Michele Cosi

Data Science Institute, BSRL building,
University of Arizona
1239 N Cherry Avenue
Tucson, AZ 85721-0240

Website: cosimichele.github.io
GitHub: CosiMichele
Email: cosi@arizona.edu

EMPLOYMENT & RESEARCH EXPERIENCE

Data Science Institute, University of Arizona Sept 2023 – Present
Data Scientist *Tucson, AZ*

- Supported the cyberinfrastructure of national and international workshops (ESIIL Innovation Summit 2024, CompBio Asia 2023) in their respective fields of environmental data science and genomics.
- Held multiple data science related workshops (Foundational of Open Science Skills, DataLab workshop series) as part of the Data Science Institute and DataLab, with a target audience of graduate students, staff and faculty members.
- Instructed data science fundamentals at PHIRE and KEYS internships as wells as supporting PHIRE interns
- Expanded knowledge and experience of CyVerse and Cloud research platforms such as JetStream2/ACCESS.

CyVerse, University of Arizona Jan 2022 – Present
Science Analyst *Tucson, AZ*

- Earned certification as a *Carpentries* instructor and contributed to *Software* and *Data Carpentries* lectures as an educator
- Developed teaching materials and assumed an instructor role in Open Science and containerization technology for CyVerse hosted workshops (Foundational Open Science Skills, Container Camp & Cloud Native Camp)
- Provided cyberinfrastructure assistance by executing and monitoring large scale deployments of personalized virtual machines engineered for molecular docking simulations in international workshop settings (CompBio 22-23 and the 6th Uppsala Transposon Symposium)
- Extensive experience working with CyVerse including creating and maintaining CyVerse Apps, curating CyVerse hosted data, and providing support for users using the

CyVerse platform
- Gained project management experience by overseeing the development of OpenDendro, an open source analytic software for dendrochronologists

University of Arizona, Lyons Lab Sept 2019 – Dec 2021
Research Assistant *Tucson, AZ*

- Developed a distributive, scalable computing workflow system (StarBLAST, PhytoOracle) cable of processing large volumes of high-throughput phenotyping data using the CyVerse platform as key component for data storage, sharing and visualization
- Gained proficiency in BASH, Python, R computing languages as well as computer vision and machine learning methods
- Familiarized with cloud-based computing infrastructure and High-Performance Computing (HPC) systems
- Developed a PyTorch-based deep learning algorithm capable of detecting and segmenting charcoal rot disease in *Sorghum bicolor*
- Utilized image segmentation techniques to detect *Lactuca sativa* varieties and extract phenotypical data (median canopy temperature) throughout the growth season

Nagoya University, Laboratory of Plant Gene Function Apr 2017 – Aug 2019
Research & Teaching Assistant *Nagoya, Japan*

- Assembled and annotated genomes of rice species *Oryza longistaminata* and *Oryza glaberrima* combining PacBio long-reads sequencing and Illumina short-read sequencing
- Conducted Principal Component Analysis (PCA) to identify outliers within semi-cultivated *Oryza glaberrima* accessions
- Performed Genotyping by Sequencing (GBS) in order to accomplish Genome Wide Association Study (GWAS) across an *Oryza glaberrima* population to identify potential key genetic loci related to deep water stress response

EDUCATION

University of Arizona Aug 2019 – Dec 2021
MS, Plant Sciences, School of Plant Sciences *Tucson, AZ*

- 1 scientific publication

Nagoya University Oct 2017 – Sept 2019
MS, Applied Biosciences, Graduate School of Bioagricultural Sciences *Nagoya, Japan*

- Thesis: Genome-Wide Association Study of growth-related genes in *de novo* genome assembly of *Oryza glaberrima*
- 1 scientific publication
- Recipient of *JASSO Scholarship*

Nagoya University Oct 2013 – Sept 2017
BS, Applied Biosciences, School of Agricultural Sciences *Nagoya, Japan*

- Recipient of *Nagoya University Global 30 Undergraduate Scholarship* and *JASSO Scholarship*

