

Datenanalyse zum Coronavirus

Die Plattform Galaxy ermöglicht die freie und transparente Ansicht von Genominformationen zu COVID-19

Dr. Wolfgang Maier und Dr. Björn Grüning von der Albert-Ludwigs-Universität haben zusammen mit Forschenden von Universitäten in Belgien, Australien und den USA die bisher verfügbaren Daten zu Sequenzen des neuartigen Coronavirus überprüft und auf der Open-Source-Plattform Galaxy veröffentlicht. Die beiden Freiburger Bioinformatiker wollen damit den Datenaustausch zwischen den Behörden, Instituten und Laboren, die sich mit dem Virus beschäftigen, vereinfachen. Ihr Vorgehen und die Ergebnisse haben die Freiburger Forschenden auf dem Portal bioRxiv dokumentiert.

Die Plattform Galaxy eignet sich zur Analyse von Big Data in den Lebenswissenschaften. Über öffentliche Server erhalten Wissenschaftlerinnen und Wissenschaftler unter anderem freien Zugang zu Analysewerkzeugen und reproduzierbaren Auswertungsverfahren. Maier und Grüning haben zusammen mit Kolleginnen und Kollegen für ihre Studie alle bisher öffentlich verfügbaren COVID-19-Genomdaten mit Galaxy neu analysiert. Bisherigen Veröffentlichungen habe es oft an Transparenz bezüglich der Datenanalyse gefehlt, erklärt Grüning. So enthalte beispielsweise nur eine von vier bis Anfang Februar publizierten Studien zum COVID-19-Genom eindeutige Angaben zu den verwendeten Rohdaten, sagt Grüning: „Und die Analysen waren zudem nicht gut dokumentiert und nicht reproduzierbar.“ Dadurch war es nicht möglich, die jeweiligen Aussagen nachzuvollziehen oder zu überprüfen.

Innerhalb weniger Tage gelang es dem Team, auf die vorliegenden Sequenzen jeweils identische Arbeitsabläufe anzuwenden und mittels Galaxy öffentlich zugänglich zu machen. Forschenden steht damit nun weltweit über ein Netzwerk von Galaxy-Servern in Europa, den USA und Australien nicht nur die Auswertung der Daten, sondern gleichzeitig auch die wissenschaftliche Infrastruktur für eigene Analysen von COVID-19-Daten zur Verfügung. Damit lassen sich in Zukunft neuveröffentlichte Daten innerhalb von Stunden neu analysieren und mit den bisherigen Daten vergleichen.

Die Forschenden sind sich einig, dass es derzeit bei der Forschung zu COVID-19 an Datenaustausch mangle, so Maier. Das solle sich mit den Veröffentlichungen auf Galaxy ändern: „Eine globale Zusammenarbeit, die für die Bewältigung von Notfällen im Bereich der öffentlichen Gesundheit wie dem Ausbruch von COVID-19 notwendig ist, erfordert schließlich einen ungehinderten Zugang zu Daten, Analysewerkzeugen und zur Berechnungsinfrastruktur.“

Galaxy wurde an der US-amerikanischen Penn State University initiiert und an der Universität Freiburg im Sonderforschungsbereich „Medizinische Epigenetik“ sowie als Teil des Deutschen Netzwerks für Bioinformatik-Infrastruktur (de.NBI) weiterentwickelt. Der Europa-Server befindet sich im Rechenzentrum der Universität Freiburg und ist als Community-Projekt angelegt. Die Daten sind online frei zugänglich. Wissenschaftler, die den Server nutzen möchten, brauchen keine Kenntnisse im Programmieren: Alle Einstellungen lassen sich über eine grafisch aufbereitete Oberfläche vornehmen. Federführend in der Weiterentwicklung von Galaxy ist das Team der Albert-Ludwigs-Universität um Grüning aus der Arbeitsgruppe von Prof. Dr. Rolf Backofen am Institut für Informatik.

Anmerkung

Die auf Galaxy analysierten Genomdaten gehören zu dem Virus SARS-CoV-2, der die Krankheit COVID-19 auslöst. Da die auf der Plattform bisher veröffentlichten Daten von Personen stammen, bei denen COVID-19 bereits ausgebrochen war, werden im Text verkürzt die Begriffe COVID-19-Genom und COVID-19-Genomdaten verwendet.

Pressemitteilung

05.03.2020

Quelle: Albert-Ludwigs-Universität Freiburg

Weitere Informationen

Dr. Björn Grüning
Institut für Informatik
Albert-Ludwigs-Universität Freiburg
Tel.: +49 (0)761/203-54130
E-Mail: [gruening\(at\)informatik.uni-freiburg.de](mailto:gruening(at)informatik.uni-freiburg.de)

- ▶ [Albert-Ludwigs-Universität Freiburg](#)
- ▶ [Originalpublikation](#)