

# Mapping innovation trajectories on SARS-CoV-2 and its variants

**To the Editor** — SARS-CoV-2, the etiological agent of the COVID-19 pandemic, was discovered in late 2019 and its sequence made public<sup>1</sup> on 10 January 2020. Recently, a number of viral variants have been identified, such as B.1.1.7 in the United Kingdom, B.1.351 in South Africa and P.1 in Brazil, with the potential for increased transmissibility and pathogenicity, potentially exacerbating the crisis. Although papers and preprints concerning these variants are being published rapidly, much

information about sequences of the virus variants and their associated scientific knowledge is published in the patent literature rather than the academic literature or other online sources. The Lens, an open platform run by Cambia, a global non-profit social enterprise (<https://cambia.org/>), provides a freely available, comprehensive resource that links different sources of information. With over 127 million global patent records from over 100 countries, over 225 million non-patent research publications

and over 370 million sequences from patent records, the Lens can provide information on patent rights related to SARS-CoV-2 and its variants, as well as the underlying scientific understanding and research, and the people and institutions behind the work.

When derived from publicly funded or academic research, DNA, RNA and protein sequences are often readily accessible in public repositories, such as GenBank. However, millions of naturally occurring and artificial biological sequences have

The screenshot displays the PatSeq Finder interface. On the left, there are filter controls for 'Filters' and 'Sort', including options for 'Apply these filters/search terms', 'Minimum coverage' (0-100), 'Minimum similarity' (0-100), 'Blast score' (40-1920), and 'Maximum E-value' (1). Below these are date filters ('BEFORE' and 'AFTER') and 'Sequence length' and 'Jurisdiction' dropdowns. A 'Download results as...' and 'Export results...' section is also present. The main search area shows the query 'SARS-COV-2:RNA-dependent RNA polymerase' and a list of 12 filtered hits. One hit is selected, showing details for 'Patent Application US 2011/0097730 A1'. This includes sequence listing information (AA, 936 length, Artificial organism), an alignment with an E-value of 0 and 96.2% similarity, and a list of claims. Claim 1 describes a recombinant SARS-CoV-2 nsp12 with nine amino acids deleted from the N-terminus. Other claims describe the use of the recombinant nsp12 and the expression vector. The interface also shows 'Your Search Params' and 'How to use' sections.

**Fig. 1 | PatSeq Finder enables analysis of sequence similarity search results, patent claims and sequence alignments comparison, and results verification by linking to original data sources.** To capture the emerging patents related to SARS-CoV-2 in 2020, we used the following broad search in PatSeq Finder (<https://www.lens.org/lens/bio/patseqfinder>): “SARS-CoV-2” OR “2019 new coronavirus” OR “2019 nCoV” OR “COVID 19”. We saved the result as the dynamic collection SARS-CoV-2: Emerging Patents\_2020 (<https://link.lens.org/4bF57em2sVd>). Filtering by sequence type allows sequence disclosure analysis within each patent record and the discovery of similar sequences either in the PatSeq database (using PatSeq Finder) or in GenBank. Flagging patents that cited scholarly works also allows tracking of such works. The collection can be further explored in a dashboard (<https://link.lens.org/6UjwhB19P5c>) in terms of who or which entity is pursuing them and which institutions or companies may have capabilities to make the products and practice the changes we need.

been disclosed only in patents, and these can be fragmented, obscure and often inaccessible. Better public disclosure of such biological sequences, as well as any associated knowledge, is critical not only for enabling future innovations, but also for marking the boundaries of what has already been claimed. Patented sequences may become associated with monopoly rights after examination, potentially restricting the freedom to operate of enterprises or researchers either through onerous licensing or the threat of litigation. Filing patents before public sequence disclosure is typical, so those who publish sequences early could become dominant applicants. For example, there is already a substantial corpus of relevant patent-disclosed knowledge (<https://www.lens.org/lens/report/view/Human-Coronaviruses-Patent-and-Research-Landscape/1083/page/1089>)—including viral and host sequences—from the previous severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS) coronavirus outbreaks, not to mention critical platform technologies associated with vaccines, therapeutics and diagnostics. For SARS-CoV-2, variants found to date may differ only slightly from canonical published sequences, and so open analysis of patent applications disclosing variant sequences or claiming rights to detect or specifically treat such variants is urgently needed. But there is no single comprehensive and harmonized public patent sequence dataset or facility that would make such sequences and associated technology accessible for researchers interested in SARS-CoV-2.

The Lens has been working to remedy this shortcoming. In collaboration with the European Patent Office (EPO), the US Patent and Trademark Office (USPTO) and other patent offices, we have spent the past decade extracting sequences from the full text of patents, from their claims, and from associated files and disclosures, creating a publicly available resource and toolset to explore patent-derived sequences<sup>2</sup>. While the resource is open to the public research community, we broadly license these data to the private sector as well to defray the costs of maintaining a public resource.

