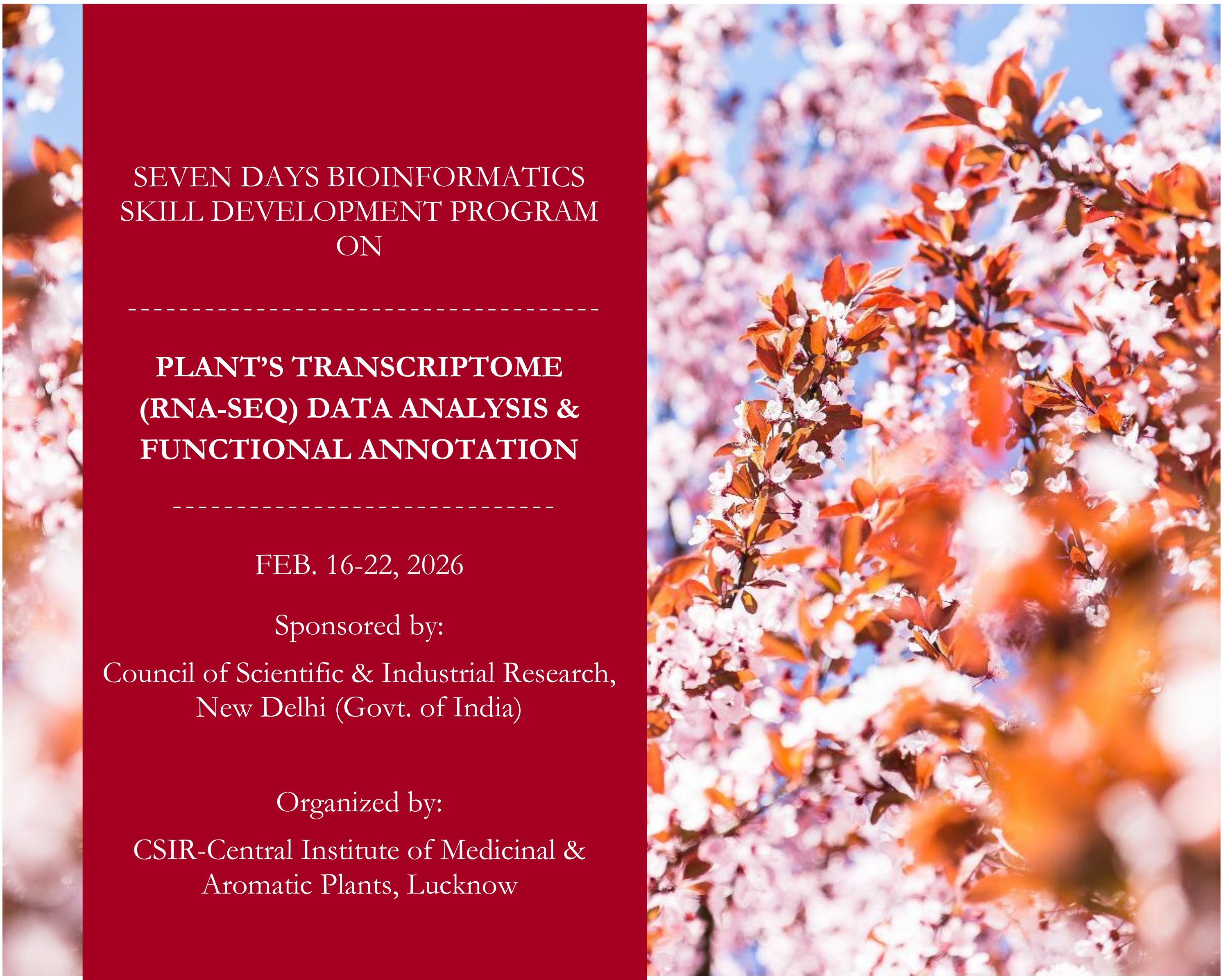

SEVEN DAYS BIOINFORMATICS SKILL DEVELOPMENT PROGRAM ON

PLANT'S TRANSCRIPTOME (RNA-SEQ) DATA ANALYSIS & FUNCTIONAL ANNOTATION

FEB. 16-22, 2026

Sponsored by:

Council of Scientific & Industrial Research, New Delhi (Govt. of India)

Organized by:

