Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The application of DNA nucleotide footprint plotting in coronavirus

Samuel Krem

Buckingham Browne & Nichols School, 37 Thayer Road, Belmont, MA, 02478, USA

ARTICLE INFO

Keywords:
Coronavirus
Visualization
Genome
DNA
Footprint

ABSTRACT

Direct visualization of the key features of coronavirus genomes can lead to a better understanding of this virus as well as a method to distinguish this type of virus from others. The DNA nucleotide footprint plotter is a tool that makes it possible for straightforward visualization of the characteristics of viral genomes. It can also distinguish different gene types and gene structures. The current project provides a novel tool for biological studies that can contribute to breakthroughs in coronavirus diagnosis, treatment, and prevention.

1. Introduction

Coronaviruses are new viruses. Although there have been many timely discoveries for this virus, there are still many things unknown [1–3]. I designed a novel tool named the DNA nucleotide footprint plotter. It can distinguish different types of viruses such as Coronavirus (SARS-CoV-2), SARS-Cov virus, HIV virus, and influenza virus, as indicated in Figs. 1–4. It was also used to compare the complete sequence of COVID-19 sub-species and various corona strains such as OC43, HKU1, MERS-CoV (Middle East Respiratory Syndrome), SARS-CoV (Fig. 5 and Figs. 12–16). The characteristics of hundreds of virus genomes and gene segments have been displayed by DNA footprint plotter (data not all shown). This application has the capability to display the exact changes that occur in nucleotide position of nonstructural and structural regions of the novel different complete genome of COVID-19 corona strains (Fig. 6). Even if the method is designed to observe structural changes at the genomic level that single nucleotide polymorphism calling tools cannot identify, it is possible to show genotype changes and amino acid changes at a single nucleotide resolution. In addition, DNA footprint plotter is a useful tool to determine the evidence of recombination. A recombination between Corona virus MT121215 and Influenzavirus HC668016 was simulated by making a FASTA file containing the beginning of the flu genome and the end of the corona virus genome. As demonstrated in Fig. 7, the simulated recombination created a chimera of partial flu virus and partial corona virus.

2. Methods

The genome sequences of coronavirus (SARS-CoV-2), SARS-CoV virus, OC43, HKU1, MERS-CoV (Middle East Respiratory Syndrome), HIV virus, and influenza virus were downloaded from NCBI virus data hub [4]: https://wwwncbi.nlm.nih.gov/labs/virus/vssi/#/.

The figures have an X axis and a Y axis.

There are three ways to draw DNA footprint plots. They are ACat, AGac, and ATag. ACat is used as an example. If the nucleotide is A, then x value increments by 1 (X+1). If the nucleotide is T, then x value decrements by 1 (X-1). If the nucleotide is C, then y value plusses 1 (Y+1). If the nucleotide is G, then y value minuses 1 (Y-1). AGac and ATag use the same logic. The three types of footprint of Coronavirus2_China_Homo_sapiens_NC_045512 are displayed in Figs. 8–10 as examples. The framework of the script is shown in Fig. 11. The program was written in JAVA. The three types of footprint plots look different. I only use ACat plot in the virus genome structure comparisons in the current paper.

The method aims to demonstrate structural variations at higher levels that single nucleotide mutation callers cannot identify. However, the tool can precisely demonstrate single nucleotide information (Fig. 6). The method itself has no limit to plot the number of nucleotide bases. But due to the size of computer screens and paper, too many bases will cause the lines to overlay each other. Thus the sizes of virus genome or segments of eukaryote genome are recommended for one particular plot. One plot can display a complete virus genome, and users can combine multiple plots together to visualize more complicated genomes such as the genomes of eukaryotes. Or, if only a partial virus genome or a particular gene is of interest, the corresponding regions can be displayed by DNA foot plotter.
Fig. 1. ACat plot of Coronavirus_Rhinolophus_China_CKY417151 (30307bp).

Fig. 2. ACat plot of SARS_coronavirus_KF514411 (29687bp).

Fig. 3. ACat plot of HIV_DM461230 (15524bp).
3. Results

3.1. Recognition of various types of viruses

Coronavirus demonstrates a unique footprint (Fig. 1), and so do other types of viruses (Figs. 2–5). The DNA nucleotide footprints are different in different types of viruses, and it makes it convenient to distinguish one type of virus from another.

3.2. Comparisons within the same type of viruses

To compare two coronavirus subtypes, I overlapped their footprint plots and let one cancel another if they are the same. In this way, we can see the differences between them. The coronavirus in Figs. 12–15 are from mainland China (MT121215), Taiwan (MT066176), Japan (LC529905), and Italy (MT066156). They were compared with one type of coronavirus from the USA (MT188340). Another intuitive way to compare virus genomes is to color them in different colors. In Fig. 16, Corona_virus_China_MT121215 (black) is compared with SARS-Urbani_AY278741 (green).