The Lens Project (<https://about.lens.org/>) provides context for inventions described in patents by linking the inventions to the scientific research cited in the patents and, in collaboration with Microsoft Academic (<https://www.microsoft.com/en-us/research/project/academic/articles/sharpening-insights-into-the-innovation-landscape-with-a-new-approach-to-patents/>), to scholarly works by use of machine learning. We have populated

resulting patent data in an open facility, called Lens Labs (<https://www.lens.org/lens/labs>), and in collaboration with MIT Knowledge Futures Group (<https://www.knowledgefutures.org/>), improvements in data quality are being developed with several institutions (<https://iii.pubpub.org/>).

The newest release of the Lens Patent MetaRecord architecture and its application programming interface (API; <https://www.lens.org/lens/user/subscriptions#patents>) also provides the legal status and events of patents and applications in dozens of countries. This means that jurisdictions in which patents have not been filed or in which patents have been abandoned, have lapsed, or have been challenged, rejected, acquired or sold can be readily examined, for example, to inform strategies that involve different markets or manufacturing jurisdictions.

The Lens Report Builder (<https://www.lens.org/lens/reports>), currently in beta release, foreshadows our approach to bridging the gulf between science and social outcomes with innovation cartography<sup>3</sup>. To illustrate the utility of our platform, we present a dynamic collection, SARS-CoV-2 genetic variants (<https://link.lens.org/IdHkLFwWMh>), highlighting emerging scholarly works, those cited in patents, and those citing patents. Dynamic collections are automatically updated when new works matching the linked saved query are added to the search index, and they enable live dashboards (<https://link.lens.org/22YcK2hJlgf>). The platform also has the option to provide customized alert notifications for newly added works. Published works can manually be mined and split by specific geographic regions, countries or selected research disciplines. The resulting subcollections are publicly available to the community on the Lens Labs portal.

An examination of patent sequence disclosures from viruses similar to SARS-CoV-2 that are hotspots for viral recombination and mutation—spike protein (<https://www.lens.org/lens/bio/patseqfinder#results/275d8acd-84af-49df-9d2d-5ac98ae68560>), ORF1ab (<https://www.lens.org/lens/bio/patseqfinder#results/bd2b06a8-2821-4010-bad3-d4fd563eafcd>) and RdRP (<https://www.lens.org/lens/bio/patseqfinder#results/a3ca66be-61a6-4870-a360-b8a64f51bcee>)—reveals the presence of a few granted patents referencing these sequences in their claims and several pending patent applications related to other coronavirus sequences. The search results also allow the discovery of similar sequences that have been referenced in the claims or simply disclosed in the patent specification and to what extent they support the invention and its scope. Through the PatSeq Finder application, Lens users can

also explore confidentially and securely their query sequences and compare patent claims and sequence alignments from resulting patents side by side, with the option to embed the findings in online reports (Fig. 1).

The ongoing COVID-19 crisis has highlighted the difficulties of developing and implementing evidence-driven public policy and a fair and rapid access to outcomes, within the context of a competitive innovation ecosystem, a glut of information of varying quality, and rising vaccine nationalism. To deliver outcomes, diverse capabilities running the gamut from science to intellectual property to business, law, policy, regulation, manufacturing and beyond need to be coordinated. Patents and their metadata can provide insights into the potential partners and their capabilities that must be found and engaged. But there is concern that patents, if insufficiently understood and/or inappropriately used or licensed, could create a crisis within a crisis, impair coronavirus research, accelerate private capture of public work products, and slow access to medical products and outcomes across the globe. Already the differential access to first-generation vaccines is having a destabilizing effect politically and economically<sup>4</sup>.

Our Lens platform, with its open, comprehensive and aggregated corpus of patent and scholarly data enriched with sequences, will not only help scientists gain rapid access to evolving works on SARS-CoV-2 and its variants but also help the wider research and policy community keep one step ahead of proprietary information that threatens to impair our ability to create and access interventions against the virus that will bring the pandemic under control across the globe and set the stage for a more prepared forward-looking global health system. □

Osmat Azzam Jefferson , Deniz Koellhofer , Ben Warren, Thomas Ehrich , Simon Lang , Kenny Williams , Aaron Ballagh , Bernard Schellberg, Roshan Sharma  and Richard A. Jefferson 

*Cambia, Canberra, Australian Capital Territory, Australia.*

✉e-mail: [osmat@cambia.org](mailto:osmat@cambia.org)

Published online: 22 February 2021  
<https://doi.org/10.1038/s41587-021-00849-z>

#### References

1. Novel 2019 coronavirus genome. <https://virological.org/t/novel-2019-coronavirus-genome/319> (2020).
2. Jefferson, O. A., Köllhofer, D., Ajikuttira, P. & Jefferson, R. A. *World Patent Inf.* **43**, 12–24 (2015).
3. Jefferson, R. *Nature* **548**, S8 (2017).

