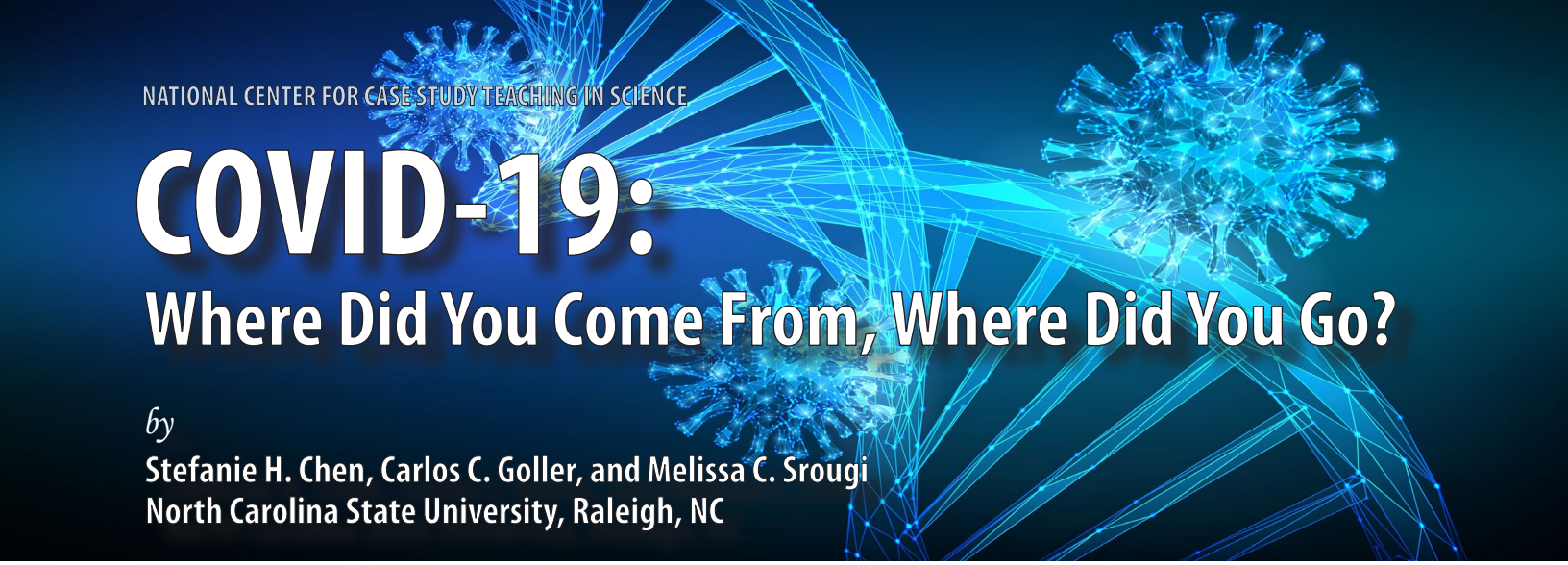

NATIONAL CENTER FOR CASE STUDY TEACHING IN SCIENCE

COVID-19:
Where Did You Come From, Where Did You Go?

by
Stefanie H. Chen, Carlos C. Goller, and Melissa C. Srougi
North Carolina State University, Raleigh, NC

Introduction

As one of the leading virologists in the Epidemic Intelligence Service (EIS) branch of the Centers for Disease Control (CDC), you have been tapped to lead a team of scientists in tracking the spread of the novel coronavirus (COVID-19) pandemic in the United States. Almost immediately, two new cases of COVID-19 appear on your desk from individuals residing in metropolitan Washington, D.C. The coronavirus disease spreads easily, such as by contact with an infected person or object, and is then transfered by touching of eyes, nose or mouth. It is a race against the clock to track the origins of these infections to prevent the further spread of the disease in one of the busiest cities in the United States. After interviewing the patients, you know both individuals recently flew abroad and are unsure where they might have contracted COVID-19. Based on your expertise, you know that the SARS-CoV-2 (the virus that causes COVID-19) genome will begin to accumulate changes (mutations) as it passes from person to person. You reason that if you can identify the genomic sequences of each viral strain, perhaps you can track the source of the infection. With this in mind, you mobilize your team to perform a series of sequence comparisons using bioinformatics tools to compare the viral sequences from the two new COVID-19 cases to existing sequences of SARS-CoV-2. Before you begin, you must first identify the existing sequences of SARS-CoV-2 from patient zero (or the first patient known to have contracted COVID-19).