CSIR-Central Institute of Medicinal & Aromatic Plants, Lucknow

Training Program Contents

- Transcriptome data retrieval (SRA/ENA)
- Data quality check (FastQC)
- Trimming of garbage sequences (Cutadapt)
- *De-novo* assembly of high-quality reads (TRINITY)
- Assessment of assembled reads (BUSCO, N50)
- Non-redundant data clustering using Hierarchical clustering (CD-HIT)
- Gene expression quantification (RSEM)
- Differentially Expressed Gene (DEG) analysis (EdgeR)
- Functional Annotation through BLASTp, InterProScan, Gene Ontology, Sec. Metabolic Pathways (KEGG) & Enrichment studies.
- R programming and plots generation (Heatmap and Volcano plot)

Convener
Dr. Feroz Khan

Coordinator
Dr. Laiq-ur Rahman

Chairman
Dr. Prabodh K. Trivedi
Director, CSIR-CIMAP
Lucknow

About CSIR-CIMAP, Lucknow

CSIR-Central Institute of Medicinal & Aromatic Plants (CSIR-CIMAP) is a premier multidisciplinary research institute of Council of Scientific & Industrial Research (CSIR), New Delhi, India, with its major focus on exploiting the potential of medicinal and aromatic plants (MAPs) by cultivation, bioprospection, chemical characterization, extraction, and formulation of bioactive phytomolecules. With a strength of 100 scientists, 162 technical officers, 129 support staff, and nearly 300 doctoral and post-doc scholars at its headquarters in Lucknow and research centers at Bengaluru, Hyderabad, Pantnagar, and Purara. CSIR-CIMAP has played a key role in positioning India as a global leader in the production of mints, vetiver, and other aromatic grasses, and in ensuring the indigenous production of artemisinin – a WHO-approved antimalarial. CSIR-CIMAP houses a National Gene Bank on MAPs, which is one of the three of its kind in India. CSIR-CIMAP has played a key role in successfully commercializing an ayurvedic herb-based antidiabetic formulation, which has now benefited millions. The institute is presently accredited by ICS-UNIDO and the Indian-Ocean Rim Association (IORA) as a focal point for research and training on Medicinal Plants among 21 participating member countries. For more details, please see the CSIR-CIMAP website www.cimap.res.in

About Bioinformatics Skill Development Program

'Omics' technologies cover universal detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics) in a specific biosample. Data analysis is complicated as a massive amount of data is generated, and bioinformatician involvement in the process is essential. Transcriptomics data mining is an efficient way to discover genes or gene families encoding enzymes involved in various metabolic pathways. High-throughput next-generation sequencing (NGS) technologies have revolutionized transcriptomics, especially with the advent of RNA-sequencing (RNA-seq). This technology can be used to obtain RNA sequences on a massive scale with enormous sequencing depth. Plants produce a vast array of specialized metabolites, many of which are used as pharmaceuticals, flavors, fragrances, and other high-value fine chemicals. Most of these compounds occur in non-model plants for which genomic sequence information is not yet available. The production of a large amount of nucleotide sequence data using next-generation technologies is now relatively fast and cost-effective, especially when using the latest Roche-454 and Illumina sequencers with enhanced base-calling accuracy. To investigate specialized metabolite biosynthesis in plants, the establishment of a data-mining framework is required by employing next-generation sequencing and computational algorithms to construct and analyze the transcriptomes of plants that produce compounds of interest for biotechnological applications. After sequence assembly, an extensive annotation approach is required to assign functional information to transcripts. The annotation is based on direct searches against public databases, e.g., RefSeq, InterPro, GO, EC, and associated KEGG pathway maps. This study aims to identify biosynthetic gene candidates related to specific metabolic pathways. These assembled transcriptome data are accessed through a web-based BLAST server. Transcriptomes are studied for interpreting functional elements of the genome and revealing molecular constituents of cells and tissues.

The Aim of Bioinformatics Skill Development Training Program

To familiarize students/researchers/academicians/industry experts with the basics of machine learning methods, e.g., hierarchical clustering and its use in RNA-Seq data analysis, especially Heat-map/Dendrogram tree representation of Differentially Expressed Genes (DEGs). Participants may understand the role of Hierarchical Clustering in Dendrogram tree generation and the interpretation of the Heatmap. In parallel, practical exercises/example demos for technical skill development will be scheduled after introductory lectures. Participants need to follow the instructions and perform the different steps during Hands-On training. Live troubleshooting will assist the participants in a smooth learning of tools and techniques. The training program will cover an invited expert lecture, a training program theme lecture, and a demo presentation/practical exercise session. The training program would cover the following aspects:

