kasif S, White O, Salzberg SL.** 1999. Improved microbial gene identification with GLIMMER. Nucleic Acids Res **27:**4636–4641. http://dx.doi.org/10.1093/nar/27.23.4636.
6. **Hardies SC, Comeau AM, Serwer P, Suttle CA.** 2003. The complete sequence of marine bacteriophage VpV262 infecting *Vibrio parahaemolyticus* indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment. Virology **310:**359–371. http://dx.doi.org/10.1016/S0042-6822(03)00172-7.
7. **de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N.** 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. Nucleic Acids Res **34:**W362–W365. http://dx.doi.org/10.1093/nar/gkl124.
8. **Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M.** 2014. Pfam: the protein families database. Nucleic Acids Res **42:**D222–D230. http://dx.doi.org/10.1093/nar/gkt1223.

9. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J, Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJ, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C. 2009. Interpro: the integrative protein signature database. Nucleic Acids Res 37:D211–D215. http://dx.doi.org/10.1093/nar/gkn785.

10. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS Web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:W686–W689. http://dx.doi.org/10.1093/nar/gki366.

11. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325. http://dx.doi.org/10.1093/nar/gki008.

12. Zhou CE, Smith J, Lam M, Zemla A, Dyer MD, Slezak T. 2007. MvirDB—a microbial database of protein toxins, virulence factors and antibiotic resistance genes for bio-defence applications. Nucleic Acids Res 35:D391–D394. http://dx.doi.org/10.1093/nar/gkl791.