Will you and your team be able to find the source of the patients' infections before they spread the disease any further?

The race is on!

Part I – Sequence Divergence from Patient Zero

First, your team will use the National Center for Biotechnology Information (NCBI) database to locate the consensus sequence of the SARS-CoV-2 genome from patient zero in Wuhan, China.

Sequence Search and Alignment

1. Navigate to <https://www.ncbi.nlm.nih.gov/> and choose "Genome" from the dropdown menu next to the search bar.
2. Search for "Severe acute respiratory syndrome coronavirus 2" (the full name of SARS-CoV-2). Identify the RefSeq number for the sequence:

   _____.
3. Next, go to <https://blast.ncbi.nlm.nih.gov/Blast.cgi> and choose Nucleotide BLAST. Copy in the RefSeq accession number you found in Step 2.
4. At the bottom under the blue BLAST button, click on the plus sign next to "Algorithm parameters." Change "Max target sequences" to 500. Then, click on BLAST to start the alignment process. Since it is aligning an entire genome, it may take a few minutes.

BLAST (basic local alignment search tool) will show similar sequences that have been deposited to the NCBI sequence database, starting with the sequence with the highest similarity. Take a look at the results, noting the source of each sequence (in the name) and the percent identity (how many nucleotides within the sequence are the same). Additional information on each of the aligned sequences is available by selecting the accession number on the right-hand side.

Questions

1. What is the source (geographical location) of the top five most similar sequences?

2. Examine the values provided (Max score, Query cover, E-value, Percent Identity). How similar are the sequences? How do you know?

3. Looking through all of the results returned, what trends do you notice?

4. Do any non-human viruses appear on the list? (*Hint:* you may have to redo the search and expand the Max target sequences to 5000 to see any non-human hits.) How is this related to the epidemiology of the virus?

Part II – Phylogenetic Trees

A more visual way to observe mutations across time is to use a phylogenetic tree, which groups related sequences and shows branching points as sequences diverge. Your EIS team will next explore phylogenetic trees constructed from known SARS-CoV-2 sequences isolated from COVID-19 patients across the globe. These data will provide valuable insight into SARS-CoV-2 viral origins and how strains mutate as they spread in the world.

Build a Tree

1. Navigate to Nextstrain, a web-based platform that allow users to learn about pathogen evolution in real-time: <https://nextstrain.org/ncov>.
2. Hover your cursor over the lines and dots of the phylogenetic tree to observe the origin of and mutations present in the various sequences.
3. Change the "tree options" on the left side of the screen to "clock."
4. Change the "branch length" on the left side of the screen to "divergence."
5. Change the "map option" on the left side of the screen to "country."
6. Under the transmissions map of the world, select "play."

Questions

For help in interpreting the phylogenetic tree, you may want to look at "How to read a tree" help page: <https://nextstrain.org/help/general/how-to-read-a-tree>.

1. What do the colors represent?

2. What do the dots and lines represent?

3. What do you notice about the general mutation frequency? Typically, how many mutations have occurred between isolates?

4. Are there similarities or differences in viral strains in a country or between countries? How is this related to the epidemiology of the virus?

Part III – Determining the Origin of Patient SARS-CoV-2 Strains

After diligently examining the phylogenetic data, your fellow EIS team member surprises you with the actual DNA sequences of specific regions of SARS-CoV-2. These sequences were obtained from the two recently arrived infected patients in Washington, D.C. Using the BLAST tool and the provided sequences, you finally have the ability to identify the source of origin of the SARS-CoV-2 that is infecting each patient. Thus, these data can provide critical clues on where the patients were infected and ways to prevent its further spread.

Patient #1 SARS-CoV-2 sequence (orf1ab):

