## Einführung in das Programmieren ILV

## Input/Ouput Biopython

**SeqIO Modul**

Ein benutzerfreundliches Modul für das Einlesen, Schreiben und Verwalten von Sequenzdaten. Ermöglicht die Verarbeitung vieler biologischer Dateiformate (FASTA, GenBank…) und stellt diese als Objekte bereit.

**Sequenzen einlesen**

Mit der Funktion **SeqIO.parse()** können große Mengen von biologischen Sequenzen aus Dateien (z.B. FASTA oder GenBank) eingelesen und analysiert werden. Diese Sequenzen werden als seq\_record Objekte, Iteratoren oder als Liste gespeichert. Diese Funktion erwartet zwei Argumente: den Dateinamen und das Dateiformat (in Kleinbuchstaben). Die Angabe des Formats ist zwingend erforderlich, da die Formate nicht automatisch erkannt werden.

**Unterschied read vs parse**

**SeqIO.parse()** wird verwendet für Dateien mit mehrere Datensätze.

**SeqIO.read()** für Dateien mit nur einem Datensatz.

**Daten Extrahieren**

Bei GeneBank Dateien kann man sich den Inhalt des Annotations Wörterbuchs mit **.annotations** ausgeben lassen. Zusätzlich können spezifische Schlüssel („Keys“) oder Werte gezielt abgerufen werden.

**Sequenzen Schreiben**

Die Funktion **SeqIO.write()** ermöglicht das Speicher von SeqRecord-Objecten, Iteratoren oder Listen in eine Datei (Auch hierfür werden alle gängigen Dateiformate unterstützt). Dies ist besonders nützlich, wenn bestimmte bioinformatische Werkzeuge nur spezifische Formate akzeptieren.

**Sequenzen Konvertieren**

Es besteht die Möglichkeit, Sequenzdateien von einem Format in ein anderes Format zu Konvertieren. Die Funktion lautet **seqIO.convert()**, diese erwartet vier Argumente: den Ursprünglicher Dateiname, das Format sowie der neue Dateiname und das Ziel-Format.

**Zusammenfassung**

SeqIO ermöglicht das benutzerfreundliche Einlesen, Verwalten und Konvertieren von Sequenzdaten. Unterstützt alle gängigen Dateiformate in der Bioinformatik.