



PRESSEINFORMATION

FORSCHUNG

Mit der Verwandtschaft in Reih und Glied – Bessere Sequenzanalysen bei Genen und Proteinen

München, 23. Februar 2009 – Seit der Sequenzierung des menschlichen Genoms vor acht Jahren wurden bei der Analyse der enthaltenen Gene und deren Funktionen große Fortschritte gemacht. Dennoch ist der größere Teil der menschlichen Gene bisher in ihrer Funktion kaum verstanden. Zur Funktionsbestimmung hilft oft ein Vergleich der Gensequenzen mit denen von hunderten anderer Organismen, die experimentell vielfach leichter zu untersuchen sind und deren Sequenzen weltweit in Datenbanken bereitgehalten werden. Diese dienen Forschern als Vergleich, um auf die unbekannte Funktion menschlicher oder tierischer Gene rückschließen zu können. Nun ist es den Bioinformatikern Dr. Johannes Söding und Andreas Biegert vom Genzentrum der Ludwig-Maximilians-Universität (LMU) München gelungen, ein Verfahren zu entwickeln, das die Datenbanksuche bei gleicher Geschwindigkeit deutlich empfindlicher macht. Dadurch lassen sich verwandte Gen- und Proteinsequenzen in den Datenbanken mit großer Genauigkeit aufspüren. Die zugrunde liegende Idee lässt sich vielseitig für alle Arten von Sequenzanalysen einsetzen. (PNAS, 20. Februar 2009)

Bei Genen wie auch Proteinen gilt: Die Funktion beruht in erster Linie auf der Sequenz, also der Abfolge der Bausteine. Ähnlich aufgebaute Gene besitzen also häufig eine entsprechende Funktion. Dies gilt in gewissem Grade auch für Proteine, bei denen aber noch die dreidimensionale Struktur, in die sie sich falten und die sich nicht ohne Weiteres aus ihrer Sequenz ableiten lässt, eine entscheidende Rolle spielt. Ähnliche Sequenzen – zumindest in Schlüsselbereichen – lassen aber auf eine Verwandtschaft schließen und legen damit ähnliche Funktionen nahe.

Die Sequenzen und Funktionen von Genen wie auch Proteinen werden daher in Datenbanken gespeichert, die Wissenschaftler auf der ganzen Welt zum Vergleich mit neuen Daten nutzen. Doch selbst die besten und am häufigsten genutzten Algorithmen wie etwa BLAST (Basic Local

Luise Dirscherl (Leitung)

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
dirscherl@lmu.de

Infoservice:
+49 (0)89 2180 - 3423

Geschwister-Scholl-Platz 1
80539 München
presse@lmu.de
www.lmu.de

Kommunikation und Presse

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
[dirtscherl@lmu.de](mailto:dirscherl@lmu.de)

Infoservice:
+49 (0)89 2180 - 3423

Alignment Search Tool) verwenden bestimmte Vereinfachungen, um eine effiziente Suche in den riesigen Datenmengen überhaupt zu ermöglichen. Schließlich muss BLAST die von einem Forscher eingegebene Sequenz – den Buchstabencode, der die Abfolge der DNA-Bausteine oder der Aminosäuren beschreibt – mit sämtlichen Sequenzen in der Datenbank in wenigen Minuten vergleichen können.

Dabei bewerten die Suchprogramme die Ähnlichkeit der Sequenzen, indem diese so untereinander angeordnet werden, dass sich möglichst ähnliche Aminosäuren in der gleichen Spalte dieses sogenannten Alignments befinden. Die Berechnung der Ähnlichkeit erfolgt dann Aminosäure für Aminosäure, wobei berücksichtigt wird, wie gut zwei Aminosäuren ohne nachteilige Folgen gegeneinander austauschbar sind. Denn manche dieser Proteinbausteine sind sich in ihren Eigenschaften ähnlich, sodass ihr Austausch in einem Protein oft keinen Effekt auf dessen Funktion hat.

BLAST ist seit seiner Entwicklung 1990 das wichtigste Instrument der Sequenzsuche und wird weltweit rund 500.000 Mal am Tag aufgerufen. Doch das Programm ist noch lange nicht optimal. So beachtet es bei der Bewertung der Ähnlichkeit zweier Aminosäuren nicht deren Kontext, also die benachbarten Aminosäuren. Johannes Söding und Andreas Biegert vom Genzentrum München und dem Exzellenzcluster „Center for Integrated Protein Science Munich (CIPSM)“ der LMU haben nun ein Verfahren entwickelt, das die Ähnlichkeitssuche deutlich verbessert: Ihr sogenanntes kontext-spezifisches BLAST, CS-BLAST, kann bei gleicher Suchgeschwindigkeit doppelt so viele entfernte „Verwandte“ von Proteinen aufspüren wie bisher.