ATTTAAACTGTCTTATGGTATTGCTACTGTACGTGAAGTGCTGTCTGACAGAGAATTACATCTTTCATGGGAAGT
TGGTAAACCTAGACCACCACTTAACCGAAATTATGTCTTTACTGGTTATCGTGTAACTAAAAACAGTAAAGTACA
AATAGGAGAGTACACCTTTGAAAAAGGTGACTATGGTGATGCTGTTGTTTACCGAGGTACAACAACTTACAAAT
TAAATGTTGGTGATTATTTTGTGCTGACATCACATACAGTAATGCCATTAAGTGCACCTACACTAGTGCCACAAG
AGCACTATGTTAGAATTACTGGCTTATACCCAACACTCAATATCTCAGATGAGTTTTCTAGCAATGTTGCAAATT
ATCAAAAGGTTGGTATGCAAAAGTATTCTACACTCCAGGGACCACCTGGTACTGGTAAGAGTCATTTTGCTATT
GGCCTAGCTCTCTACTACCCTTCTGCTCGCATAGTGTATACAGCTTGCTCTCATGCCGCTGTTGATGCACTATGTG
AGAAGGCATTAAAATATTTGCCTATAGATAAATGTAGTAGAATTATACCTGCACGTGCTCGTGTAGAGTGTTTT
GATAAATTCAAAGTGAATTCAACATTAGAACAGTATGTCTTTTGTACTGTAAATGCATTGCCTGAGACGACAGC
AGATATAGTTGTCTTTGATGAAATTTCAATGGCCACAAATTATGATTTGAGTGTTGTCAATGCCAGATTACGTGC
TAAGCACTATGTGTACATTGGCGACCCTGCTCAATTACCTGCACCACGCACATTGCTAACTAAGGGCACACTAG
AACCAGAATATTTCAATTCAGTGTGTAGACTTATGAAAACTATAGGTCCAGACATGTTCCTCGGAACTTGTCGGC
GTTGTCCTGCTGAAATTGTTGACACTGTGAGTGCTTTGGTTTATGATAATAAGCTTAAAGCACATAAAGACAAAT
CAGCTCAATGCTTTAAAATGTTTTATAAGGGTGTTATCACGCATGATGTTTCATCTGCAATTAACAGGCCACAAA
TAGGCGTGGTAAGAGAATTCCTTACACGTAACCTTGCTTGGAGAAAAGCTGTCTTTATTTCACCTTA

Patient #2 SARS-CoV-2 sequence (N):

AGCAGTCCAGATGATCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAAATGA
AAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCTGGACTTCCCTATGGTGCTAAC
AAAGACGGCATCATATGGGTTGCAACTGAGGGAGCCTTGAATACACCAAAAGATCACATTGGCACCCGCAATC
CTGCTAACAATGCTGCAATCGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGG
AGCAGAGGCGGCAGTCAAGCCTCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAGAAATTCAACTCCAGG
CAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCAATGGCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTG
ACAGATTGAACCAGCTTGAGAGCAAAATGTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAA
ATCTGCTGCTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTT
TCGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTTTGGGGACCAGGAACTAATCAGACAAGGAACTGATTA
CAAACATTGGCCGCAAATTGCACAATTTGCCCCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGG
AAGTCACACCTTCGGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTTCAAA
GATCAAGTCATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCAACAGAGCCTAAAAAGGACAA
AAAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCAGAGACAGAAGAAACAGCAAACTGTGACTCTTCTTCCT
GCTGCAGATTTGGATGATTTCTCCAAACAATTGCAACAATCCATGAGCAGTTCTGACTCAACTCAGGCCTAAACT
CATG

Back to BLAST

1. Navigate to <https://blast.ncbi.nlm.nih.gov/Blast.cgi> and choose "Search Betacoronavirus Database" at the top right. (If the direct link for betacoronavirus searches is no longer available, choose Nucleotide BLAST on the left and select Betacoronavirus from the Database under "Choose Search Set.")
2. Copy and paste in the sequence from the first patient. Keep all of the default settings, then choose BLAST at the bottom of the page.
3. In a new window, repeat the previous two steps for the sequence from the second patient.

Questions

1. Observe the top five hits (i.e., the most closely related strains) for each. Where did each patient most likely obtain the virus?

2. Given that these patients both presented in Washington, D.C. after flying in from unidentified international locations, what recommendations would you make for limiting further spread of the virus? *Hint:* Use outside sources to help guide your recommendations and be sure to list the references you used.

3. Based on your recommendations, reflect on the possible outcomes, both good and bad, that your advice will have on the human population. You will want to think beyond just physical health benefits.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–10.

Centers for Disease Control and Prevention. 2020. Coronavirus (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/index.html> [accessed 4/2/2020].

Hadfield, J., C. Megill, S.M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, and R.A. Neher. 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34(23): 4121–3. <https://doi.org/10.1093/bioinformatics/bty407>.

World Health Organization. Coronavirus disease (COVID-19) Pandemic. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> [accessed 4/2/2020].