Rifki Fahrial Zainal, S.T., M.Kom

BIOINFORMATIKA KOMPUTASIONAL

Dasar Biologi, Representasi Data, dan Pemrograman Python



BIOINFORMATIKA KOMPUTASIONAL

Dasar Biologi, Representasi Data, dan Pemrograman Python

Rifki Fahrial Zainal, S.T., M.Kom

Penerbit Widina

BIOINFORMATIKA KOMPUTASIONAL:
DASAR BIOLOGI, DATA, DAN PEMROGRAMAN PYTHON

Penulis:
Rifki Fahrizal Zainal, S.T., M.Kom.

Desain Cover:
Helmaria Ulfa

Sumber Ilustrasi:
www.freepik.com

Tata Letak:
Handarini Rohana

Proofreader:
Aas Masruroh

ISBN:
978-634-317-538-4

Cetakan Pertama:
November, 2026

Hak Cipta Dilindungi Oleh Undang-Undang
by Penerbit Widina Media Utama

Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.

PENERBIT:
WIDINA MEDIA UTAMA
Komplek Puri Melia Asri Blok C3 No. 17 Desa Bojong Emas
Kec. Solokan Jeruk Kabupaten Bandung, Provinsi Jawa Barat

Anggota IKAPI No. 519/JBA/2025
Website: www.penerbitwidina.com
Instagram: @penerbitwidina
Telepon (022) 87355370

KATA PENGANTAR

Puji syukur ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga buku **"Bioinformatika Komputasional: Dasar Biologi, Data, dan Pemrograman Python"** ini dapat diselesaikan. Buku ini merupakan **Buku 1 dari Seri Bioinformatika Komputasional** yang disusun untuk memberikan landasan awal bagi mahasiswa, khususnya mahasiswa Informatika dan Ilmu Komputer, dalam memahami keterkaitan antara ilmu biologi dan pendekatan komputasional.

Perkembangan teknologi biologi modern telah menghasilkan data dalam jumlah yang semakin besar dan beragam. DNA, RNA, protein, serta berbagai informasi biologis kini dapat direpresentasikan sebagai data digital yang dapat diolah menggunakan algoritma dan pemrograman. Kondisi tersebut menjadikan bioinformatika sebagai bidang interdisipliner yang mempertemukan biologi dengan ilmu komputer, matematika, statistika, dan data science.

Buku ini disusun dengan pendekatan yang menempatkan **masalah biologis, data, representasi komputasional, algoritma, implementasi Python, dan interpretasi** sebagai satu rangkaian pembelajaran. Pendekatan tersebut diharapkan membantu mahasiswa berlatar belakang Informatika memahami konsep biologi yang diperlukan tanpa harus terlebih dahulu memiliki pengetahuan biologi yang mendalam.

Materi dalam buku terdiri atas lima bab. Pembahasan dimulai dengan pengenalan bioinformatika komputasional, dilanjutkan dengan dasar biologi molekuler untuk Informatika, representasi data biologi dan biomedis, database biologi dan *biomedical data resources*, serta pengolahan *biological sequence* menggunakan Python. Pada bagian implementasi, pembaca diperkenalkan pada berbagai operasi sequence, seperti validasi DNA, komposisi nukleotida, GC content, *complement*, *reverse complement*, transkripsi, translasi, *Open Reading Frame* (ORF), motif, k-mer, serta penggunaan Biopython.

Python digunakan sebagai bahasa pemrograman utama karena memiliki sintaks yang relatif mudah dipahami serta didukung oleh ekosistem yang luas untuk *scientific computing* dan bioinformatika. Namun, tujuan pembelajaran tidak sekadar menguasai sintaks atau pustaka tertentu. Pembaca diarahkan untuk mengembangkan *computational thinking*, yaitu kemampuan memahami masalah, menentukan representasi data, merancang proses komputasional, serta menginterpretasikan hasil dalam konteks biologis.

Sebagai bagian pertama dari **Seri Bioinformatika Komputasional**, buku ini menjadi fondasi bagi pembahasan pada buku-buku berikutnya. Setelah memahami bagaimana data biologis direpresentasikan, diperoleh, dan diproses, pembaca dapat melanjutkan ke analisis sequence, data biomedis, Artificial Intelligence untuk bioinformatika, serta *research workflow* dan proyek bioinformatika komputasional.

Penulis menyadari bahwa buku ini masih memiliki ruang untuk pengembangan dan penyempurnaan. Oleh karena itu, kritik dan saran dari pembaca sangat diharapkan untuk perbaikan pada edisi berikutnya. Penulis juga menyampaikan terima kasih kepada semua pihak yang telah memberikan dukungan dalam proses penyusunan buku ini.

