EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

**Hochschul-
kommunikation**

Dr. Karl Guido Rijkhoek
Leiter

Janna Eberhardt
Forschungsredakteurin

Telefon +49 7071 29-76788
+49 7071 29-77853
Telefax +49 7071 29-5566
karl.rijkhoek[at]uni-tuebingen.de
janna.eberhardt[at]uni-tuebingen.de

www.uni-tuebingen.de/aktuell

Pressemitteilung

Die Rechenzeit für genetische Großprojekte reduziert sich von Jahren auf Tage

Tübinger Bioinformatiker entwickeln das Programm DIAMOND, das die Daten vor der Analyse besser als bisherige Tools strukturiert und schnell bearbeiten kann

Tübingen, den 18.11.2014

Wissenschaftlern stehen immer bessere Methoden zur Verfügung, um die DNA von Lebewesen zu sequenzieren, also auszulesen. Längst wird nicht mehr nur das Erbgut eines einzelnen Organismus untersucht, sondern zum Beispiel die DNA einer ganzen Bodenprobe, in der eine Vielzahl von Bakterien, Pilzen und Insekten leben. Oder Mediziner möchten anhand einer Stuhlprobe über die enthaltene DNA herausbekommen, wie sich die Darmflora des Menschen zusammensetzt. Bei solchen Projekten erhalten die Wissenschaftler gigantische Datenmengen, die analysiert werden müssen. Dazu werden die gewonnenen DNA-Daten mit den Einträgen großer Datenbanken zu bereits untersuchten Lebewesen verglichen, genauer gesagt mit deren Proteinen. So lassen sich die Organismen aus der Probe identifizieren. Bisher verwenden Bioinformatiker für solche Aufgaben Programme der Blast-Familie, die dafür viel Rechenzeit benötigen – und dadurch den sprichwörtlichen Flaschenhals der ehrgeizigen Projekte bilden. Nun ist es Benjamin Buchfink und Professor Daniel Huson vom Zentrum für Bioinformatik der Universität Tübingen in Zusammenarbeit mit Chao Xie von der National University in Singapur gelungen, dieses Verfahren um das 20.000-Fache zu beschleunigen. Die Rechenzeit von Jahren wird dabei auf einzelne Tage reduziert. DIAMOND (double index alignment of next-generation sequencing data) haben die Wissenschaftler ihr neues Programm genannt, das die früheren Tools sehr schnell ablösen könnte. Ihr neues leistungsfähiges Programm stellen die Wissenschaftler in der Fachzeitschrift *Nature Methods* vor.

Die Programme arbeiten alle nach dem gleichen Prinzip: Jeweils ein kurzes Stück der DNA, eine Abfolge von mehreren Basen, vergleichbar mit einem Wort aus einigen Buchstaben, sucht in der Datenbank sein Gegenstück. Dann wird die Suche nach links und rechts ausgeweitet, um zu sehen, ob es sich um eine kurze zufällige Übereinstimmung handelt oder tatsächlich das entsprechende Element gefunden worden ist. „Das Blast-

Programm arbeitet mit einem einfachen Index, was man sich so als Wörterbuch vorstellen kann. Im Computer müssen dabei immer wieder Daten vom Hauptspeicher in den Arbeitsspeicher übertragen werden, das kostet viel Zeit“, erklärt Daniel Huson, „eigentlich ist viel mehr Rechenkapazität vorhanden, aber die Prozessoren müssen warten, bis sie wieder etwas tun können.“ Sein Mitarbeiter Benjamin Buchfink hat verschiedene Ideen und Möglichkeiten getestet, wie sich das Verfahren beschleunigen ließe. Den Durchbruch brachte die doppelte Indizierung: „DIAMOND sortiert sowohl die DNA-Daten aus der Probe als auch die Proteindaten aus der Datenbank, und diese beiden Listen werden miteinander abgeglichen. Wir waren selbst überrascht, dass sich das Verfahren dadurch so stark beschleunigen lässt“, sagt Huson. Die Genauigkeit der Ergebnisse sei mit der des früheren Verfahrens vergleichbar.

Für drei Milliarden Abgleiche von Proben-DNA mit der Datenbank würde ein einzelner Computerprozessor mit dem bisherigen BlastX-Programm 29 Jahre lang rechnen. „Mit DIAMOND dauert die Abarbeitung der gleichen Aufgabe gerade einen Tag“, sagt der Wissenschaftler. „Erst mit dem leistungsfähigeren Tool können wir überhaupt neue anspruchsvolle Projekte angehen. Wir wollen zum Beispiel in Zusammenarbeit mit Medizinern Therapien für einzelne Patienten entwickeln, die auf den individuellen Genen beruhen. Wir schätzen, dass wir dabei 15 Milliarden DNA-Abgleiche benötigen.“

Originalpublikation:
Benjamin Buchfink, Chao Xie, Daniel H Huson: Fast and sensitive protein alignment using DIAMOND. Nature Methods, Online-Veröffentlichung am 17. November 2014, DOI: 10.1038/NMETH.3176; www.nature.com/nmeth/journal/vaop/ncurrent/full/nmeth.3176.html

Kontakt:
Prof. Dr. Daniel Huson
Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Wilhelm-Schickard-Institut für Informatik – Bioinformatik
Telefon +49 7071 29-70450
daniel.huson[at]uni-tuebingen.de