

YUQIAO TANG

Boston, Massachusetts|tang.yuqi@northeastern.edu |<https://yuqiaotang.github.io/> |(857)7637493

EDUCATION

Master of Bioinformatics, Northeastern University

Boston, MA

Bioinformatics Major, GPA: 3.8/4.0

Expected May 2025

Coursework: Bioinformatics programming, Algorithms, Statistics for Bioinformatics, Machine Learning

Bachelor of Science in Plant Science, Southwest University

Chongqing, China

Agriculture Major, GPA: 3.5/4.0

June 2023

Coursework: Biotechnology, Cell Biology, Genetics, Linear Algebra

TECHINICAL SKILLS

Programming Language: R, Python, Bash Shell, Git, Nextflow

Bioinformatics Tools: HISAT2, STAR, vcftools, SAMtools, DESeq2, KEGG, Seurat, ggplot2, Scanpy, SnpEff

Wetlab Skills: Gene Cloning, Vector Construction, Transformation, qPCR, Western Blotting, GUS Staining

Statistics Tools: Matlab, SPSS

PUBLICATION

Mengna Yu, ..., *Yuqiao Tang*, etc. "Quantitative Trait Locus Mapping and Transcriptome Analysis Reveal Candidate Genes for a Stem Bending Mutant in Rapeseed (*Brassica napus*). " *Industrial Crops and Products* (0926-6690). 2022: e177, 114456

Mengqi Li†, *Yuqiao Tang†*, etc. "Systematic Characterization of *Brassica napus* HIR Gene Family Reveals a Positive Role of BnHIR2.7 in *Sclerotinia sclerotiorum* Resistance." *Horticulturae* (2311-7524). 2022: 8(10), 874

Research Experience

Upcoming Bioinformatics and Machine Learning Co-Op, Institute for Experiential AI at Northeastern University, US

Except Jan. 2025-Jun. 2025

- Developed ETL Pipelines: Designed and implemented ETL pipelines to process genomics, transcriptomics, and proteomics data from public databases. Utilized Python and R for data extraction, transformation, and analysis.
- Maintained Reproducible Workflows: Documented all processes and code with detailed annotations and regular commits to a GitHub repository, ensuring workflow reproducibility and collaborative accessibility.
- Implemented Machine Learning Models: Applied machine learning techniques to automate data processing and derive insights from complex biological datasets.
- Produced Scientific Outputs: Authored technical reports and slide decks summarizing the analysis, methodologies, and results for internal and external stakeholders.

Genomic analysis in neuromuscular patients, Brigham and Women's Hospital, US

May. 2024-Aug. 2024

- Developed Variant Filtering Pipeline: Designed and implemented a comprehensive pipeline for variant filtering and annotation, utilizing tools like SnpEff and SnpSift. Enhanced data processing accuracy by integrating advanced filtering criteria and quality control measures.
- Conducted Ancestry Analysis: Analyzed genomic data to determine sample ancestry and removed non-European samples. Leveraged admixture analysis techniques to ensure dataset homogeneity, improving the reliability of subsequent analyses.
- Performed Statistical Disease Testing: Executed statistical disease testing using custom scripts to identify potential associations. Analyzed mutation data to detect significant findings, contributing to the understanding of genetic influences on disease.

Analysis of Gene Families, Southwest University, China

May. 2021-Sep. 2022

- Key Genes Selection and Verification: Identified genes influencing silique length in *Brassica napus* from fox-hunting database; applied qPCR and gene analytical methods for preliminary function verification
- Comparative Genomic Analysis: Utilized blast for comparing target gene sequences with homologous sequences in *Arabidopsis thaliana* and rapeseed; identified motifs, and visualized gene families' evolutionary relationships through MEGA and phylml models
- Gene Structure and Expression Analysis: Utilized GSDS for gene structure construction, analyzed promoter elements with PlantCare, and mapped expression and resistance patterns using R Language based on extracted gene location data
- Experimental Infection and Analysis: Infected specific plants with *Sclerotinia sclerotiorum*, conducted RNA extraction for qPCR analysis, and interpreted experimental outcomes utilizing SPSS

Workflow of Transcriptome Analysis Results , Southwest University, China

Aug. 2021-Jan. 2022

- Processed and Mapped Data: Utilized Fastp to process raw sequencing data and employed STAR to map the cleaned data against the Brassica napus reference genome.
- Generated Gene Counts and Detected DEGs: Used FeatureCounts for generating gene counts data and employed the DESeq2 R package for detecting differentially expressed genes (DEGs). Visualized data analysis results with the ggplot2 R package.
- Integrated Analysis Workflow: Integrated the above analyses into a unique workflow for this study, enhancing reproducibility and convenience for subsequent related research in the laboratory.
- Conducted Experimental Infection and Analysis: Infected specific plants with Sclerotinia sclerotiorum, performed RNA extraction for qPCR analysis, and analyzed experimental outcomes with SPSS.

Project Experience

Brain CT Image Hemorrhage Segmentation , Northeastern University, US

Sep. 2024-Jan.2025

- Developed Machine Learning Models: Designed and implemented machine learning pipelines for classification, regression, and brain hemorrhage segmentation in CT images. Leveraged Python and TensorFlow to create robust and scalable solutions.
- Applied Neural Network Techniques: Engineered and fine-tuned artificial neural networks to enhance segmentation accuracy. Incorporated advanced techniques such as transfer learning and model regularization to optimize performance.
- Analyzed Labeled Datasets: Investigated labeled medical datasets, focusing on preprocessing, feature extraction, and data augmentation. Enhanced model robustness by addressing class imbalances and noise in CT image data.
- Achieved Brain Hemorrhage Segmentation: Successfully segmented brain hemorrhages in CT scans, delivering insights critical for clinical applications and improving diagnostic workflows.

Update of qPrimerDB - qPCR Primer Database , Northeastern University, US

Sep. 2023-Jan.2024

- Expanded qPCR Primer Databases: Integrated information to expand qPCR primer database, such as unanalyzed species, recent sequencing results, pan-genomic data, etc
- Optimized Primer Design: Optimized the primer design process (including plant genome and pan-genome primer design), updated software to improve precision and efficiency of data analysis, and completed all primer analysis from JGI database

Teaching Experience

Graduate Teaching Assistant, Northeastern University, US

Sep. 2024-Dec. 2024

- Supported Course Instruction: Assisted in teaching core bioinformatics concepts, including sequence alignment, variant analysis, and functional annotation. Conducted lab sessions and tutorials to reinforce theoretical knowledge with hands-on practice.
- Provided Student Support: Guided students in applying bioinformatics tools and methodologies, offering individualized support for projects and assignments. Helped troubleshoot technical issues and clarify course concepts.
- Assessed Performance: Evaluated assignments and examinations, providing constructive feedback to improve student understanding and skills in bioinformatics techniques.
- Promoted Collaborative Learning: Fostered a collaborative learning environment by organizing office hours and discussions on emerging trends and applications in bioinformatics.