Genome Sequence of a Novel Archaeal Fusellovirus Assembled from the Metagenome of a Mexican Hot Spring

Luis E. Servín-Garcidueñas,[a] **Xu Peng,**[b] **Roger A. Garrett,**[b] **Esperanza Martínez-Romero**[a]

Department of Ecological Genomics, Center for Genomic Sciences, National University of Mexico, Cuernavaca, Morelos, Mexico[a]; Department of Biology, Archaea Centre, University of Copenhagen, Copenhagen, Denmark[b]

The consensus genome sequence of a new member of the family *Fuselloviridae* designated as SMF1 (*Sulfolobales* Mexican fusellovirus 1) is presented. The complete circular genome was recovered from a metagenomic study of a Mexican hot spring. SMF1 exhibits an exceptional coding strand bias and a reduced set of fuselloviral core genes.

Address correspondence to Luis E. Servín-Garcidueñas, luis.e.servin@gmail.com.

Members of the *Fuselloviridae* family from the crenarchaeal order *Sulfolobales* have been characterized, and they are abundant in extreme geothermal environments (1, 2). They carry circular double-stranded DNA (dsDNA) genomes and exhibit spindle-shaped morphologies. Here, we report the consensus genome sequence of a novel fusellovirus recovered from aqueous sediments from Los Azufres, Mexico.

Samples were collected from a hot spring with a pH of 3.6 and a temperature of 65°C. DNA was purified using the UltraClean microbial and the UltraClean Mega soil DNA kits (MoBio Laboratories, Inc., Carlsbad, CA). Sequencing was performed on an Illumina GAIIx platform, producing 36-bp paired-end reads with 300-bp inserts representing 216 Mb. Reads were assembled using Velvet 1.2.07 (3). A set of contigs were predicted by BLASTX searches to be of fuselloviral origin. Gaps were closed iteratively by mapping and reassembling reads to these contigs using Maq 0.7.1 (4) and Velvet. Open reading frames (ORFs) were predicted using GeneMark.hmm2.0 (5) and were manually verified using Artemis (6).

The average sequence coverage of the 14,847-bp circular dsDNA genome was 1,257-fold. We detected 57 candidate single nucleotide polymorphisms by Maq. The G+C content was 45.43%, higher than the 37.5 to 39.7% content of other fuselloviral genomes (1, 2, 7–9).

The genome has a strong coding-strand bias, not previously seen for fuselloviruses, with only the ubiquitous fuselloviral integrase encoded on one strand. The gene organization is also exceptional for fuselloviruses, with a high incidence of genes arranged in operons, which are also likely to encode cofunctional proteins.

Twenty-four genes were predicted, 22 of which are arranged in five operons. Fourteen genes have putative fuselloviral homologs, consistent with SMF1 being a member of the *Fuselloviridae* family. Most gene products show 30 to 70% amino acid sequence similarity to the best fuselloviral matches. Previous studies identified thirteen genes conserved in all fusellovirus genomes (2),

and nine of these were localized in a "core" genomic region of SMF1. The core genes encode a DnaA-like protein, the integrase, one VP1-like structural protein, a putative helix-turn-helix (HTH) transcriptional regulator, and five proteins with unknown functions.

Five additional putative gene products shared with other fuselloviruses include a second VP1-like protein, a VP2-like structural protein, a putative end-filament protein, a regulatory protein, and a hypothetical protein. Three further nonconserved ORF products showed sequence similarities to putative regulatory proteins.

The host of SMF1 is likely to be a member of the order *Sulfolobales*. Fuselloviruses can replicate in both *Sulfolobus* and *Acidianus* species of the order *Sulfolobales* (2), and they are predicted to have an extended host range that may include as-yet-uncultured species (10).

In conclusion, the SMF1 genome was recovered from a site widely separated geographically from the locations of other sequenced fuselloviruses. The SMF1 genome shows exceptional properties, including a coding-strand bias and a high incidence of genes organized in operon structures, but nevertheless, it retains a large set of conserved fusellovirus genes, which lends further support to the exchange of genetic material over intercontinental distances (2, 10).

Nucleotide sequence accession number. The genome sequence was deposited in GenBank under the accession no. KC618393.

ACKNOWLEDGMENTS

REFERENCES

1. **Wiedenheft B, Stedman K, Roberto F, Willits D, Gleske AK, Zoeller L, Snyder J, Douglas T, Young M.** 2004. Comparative genomic analysis of hyperthermophilic archaeal *Fuselloviridae* viruses. J. Virol. **78**:1954–1961.

2. **Redder P, Peng X, Brügger K, Shah SA, Roesch F, Greve B, She Q, Schleper C, Forterre P, Garrett RA, Prangishvili D.** 2009. Four newly isolated fuselloviruses from extreme geothermal environments reveal unusual morphologies and a possible interviral recombination mechanism. Environ. Microbiol. **11**:2849–2862.

3. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.

4. **Li H, Ruan J, Durbin R.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Res. **18**:1851–1858.

5. **Besemer J, Borodovsky M.** 1999. Heuristic approach to deriving models for gene finding. Nucleic Acids Res. **27**:3911–3920.

6. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21**:3422–3423.

7. **Palm P, Schleper C, Grampp B, Yeats S, McWilliam P, Reiter WD, Zillig W.** 1991. Complete nucleotide sequence of the virus SSV1 of the archaebacterium *Sulfolobus shibatae*. Virology **185**:242–250.

8. **Stedman KM, She Q, Phan H, Arnold HP, Holz I, Garrett RA, Zillig W.** 2003. Relationships between fuselloviruses infecting the extremely thermophilic archaeon *Sulfolobus*: SSV1 and SSV2. Res. Microbiol. **154**:295–302.

9. **Peng X.** 2008. Evidence for the horizontal transfer of an integrase gene from a fusellovirus to a pRN-like plasmid within a single strain of *Sulfolobus* and the implications for plasmid survival. Microbiology **154**:383–391.

10. **Snyder JC, Wiedenheft B, Lavin M, Roberto FF, Spuhler J, Ortmann AC, Douglas T, Young M.** 2007. Virus movement maintains local virus population diversity. Proc. Natl. Acad. Sci. U. S. A. **104**:9102–19107.