



September 8, 2021

The Honorable Roger Marshall  
United States Senate  
Washington, DC 20510

Dear Senator Marshall:

Thank you for your June 28, 2021, letter regarding the National Library of Medicine's (NLM) Sequence Read Archive (SRA).

Early in the pandemic, the National Institutes of Health (NIH) and other federal agencies moved quickly to make COVID-19 open-access data and computational resources freely available to researchers. NIH's National Library of Medicine (NLM) has a broad portfolio of open-access databases, including the Sequence Read Archive (SRA), the world's largest publicly available repository of high-throughput sequencing data. In the past year, SRA received approximately 2.4 million submissions of sequence data.

SRA is managed by NLM's National Center for Biotechnology Information (NCBI), which is the U.S. participating member of the International Nucleotide Sequence Database Collaboration (INSDC) since 1987. NCBI follows the INSDC [policies](#)<sup>1</sup> and [guidelines for data submission and change requests](#),<sup>2</sup> and collaborates with participating organizations in updating policies and guidelines as described in this 2018 [article](#).<sup>3</sup>

The guidelines describe the criteria for which submitting researchers can request a change in data status (for example, if the data have been corrupted) and actions taken if the criteria are met. Submitting institutions must contact NLM/NCBI to request that data be withdrawn. NLM/NCBI staff review the request against the INSDC guidelines.

In March 2020, the SARS-CoV-2 sequences you mention in your letter were submitted by a researcher at Wuhan University for public release status via SRA. The researcher from Wuhan University published relevant information about these sequences [by preprint in March 2020](#)<sup>4</sup> and in an international scientific [journal in June 2020](#).<sup>5</sup> In June 2020, NCBI received a request to withdraw the sequences from the same researcher. The reason given by the researcher was they were depositing updated data in a different database and they wanted to prevent version confusion. In response, NCBI withdrew the sequences from the SRA but they remained

---

<sup>1</sup> <https://www.insdc.org/policy.html>

<sup>2</sup> <https://www.insdc.org/documents/insdc-status-document>

<sup>3</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5753279/>

<sup>4</sup> <https://www.medrxiv.org/content/10.1101/2020.03.04.20029538v1>

<sup>5</sup> <https://onlinelibrary.wiley.com/doi/full/10.1002/sml.202002169>

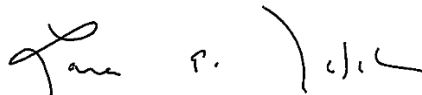
available to the world's scientists and researchers through the publication and on other platforms. No conditions were made on that request and no conditions were granted.

NIH subsequently conducted an analysis of withdrawal requests from January 2020 through June 2021. Withdrawal requests were generally submitted for scientific reasons. Examples include problems with the quality of the data, data were not correctly attributed to their owner, or data were not correctly identified. In that time, six institutions requested withdrawal of SARS-CoV-2 submission packages through NLM/NCBI services. This included the one requested by the researcher at Wuhan University and the rest from researchers at institutions from other countries, predominantly the U.S. In addition, NCBI received withdrawal requests from five institutions through INSDC collaborators for sequence data that were in the collaborating repositories and were replicated within the SRA.

While NIH considers the policies and guidelines of the INSDC sound, NCBI has initiated an independent review of SRA processes and standard operating procedures to determine whether the appropriate steps were taken to assess this withdrawal request. Withdrawal makes the data undiscoverable but does not erase it. Per the INSDC guidelines, NCBI retains withdrawn data for the scientific record and for disaster recovery. Pending outcome of the review, NCBI will work with INSDC to assign the data to the appropriate status.

Thank you for your interest in NIH's National Library of Medicine's broad portfolio of open-access databases. I hope this information is helpful to you. I have also sent an identical letter to the co-signers of your letter.

Sincerely,

A handwritten signature in black ink, appearing to read "Lawrence A. Tabak". The signature is fluid and cursive, with the first name being the most prominent.

Lawrence A. Tabak, D.D.S., Ph.D.  
Principal Deputy Director