

Algorithmen in JAguc

Sebastian Wild
s_wild@cs.uni-kl.de

Markus E. Nebel
nebel@cs.uni-kl.de

Fachbereich Informatik — TU Kaiserslautern

Im Rahmen eines Bachelor-Praktikums entstand JAGUC, ein Tool zur Analyse der Biodiversität von Ökosystemen auf Basis von Sequenzdaten, üblicherweise SSU rRNA (small subunit ribosomal RNA). Dabei gruppiert man ähnliche Sequenzen, sodass jeder Cluster für eine Art steht (operational taxonomic units). Für die Bestimmung der Cluster benutzt man paarweise Alignments aller Sequenzen und wendet darauf eine hierarchische bottom-up Cluster-Methode an, z. B. UPGMA.

Da eine typische Probe mehrere hunderttausend RNA-Sequenzen enthält, stoßen naive Implementierungen schnell an Grenzen. In diesem Vortrag werde ich vorstellen, welche Optimierungen der Algorithmen und Datenstrukturen wir für JAGUC entworfen haben und zeigen, dass sich damit Eingaben gemäß der typischen Verteilung für SSU rRNA effizient verarbeiten lassen.