- Installation and setup of required software and packages, e.g., TRINITY and its associated packages on Linux Ubuntu OS.
- Transcriptome data fetching
    - Sequence Read Archive (SRA)
    - European Nucleotide Archive (ENA)
- Quality check of selected transcriptome data using FastQC software.
- Pre-processing of raw read files (FASTQ) in order to remove low-quality reads, noise sequences, etc, using the Cutadapt tool.
- *De-novo* transcriptome assembly of high-quality reads using TRINITY software.
- Assessment of assembled high-quality reads
    - Benchmarking Universal Single-Copy Orthologs (BUSCO)
    - N50 (Trinity stats)
    - Total alignment rate (Bowtie2)
- Hierarchical clustering of high-quality assembled reads using the CD-HIT pipeline, to generate non-redundant reference transcripts.
- Introduction to Hierarchical Clustering and its application in Dendrogram tree generation.
- RSEM pipeline for abundance estimation based on the mapping of RNA-seq reads to TRINITY assembled contigs.
- The differential gene expression (DEG) analysis of selected samples using the EdgeR (Bioconductor package).
- Functional annotation using the Standalone BLAST+ pipeline against the UniProt database.
- Metabolic pathway mapping (KEGG) and gene ontology (GO) enrichment analysis.
- Visualization of DEG result by making Hierarchical clustering Heatmap and Volcano plot.

The skill development training gained through this program may help in making a career in Biotechnology, Bioinformatics, Functional Genomics, Machine Learning, Big Data, and Data Science.

Eligibility

UG/PG students (Biological Sciences/Engg./Pharmacy/Agriculture/Data Sciences discipline), Ph.D. scholars, Post-doc/RA fellows/Scientist fellows/Project fellows/Technical Officer/Scientist/Academicians & Industry Professionals/Entrepreneurs/Start-ups persons. Basic knowledge of Biology, Chemistry, Statistics, and working experience with a computer (Windows/Linux OS) is required. For fresher’s, a tutorial will be provided for basic commands of Linux.

Certification

Training program participants will receive a certificate and a digital certificate of participation from the CSIR-CIMAP, Lucknow, after successful completion of the skill development program. The digital certificate will be emailed after the successful completion of training program.

Feedback

After the training program, participants may be asked to submit the provided feedback form. Participants may be asked to express their training experiences and suggestions for further improvement or expected training areas.

Training mode: Offline & Online (hybrid) mode. Online mode training will be done through MS Teams/Google Meet/Zoom, or similar apps.

Registration Fee: Rs.5,000/- for each participant (without accommodation) and **Rs.10,000/-** for each participant (with meal & accommodation on shared room basis (2 beds).

The registration fee includes Registration laptop bag with registration kit which includes training brochure, program schedule, stationary items in a folder, tutorials/practical exercises, feed-back form, a group photo (printed & digital copy) and a certificate (printed & digital copy through e-mail). Registration fee includes morning & evening tea & snacks, and lunch.

Registration fee can be pay through Online/UPI mode to SBI bank A/c No. 00000030267691783, SBI Main branch, Hazratganj, Lucknow (IFSC code: SBIN0000125) or through Demand Draft in favor of '**Director, CIMAP**', payable to Lucknow. Complete application cum registration form along with the fee details should reach us (fkhan.cimap@csir.res.in) before the deadline, i.e., Feb. 12, 2026, up to 5:00 PM. Registration for the skill development training program will be on a 'first-come, first-served' basis. Seats are limited.

For any query related to this skill development training program, kindly contact:

Dr. Feroz Khan, Senior Principal Scientist & Convener, SDP-2025-26 (Bioinformatics), (E-mail: fkhan.cimap@csir.res.in), Mobile No.: 9415538701/ Phone (O) +91 522 2718668

Dr. Laiq-Ur Rahman, Chief Scientist & Coordinator, SDP-2025-26 (CSIR-CIMAP), (E-mail: l.rahman@cimap.res.in), Mobile No.: 9412506278

For any further details, please contact:

Dr. Prabodh Kumar Trivedi,

Director

CSIR-Central Institute of Medicinal & Aromatic Plants,

P.O.-CIMAP, Kukrail Picnic Spot Road, Lucknow-226015, INDIA

Ph.: +91 522 2718639, 2718641, 2718505

E-mail: director@cimap.res.in , Website: www.cimap.res.in