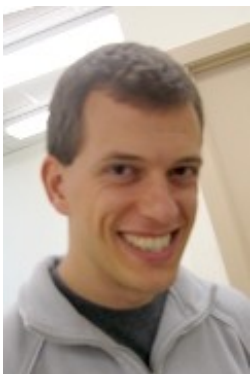


Tancred Frickey – Von Proteinen und Programmen

Ein sehr wichtiges Einsatzgebiet der Bioinformatik ist die Klassifizierung von Proteinsequenzen. Prof. Dr. Tancred Frickey, Professor für Angewandte Bioinformatik an der Universität Konstanz, hat ein Programm namens CLANS entwickelt, mit dem er zum ersten Mal die Protein-Familie der AAA-ATPasen klassifizieren konnte. Die Software eignet sich überraschenderweise auch für die Visualisierung der Ähnlichkeiten von Schauspielern anhand des Genres, in dem sie mitgespielt haben.



Prof. Dr. Tancred Frickey ist Inhaber des Lehrstuhls für Angewandte Bioinformatik an der Universität Konstanz.
© privat

Mit computergestützten Methoden Lösungen für Probleme aus den Lebenswissenschaften finden – diese Aufgabe steht im Mittelpunkt der Forschung von Professor Tancred Frickey, der nach fünf Jahren Tätigkeit am Australian National University (ANU) College of Science in Australien wieder nach Konstanz zurückgekehrt ist. Ein großer Forschungszweig, mit dem er sich beschäftigt, ist die Analyse von Proteinen und Gen-Expressionsdaten. „Das Ermitteln von Verwandtschaftsgraden von Proteinen ist deshalb von großer Bedeutung, da man bei Kenntnis der Verwandtschaftsverhältnisse abschätzen kann, welche Eigenschaften von beschriebenen Proteinen auf bisher unbeschriebene oder unbekannte Proteine übertragen werden können“, erklärt Professor Frickey, der selbst schon mehrere Software-Anwendungen geschrieben hat, allen voran CLANS, das dafür geschaffen wurde, Daten mit paarweiser Sequenzähnlichkeit im dreidimensionalen Raum aufzuzeigen.

