



Course/Workshop: Natural Product Discovery: bioinformatic analyses for beginners

Plan of activities

Infrastructure Needed: Participants may bring their laptops.

Period: May 19th – May 23rd, 2025.

Objective: To equip students with the skills to use simple, free bioinformatics tools for integrative analysis of genes, proteins, and genomes, focusing on the genetic components of metabolite discovery in prokaryotes.

Course Outcomes: By the end of the course, students will be able to:

- Identify the function of unknown genes.
- Annotate bacterial genomes.
- Compare genes and genomes.
- Build phylogenetic trees.
- Model proteins based on sequences using freely available bioinformatics tools.

Specific Objectives:

1. Analyze nucleotide sequences using bioinformatics tools (database queries, sequence alignment, promoter sequence prediction, phylogenetic analysis, oligonucleotide design, etc.).
 2. Analyze amino acid sequences using bioinformatics tools (database queries, sequence alignment, prediction of conserved motifs, hydrophobicity, post-translational modifications, construction of function-based phylogenetic trees, etc.).
 3. Perform gene and genome/metagenome mining for comparative genomics, including metabolism mapping, genome comparisons, and proteome comparisons.
 4. Conduct metabolomic analyses, including mapping metabolic pathways and specialized metabolites pathways for novel compound discovery.
-



Thematic Content and Plan of Activities

1. Monday 19th May; Introduction to Bioinformatics (4 hours, 9 am-1 pm) and project designs (1 hour, 2.3 pm-3.3 pm)

- Computer Science Applied to Biochemistry and Biotechnology
- Sequencing of Macromolecules: DNA and Proteins
- Algorithms and Platforms for Gene and Genome Annotations (RAST, PGAP)
- Biological Databases
 - Search Engines for Bioinformatics Analysis
 - Features and Formats of FASTA, PDB, etc.
 - Search for Nucleic Acid and Protein Sequences (ExPASy, NCBI, EMBL)
- Sequencing Technologies and their use for natural products discovery
- Practical exercises

2. Tuesday 20th May: Sequence Alignment (4 hours, 9 am-1 pm)

- Similarity, Homology, and Identity
- Alignment Between Pairs of Sequences
 - Data Matrices
 - Statistical and Biological Interpretation
 - Using the BLAST Tool
 - Search Parameters
 - Interpretation of E-values, Identity, and Coverage
- Alignment of Multiple Sequences
 - Algorithms (MUSCLE, Clustal W)
 - Markov Hidden Models
 - Relationship Between E-values and Probability Index
- Preparation and Interpretation of Clustering Trees (MEGA)
 - Distance Methods (UPGMA, Neighbour Joining)



- Maximum Likelihood
- Maximum Parsimony
- Practical exercises

3. Wednesday 21st May: Protein Analysis (4 hours, 9 am-1 pm)

- Protein Motifs and Domains (HMMR, ProtParam, InterProScan, SignalP, PsortB, SwissProt/UniProt, CAZy, Pfam)
- Prediction of Secondary Structures
- Prediction of Tertiary Structures
 - Ab Initio Modeling (Rosetta, Folding, Zhang Lab)
 - Homology Modeling (Phyre2, I-TASSER, SwissModel, QUARK)
 - Threading (MUSTER, SEGMENT)
- Validation of Built Models (ProSA, PROCHECK)
- Molecular Visualization (PyMOL, UCSF Chimera)
- *In Silico* Protein Purification
- Calculation of Physicochemical Parameters (Molecular Weight, Isoelectric Point, Molar Extinction Coefficient)
- Prediction of Hydrophobicity, Post-Translational Modifications, Antigenic Regions, Signal Peptide
- Practical exercises

4. Thursday 22nd May: Analysis of Whole Genomes and Metagenomes and Their Practical Applications (4 hours, 9 am-1 pm) and project advancements (1 hour, 2.30 pm-3.30 pm)

- Genome Browsers (Integrated Genome Viewer, EMBL-EBI, PATRIC)
- Free Online Sites with Multiple Tools (PATRIC, Expasy, EMBL-EBI)
- Metabolic Mapping of Genomes (KEGG, Biocyc)
- Proteome Comparisons (RAST)
- Secondary Metabolite Detection (AntiSMASH, Prism, Bagel)
- Practical exercises



5. **Friday 23rd May: Project presentations and final questions and answers about the content and practical exercises (4 hours, 9 am-1 pm).**