4. Mueller, B. & Stevis-Gridneff, M. E.U. and U.K. fighting over scarce vaccines. *The New York Times* <https://www.nytimes.com/2021/01/27/world/europe/eu-uk-covid-vaccine.html> (27 January 2021).

#### Acknowledgements

This work was funded by Bill & Melinda Gates Foundation grant 015897, Rockefeller Foundation grant 2020 FOD

006 and Alfred P. Sloan Foundation grant G-2019-12326, “Innovation Information Initiative.” We are grateful to Amazon Web Services for a grant from their COVID Emergency Response team for support of cloud-based computing and platform expenses. We thank Adrian Gibbs, Gilbert Faure and Marie-Christine Béné for their edits, review and constructive comments on the earlier version of the SARS-CoV-2 report prototype. The extended

online version can be accessed at <https://link.lens.org/tk10f5UfAbb>.

#### Competing interests

All authors except T.E. are employed by Cambia, a non-profit with a community-funded infrastructure that receives public and private funds. The Lens is a project of Cambia.

# Democratizing biotechnology requires more than availability

**To the Editor** — In their Correspondence “The accelerating pace of biotech democratization,” Jackson et al.<sup>1</sup> find evidence for an accelerated pace at which novel biotechnologies become widely available. They state that biotechnologies that are initially available to only a handful of well-resourced specialists currently take about 4.5 years to spread to people with relatively low levels of technical skills and financial resources. By the end of the decade, the authors predict, this transition is expected to take less than 3.5 years. The authors conclude that the ‘democratization’ of biotechnology is accelerating.

Jackson et al. provide valuable insights into the timeframes for the spread of biotechnologies and raise interesting questions about the technical and societal factors that can explain these changes. However, the conclusion that this increased accessibility amounts to democratization of biotechnology is premature at best. The democratization of science and technology has been extensively studied by scholars in social sciences and humanities over the last thirty years<sup>2–4</sup>. Without exception, these studies conceptualize democratization not as the need to spread technologies faster and further<sup>5</sup>, but to involve society in the development and governance of technologies, from identifying issues that merit solving in the first place to deliberating the particular technologies used to solve those problems. Equating



accessibility to democracy instead assumes that the technologies in question are by definition desirable and beneficial, while this is exactly what is in question in societal controversies regarding new technologies, including various biotechnologies and their applications<sup>6</sup>.

In addition, by assuming the unconditional desirability of the wide availability of biotechnologies, democratic publics are portrayed as passive recipients of scientific work. Instead, democratic theory invites us to take publics seriously:

to actively listen to publics and to include their interests and concerns in the design and application of new biotechnologies<sup>7,8</sup>, so that science and society to become mutually responsive to each other<sup>9</sup>. The timeframe from first usage to widespread availability certainly matters, but the praiseworthy objective of democratizing biotechnology invites us to move beyond that. □

Koen Beumer  

*Copernicus Institute of Sustainable Development–Innovation Studies, Utrecht University, Utrecht, the Netherlands.*

*✉e-mail: k.beumer@uu.nl*

Published online: 26 March 2021  
<https://doi.org/10.1038/s41587-021-00890-y>

#### References

- Jackson, S. S. et al. *Nat. Biotechnol.* **37**, 1403–1408 (2019).
- Ezrahi, Y. *The Descent of Icarus: Science and the Transformation of Contemporary Democracy* (Harvard Univ. Press, 1990).
- Bijker, W. E. *Democratizing Technological Culture* (Schrijen-Lippertz, 1995).
- Nahuis, R. & Van Lente, H. *Sci. Technol. Human Values* **33**, 559–581 (2008).
- Jasanoff, S. *Nature* **450**, 33 (2007).
- Jasanoff, S. *Designs on Nature: Science and Democracy in Europe and the United States* (Princeton Univ. Press, 2005).
- Beumer, K. *Nat. Nanotechnol.* **14**, 510–512 (2019).
- Veak, T. J. *Democratizing Technology: Andrew Feenberg’s Critical Theory of Technology* (SUNY Press, 2006).
- von Schomberg, R. in *Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society* (eds Owen, R., Bessant, J. & Heintz, M.) 51–74 (Wiley, 2013).

#### Competing interests

The author declares no competing interests.

# Reply to: Democratizing biotechnology requires more than availability

**S. S. Jackson et al. reply** — Our Correspondence was written with a limited definition of “democratize” in

mind: according to the [Merriam-Webster](#) definition, “to make (something) available to all people.” That is, we were referring

simply to the widespread availability of the biotechnologies studied to those with limited resources and training, and