

Ein Code für die Evolutionsbiologie

Alexandros Stamatakis ist Leiter der neuen Forschungsgruppe „Scientific Computing“ am Heidelberger Institut für Theoretische Studien (HITS) – Software und Supercomputer zur Auswertung großer Datenmengen in der Biologie.



Dr. Alexandros (Alexis) Stamatakis im Garten des HITS
© HITS

Sein eigener Stammbaum ist international: die Mutter Deutsche, der Vater Grieche, in Saarbrücken geboren, in Athen Abitur. Dass sich Dr. Alexandros Stamatakis für Stammbäume von Pflanzen interessiert, liegt daran, dass ihre Auswertung ihn als Informatiker reizt: „Die Evolutionsbiologie produziert ungeheuer viele Daten, die selbst Höchstleistungsrechner nicht ohne Weiteres analysieren können“, sagt der 34-jährige. „Eine Herausforderung für die Informatik besteht darin, Programme und Methoden zu entwickeln, um aus den riesigen molekularen Datenmengen z.B. Stammbäume zu berechnen, bzw. allgemein Wissen zu extrahieren.“

An Software zur Berechnung evolutionärer, sogenannter phylogenetischer Bäume sowie an der Entwicklung neuer Rechnerarchitekturen und Schaltungen zur Stammbaumberechnung arbeitet Stamatakis seit Oktober dieses Jahres am Heidelberger Institut für Theoretische Studien (HITS) als Leiter der neu eingerichteten Forschungsgruppe „Scientific Computing“ (SCO). Er baut zudem den neuen Parallelrechner am HITS auf und stellt den anderen fünf Forschungsgruppen seine Expertise in den Bereichen der parallelen Rechnerarchitekturen und der parallelen Programmierung zur Verfügung. Derzeit hat die SCO-Gruppe zehn Mitglieder, nächstes Jahr kommen weitere Doktoranden und Gastwissenschaftler hinzu.

„Mit Alexandros Stamatakis konnten wir einen jungen Forscher gewinnen, der die Ziele des HITS beispielhaft verkörpert, denn computerbasierte Methoden helfen, die Datenflut in den Lebenswissenschaften zu bewältigen und neue wissenschaftliche Erkenntnisse zu gewinnen“, so HITS-Gründer und Geschäftsführer Klaus Tschira.

Alexandros Stamatakis studierte Informatik in München, Lyon (École Normale Supérieure), Paris und Madrid und promovierte 2004 an der TU München. Nach zwei Postdoc-Stationen auf Kreta und an der ETH Lausanne kehrte Stamatakis 2008 nach München zurück, zunächst an die LMU, dann an die TU München, als Leiter einer Nachwuchsforschungsgruppe. Zuvor hatte er zwei Rufe auf Assistant Professorships in den USA abgelehnt.

Primäres Forschungsziel des Informatikers ist es, Programme zu entwickeln, um die evolutionäre Entwicklung aller Lebewesen zu rekonstruieren, für die genetische Daten vorliegen. „Die Evolutionsbiologen stoßen hier auf ein fundamentales Problem, das in der Informatik als „NP-hard“ („non deterministic polynomial time hard“) bezeichnet wird. Ein Beispiel: Wenn man die DNA von fünfzig Organismen mit einer Bewertungsfunktion untersucht und herausfinden will, wie gut die einzelnen Daten zu einem bestimmten Stammbaum passen, dann ist es unmöglich, alle Bäume mit dieser Funktion zu bewerten, weil es einfach zu viele sind. „Selbst mit der gesamten Rechenleistung dieser Erde müssten wir zu lange warten, bis wir den optimalen Baum finden“, erklärt Alexandros Stamatakis. „Die aktuell veröffentlichten phylogenetischen Bäume umfassen jedoch nicht fünfzig, sondern mehrere tausend bis zehntausend Organismen.“

Um diesem Problem zu begegnen, hat der Informatiker das Programm RAXML entwickelt, das gegenwärtig die Analyse von Stammbäumen mit bis zu 120.000 Organismen ermöglicht. Die Software gehört mittlerweile zu den zehn weltweit am weitesten verbreiteten Anwendungen zur Rekonstruktion von Stammbäumen. Die wissenschaftliche Publikation zu dieser Software zählt zu den am häufigsten zitierten Informatik-Veröffentlichungen der letzten fünf Jahre. „RAXML ist ein open source code und damit uneingeschränkt für alle zugänglich und modifizierbar“, sagt Alexandros Stamatakis. „Wir stellen damit Biologen ein Werkzeug zur Verfügung, anhand dessen sie ihre Daten selbst analysieren können.“ RAXML wurde übrigens in diesem Jahr auch in die Liste der SPEC-Benchmark-Programme aufgenommen, mit denen die Leistungsfähigkeit von Hoch- und Höchstleistungsrechnern analysiert wird.

Außerdem arbeitet der Informatiker an einem großen internationalen Projekt mit: „iPlant“ wurde von der US-amerikanischen National Science Foundation (NSF) initiiert und soll eine Art Cyber-Infrastruktur für die Pflanzenwissenschaften schaffen. Alexandros Stamatakis ist hier als einziger Europäer beteiligt, gefördert von der Deutschen Forschungsgemeinschaft (DFG). Ebenfalls einzigartig ist, dass Stamatakis als bislang einziger Informatiker zum Mitglied des Councils der „Society of Systematic Biologists“ (Gesellschaft für Systematik in der Biologie) gewählt wurde.

In den kommenden Monaten setzt Alexandros Stamatakis seine Kooperation mit verschiedenen Institutionen fort, zum Beispiel mit dem Evolutionsbiologen Stephen Smith am Dunn Lab der Brown University, Rhode Island/USA. Ein Austauschprogramm mit dem Imperial College London kommt seinen Doktoranden zugute. Und gemeinsam mit Wissenschaftlern des Europäischen Molekularbiologischen Laboratoriums (EMBL) und des European Bioinformatics Institute (EBI) organisiert Stamatakis den Kurs „Computational Molecular Evolution“ vom 10. bis 21. April 2011 in Hinxton bei Cambridge (Großbritannien). Dort sollen junge

Biologen das theoretische Wissen und das praktische Können erlernen, molekulare Evolutionsanalysen selbst durchzuführen.