Akhir kata, semoga buku ini dapat memberikan manfaat bagi mahasiswa, dosen, peneliti, dan pembaca yang ingin mempelajari bioinformatika dari perspektif komputasional serta menjadi landasan untuk mempelajari bidang bioinformatika yang lebih lanjut.

Surabaya, 2026
Penulis

DAFTAR ISI

KATA PENGANTAR ..... **iii**
DAFTAR ISI ..... **v**
DAFTAR GAMBAR ..... **xii**
DAFTAR TABEL ..... **xiv**
BAB 1 PENGANTAR BIOINFORMATIKA KOMPUTASIONAL **1**
1. Pendahuluan 2
2. Apa Itu Bioinformatika? 3
3. Bioinformatika sebagai Ilmu Interdisipliner 5
4. Dari Eksperimen Biologi menjadi Data 6
5. Biological Data sebagai Computational Data 7
6. Computational Thinking dalam Bioinformatika 9
7. Computational Bioinformatics Workflow 10
8. Peran Algoritma dalam Bioinformatika 11
9. Mengapa Python? 12
10. Eksperimen Python Pertama: DNA sebagai Data 12
11. Dari Pemrograman Konvensional menuju Machine Learning 13
12. Artificial Intelligence dalam Bioinformatika Modern 14
13. Dari Single-Modality menuju Multimodal Bioinformatics 15
14. Bioinformatika, Computational Biology, Biomedical Informatics, dan Health Informatics 16
15. Studi Kasus: dari DNA hingga Artificial Intelligence 16
16. Hands-on Laboratory: Exploring DNA with Python 18
17. Mini Project: Biological Data Explorer 18

BAB 2 DASAR BIOLOGI MOLEKULER UNTUK INFORMATIKA **21**
1. Pendahuluan 21
2. Sel sebagai Sistem Biologis 22
3. DNA sebagai Penyumpan Informasi Genetik 23
4. Complement dan Reverse Complement 24
5. Gen 26
6. Genome 26
7. Kromosom 27
8. RNA 28
9. Jenis-Jenis RNA 28

10. Transcription 29
11. Protein 29
12. Kodon 30
13. Translation 30
14. Central Dogma 31
15. Gene Expression 32
16. Mutation 33
17. Mutation sebagai String Comparison Problem 34
18. Reading Frame 34
19. Start Codon dan Stop Codon 35
20. Open Reading Frame 36
21. Dari Biological Molecule menuju Computational Object 36
22. Sequence sebagai Struktur Data 37
23. Validasi DNA Sequence 37
24. Menghitung Komposisi Nucleotida 38
25. GC Content 38
26. Biological Sequence sebagai Feature Vector 39
27. One-Hot Encoding Sequence 39
28. Dari Sequence ke Numerical Representaion 40
29. Mini Studi Kasus: Deteksi Perubahan Sequence 41
30. Hands-on Laboratory: DNA, RNA dan Protein Representation 41
31. Mini-Project: DNA Sequence Analyzer 42
32. Tantangan Pemrograman 42

BAB 3 REPRESENTASI DATA BIOLOGIS DAN BIOMEDIS 43

1. Pendahuluan 44
2. Mengapa Representasi Data Penting? 45
3. Biological Object dan Computational Object 45
4. Sequence Data 46
5. Sequence sebagai String 46
6. Numerical Encoding Sequence 47
7. One-Hot Encoding 47
8. k-mer Representation 48
9. Tabular Data 49
10. Vector Representation 49
11. Feature 50
12. Label dan Target 50

13. Metadata 51
14. Matrix Data 51
15. High-Dimensional Biological Data 52
16. Dimensionality 53
17. Feature Space 53
18. Time-Series Data 53
19. Physiological Time Series 53
20. Sampling Frequency 54
21. Windowing 54
22. Multichannel Time Series 55
23. Signal Data 56
24. Raw Signal dan Feature 56
25. Image Data 57
26. Pixel dan Voxel 57
27. Tensor Representaion 57
28. Graph Data 57
29. Biological Network 58
30. Data Multimodal 58
31. Representasi Multimodal 58
32. Data Fusion 59
33. Structured dan Unstructured Data 60
34. Static dan Temporal Representation 60
35. Raw Data dan Derived Data 61
36. Feature Engineering 61
37. Representation Learning 62
38. Missing Data 62
39. Noise 63
40. Normalization 63
41. Standardization 63
42. Tipe Data dan Presisi 64
43. Representasi Data dan Pemilihan Algoritma 64
44. Studi Kasus: Satu Pasien, Banyak Representasi 65
45. Hands-on Laboratory: Mengenali Representasi Data 65
46. Praktikum Python dengan NumPy 66
47. Mini Project: Biological Data Representation Explorer 67
48. Tantangan Mini Project 67

