

Science: Big Data Explain Evolution of Birds

Bird Tree of Life Reproduced by Gene Analysis and Supercomputing / New Findings about Basics of Birdsongs, Feathers, Biodiversity, and Bird Evolution



Monika Landgraf
Chief Press Officer

Kaiserstraße 12
76131 Karlsruhe, Germany
Phone: +49 721 608-47414
Fax: +49 721 608-43658
E-mail: presse@kit.edu

For the family of bee-eaters (on the photo Merops bullocki), the study revealed a close relationship to oscine birds, parrots, and birds of prey. (Photo: Peter Houde)

About 95 percent of the more than 10,000 bird species known only evolved upon the extinction of dinosaurs about 66 million years ago. According to computer analyses of the genetic data, today's diversity developed from a few species at a virtually explosive rate after 15 million years already. KIT scientists designed the algorithms for the comprehensive analysis of the evolution of birds. To obtain the results that are now presented in the Science journal, a computing capacity of 300 processor-years was required. (DOI 10.1126/science.1253451)

"Computation of these trees of life for evolution research is impossible without adequate algorithms and supercomputers," Alexandros Stamatakis, Professor for High-performance Computing in Life Sciences of Karlsruhe Institute of Technology and Head of the Research Group "Scientific Computing" of the Heidelberg Institute for Theoretical Studies (HITS), explains. "Today, modern sequence analyses supply comprehensive genetic data for numerous species. So far, however, computer programs even on supercomputers have been overstrained by the task of generating evolutionary knowledge from these large and complex data volumes."

**For further information,
please contact:**

Kosta Schinarakis
PKM – Science Scout
Phone: +49 721 608 41956
Fax: +49 721 608 43658
E-mail: schinarakis@kit.edu

Although supercomputers have meanwhile been equipped with thousands of processors, the software for analyzing trees of life was limited to about 500 processors. "We therefore had to redesign and redevelop the communication scheme between the program components on various processors," Stamatakis says. The new approach accelerated the software by a factor of 3 and now allows to efficiently distribute the computations to 4000 processors. Computer scientists speak of a parallelization of algorithms. "Instead of 24 months, we are now waiting one month for the results."

The computation of trees of life is an extremely computation-intensive (NP-hard) mathematical problem. "For 50 species, more than 10^{76} possible trees of life exist. Of these, the right one has to be found," Andre J. Aberer, doctoral student of KIT, explains. He works at the HITS and performed the computer analyses. "For comparison: About 10^{78} atoms exist in the universe." First, the algorithms coarsely filter out the improbable evolution scenarios. Then, based on the data of 14,000 genes of 48 representative bird species, the evolutionary tree of life is calculated, which plausibly explains the data. The new parallel software was run on the high-performance computer "SuperMUC" of the Leibniz Computing Center of the Bavarian Academy of Sciences and at two computing centers in the US. The computing power needed corresponds to a computing time of 300 years on a single processor.

"The methods we developed for the computation of the tree of life can be applied to all types of creatures," Stamatakis says. They already allowed for a comprehensive study of the tree of life of insects with 144 species, which was published recently in the Science journal. In addition, it is possible to reproduce the origin and abundance of viruses and bacteria in order to better fight pathogens. Analysis of the genetic relationship of Australian poisonous snakes helped identify the still lacking antidotes for some snake species.

Apart from the tree of life and the evolution of birds, the following new findings relating to birds are reported simultaneously by a total of 23 scientific publications in Science and other expert journals:

- Genetic fundamentals of biodiversity.
- Genetic fundamentals of the brain regions controlling the evolution of birdsong.
- The loss of teeth of birds about 100 million years ago.
- The relationship between dinosaurs and birds.

- The evolution of colored feathers.

The study of the evolution of birds was made by the "Avian Phylogenomics Consortium", consisting of 200 scientists of 80 institutes in 20 countries. The project was coordinated by Guojie Zhang, BGI, China, Erich D. Jarvis, Duke University, USA, and M. Thomas P. Gilbert, Natural History Museum, Denmark. Tandy Warnow of the University of Illinois and Alexandros Stamatakis coordinated the computer analyses. Alexandros Stamatakis and Andre J. Aberer are the only German co-authors. The study represents the most comprehensive genomic analysis of a vertebrate class and covers ducks falcons, woodpeckers, and other representatives of all branches of modern birds, among others. All data and methods will now be made available cost-free to researchers worldwide for further studies.

Website of the research group of Alexandros Stamatakis:

www.exelixis-lab.org

http://www.informatik.kit.edu/309_6333.php

KIT interview of Alexandros Stamatakis:

http://www.youtube.com/watch?v=_6Laa58M-vk

The Karlsruhe Institute of Technology (KIT) is a public corporation according to the legislation of the state of Baden-Württemberg. It fulfills the mission of a university and the mission of a national research center of the Helmholtz Association. Research activities focus on energy, the natural and built environment as well as on society and technology and cover the whole range extending from fundamental aspects to application. With about 9400 employees, including more than 6000 staff members in the science and education sector, and 24,500 students, KIT is one of the biggest research and education institutions in Europe. Work of KIT is based on the knowledge triangle of research, teaching, and innovation.

This press release is available on the internet at www.kit.edu.