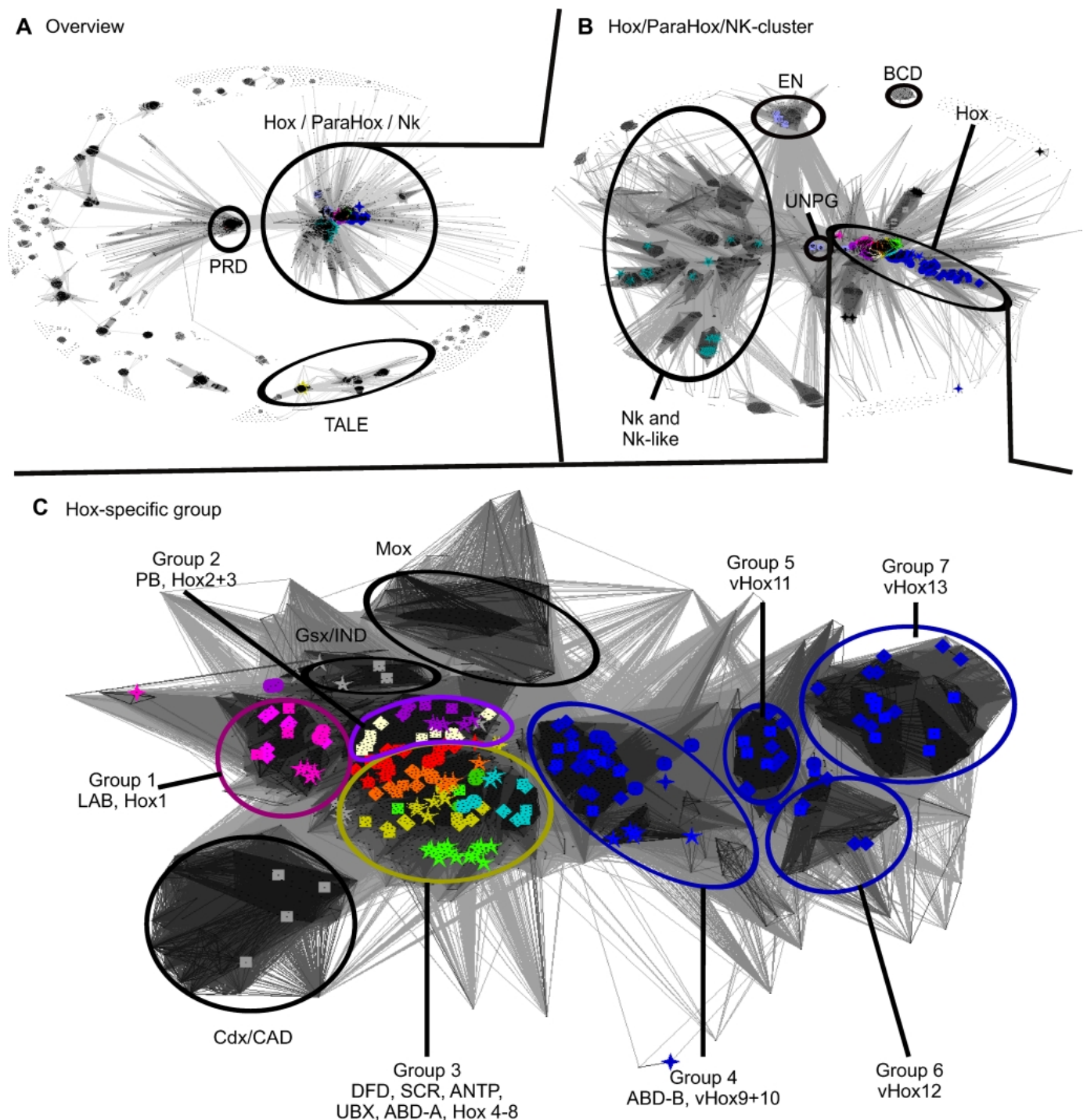
Eine ganz besondere Proteinfamilie, die Professor Frickey dank CLANS erstmalig klassifizieren konnte, sind AAA-ATPasen, die in allen Lebewesen vorkommen und meist an der Entfaltung von Proteinen beteiligt sind. „Bis vor ein paar Jahren war es nicht möglich, eindeutig zu definieren,

welche Sequenzen der Familie AAAs waren und welche nicht, da öffentliche Protein-Datenbanken über 5.000 Einträge zu AAA- und AAA-verwandten Sequenzen enthielten“, stellt der Bioinformatiker fest. Auch die Mutationssättigung der berechneten multiplen Sequenz-Alignierungen stand der Klassifikation der Superfamilie im Weg. „Mithilfe von CLANS ist es erstmals möglich, potenzielle AAA-ATPasen objektiv zu beschreiben und einzuordnen, indem man die AAA-Sequenzen anhand ihrer Anordnung im Sequenzraum identifiziert“, beschreibt dies der Bioinformatiker. In Kombination mit phylogenetischen Methoden können so auch die Verwandtschaftsverhältnisse innerhalb der AAA-Subfamilien entschlüsselt werden.

Unvorhergesehenes Potenzial: Kategorisierung von Spezies und Schauspielern

Viele der in der Bioinformatik verwendeten Technologien, etwa Programme, die zur Interpretation von Gemüse-Sequenzdaten entwickelt wurden, haben einen nicht immer auf den ersten Blick erkennbaren weiteren Nutzen. Mit nur geringfügigen Änderungen können diese Anwendungen zur Analyse des Transkriptoms oder des menschlichen Genoms verwendet werden. Auch Professor Frickeys Software CLANS ist vielseitig einsetzbar: So kann man damit beispielsweise mehrere Proteome vergleichen, um herauszufinden, welche Proteine in der einen Spezies, aber nicht der anderen vorhanden sind. Dazu werden Daten, die alle Proteine der jeweiligen Spezies enthalten, in das Programm eingepreist, das dann die paarweisen Sequenzähnlichkeiten grafisch darstellt. „Wenn man eine Spezies grün und die andere rot einfärbt, kann man die einfarbigen Proteingruppen und -familien identifizieren, die nur in einer Spezies vorkommen“, erklärt der Bioinformatiker die Funktionsweise.