Um die Ähnlichkeit einer Aminosäure mit den Referenzdaten zu bestimmen, wird bei CS-BLAST auch der Sequenzkontext jeder Aminosäure, nämlich deren sechs linke und sechs rechte Nachbarn, in die Analyse mit einbezogen. „Die Idee ist, dass der Kontext sehr viel darüber aussagt, als wie ähnlich zwei Aminosäuren zu bewerten sind“, erläutert Söding, der die Arbeitsgruppe „Protein Bioinformatics and Computational Biology“ am Genzentrum München leitet. „So gibt es beispielsweise bei Proteinen gefaltete und ungefaltete Bereiche. In einem ungefalteten Bereich kann etwa die Aminosäure Valin oft ohne nachteiligen Effekt in die anderen 19 Aminosäuren mutieren. In einem gefalteten Bereich geht sie dagegen mit hoher Wahrscheinlichkeit in Aminosäuren über, die ebenfalls hydrophob, also wasserabstoßend, sind.“

Dem Programm liegt eine sehr allgemeine Idee zugrunde, die sich auf alle Arten der Sequenzsuche und der Alignmentmethoden übertragen lässt. Dies zeigten die Forscher am Beispiel von PSI-BLAST, einem Algorithmus, bei dem bereits aufgefundene verwandte Sequenzen in einem sogenannten multiplen Alignment untereinander angeordnet werden. Dadurch lassen sich Gruppen von Aminosäuren identifizieren, die immer an der gleichen

Kommunikation und Presse

Telefon +49 (0)89 2180 - 2706
Telefax +49 (0)89 2180 - 3656
[dirtscherl@lmu.de](mailto:dirscherl@lmu.de)

Infoservice:
+49 (0)89 2180 - 3423

Stelle der Sequenz vorkommen und so Hinweise auf die Funktion dieses Abschnitts geben. „Uns ist es gelungen, mit der Idee des Sequenzkontextes auch PSI-BLAST deutlich sensitiver zu machen. Dadurch liefern zwei hintereinandergeschaltete Suchen mit unserer kontext-spezifischen Version von PSI-BLAST bessere Resultate als fünf Suchen des herkömmlichen Programms“, sagt Söding.

Dass die neue Methode trotz besserer Empfindlichkeit gleich schnell ist, erklärt der Forscher damit, dass die Sequenzsuche in zwei Schritten erfolgt: „Sowohl beim herkömmlichen BLAST als auch bei unserer Methode wird zunächst eine Suchmatrix berechnet“, erklärt Söding. „Dieser Schritt ist bei uns zwar aufwendiger, aber mit einer Sekunde immer noch sehr schnell. Erst der zweite Schritt, die Datenbanksuche mithilfe der Suchmatrix, nimmt viel Zeit in Anspruch – und dieser Schritt ist bei beiden Ansätzen gleich.“

In Zukunft wollen die Biowissenschaftler den neu entwickelten Algorithmus auch auf genomische Alignments anwenden, bei denen nicht nur einzelne Gene, sondern ganze Abschnitte des Erbguts verglichen werden. „Ähnlich wie bei Proteinen gibt es in der DNA bestimmte Schlüsselstellen, die besonders wichtige Funktionen erfüllen“, erläutert Söding. „Diese auch für das tiefere Verständnis vieler Krankheiten wichtigen regulatorischen Bereiche kann man identifizieren, wenn man das menschliche Genom denen von anderen Säugetieren gegenüberstellt.“ Durch eine kontext-spezifische Methodik hoffen die LMU-Forscher, die Qualität solcher genomischer Alignments und damit die Identifikation regulatorischer Bereiche wesentlich zu verbessern. „Wir glauben, dass sich unsere Methode im gesamten Bereich der biologischen Sequenzanalyse durchsetzen könnte“, schließt Söding. (ca/sw)

CS-BLAST Server und Software:

http://toolkit.lmb.uni-muenchen.de/cs_blast (frei für alle)
<ftp://toolkit.lmb.uni-muenchen.de/csblast/> (frei für Akademiker)

Publikation:

“Sequence context-specific profiles for homology searching”;
Andreas Biegert und Johannes Söding;
Proceedings of the National Academy of Sciences (PNAS) online;
20. Februar 2009;
DOI: 10.1073/pnas.0709640104

Ansprechpartner:

Dr. Johannes Söding
Genzentrum der Ludwig-Maximilians-Universität (LMU) München und
Department für Chemie und Biochemie
Tel.: 089 / 2180 – 76742
E-Mail: soeding@lmb.uni-muenchen.de
Web: <http://www.lmb.uni-muenchen.de/soeding/>