BAB 4 DATABASE BIOLOGI DAN BIOMEDICAL DATA RESOURCES 69
1. Pendahuluan 69
2. Apa yang Dimaksud dengan Biological Database? 70
3. Mengapa Biological Database Diperlukan? 71
4. Dari Data ke Informasi 72
5. Primary Database 73
6. Secondary Database 73
7. Specialized Database 74
8. NCBI 75
9. Ekosistem NCBI 75
10. Entrez 76
11. Query dalam Biological Database 76
12. Search Strategy 77
13. Identifier dalam Bioinformatika 78
14. Accession Number 78
15. Versioned Identifier 78
16. GenBank 79
17. Struktur Konseptual GenBank Record 79
18. FASTA dan GenBank: Apa Bedanya? 79
19. Metadata pada Sequence Database 80
20. Annotation 80
21. Reference Sequence 81
22. Gene Database 81
23. Protein Database 82
24. UniProt 82
25. Protein Sequence 81
26. Database Cross-Reference 83
27. Protein Data Bank 83
28. Mengapa Struktur 3D Penting? 85
29. PDB ID 85
30. Representasi Struktur Protein 85
31. Dari Sequence ke Structure 86
32. PubMed sebagai Literature Resource 86
33. Data Retrieval Manual 87
34. Programmatic Data Retrieval 87
35. API dalam Bioinformatika 88

36. NCBI E-utilities 88
37. Biopython dan Entrez 89
38. Mencari Record dengan Python 89
39. Memahami Hasil Search 90
40. Mengambil Sequence 91
41. Mengambil GenBank Record 92
42. Biopython SrqIO 93
43. Membaca FASTA 93
44. Membaca Banyak Sequence 93
45. SeqRecord 94
46. Data Parsing 94
47. Manual Parsing versus Library 94
48. Batch Data Retrieval 95
49. Rate Limiting 95
50. Caching 96
51. Data Provenance 94
52. Reproducibility 97
53. Database Versioning 97
54. Data Quality 97
55. Curated dan Unreviewed Data 98
56. Protein Structure Resources 98
57. Experimental Structure dan Computed Structure 98
58. Biological Data Integration 98
59. Identifier Mapping 99
60. Database Integration sebagai Graph 99
61. Data Retrieval Workflow 101
62. Studi Kasus: Menelusuri Gene dari DNA Struktur Protein 101
63. Hands-on Laboratory: Eksplorasi NCBI 102
64. Hands-on Laboratory II: FASTA Retrieval 102
65. Hands-on Laboratory III: Programmatic NCBI Search 103
66. Hands-on Laboratory IV: Sequence Retrieval 103
67. Hands-on Laboratory V: Menggabungkan Entrez dan SeqIO 104
68. Menghitung GC Content Data Online 104
69. Error Handling 105
70. Validation setelah Retrieval 105
71. Etika Biological Data 106

72. Public Data tidak selalu Berarti Tanpa Resiko 106
73. Mini-Project: Biological Database Explorer 106
74. Tantangan Mini-Project 107
75. Integrative Challenge 107
BAB 5 BIOLOGICAL SEQUENCE PROCESSING DENGAN PYTHON 109
1. Pendahuluan 109
2. Biological Sequence sebagai String 110
3. Sequence Order 113
4. DNA Sequence 112
5. RNA Sequence 112
6. Protein Sequence 112
7. Sequence Validation 113
8. Membuat Fungsi Validasi DNA 113
9. Normalisasi Input Sequence 113
10. Sequence Length 114
11. Nucleotide Composition 114
12. Nucleotide Frequency 115
13. GC Content 115
14. Mengapa GC Content Penting? 115
15. AT Content 116
16. Complement DNA 116
17. Reverse Complement 117
18. Membuat Fungsi Reverse Complement 117
19. Transcription 118
20. Fungsi Transcription 118
21. Codon 118
22. Inclomplete Codon 119
23. Start Codon 119
24. Stop Codon 119
25. Translation 120
26. Stop Translation 120
27. Reading Frame 121
28. Tiga Reading Frame pada satu Strand 121
29. Open Reading Frame 121
30. Pencarian ORF Sederhana 122
31. Motif 122