PUBLICATIONS

Swetnam, T. L., Antin, P. B., Bartelme, R., Bucksch, A., Camhy, D., Chism, G., Choi, I., Cooksey, A. M., **Cosi, M.**, Cowen, C., Culshaw-Maurer, M., Davey, R., Davey, S., Devisetty, U., Edgin, T., Edmonds, A., Fedorov, D., Frady, J., Fonner, J., Gillan, J. K., Hossain, I., Joyce, B., Lang, K., Lee, T., Littin, S., Mcewen, I., Merchant, N., Micklos, D., Nelson, A., Ramsey, A., Roberts, S., Sarando, P., Skidmore, E., Song, J., Sprinkle, M. M., Srinivasan, S., Strootman, J. D., Stryeck, S., Tuteja, R., Vaughn, M., Wali, M., Wall, M., Walls, R., Wang, L., Wickizer, T., Williams, J., Wregglesworth, J., Lyons., E. (2024, Feb 7) CyVerse: Cyberinfrastructure for Open Science. *PLOS Computational Biology.* doi: 10.1371/journal.pcbi.1011270

Skidmore, E., **Cosi, M.**, Swetnam, T. L., Merchant, N., Xu., Z., Choi, I., Davey., S, Frady, J., Wall, M., Yung, M. (2023-07-23). Cloud Computing for Research and Education Gets a Sweet Upgrade with CACAO. *PEARC '23*. doi: 10.1145/3569951.3597555

Gonzalez, E. M., Zarei, A., Hendler, N., Simmons, T., Zarei, A., Demieville, J., Strand, R., Rozzi, B., Calleja, S., Ellingson, H., **Cosi, M.**, Davey, S., Lavelle, D. O., Truco, M. J., Swetnam, T. L., Merchant, N., Michelmore, R. W., Lyons, E., Pauli, D. (2023, February 14). Phytooracle: Scalable, Modular Phenomics Data Processing Pipelines. *Frontiers*. doi: 10.3389/fpls.2023.1112973

Cosi, M., Forstedt, J., Gonzalez, E., Xu, Z., Peri, S., Tuteja, R., Blumberg, K., Campbell, T., Merchant, N., & Lyons, E. (2021, April 27). StarBLAST: A scalable blast+ solution for the classroom. *Journal of Open Source Education*. doi: 10.21105/jose.00102

Reuscher, S., Furuta, T., Bessho-Uehara, K., **Cosi, M.**, Jena, K. K., Toyoda, A., Fujiyama, A., Kurata, N., & Ashikari, M. (2018, October 5). Assembling the genome of the African wild rice *Oryza longistaminata* by exploiting synteny in closely related *Oryza* species. *Nature News*. doi: 10.1038/s42003-018-0171-y

WORKSHOPS & TEACHING EXPERIENCE

NASA SBG and ESIIL: HYR-SENSE June 2024
Boulder, CO

- Supported cyberinfrastructure and cloud platform access
- Led discussion on data sovereignty (CARE principals) and teaching computational skills (Unix shell)

KEYS Research Internship June 2024

- Instructed interns on the access and use of the University of Arizona's high-performance computing (HPC) clusters, reproducibility and version control (git/GitHub) *Virtual & Tucson, AZ*

PHIRE May 2024

- Led instructional sessions on version control and open science skills (git/GitHub, Hugging Face, R/RStudio) *Tucson, AZ*

ESIIL Innovation Summit 2024 May 2024

- Leading discussions on cyberinfrastructure and data sovereignty whilst supporting the attendees with computational skills and cloud infrastructure *Boulder, CO*

Data Science Institute's DataLab Workshop Series Sept 2023 - present

- Data Science instructor covering topics ranging from Open Science skills, reproducibility and cloud cyberinfrastructure *Virtual & Tucson, AZ*

CyVerse: Introduction to LLMs & ChatGPT Aug 2023

- Prompt engineering educator for the Large Language Models ChatGPT, Bard and Bing Chat *Tucson, AZ*

CompBio Asia 2023 Jul 2023

- Curated virtual machines with specific molecular docking software (Amber22, Gromacs, AlphaFold) *Singapore*
- Provided cyberinfrastructure support by executing and monitoring large scale deployments of personalized virtual machines
- Monitored data production and assisted with data maintenance
- Instructed attendees in utilization of the Unix system and shell

ESIIL Innovation Summit 2023 May 2023

- Assisted primary event coordinators, providing management and mentorship within subgroups whilst encouraging innovative, scientifically-driven thinking. *Boulder, CO*

