idw - Informationsdienst Wissenschaft
Nachrichten, Termine, Experten

Press release

Universität Bielefeld

Jörg Heeren

03/30/2022

http://idw-online.de/en/news791092

Research projects, Research results
Biology, Information technology, Medicine
transregional, national

Neue Methode für bessere Analyse von Virengenomen

Forschende stellen Ansatz zur Rekonstruktion von Gensequenzen vor Oft unterscheiden sich Varianten von SARS-CoV-2 nur in winzigen Details. Deshalb ist eine möglichst exakte Darstellung des Genoms wichtig, um Virenstämme miteinander vergleichen zu können. Bei herkömmlichen Methoden treten häufig Ablesefehler auf, die das Ergebnis verfälschen können. Dieses Problem wollen der Bioinformatiker Professor Dr. Alexander Schönhuth von der Universität Bielefeld und sein Team mit ihrem neuen Verfahren lösen. Die Methode Strainline ermöglicht es, Sequenzen von Virengenomen auf neue Weise zu rekonstruieren. Unter anderem werden Ablesefehler frühzeitig erkannt und korrigiert.

Als weitere Methode mit ähnlichem Ansatz hilft Phasebook darüber hinaus dabei, auch menschliche Chromosomensätze besser analysieren zu können. Studien zu den beiden Verfahren haben die Forschenden in der Fachzeitschrift Genome Biology veröffentlicht.

Wie lassen sich neue Virenstämme von SARS-CoV-2 frühzeitig erkennen? Eine Möglichkeit ist es, das Abwasser einer Stadt zu untersuchen, in dem Coronaviren ihre Spuren hinterlassen. Kurz gesagt wird dazu die Viren-RNA mit einem Gerät in kleine Stücke zerschnitten und analysiert. Aus diesen Stücken werden anschließend die Genome der Viren rekonstruiert – bislang allerdings mit einer recht hohen Fehlerrate.

Solche genetischen Analysen werden in der Regel mit sogenannten Nanopore-Sequenziergeräten durchgeführt. Der Vorteil: Sie sind günstig und handlich – und können lange Gensequenzen ausgeben. „Eine Sequenz kann bis zu 10.000 Basen umfassen“, sagt Professor Dr. Alexander Schönhuth von der Technischen Fakultät und dem Institut für Bioinformatik-Infrastruktur (BIBI) der Universität Bielefeld. Das Genom von SARS-CoV-2 besteht aus etwa 30.000 Basen – entsprechend entstehen durch die Sequenzierung nur wenige Teile, die aber viele Informationen umfassen.

Fehlerhafte Gensequenzen müssen korrigiert werden

„Das Problem ist, dass die Methode sehr fehleranfällig ist“, sagt Schönhuth, Erstautor der Studie. Um etwa neue Varianten zu erkennen, kommt es auf Details an – und die stimmen nicht immer: Vereinfacht gesagt kann es zum Beispiel passieren, dass Basen bei der Analyse verzögert und dadurch doppelt abgelesen werden. „Die Fehlerrate beträgt bis zu zehn Prozent.“ Das ist viel, wenn es auf winzige Unterschiede ankommt.

Für eine Analyse müssen die Teile deshalb nicht nur wieder zusammengesetzt, sondern auch korrigiert werden. Für Forschende ist die Arbeit mit den Gensequenzen der Viren-Varianten wie das Lösen eines Puzzles: „Wir haben viele große Teile, aber wir wissen gar nicht, wie viele Puzzles es eigentlich gibt“, sagt Alexander Schönhuth. Das liegt daran, dass nicht klar ist, wie viele Varianten von SARS-CoV-2 das Material überhaupt enthält. „Manche Teile sind zudem verschwommen oder haben Fehler im Bild.“

Die Methode Strainline korrigiert solche Fehler frühzeitig. Schönhuth hat sie gemeinsam mit den Doktorand*innen Xiao Luo und Xiongbin Kang aus seiner Arbeitsgruppe Genominformatik entwickelt. Zugrunde liegt der Methode die Idee, die Genome nicht als Buchstabenketten, sondern in Form einer grafischen Anwendung darzustellen, die Verbindungen zwischen Genomen als Knotenpunkte zeigt und auch Auffälligkeiten schnell herausfiltert.

Algorithmus ermittelt, welche Genabschnitte zueinander gehören

Auf ähnliche Weise funktioniert auch die Methode Phasebook, die das Forschungsteam ebenfalls entwickelt hat. Sie eignet sich für sogenannte diploide Chromosomensätze, bei denen in den Zellen jedes Chromosom doppelt vorhanden ist. Menschen und auch andere Wirbeltiere besitzen einen solchen doppelten Chromosomensatz, bei dem je ein Teil von der Mutter und ein Teil vom Vater stammt.