32. Menghitung Motif 123
33. Overlapping Motif 123
34. Pattern Matching 124
35. k-mer 124
36. k-mer Frequency 125
37. Sequence Similarity Sederhana 125
38. Hamming Distance 125
39. Keterbatasan Direct Comparison 126
40. Pengenalan Bio.Seq 125
41. Complement dengan BioPython 127
42. Reverse Complement dengan BioPython 127
43. Transcription dengan Biopython 127
44. Translation dengan Biopython 127
45. Stop Codon dalam Translation 128
46. Seq dan String 128
47. SeqRecord 128
48. Designing Reusalbe Functions 129
49. Fungsi Analisis DNA 130
50. Structured Output 130
51. Batch Sequence Processing 130
52. Menggunakan PANDAS 131
53. Feature Extraction dari Sequence 131
54. Hands-on Laboratory I: DNA Sequence Analyzer 131
55. Hands-on Laboratory II: Motif Finder 132
56. Hands-on Laboratory III: k-mer Explorer 132
57. Hands-on Laboratory IV: Compare Two Sequences 132
58. Hands-on Laboratory V: From Scratch vs Biopython 133
59. Mini-Project: Sequence Analysis Toolkit 133
60. Tantangan Mini-Project 134
61. Integrative Challenge 134
DAFTAR PUSTAKA 135

DAFTAR GAMBAR

Gambar 1 Paradigma Bioinformatika....3
Gambar 2 Interdisipliner dalam Bioinformatics....6
Gambar 3 Formulasi Permasalahan Algoritmik....7
Gambar 4 Contoh Graph Data....8
Gambar 5 Proses Decomposition DNA....9
Gambar 6 Workflow dari Computational Bioinformatics....10
Gambar 7 Konsep Bioinformatika dengan AI....14
Gambar 8 Konsep Multimodal....15
Gambar 9 Workflow Penelitian....16
Gambar 10 Model Hasil Penelitian....17
Gambar 11 Model Gen Sederhana....24
Gambar 12 Struktur Hierarkis DNA....26
Gambar 13 Konsep Transformasi Informasi....30
Gambar 14 Representasi Coding Region....33
Gambar 15 Proses Encoding....38
Gambar 16 Proses Perubahan Biologis....42
Gambar 17 Proses Windowing....53
Gambar 18 Proses Pembacaan Sinyal ECG....54
Gambar 19 Graph Protein....55
Gambar 20 Early Fusion....57
Gambar 21 Proses Perubahan Data Mentah ke Fitur Utama....59
Gambar 22 Proses Pembelajaran pada
Machine Learning (atas) dan Deep Learning (bawah)....60
Gambar 23 Proses Data untuk Prediksi Penyakit....63
Gambar 24 Hubungan antar Objyek Biologis....68
Gambar 25 Kombinasi Biological Database....69
Gambar 26 Proses Kombinasi Biological Database....70
Gambar 27 Jaringan Informasi Biologis....71
Gambar 28 Proses Perubahan Primary Database....72
Gambar 29 Proses Kerja Entrez....74
Gambar 30 Konseptual Gene Database....80
Gambar 31 Konseptual UniProt....81
Gambar 32 Konseptual Database Cross-Reference....82

Gambar 33 Level Bioinformatika 84
Gambar 34 Penanganan Level Database 84
Gambar 35 Proses Pencarian Literatur 84
Gambar 36 Workflow Data Retrieval 85
Gambar 37 Proses Programmatic Access 85
Gambar 38 Proses Komunikasi API 86
Gambar 39 Konsep Dasar E-utilities 86
Gambar 40 Proses Kerja Pencarian di Python 88
Gambar 41 Tahapan Pengambilan Record 88
Gambar 42 Contoh Data Parsing 92
Gambar 43 Contoh Integrasi Database 97
Gambar 44 Graph Database UniProt 98
Gambar 45 Workflow Data Retrieval 99
Gambar 46 Workflow Pembelajaran Human Gene 99
Gambar 47 Workflow validasi 103

DAFTAR TABEL

Tabel 1 Representasi Obyek Biologis 4
Tabel 2 Pendekatan Komputasional Permasalahan Bioinformatika 11
Tabel 3 Analogi komoputasional untuk Sistem Biologis 21
Tabel 4 Contoh Matriks Ekspresi Gen 30
Tabel 5 Pemetaan Biological Object ke dalam Computational Object 43
Tabel 6 Contoh Data Eksperimen 47
Tabel 7 Contoh Ekspresi Gen 49
Tabel 8 Detail Pendekatan Fusion 57
Tabel 9 Data Hilang 60
Tabel 10 Representasi Data dan Algoritma 62
Tabel 11 Contoh Data Pasien 63
Tabel 12 Perbedaan FASTA dan GenBank 78