BIBLIOGRAPHY

General Bioinformatics

1. **Bioinformatics: Sequence and Genome Analysis** by David W. Mount
2. **Bioinformatics for Dummies** by Jean-Michel Claverie and Cedric Notredame
3. **Introduction to Bioinformatics** by Arthur Lesk

Nucleotide Sequence Analysis

1. **Bioinformatics Tools for DNA Sequence Analysis** - Bioinformatics Home[1]
2. **Bioinformatics Databases, Software, and Tools with Uses** - Microbe Notes[2]
3. **Bioinformatics Tools for Sequence Analysis** - Omics Tutorials[3]

Amino Acid Sequence Analysis

1. **Expasy - ProtParam** - Expasy[4]
2. **SIB Swiss Institute of Bioinformatics | Expasy** - Expasy[5]

Gene and Genome Mining for Comparative Genomics

1. **The Integration of Genome Mining, Comparative Genomics, and Functional Genetics for Biosynthetic Gene Cluster Identification** - Frontiers in Genetics[6]
2. **Comparative Genomics** - Pennsylvania State University[7]

Metabologenomics Analyses

1. **Genome-scale models in human metabologenomics** - Nature Reviews Genetics[8]
2. **Bioactivity-driven fungal metabologenomics identifies antiproliferative stemphone analogs and their biosynthetic gene cluster** - Metabolomics[9]
3. **References for the metabologenomics tools** [10-16]



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
INSTITUTO DE QUÍMICA

CIRCUITO EXTERIOR S/N, CIUDAD UNIVERSITARIA, ALCALDÍA COYOACÁN. C.P. 04510, CIUDAD DE MÉXICO, TEL. 55 5622 4442



References

- [1] [67 Free DNA Sequence Analysis Tools - Software and Resources](#)
- [2] [Bioinformatics Databases, Software, and Tools with Uses - Microbe Notes](#)
- [3] [Bioinformatics Tools for Sequence Analysis - Omics tutorials](#)
- [4] [Expasy - ProtParam](#)
- [5] [SIB Swiss Institute of Bioinformatics | Expasy](#)
- [6] [The Integration of Genome Mining, Comparative Genomics, and Functional ...](#)
- [7] [Comparative Genomics - Pennsylvania State University](#)
- [8] [Genome-scale models in human metabologenomics - Nature](#)
- [9] [Bioactivity-driven fungal metabologenomics identifies antiproliferative ...](#)
- [10] Caesar, L. K., Montaser, R., Keller, N. P., & Kelleher, N. L. (2021). Metabolomics and genomics in natural products research: Complementary tools for targeting new chemical entities. In *Natural Product Reports* (Vol. 38, Issue 11, pp. 2041–2065). Royal Society of Chemistry. <https://doi.org/10.1039/d1np00036e>
- [11] Aron, A. T., Gentry, E. C., McPhail, K. L., Nothias, L. F., Nothias-Esposito, M., Bouslimani, A., Petras, D., Gauglitz, J. M., Sikora, N., Vargas, F., van der Hoof, J. J. J., Ernst, M., Kang, K. bin, Aceves, C. M., Caraballo-Rodríguez, A. M., Koester, I., Weldon, K. C., Bertrand, S., Roullier, C., Dorrestein, P. C. (2020). Reproducible molecular networking of untargeted mass spectrometry data using GNPS. *Nature Protocols*, 15(6), 1954–1991. <https://doi.org/10.1038/s41596-020-0317-5>
- [12] Wang, M., Carver, J. J., Phelan, V. v., Sanchez, L. M., Garg, N., Peng, Y., Nguyen, D. D., Watrous, J., Kapon, C. A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A. v., Meehan, M. J., Liu, W. T., Crüsemann, M., Boudreau, P. D., Esquenazi, E., Sandoval-Calderón, M., ... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. In *Nature Biotechnology* (Vol. 34, Issue 8, pp. 828–837). Nature Publishing Group. <https://doi.org/10.1038/nbt.3597>
- [13] Caesar, L. K., Montaser, R., Keller, N. P., & Kelleher, N. L. (2021). Metabolomics and genomics in natural products research: Complementary tools for targeting new chemical entities. In *Natural Product Reports* (Vol. 38, Issue 11, pp. 2041–2065). Royal Society of Chemistry. <https://doi.org/10.1039/d1np00036e>