„Um im Bild zu bleiben: Wir haben hier einen großen Karton mit sehr vielen Puzzleteilen“, sagt Schönhuth. „Wir wissen, dass sich daraus zwei Puzzles ergeben, aber wir wissen nicht, welches Teil zu welchem Puzzle gehört, weil wir keine Abbildung haben, die uns das zeigen würde.“ Die Teile genau zuordnen zu können, ist aber entscheidend – etwa für die Funktionelle Genomik, die Präzisionsmedizin und viele andere Disziplinen.

Ähnlich wie Strainline setzt Phasebook darauf, Analysefehler in langen Sequenzen schon frühzeitig zu erkennen und zu korrigieren. Außerdem ermöglicht die Methode durch einen Algorithmus eine bessere Zuordnung der Teile. „Viele andere Methoden verwenden stattdessen ein sogenanntes Referenzgenom“, sagt Schönhuth. „Dieses stammt aber von einem Europäer und funktioniert nicht so gut, wenn man beispielsweise das Genom von Menschen in Afrika oder Südamerika analysieren möchte.“ Mit Phasebook lässt sich laut Schönhuth auch ohne ein solches Referenzgenom rekonstruieren, welcher Teil der genetischen Ausstattung von der Seite der Mutter und welche von der Seite des Vaters stammt.

Verfahren sind in frei verfügbarer Software integriert

Strainline und Phasebook sind als Open-Source-Anwendungen konzipiert. Dadurch haben Forschende und Interessierte kostenlosen Zugriff auf die Softwares. Beide Methoden sind in Zusammenhang mit dem von der EU geförderten Promotionsnetzwerk Alpaca zur Pangenomik und dem ebenfalls von der EU geförderten Verbundprojekt Pangaia entstanden. Pangaia hat das Ziel, eine computergestützte Analyse großer Datensätzen zur Genomanalyse zu entwickeln. Sowohl das Promotionsnetzwerk wie auch das Verbundprojekt werden von der Universität Bielefeld geleitet. Schönhuth ist Koordinator von Alpaca und ist mit seiner Arbeitsgruppe Genominformatik an Pangaia beteiligt.

Alexander Schönhuth ist seit 2020 als Professor für Genome Data Science (Genom-Datenwissenschaft) an der Technischen Fakultät tätig. Außer am Institut für Bioinformatik-Infrastruktur (BIBI) forscht er auch am Centrum für Biotechnologie (CeBiTec) der Universität Bielefeld. Ein besonderer Schwerpunkt seiner Arbeit ist der Einsatz von Künstlicher Intelligenz zur Erforschung von Krankheiten, deren Ursachen bislang unklar sind. Mit dieser Forschung will er ermöglichen, Krankheiten gezielter und individueller behandelt zu können.

contact for scientific information:

Prof. Dr. Alexander Schönhuth, Universität Bielefeld
Technische Fakultät
Telefon: 0521 106-3793
E-Mail: gds@cebitec.uni-bielefeld.de

Original publication:

- Xiao Luo, Xiongbin Kang, Alexander Schönhuth: Strainline: full-length de novo viral haplo-type reconstruction from noisy long reads, Genome Biology, https://doi.org/10.1186/s13059-021-02587-6 online erschienen am 20. Januar 2022

- Xiao Luo, Xiongbin Kang, Alexander Schönhuth: phasebook: haplotype-aware de novo assembly of diploid genomes from long reads. Genome Biology, https://doi.org/10.1186/s13059-021-02512-x, online erschienen am 27. Oktober 2021

URL for press release: https://cordis.europa.eu/project/id/872539/de Website zum Pangaia-Projekt

URL for press release: https://alpaca-itn.eu/ Website des Promotionsnetzwerks Alpaca

URL for press release: https://www.uni-bielefeld.de/(de)/ZiF/AG/2019/09-30-Stoye.html Website zur ZiF-Arbeitsgemeinschaft „Computergestützte Pangenomik“

URL for press release: https://github.com/HaploKit/Strainline Zugriff auf die Open-Source-Anwendung Strainline

URL for press release: https://github.com/phasebook/phasebook Zugriff auf die Open-Source-Anwendung Phasebook

Prof. Dr. Alexander Schönhuth befasst sich an der Technischen Fakultät der Universität Bielefeld mit datenwissenschaftlichen Methoden für die Analyse von Virenerbgut und anderen Genomen.
Foto: Universität Bielefeld/S. Jonek