6th Uppsala Transposon Symposium Oct 2022

- Curated virtual machines with specific transposon-detection software *Uppsala, Sweden*
- Provided cyberinfrastructure support by executing and

monitoring large scale deployments of personalized virtual machines

CyVerse: Foundational Open Science Skills (FOSS) Sept 2022 – present
Virtual

- Organizer, educator and curator of teaching materials
- Instructed attendees of Open Science related technologies (git, GitHub, Unix shell, containerization) and best practices (project management, FAIR and CARE data principles, documentation, reproducibility)

The Carpentries Sept 2022 – present
Virtual

- Certified Carpentries instructor (Data, Library and Software Carpentry) teaching version control (git, GitHub), Unix shell, Python and R.

CompBio Asia 2022 Jun 2022
Bangkok, Thailand

- Curated virtual machines with specific molecular docking software (Amber22, Gromacs, AlphaFold)
- Provided cyberinfrastructure support by executing and monitoring large scale deployments of personalized virtual machines
- Monitored data production and assisted with data maintenance
- Instructed attendees in utilization of the Unix system and shell

CyVerse: Container Camp & Cloud Native Camp May 2022 – present
Virtual & Tucson, AZ

- Organizer, educator and curator of teaching materials
- Instructed on containerization technologies (Docker, Singularity), deployment software (Terraform) and orchestration platforms (Kubernentes, Docker Compose)

PROFESSIONAL MEMBERSHIPS

The Environmental Data Science Innovation & Inclusion Lab (ESIIL) Jan 2023 – present
Virtual & Boulder, CO

- Advisor and collaborator, supporting the ESIIL community with the necessary computational power through CyVerse.

The Carpentries Sept 2022 – present
Virtual & Tucson, AZ

- Member of the Carpentries as an educator and

contributor to *Software* and *Data Carpentries* materials

Biosciences Toastmasters Feb 2021 – present
- Member of the University of Arizona's Biosciences Toastmasters, a professional speech development group, affiliated to Toastmasters International *Virtual & Tucson, AZ*

CONFERENCES, SEMINARS & WEBINARS

Conferences

Practice and Experience in Advanced Research Computing (PEARC 23) Jul 2023
- Contributor to the short publication *Cloud Computing for Research and Education Gets a Sweet Upgrade with CACAO* (Skidmore *et al., 2023*) to be published in the PEARC 2023 Proceeding (*ACM Digital Library*)

Plant & Animal Genome Conference (PAG 30) Jan 2023
- Speaker for Advanced Computational Methods - CyVerse for Machine Learning, Containers, and Clouds workshop: *CACAO: A Workshop-Friendly Deployment System*

Phenome 2020 Feb 2020
- Poster: *PhytoOracle: A scalable, modular data processing pipeline*

Seminars & Webinars

Agricultural Genome to Phenome Initiative (AG2PI) Dec 2020
- Webinar: *PhytoOracle: A Case Study in Automating Phenotyping*

CyVerse Webinar Series Nov 2020 - present
- Webinar: *Get Started With CyVerse*
- Webinar: *PhytoOracle - A Workflow Manager for Phenomic Data Processing*

STUDENT ADVISING, MENTORING, SUPERVISORY

Undergraduate Advising: I. Ale (Software Engineering, 2022-2024), E. Hagyard (Software Engineering, 2023), A. Noor (KEYS, Computer Science, 2023)

High School Mentoring: Tenmay Dewangan (KEYS, Paradise Valley High School, 2024), D. Patel (KEYS, BASIS Scottsdale, 2024) E. Dorland (KEYS, City High School, 2023)

SKILLS & ABILITIES

Computational and programming languages: Python, BASH, R, JavaScript

Operating Systems and platforms: Unix (MacOS, Linux, Ubuntu), HPC (SLURM, PBS schedulers for HPC), Jupyter (notebooks/lab/hub), RStudio, CyVerse, ACCESS, JetStream2

Applied computational skills: containerization (Docker, Singularity), cloud orchestration (Kubernentes, Terraform), machine learning (PyTorch), computer vision (image segmentation, object detection), pipeline management (Work Queue, Makeflow, Nextflow)

Languages spoken: Italian (native), English (excellent), Japanese (conversational), Thai (conversational)

Teaching abilities: able to communicate scientific information and computational concepts to diverse audiences, ranging from the general public to various academic levels.