BAB 1

PENGANTAR BIOINFORMATIKA KOMPUTASIONAL

CAPAIAN PEMBELAJARAN

Setelah mempelajari bab ini, pembaca diharapkan mampu:

1. Menjelaskan pengertian, ruang lingkup, dan perkembangan bioinformatika
2. Menjelaskan hubungan antara biologi, ilmu komputer, matematika, statistika, dan kecerdasan artifisial dalam bioinformatika
3. Mengidentifikasi berbagai jenis data biologi dan biomedis, serta representasi komputasionalnya
4. Membedakan bioinformatika, *computational biology*, *biomedical informatics*, dan *health informatics*
5. Menjelaskan tahapan umum *computational bioinformatics workflow*
6. Mengidentifikasi peran algoritma, pemrograman *Python*, *Machine Learning*, dan *Artificial Intelligence* dalam analisis data biologis
7. Merancang penyelesaian sederhana terhadap permasalahan bioinfromatika menggunakan pendekatan *computational thinking*

BAB 2

DASAR BIOLOGI MOLEKULER UNTUK INFORMATIKA

CAPAIAN PEMBELAJARAN

Setelah mempelajari bab ini, pembaca diharapkan mampu:

1. Menjelaskan struktur data sel sebagai unit kehidupan
2. Menjelaskan struktur dan fungsi DNA, RNA, dan protein
3. Memahami konsep gen, genom, kromosom, dan ekspresi gen
4. Menjelaskan proses *transcription* dan *translation*
5. Membedakan DNA, RNA, dan protein dari perspektif biologis dan komputasional
6. Merepresentasikan DNA, RNA, dan protein sebagai data komputasional
7. Menjelaskan hubungan *central dogma* dengan aliran informasi biologis
8. Melakukan manipulasi *sequence* sederhana menggunakan *Python*

BAB 3

REPRESENTASI DATA BIOLOGIS DAN BIOMEDIS

CAPAIAN PEMBELAJARAN

Setelah mempelajari bab ini, pembaca diharapkan mampu:

1. Menjelaskan konsep representasi data dalam bioinformatika
2. Membedakan berbagai jenis data biologis dan biomedis
3. Merepresentasikan *sequence*, tabular data, matriks, *time series*, sinyal, *image*, *graph*, dan multimodal data
4. Menjelaskan hubungan antara *biological object* dan *computational representation*
5. Memahami konsep *feature*, label, metadata, dimensionality, dan *feature space*
6. Menjelaskan pentingnya *encoding* dalam *machine learning bioinformatics*
7. Memilih bentuk representasi data yang sesuai berdasarkan karakteristik permasalahan bioinformatika
8. Melakukan eksplorasi sederhana berbagai bentuk data menggunakan *Python*

BAB 4

DATABASE BIOLOGI DAN *BIOMEDICAL DATA RESOURCES*

CAPAIAN PEMBELAJARAN

Setelah mempelajari bab ini, pembaca diharapkan mampu:

1. Menjelaskan peran database dalam bioinformatika
2. Membedakan *primary database*, *secondary database*, dan *specialized database*
3. Menjelaskan fungsi NCB dan berbagai sumber daya yang tersedia di dalamnya.
4. Menjelaskan fungsi *GenBank* sebagai *database sequence nucleotide*
5. Menjelaskan penggunaan *UniProt* sebagai sumber informasi protein
6. Menjelaskan penggunaan *Protein Data Bank* untuk memperoleh data struktur tiga dimensi biomolekul
7. Memahami konsep *accession number identifier*, *annotation*, *metadata*, dan *database cross-reference*
8. Melakukan pencarian data biologis menggunakan *query*
9. Memahami system *Entrez* dan *programmatic data retrieval*
10. Mengambil dan memproses data biologis menggunakan *Python* dan *Biopython*
11. Menjelaskan konsep *reproducibility*, *provenance*, dan etika penggunaan data biologis

BAB 5

BIOLOGICAL SEQUENCE PROCESSING DENGAN *PYTHON*

CAPAIAN PEMBELAJARAN

Setelah mempelajari bab ini, pembaca diharapkan mampu:

1. Menjelaskan *biological sequence* sebagai obyek komputasional
2. Membedakan *sequence* DNA, RNA, dan protein
3. Melakukan validasi *sequence* biologis
4. Menghitung panjang dan komposisi *nucleotide*
5. Menghitung GC *Content*
6. Menghasilkan *complement* dan *reverse complement*
7. Melakukan *transcription* dari DNA menjadi RNA
8. Melakukan *translation* dari RNA menjadi protein
9. Menjelaskan *codon*, *start codon*, *stop codon* dan *reading frame*
10. Melakukan pencarian motif pada *sequence*
11. Memahami konsep k-mer dan penggunaannya
12. Membangun fungsi *reusable* untuk *sequence processing*
13. Membandingkan implementasi manual dengan *Biopython*
14. Merancang *pipeline* analisis *sequence* sederhana menggunakan *Python*

DAFTAR PUSTAKA

Alberto Anguita, Miguel García-Remesal, Diana de la Iglesia, Victor Maojo (2013). NCBI2RDF: Enabling Full RDF-Based Access to NCBI Databases. BioMed Research International. https://doi.org/10.1155/2013/983805.

