

DeCIFR: an integrated biological data informatics platform

Ignazio Carbone and James B. White, 2025

Center for Integrated Fungal Research, Department of Entomology and Plant Pathology,
North Carolina State University, Raleigh, NC 27695, USA

Corresponding author: Ignazio Carbone (icarbon@ncsu.edu)

DOI: <https://doi.org/10.52750/845633>

Platform URL: <https://decifr.cifr.ncsu.edu/>

Cite as: Carbone, I., & White, J.B. (2025). DeCIFR: an integrated biological data informatics platform [Platform record]. DeCIFR. <https://doi.org/10.52750/845633>

Abstract

DeCIFR is an integrated biological data informatics platform that brings together phylogenetic, taxonomic, microbiome, and population-genetic workflows across animals, plants, fungi, protists, bacteria, archaea, and viruses. Rather than functioning only as a collection of software notes or individual tools, DeCIFR serves as a program platform for delivering curated analytical workflows, reproducible execution environments, interactive visualization, and centralized metadata tracking within a unified web-based system. Housed within the Center for Integrated Fungal Research (CIFR) at NC State University, the platform is maintained on institutionally managed compute and storage infrastructure to provide a sustainable and scalable foundation for research, teaching, and collaborative biodiversity informatics.

Document type: Platform record

This document provides descriptive documentation for the DeCIFR platform. The canonical platform citation is: Carbone, I., & White, J.B. (2025). DeCIFR: an integrated biological data informatics platform [Platform record]. DeCIFR. <https://doi.org/10.52750/845633>

Features and capabilities

DeCIFR provides a program platform for biodiversity informatics by integrating sequence processing, tree inference, taxonomic assignment, phylogenetic placement, comparative analysis, and visualization within a single operational environment. Users can run multiple sequence alignment, model testing, and maximum-likelihood or Bayesian tree inference, as well as species-tree reconstruction and placement workflows, with outputs available for direct interpretation or export to companion resources such as the T-BAS reference tree repository.

The platform supports marker-gene, metabarcoding, ASV, and genome-scale analyses through curated workflows rather than stand-alone software descriptions. Standard pipelines for loci such as ITS and 16S connect raw-read processing to taxonomic assignment and ecological summaries, while integrated population-genetic workflows support multilocus genotyping, clustering and structure inference, demographic analyses, and Isolation-with-Migration (IM) modeling. IMA3 workflows in DeCIFR use input files generated by the Pop-Gen Pipeline Platform

(Webb et al. 2021) and can also create IMA3-formatted inputs internally through the IMgc pre-processing step (Woerner et al. 2007).

Reproducibility is a core platform function in DeCIFR. Galaxy-based workflows (Afgan et al. 2018), the Pop-Gen Pipeline Platform (Webb et al. 2021), and Nextflow-managed pipelines (Di Tommaso et al. 2017) executed through Seqera Platform (Seqera Labs 2024) capture provenance, parameters, workflow definitions, histories, and intermediate files so analyses can be revisited, compared, and shared across research teams. This framework positions DeCIFR as a maintained informatics platform for program delivery, not simply as a static description of available software.

All DeCIFR source code is maintained in a public GitHub repository, allowing the platform to be redeployed or migrated in the event of server failure at NC State. The site and tools are written in Python, using Flask for the web front end and Celery to manage jobs on local worker servers. Docker-based deployments can also be hosted in external cyberinfrastructure such as the CyVerse Discovery Environment (Swetnam et al. 2024), providing mirrored instances that further enhance redundancy and long-term resilience.

Availability

DOI: <https://doi.org/10.52750/845633>

Source code and documentation: <https://github.com/ncsu-decifr>

Platform: <https://decifr.cifr.ncsu.edu/>

References

Afgan, E., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. *Nucleic Acids Research* 46(W1): W537–W544.

Woerner, A.E., Cox, M.P., and Hammer, M.F. (2007). Recombination-filtered genomic datasets by information maximization. *Bioinformatics* 23(14): 1851–1853. doi:10.1093/bioinformatics/btm253.

Webb, A., Knoblauch, J., Sabankar, N., Kallur, A.S., Hey, J., and Sethuraman, A. (2021). The Pop-Gen Pipeline Platform: A software platform for population genomic analyses. *Molecular Biology and Evolution* 38(8): 3478–3485. doi:10.1093/molbev/msab113.

Di Tommaso, P., et al. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology* 35: 316–319.

Seqera Labs (2024). Seqera Platform: centralized monitoring and orchestration for Nextflow workflows. <https://seqera.io/platform/>.

Swetnam, T.L., et al. (2024). CyVerse: Cyberinfrastructure for open science. *PLOS Computational Biology* 20(1): e1011270.