Syn-CpG-Spacer: A Panel web app for synonymous recoding of viral genomes with CpG dinucleotides

Aleksander Sulkowski [1]¶, **Clément Bouton** [1], and **Chad Swanson** [1]

1 Department of Infectious Diseases, King's College London, London, United Kingdom ¶ Corresponding author

Summary

Vertebrate genomes contain lower than expected frequencies of the CpG dinucleotide. Consequently, many vertebrate viruses have evolved to mimic this composition, possibly in order to evade host antiviral defences (Greenbaum et al., 2008). For example, the antiviral protein ZAP binds CpGs in viral single stranded RNA with specific spacing requirements (Gonçalves-Carneiro et al., 2022), though CpGs are also likely depleted in viral genomes due to other selective pressures (Forni et al., 2023). Increasing CpG abundance by synonymous recoding could facilitate attenuation of viruses without compromising their epitope antigenicity by changing non-CpG codons to alternatives containing CpG without changing the overall amino acid sequence (Gonçalves-Carneiro et al., 2022; Le Nouën et al., 2019; Sharp et al., 2023). There are three ways CpGs can be synonymously introduced in codons: at positions 1-2 for arginine (e.g. AGA → CGA), 2-3 for several amino acids (e.g. ACA → ACG), or in a 3-1 split configuration, if a subsequent codon begins with a G (e.g. ATA-GCA → ATC-GCA).

Syn-CpG-Spacer is a Python progressive web app (PWA) made with the Panel library that allows for consistent recoding of viral sequences and applying biologically relevant constraints. These include setting a minimum gap between CpG's, optimizing for an average CpG gap, protecting cis-acting regulatory signals from modification, and modulating the A-content in the overall sequence. The app features a sequence viewer made with the Bokeh library that highlights CpG dinucleotides, allowing for efficient analysis of the resulting distribution of CpGs. This is complemented by a statistical data table. Utilizing Biopython modules, the user can load their sequence as a FASTA file and download the outputs as an alignment in the same format. As a PWA running on Pyodide, the code is only executed in the user's browser and they can install the app onto their machine for offline use.

Statement of need

There are currently no published tools that allow specifying the spacing of CpG dinucleotides when synonymously recoding sequences. SSE can recode a sequence to a defined CpG frequency, and has a graphical interface (Simmonds, 2012). However, it cannot control the spacing of the CpGs. Newer tools require the knowledge of R, such as SynMut (Gu & Poon, 2023). A problem with bioinformatics packages is a high access barrier for users who are not familiar with programming languages. Here, the installation and update processes are streamlined thanks to the PWA approach.

With the current lack of CpG recoding tools, researchers may turn to in-house solutions which can hamper the reproducibility of their results, while also introducing the room for error. Syn-CpG-Spacer makes it more efficient to synonymously recode a sequence compared to doing so without any support tools.

Research applications

In recent years, there has been an increased research focus on introducing CpGs into viral genomes as a mechanism to create live attenuated virus vaccines, such as influenza A virus (Gaunt et al., 2016; Sharp et al., 2023). However, the mechanism of action for how CpGs restrict the virus is unclear and it could be due to sensitising it to antiviral proteins such as ZAP (Ficarelli et al., 2021) or other poorly characterised effects on viral gene expression. This tool can be used to introduce CpGs into different viral genomes with specific spacing to determine if this attenuates the virus in vitro or in vivo and characterise the mechanism of attenuation, which will aid the development of live attenuated viral vaccines (Le Nouën et al., 2019).

Another potential application for the software is creation of CpG islands, which are long stretches of DNA rich in CpG dinucleotides that allow for epigenetic control of transcription. While most vertebrate CpGs are methylated, and thus transcriptionally silent, the DNA in CpG islands is hypomethylated, facilitating transcription factor binding (Angeloni & Bogdanovic, 2021).

Acknowledgements

References

Angeloni, A., & Bogdanovic, O. (2021). Sequence determinants, function, and evolution of CpG islands. *Biochemical Society Transactions*, *49*(3), 1109–1119. https://doi.org/10.1042/BST20200695

Ficarelli, M., Neil, S. J. D., & Swanson, C. M. (2021). Targeted restriction of viral gene expression and replication by the ZAP antiviral system. *Annual Review of Virology*, *8*(1), 265–283. https://doi.org/10.1146/annurev-virology-091919-104213

Forni, D., Pozzoli, U., Cagliani, R., Clerici, M., & Sironi, M. (2023). Dinucleotide biases in RNA viruses that infect vertebrates or invertebrates. *Microbiology Spectrum*, *0*(0), e02529–23. https://doi.org/10.1128/spectrum.02529-23

Gaunt, E., Wise, H. M., Zhang, H., Lee, L. N., Atkinson, N. J., Nicol, M. Q., Highton, A. J., Klenerman, P., Beard, P. M., Dutia, B. M., Digard, P., & Simmonds, P. (2016). Elevation of CpG frequencies in influenza a genome attenuates pathogenicity but enhances host response to infection. *eLife*, *5*, e12735. https://doi.org/10.7554/eLife.12735

Gonçalves-Carneiro, D., Mastrocola, E., Lei, X., DaSilva, J., Chan, Y. F., & Bieniasz, P. D. (2022). Rational attenuation of RNA viruses with zinc finger antiviral protein. *Nature Microbiology*, *7*(10), 1558–1567. https://doi.org/10.1038/s41564-022-01223-8

Greenbaum, B. D., Levine, A. J., Bhanot, G., & Rabadan, R. (2008). Patterns of evolution and host gene mimicry in influenza and other RNA viruses. *PLOS Pathogens*, *4*(6), e1000079. https://doi.org/10.1371/journal.ppat.1000079

Gu, H., & Poon, L. L. M. (2023). *SynMut: SynMut: Designing synonymously mutated sequences with different genomic signatures* (Version 1.16.0). Bioconductor version: Release (3.17). https://doi.org/10.18129/B9.bioc.SynMut

Le Nouën, C., Collins, P. L., & Buchholz, U. J. (2019). Attenuation of human respiratory viruses by synonymous genome recoding. *Frontiers in Immunology*, *10*. https://www.frontiersin.org/articles/10.3389/fimmu.2019.01250

84 Sharp, C. P., Thompson, B. H., Nash, T. J., Diebold, O., Pinto, R. M., Thorley, L., Lin,
85 Y.-T., Sives, S., Wise, H., Hendry, S. C., Grey, F., Vervelde, L., Simmonds, P., Digard,
86 P., & Gaunt, E. R. (2023). CpG dinucleotide enrichment in the influenza a virus genome
87 as a live attenuated vaccine development strategy. *PLOS Pathogens*, *19*(5), e1011357.
88 https://doi.org/10.1371/journal.ppat.1011357

89 Simmonds, P. (2012). SSE: A nucleotide and amino acid sequence analysis platform. *BMC*
90 *Research Notes*, *5*(1), 1–10. https://doi.org/10.1186/1756-0500-5-50