Anne Morgat, Rolf Apweiler, Maria Jesus Martin, Claire O'Donovan, Michele Magrane, Yasmin Alam-Faruque, et al. (2011). Ongoing and future developments at the Universal Protein Resource. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKQ1020.

Barbara A. Rapp, David L. Wheeler (2005). Bioinformatics resources from the National Center for Biotechnology Information: An integrated foundation for discovery. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/ASI.20142.

Barbara A. Rapp, David L. Wheeler (2005). Bioinformatics resources from the National Center for Biotechnology Information: An integrated foundation for discovery: Research Articles. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/ASI.V56:5.

Berk Ekmekci, Charles E. McAnany, Cameron Mura (2016). An Introduction to Programming for Bioscientists: A Python-Based Primer. PLOS Computational Biology. https://doi.org/10.1371/JOURNAL.PCBI.1004867.

Boris A. Vinatzer, Lenwood S. Heath, Hussain M. J. Almohri, Michael J. Stulberg, Christopher Lowe, Song Li (2019). Cyberbiosecurity Challenges of Pathogen Genome Databases. Frontiers in Bioengineering and Biotechnology. https://doi.org/10.3389/FBIOE.2019.00106.

C.J.H. Mann (2010). Python for Bioinformatics. Kybernetes. https://doi.org/10.1108/K.2010.06739HAE.004.

Christian Brueffer, Tiago Antao, Peter J. A. Cock, Eric Talevich, Michiel J. L. de Hoon, Wibowo Arindrarto, et al. (2016). Biopython project update 2016. F1000Research. https://doi.org/10.7490/F1000RESEARCH.1112611.1.

Chuming Chen, Hongzhan Huang, Cathy H. Wu (2017). Protein bioinformatics databases and resources. Methods of Molecular Biology. https://doi.org/10.1007/978-1-4939-6783-4_1.

David J. Winter (2017). rentrez: An R package for the NCBI eUtils API. R Journal. https://doi.org/10.7287/PEERJ.PREPRINTS.3179V2.

Dong Xu, Zhuchou Lu, Kangming Jin, Wenmin Qiu, Guirong Qiao, Xiaojiao Han, et al. (2020). SPDE: A Multi-functional Software for Sequence Processing and Data Extraction. bioRxiv. https://doi.org/10.1101/2020.11.08.373720.

Eric W Sayers, Evan Bolton, Anna M Fine, Christopher E. Kelly, Sunghwan Kim, Melissa Landrum, et al. (2025). Database resources of the National Center for Biotechnology Information in 2026. Nucleic Acids Research. https://doi.org/10.1093/nar/gkaf1060.

Eric W Sayers, Jeffrey Beck, Evan Bolton, J. Rodney Brister, Jessica Chan, Ryan Connor, et al. (2024). Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Research. https://doi.org/10.1093/nar/gkae979.

Eric W Sayers, Mark Cavanaugh, Linda Frisse, Kim D Pruitt, Valerie A Schneider, Beverly A Underwood, et al. (2024). GenBank 2025 update. Nucleic Acids Research. https://doi.org/10.1093/nar/gkae1114.

Eric W. Sayers, Evan E Bolton, J. Rodney Brister, Kathi Canese, Jessica Chan, Donald C. Comeau, et al. (2022). Database resources of the National Center for Biotechnology Information in 2023. Nucleic Acids Research. https://doi.org/10.1093/nar/gkac1032.

Eric W. Sayers, Mark Cavanaugh, Karen Clark, Kim D Pruitt, Stephen T Sherry, Linda Yankie, et al. (2023). GenBank 2024 Update. Nucleic Acids Research. https://doi.org/10.1093/nar/gkad903.

Garth Brown, Vichet Hem, Kenneth S. Katz, Michael Ovetsky, Craig Wallin, Olga Ermolaeva, et al. (2015). Gene: a gene-centered information resource at NCBI. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKU1055.

