JVI Journals.ASM.org

Codon Usage Frequency of RNA Virus Genomes from High-Temperature Acidic-Environment Metagenomes

Kenneth M. Stedman, Nicholas R. Kosmicki, Geoffrey S. Diemer

Center for Life in Extreme Environments, Biology Department, Portland State University, Portland, Oregon, USA

Bolduc et al. recently described tantalizing partial genomes of RNA viruses from amplified metagenomes from acidic hot springs. Parallel analyses of DNA sequences from these environments indicate that the majority of potential hosts are *Archaea*. Thus, the authors concluded that these sequences may represent the genomes of the first archaeal RNA virus (1).

All viruses are dependent on host translational machinery. In particular, viruses that do not encode their own tRNAs are completely dependent on host tRNAs for translation. Analysis of the partial genome sequences deposited by Bolduc et al. (1), contig0002 and contig0028, using tRNAscanSE indicates that neither of the contigs encodes a tRNA (2). In order to ascertain possible hosts for these viruses, we analyzed their codon usage frequency relative to known nonviral and nonplastid codon usage frequencies tabulated from GenBank (CUTG) (3) (http://www.kazusa.or.jp/codon). The CUTG database contains codon usage data from sequences from 11,775 organisms, of which 58.5% are eukaryal, 39.9% are bacterial, and 1.6% are archaeal. Mean codon frequency usage difference (MCUFD) was calculated by determining the mean of the differences in codon usage frequency for each amino acid for viral protein-coding genes relative to host genes.

All one hundred organisms with codon usage most similar to that of contig0002 were eukaryal. The MCUFD of these sequences were 7.02 to 9.16%. Similarly, 93 of the 100 organisms with codon usage most similar to that of contig0028 were eukaryal. The remaining 7 were bacterial (MCUFD, 8.78 to 10.03%). The most similar archaeal codon usage relative to that of contig0002 was that of *Methanosarcina acetivorans*, a mesophilic anaerobic methanogen, with 11.13% MCUFD and 887th best match.

In contrast, and in spite of their 1.6% representation in the CUTG database, the 2 organisms with codon usage most similar to that of the archaeal DNA virus SSV1 were the hyperthermophilic archaea *Sulfolobus tengcongensis* and *Sulfolobus islandicus*. Moreover, 13 of the 100 organisms with the most similar codon usage were thermophilic archaea (MCUFD, 5.59 to 8.29%), 3 were bacteria (MCUFD, 8.13 to 8.39%), and the remainder were eukaryotes (MCUFD, 6.19 to 8.45%). This probably reflects either eukaryotic bias of the database or evolutionary similarities between the translational machineries of archaea and eukarya (4).

For the eukaryotic RNA poliovirus, 99 of the 100 organisms with the most similar codon usage were eukaryal (MCUFD, 6.55 to 8.41%); the lone exception was the bacterium *Anaplasma marginale*, with an MCUFD of 8.78%. Curiously, for the RNA bacteriophage Qβ, only 10 of the top 100 organisms with the most similar codon usage were bacterial (MCUFD, 7.41 to 8.44%) and the remainder were eukaryal (MCUFD, 6.81 to 8.44%).

The lack of congruence between archaeal codon usage and that of contig0002 and contig0028 (MCUFD, 11.13%) relative to that of the known archaeal virus SSV1 (MCUFD, 5.59%) suggests that these novel viruses are unlikely to replicate in archaea. However, the high minimum MCUFD (7.02%) indicates that the host for these new viruses is probably novel. These analyses reinforce that we still have a great deal to learn about viruses in extreme environments. We look forward to future findings on the viruses represented by the fascinating genome sequences discovered by Bolduc et al. (1).

REFERENCES

1. **Bolduc B, Shaughnessy DP, Wolf YI, Koonin EV, Roberto FF, Young M.** 2012. Identification of novel positive-strand RNA viruses by metagenomic analysis of archaea-dominated Yellowstone hot springs. J. Virol. **86**:5562–5573.
2. **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.
3. **Nakamura Y, Gojobori T, Ikemura T.** 2000. Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucleic Acids Res. **28**:292.
4. **Benelli D, Londei P.** 2011. Translation initiation in Archaea: conserved and domain-specific features. Biochem. Soc. Trans. **39**(1):89–93.

Address correspondence to Kenneth M. Stedman, kstedman@pdx.edu.

For the author reply, see doi:10.1128/JVI.02883-12.

doi:10.1128/JVI.02610-12