



Medieninformation

Spezielle Software aus Greifswald beschleunigt die genaue Bestimmung aller Gene von Tieren und Pflanzen enorm

Universität Greifswald, 14.06.2024

Die neu entwickelte Software BRAKER3 bietet Wissenschaftler*innen weltweit ein leistungsstarkes Instrument für die Genomannotation, d.h. bei der Identifizierung und Kennzeichnung vieler relevanten Merkmale einer genomischen Sequenz. Die Software stellt einen bedeutenden Fortschritt in der bioinformatischen Forschung dar. Sie wurde von Forschenden der Universität Greifswald in Zusammenarbeit mit Kollegen am Georgia Tech Institute of Technology in Atlanta (USA) entwickelt. Das auf Greifswalder Seite durch Prof. Dr. Mario Stanke geleitete Bioinformatikteam stellt im Journal Genome Research (DOI 10.1101/gr.278090.123) die Software vor. BRAKER3 nutzt aus, dass verschiedene Spezies dieselben Gene in ähnlicher Form besitzen können, selbst wenn ihr gemeinsamer evolutionärer Ursprung weit zurückliegt, wie etwa bei einem Schmetterling und einer Fruchtfliege. Die Entwicklung wurde finanziert von den US National Institutes of Health.

Die genaue Bestimmung der Struktur proteinkodierender Gene in Genomsequenzen ist ein Schlüssel für das biologische Verständnis des Lebens. Der Erfolg zahlreicher Experimente hängt entscheidend von einer fehlerfreien Genomannotation ab. Die Erfassung proteinkodierender Gene in eukaryotischen Genomen ist deshalb auch eine der großen Herausforderungen des EarthBioGenome Projekts. In ihm sollen die Genome von mindestens 1,5 Millionen eukaryotischer Spezies sequenziert werden sollen. Unter Eukaryoten werden Zellen verstanden, die einen Zellkern besitzen. Zu eukaryotischen Organismen zählen Tiere, Menschen, Pflanzen und Pilze. Mögliche Anwendungen von einzelnen Genomprojekten können etwa sein: Die gezielte Bekämpfung von durch Tiere übertragenen Krankheiten, das Studium der Funktion von Genen bei Insekten oder in der Pflanzenzüchtung.

Ein zentrales Problem vieler Werkzeuge zur Genomannotation ist das sogenannte überwachte Lernen: Die zugrundeliegenden mathematischen Modelle benötigen Trainingsbeispiele bestehend aus Genen in der Zielspezies, um Parameter an diese Zielspezies anzupassen. Hier baut das BRAKER3-Team auf Erfahrungen der Vorläuferversionen der Software auf und lässt die kombinierte Evidenz aus Transkriptom- und Proteindaten in den Trainingsschritt einfließen. Im Gegensatz zu den Vorgänger-Tools können nun beide Evidenztypen simultan berücksichtigt werden.

In Benchmark-Tests mit 11 Spezies übertrifft BRAKER3 die vorherigen Versionen deutlich. Besonders offensichtlich ist die Verbesserung bei Arten mit großen und komplexen Genomen, wie z.B. bei der Maus und dem Huhn. Zudem ist neue Softwareversion auch weit genauer als andere bisher vielfach genutzte alternative Programme.

"BRAKER3 stellt einen bedeutenden Fortschritt in der Genauigkeit und Automatisierbarkeit der eukaryotischen Genomannotation dar, insbesondere für große und strukturell komplexe Genome", so Lars Gabriel vom Institut für Mathematik an der Universität Greifswald, Erstautor der Publikation. "Mit der neuen Softwareversion haben wir ein Tool mit einer bereits großen und schnell wachsenden Anzahl von Nutzern. Insbesondere die Bemühungen des Teams, die Software so zu gestalten, dass sie in isolierten Paketen läuft, die alle nötigen Bestandteile für das Programm enthalten und auf verschiedenen Computersystemen ohne zusätzliche Anpassungen funktionieren, werden von der internationalen Forschungscommunity begrüßt.

Dieses Prinzip, das man als 'Verpacken in Container' bezeichnet, wurde entscheidend durch die exzellente High Performance Computing Infrastruktur des Universitätsrechenzentrums in Greifswald beeinflusst", sagt Dr. Katharina Hoff vom Institut für Mathematik an der Universität Greifswald. Sie arbeitet seit vielen Jahren an der Entwicklung von BRAKER.

"BRAKER3 markiert einen signifikanten Fortschritt in der Bioinformatik und bietet Wissenschaftlern weltweit ein leistungsstarkes Instrument für die Genomannotation. In der weiteren Entwicklung sollen große Sprachmodelle spezifisch entwickelt und trainiert werden, da Genome eine 'Sprache' der Biologie sind, und die in ihnen kodierten Gene einer strengen Grammatik folgen", so Prof. Dr. Mario Stanke, Leiter der AG für Bioinformatik am Institut für Mathematik an der Universität Greifswald.

Weitere Informationen

Paper: genome.cshlp.org/content/early/2024/05/28/gr.278090.123.abstract (IF 9.4) Pub Date: 12.06.2024, DOI [10.1101/gr.278090.123](https://doi.org/10.1101/gr.278090.123)

Arbeitsgruppe: [AG Bioinformatik am Institut für Mathematik und Informatik](#)

Software: <https://github.com/Gaius-Augustus/BRAKER>

Ansprechpartner*in an der Universität Greifswald

PD Dr. Katharina Hoff

Institut für Mathematik und Informatik

Walther-Rathenau-Straße 47, 17489 Greifswald

Telefon 03834 420 4624

katharina.hoff@uni-greifswald.de

https://twitter.com/katharina_hoff

[@katharinahoff.bsky.social](https://bsky.app/profile/katharinahoff.bsky.social)

[@KatharinaHoff@fosstodon.org](mailto:KatharinaHoff@fosstodon.org)