Hongzhan Huang, Peter B. McGarvey, Baris E. Suzek, Raja Mazumder, Jian Zhang, Yongxing Chen, et al. (2011). A comprehensive protein-centric ID mapping service for molecular data integration. Bioinformatics. https://doi.org/10.1093/BIOINFORMATICS/BTR101.

Indah Juwita Sari, Wanwipa Vongsangnak, Pongprapan Pongsophon (2022). The Effect of Bioinformatics Module on Molecular Genetics Concepts on Senior High School Students' Computational Thinking Skills. Shanlax International Journal of Education. https://doi.org/10.34293/education.v10i2.4680.

Izdebski, Adam, Olszewski, Jan, Szczurek, Ewa (2025). seqme: a Python library for evaluating biological sequence design. https://doi.org/10.48550/arxiv.2511.04239.

Jie Dong, Jie Dong, Zhi-Jiang Yao, Lin Zhang, Feijun Luo, Qinlu Lin, et al. (2018). PyBioMed: a python library for various molecular representations of chemicals, proteins and DNAs and their interactions. Journal of Cheminformatics. https://doi.org/10.1186/S13321-018-0270-2.

Kim D. Pruitt, Tatiana Tatusova, Donna Maglott (2003). NCBI Reference Sequence Project: update and current status. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKG111.

Kim D. Pruitt, Tatiana Tatusova, Garth Brown, Donna Maglott (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKR1079.

Kim D. Pruitt, Tatiana Tatusova, William Klimke, Donna Maglott (2009). NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKN721.

Knut Dagestad Rand, Ivar Grytten, Milena Pavlović, Chakravarthi Kanduri, Geir Kjetil Sandve (2022). BioNumPy: Fast and easy analysis of biological data with Python. bioRxiv. https://doi.org/10.1101/2022.12.21.521373.

Leonardo Jose da Costa Gonzales, Emily H Bowler-Barnett, Daniel Rice, Minjoon Kim, Supun Wijerathne, Aurelien Luciani, et al. (2025). The UniProt website API: facilitating programmatic access to protein knowledge. Nucleic Acids Research. https://doi.org/10.1093/nar/gkaf394.

Lewis Y. Geer, Aron Marchler-Bauer, Renata C. Geer, Lianyi Han, Jane He, Siqian He, et al. (2010). The NCBI BioSystems database. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKP858.

Li Zhang, Yuansen Hu, Jinshui Wang, Guangle Zhang (2014). Python in ChIP-Seq data analysis. Journal of chemical and pharmaceutical research. https://www.jocpr.com/abstract/python-in-chipseq-data-analysis-3452.html.

Lys Sanz Moreta (2023). Dromi: Python package for parallel computation of similarity measures among vector-encoded sequences. bioRxiv. https://doi.org/10.1101/2023.07.05.547866.

Matthew D. Shirley, Matthew D. Shirley, Zhaorong Ma, Brent S. Pedersen, Sarah J. Wheelan, Sarah J. Wheelan (2015). Efficient "pythonic" access to FASTA files using pyfaidx. PeerJ. https://dblp.uni-trier.de/db/journals/peerjpre/peerjpre3.html#ShirleyMPW15.

Mengyi Zhang (2025). Efficient K-mer Processing in Genomics: Challenges and Innovations with Python and Greedy Algorithms. Theoretical and natural science. https://doi.org/10.54254/2753-8818/2025.ld26351.

Michele Magrane (2010). UniProt Knowledgebase: a hub of integrated data. Nature Precedings. https://doi.org/10.1038/NPRE.2010.5092.1.

Patrick Kunzmann, Kay Hamacher (2018). Biotite: a unifying open source computational biology framework in Python. BMC Bioinformatics. https://doi.org/10.1186/S12859-018-2367-Z.

Peter J. A. Cock, Tiago Antao, Jeffrey T. Chang, Brad Chapman, Cymon J. Cox, Andrew Dalke, et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. https://doi.org/10.1093/BIOINFORMATICS/BTP163.

Peter T. Habib, Alsamman M. Alsamman, Aladdin Hamwieh (2019). BioAnalyzer: Bioinformatic Software of Routinely Used Tools for Analysis of Genomic Data. Advances in Bioscience and Biotechnology. https://doi.org/10.4236/ABB.2019.103003.

Qingyu Chen, Ramona Britto, Ivan Erill, Constance J. Jeffery, Arthur Liberzon, Michele Magrane, et al. (2020). Quality Matters: Biocuration Experts on the Impact of Duplication and Other Data Quality Issues in Biological Databases. Genomics, Proteomics & Bioinformatics. https://doi.org/10.1016/J.GPB.2018.11.006.