Ein ganz interessantes Einsatzgebiet von CLANS hat nicht einmal Professor Frickey vorhersehen können: So ist es mit der Software-Anwendung möglich, die Ähnlichkeiten von Schauspielern anhand des Genres, in dem sie mitgewirkt haben, zu visualisieren. „Wenn man davon ausgeht, dass Filme von großem allgemeinem Interesse und berühmte Schauspieler menschliche Wesen sind, dann hat diese Software eine Form von menschenorientierter, allgemeiner Anwendung gefunden“, so Frickey schmunzelnd.



CLANS-Analyse von HOX-Proteinen, die das typische Clustering-Verhalten von Sequenzen zeigt
 © Hueber SD, Weiller GF, Djordjevic MA, Frickey T, 2010 Improving Hox Protein Classification across the Major Model Organisms. PLoS ONE 5(5): e10820. doi:10.1371/journal.pone.0010820

Nächste Stufe der Datenintegration

„Bioinformatik ist schnell, man kann an einem großen und abwechslungsreichen Themenfeld arbeiten und jeden Computer als Labor benutzen. Außerdem gibt es nur eine Fehlerquelle, die beim Optimieren der Programme und der Analyse berücksichtigt werden muss: dich selbst“, beschreibt Tancred Frickey seine Leidenschaft für diesen Forschungszweig. Er ist sich sicher, dass die Bioinformatik ständig wachsen wird, insbesondere im Bereich der Datenerfassung. Ein Problem sieht er darin, dass neue Technologien zur Integration dieser Daten aus den verschiedenen Datenbanken sich zu langsam entwickeln. „Die übergeordneten Zusammenhänge spielen in der Bioinformatik eine wichtige Rolle. Bei der Zusammensetzung von Genomen aus Sequenzrohdaten zum Beispiel können Vorhersagen bezüglich der enthaltenen Proteine gemacht werden und wie sie

im entstehenden Organismus arbeiten“, klärt Frickey auf.

Eine Vielzahl solcher Daten sind in den öffentlich zugänglichen Datenbanken zu finden, aber nicht alle sind in einem übergeordneten Kontext anwendbar oder gleich verlässlich: „Viele der Annotationen von Proteinsequenzen sind fehlerhaft, weil ihre Funktionen und Annotationen von ähnlichen Sequenzen übernommen wurden, ohne dass spezifische biologische Experimente durchgeführt wurden. Eine falsch annotierte Proteinsequenz kann so großen Einfluss auf andere Sequenzen haben, wenn deren Eigenschaften auf neue Proteine übertragen werden“, erklärt der Konstanzer Professor. Rückschlüsse bezüglich übergeordneter Zusammenhänge, etwa dem Interaktom von Proteinen in einer Zelle, werden dadurch erschwert. „Die Erstellung von Daten, die Zusammenhänge aufzeigen, ist sehr arbeitsintensiv. Hier ist ein technologischer Durchbruch nötig, um die Daten in der Qualität und Quantität zu produzieren, damit sie auch für ein breites Anwendungsfeld genutzt werden können“, fasst Prof. Frickey zusammen, der bereits viele weitere Projekte im Kopf hat, die aber aufgrund der Qualität der Daten noch nicht realisierbar sind.

Fachbeitrag

27.11.2010

Andreas Zangar

BioLAGO

© BIOPRO Baden-Württemberg GmbH

Weitere Informationen

Prof. Dr. Tancred Frickey

Fachbereich Biologie

Universität Konstanz

Universitätsstraße 10

78457 Konstanz

Tel.: 07531/88-2343

E-Mail: tancred.frickey(at)uni-konstanz.de

Der Fachbeitrag ist Teil folgender Dossiers



Data-Mining: Neue Chancen für Medizin und Gesundheit

Universität Konstanz

