Bioinformatics with Python Cookbook

Third Edition

Bioinformatics is an active research field that uses a range of simple to advanced computations to extract valuable information from biological data, and this book will show you how to manage these tasks using Python.

This updated edition of the *Bioinformatics with Python Cookbook* begins with a quick overview of the various tools and libraries in the Python ecosystem that will help you convert, analyze, and visualize biological datasets. As you advance through the chapters, you'll cover key techniques for next-generation sequencing, genomics, metagenomics, population genetics, phylogenetics, and proteomics with the help of real-world examples. You'll learn how to work with important pipeline systems, such as Galaxy servers and Snakemake, and understand the various modules in Python for functional programming. This book will also help you explore topics such as SNP discovery using statistical approaches under high-performance computing frameworks, including Dask and the application of machine learning algorithms to bioinformatics.

By the end of this bioinformatics Python book, you'll be equipped with the knowledge to implement the latest programming techniques and frameworks, empowering you to deal with bioinformatics data on every kind of scale.

Things you will learn:

- Become well versed in data processing libraries such as NumPy, pandas, Arrow, and zarr in the context of bioinformatic analysis
- Interact with genomic databases
- Solve real-world problems in the fields of population genetics, phylogenetics, and proteomics
- Build bioinformatics pipelines using a Galaxy server and Snakemake
- Work with functools and itertools for functional programming
- Perform parallel processing with Dask on biological data
- Explore PCA techniques with scikit-learn

ISBN 978-1-80323-642-1
9 781803 236421

packt