Rob Knight, Peter Maxwell, Amanda Birmingham, Jason Carnes, J. Gregory Caporaso, Brett C Easton, et al. (2007). PyCogent: a toolkit for making sense from sequence. Genome Biology. https://doi.org/10.1186/GB-2007-8-8-R171.

Robert Friedman (2025). EvolCat-Python: A User-Friendly Toolkit for Exploring the Language of Life. https://doi.org/10.20944/preprints202505.2059.v1.

Russell Ph, Fiddes It (2021). BioCantor: a Python library for genomic feature arithmetic in arbitrarily related coordinate systems. bioRxiv. https://doi.org/10.1101/2021.07.09.451743.

Ryan K. Dale, Leah H. Matzat, Elissa P. Lei (2014). metaseq: a Python package for integrative genome-wide analysis reveals relationships between chromatin insulators and associated nuclear mRNA. Nucleic Acids Research. https://doi.org/10.1093/NAR/GKU644.

S Samson, É Lord, V Makarenkov (n.d.). SimPlot++: a Python application for representing sequence similarity and detecting recombination.

Sebastian Bassi (2007). A primer on python for life science researchers. PLOS Computational Biology. https://doi.org/10.1371/JOURNAL.PCBI.0030199.

Shi-Ming Lin, Fei Xu, Bifang Huang, Lili Zhao, Danni Pan, Shiqiang Lin (2024). Visual codon: a user-friendly Python program for viewing and optimizing gene GC content. https://doi.org/10.22541/au.172712734.45059110/v1.

Simon Anders, Paul Theodor Pyl, Wolfgang Huber (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics. https://doi.org/10.1093/BIOINFORMATICS/BTU638.

Tamara Goldfarb, Vamsi K. Kodali, Shashikant Pujar, Vyacheslav Brover, Barbara Robbertse, Catherine M. Farrell, et al. (2024). NCBI RefSeq: reference sequence standards through 25 years of curation and annotation. Nucleic Acids Research. https://doi.org/10.1093/nar/gkae1038.

Tamás G. Szabó, Robin Palotai, Péter Antal, Itay Tokatly, Laszlo Tothfalusi, Ole Lund, et al. (2009). Critical role of glycosylation in determining the length and structure of T cell epitopes. Immunome Research. https://doi.org/10.1186/1745-7580-5-4.

Valentin Zulkower, Susan J. Rosser (2020). DNA Features Viewer: a sequence annotation formatting and plotting library for Python. Bioinformatics. https://doi.org/10.1093/BIOINFORMATICS/BTAA213.

Zeila Frade (2023). Almespar: An Open Reading Frames Detection Tool Using Python. https://doi.org/10.5281/zenodo.7818560.

BIOINFORMATIKA KOMPUTASIONAL

Dasar Biologi, Representasi Data, dan Pemrograman Python

Buku ini menyajikan pembahasan secara sistematis mengenai landasan biologis yang menjadi dasar bioinformatika, karakteristik dan jenis data biologis, serta berbagai metode komputasi yang digunakan dalam pengolahan informasi hayati. Materi diawali dengan pengenalan konsep dasar biologi molekuler, DNA, RNA, protein, dan genom, kemudian dilanjutkan dengan pembahasan mengenai basis data biologis, pengelolaan sekuens, analisis data genomik, serta pendekatan komputasional dalam memahami berbagai fenomena biologis. Selain itu, buku ini memperkenalkan pemrograman Python sebagai salah satu perangkat penting untuk mendukung proses pengolahan, visualisasi, dan analisis data biologis secara terstruktur dan efisien.

Pemanfaatan pemrograman dan metode komputasi dalam bioinformatika membuka peluang yang luas dalam pengembangan ilmu pengetahuan dan teknologi. Kemampuan mengolah data biologis dalam jumlah besar memungkinkan berbagai informasi yang sebelumnya sulit dianalisis menjadi lebih mudah dipahami dan dimanfaatkan. Melalui penggunaan Python, berbagai proses seperti manipulasi data, analisis sekuens DNA dan protein, pemrosesan informasi biologis, hingga visualisasi hasil analisis dapat dilakukan secara lebih sistematis. Penguasaan konsep dasar biologi, pemahaman karakteristik data, serta keterampilan pemrograman menjadi fondasi penting dalam menghadapi perkembangan riset berbasis data di bidang biologi dan kesehatan.

www.penerbitwidina.com
@penerbitwidina
penerbit widina
penerbitwidina@gmail.com
widina store
widina bookstore

Layanan Pembaca & Penerbitan Buku
0815-7000-699

SCANME

Teknologi & Komputer

ISBN 978-634-317-538-4

9 786343 175384