4. Discussion

DNA footprint can quickly distinguish the infection type of patients given the sequencing data. It can demonstrate the characteristics of different types of viruses and increase the confidence of diagnosis such as flu, coronavirus, or SARS-cov. In addition to specific visualization of coronavirus genome key features, it can be applied to recognize various virus genomes in general and to compare subtypes of the same virus species. The better understanding of virus genomes will lead to a more accurate diagnosis, which lays the foundation of appropriate therapeutic options. It can also help the prevention of the disease by guiding vaccine design. One example is that the tool can display which regions are similar among various virus subtypes and which regions are different. This will facilitate the design of vaccines targeting similar regions among different virus sub-species. In this way, one vaccine can fight different types of viruses.

DNA footprint plotter is useful for character identification at the genome level. It can identify key features in the genome that single nucleotide mutation callers will miss. It can also be used as an add-on tool of phylogenic trees. Phylogenic trees are very useful to group and compare viruses. But the detailed differences are not shown. If we can also apply DNA footprint plot, people will have a better idea regarding which regions are similar and which regions are different, as well as the patterns of differences.

In addition to viruses, DNA footprint plots can be applied to visualize single genes or segments of chromosomes in higher organisms to facilitate the understanding of gene structure or chromosome structure. It can also help to identify disease-causing structural variations such as repeats in diseases like Huntington’s disease.

If we digitalize the key features of the footprint shapes within genomes and genes, we can use machine learning, such as neural networks, to recognize various viruses and genes [5].

Fig. 4. ACat plot of Influenza_virus_HC668016 (6130bp).

Fig. 5. ACat plots of footprint of various viruses.

Fig. 6. ACat plot of Corona_virus_China_MT121215 with annotation (29945bp).
Fig. 7. ACat plot of simulated recombination by combining the beginning of Influenza virus with the ending of Corona virus basing on sequences in FASTA files.

Fig. 8. ATag of Coronavirus2_China_Homo_sapiens_NC_045512 (29903bp).

Fig. 9. AGac of Coronavirus2_China_Homo_sapien_NC_045512 (29903bp).
Fig. 10. ACat plot of Coronavirus2_China_Homo_sapiens NC_045512 (29903bp).

Fig. 11. The code block.
Fig. 12. ACat plot of USA_MT188340 (29845bp) versus China_MT121215 (29945bp).

Fig. 13. ACat plot of USA_MT188340 (29845bp) versus Taiwan_MT066176 (29870bp).
Fig. 14. ACat plot of USA_MT188340 (29845bp) versus Japan_LC529905 (29903bp).

Fig. 15. ACat plot of USA_MT188340 (29845bp) versus Italy_MT066156 (29867bp).
5. Conclusion

DNA footprint plotter is a novel tool to display the characteristics of virus genome, genes, and sections of eukaryote genome. It is a quick and convenient method to visualize the features of coronavirus DNA to distinguish them from other viruses, and to understand the differences of coronavirus sub species.

Ethical statement

I do not have any financial or personal relationships with other people or organizations that could inappropriately influence (bias) my work. I do not have potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, grants or other funding.

Declaration of competing author

The author(s) have no conflicts of interest

Acknowledgement

I want to thank my school Buckingham Browne & Nichols for teaching me essential knowledge in biology and Java programming skills to make this paper possible. I also want to thank my family to excuse me from house chores and to give me time to work on the project.

References

[1] Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–9. https://doi.org/10.1038/s41586-020-2008-3.
[2] Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 2020;395 (10224):565–74. https://doi.org/10.1016/S0140-6736(20)30251-8. https://pubmed.ncbi.nlm.nih.gov/32007145.
[3] Weston S, Frieman MB. COVID-19: knowns, unknowns, and questions. mSphere 2020;5(2). https://msphere.asm.org/content/5/2/e00203-20.
[4] Hatcher EL, Zhdanov SA, Bao Y, et al. Virus Variation Resource-improved response to emergent viral outbreaks. Nucleic Acids Res 2017;45:482–90. https://www.ncbi.nlm.nih.gov/pubmed/27899674.
[5] Webb SM. Deep learning for biology. Nature 2018;554:555–7. https://media.nature.com/original/magazine-assets/d41586-018-02174-z/d41586-018-02174-z.pdf.

Fig. 16. ACat plot comparing Corona_virus_China_MT121215 (black 29945bp) and SARS-Urbani_AY278741 (green 29727bp). Coloring in footprint plots is another intuitive way to distinguish different virus genomes.