

Übungsaufgaben

7 Sequenzanalyse mit BioJava

Grundlagen der Sequenzanalyse

Einführung in BioJava

Datenbanksuche

Lesen und Schreiben von Sequenzen

Eigenschaften von Sequenzen in

BioJava

Sequence Alignment

Multiple Sequence Alignment

BLAST

Next-Generation Sequencing Data

Analysis

Übungsaufgaben

Index

Einleitung

Die Biologie mit all ihren vielfältigen Teilgebieten wie Botanik, Zell- und Molekularbiologie, Genetik oder Humanbiologie ist als ein reges Forschungsfeld in stetem Wandel. Sie ist gleichzeitig, mehr als andere Domänen, ständigen Wechselwirkungen mit angrenzenden Bereichen wie Chemie, Physik oder Medizin und seit einigen Jahrzehnten auch technischeren Bereichen wie der Informatik unterworfen. Gerade durch die Interaktion mit der Informatik hat sich ein eigenes Fachgebiet

entwickelt, die Bioinformatik. In der Praxis führt das dazu, dass bei vielen Naturwissenschaftlern inzwischen weitreichende Programmiererfahrungen vorhanden sind, die deutlich über Tabellenkalkulationsprogramme hinausgehen und nicht nur die Thematik der Handhabung wissenschaftlicher Daten umfasst.

Zugleich hat sich in den letzten Jahren die Biologie gemeinsam mit der Medizin, der Chemie sowie den Pflege- und den Agrarwissenschaften zu einem spannenden, interdisziplinären Feld entwickelt: den Lebenswissenschaften (engl. *Life Sciences*). In diesem Gebiet fließen die Grenzen der »klassischen« Disziplinen zusehends ineinander. Ein Buch, das sich lediglich mit der Schnittmenge zwischen Biologie und Informatik beschäftigt, erscheint uns daher nicht weitreichend genug. Wir haben uns das Ziel gesteckt, den Bogen hier wesentlich

weiter zu spannen.

Dementsprechend geht es in diesem Buch primär darum, angewandte Probleme der Lebenswissenschaften mithilfe der Informatik zu lösen. Es wird dabei nur so weit auf biologische oder medizinische Hintergründe eingegangen, wie es für das Verständnis des Buchs nötig ist. Das Vorgehen wird mit viel Beispielcode, teilweise im Stil eines Kochbuchs, dargestellt. Wann immer möglich, sollen Übungsaufgaben und Praxisprojekte tiefer in die Materie führen.

In diesem Buch werden vor allem folgende allgemeine Themenbereiche behandelt:

- *Big Data*: die Verarbeitung einer großen Menge sich ständig ändernder Daten.
- Das Heranziehen und Adaptieren effizienter, robuster und zuverlässiger Lösungen (insbesondere *Heuristiken*

und *Algorithmen*) für bestehende Probleme.

- Informationen über *State-of-the-Art-Lösungen*, *-Bibliotheken* und *-Algorithmen*.

Somit werden alle wichtigen Grundlagen für Studierende und Praktiker in den Lebenswissenschaften, der Biologie, der Bioinformatik und anderen Fächern behandelt: Wie werden Daten und Informationen verwaltet (sogenanntes *Data Management*), wie werden sie über Algorithmen und externe Bibliotheken verarbeitet, und wie werden damit schlussendlich Probleme im Sinne der zugrunde liegenden wissenschaftlichen Fragestellung gelöst? Unser Ziel ist es, Ihnen eine Vorstellung von den verfügbaren und etablierten Softwarelösungen und Konzepten sowie von generellen Herangehensweisen bei datenbasierten Problemstellungen zu