

Center for Scientific Software Engineering Updates

April 4th, 2024

Jeff Young, presenting on behalf of SSE Leadership
(Alex Orso, Jeff Young, Irfan Essa)

Slide contributions from Head of Engineering, David Brownell



Georgia Tech College of Computing
**Center for Scientific
Software Engineering**

What's New with CSSE?

Hiring!

- Full team with 4 RSEs, Head of Engineering, Program Coordinator

External Engagements

- Continued work with Schmidt Sciences

Internal Engagements

- First round of submissions and selected projects

Outreach

- At SIAM and Supercomputing among other venues

Team Building



Faculty is First from Georgia Tech to Receive New Fellowship for AI Research

📅 THURSDAY, MARCH 14, 2024 ✉ BRYANT WINE
🏢 COLLEGE OF COMPUTING SCHOOL OF COMPUTATIONAL SCIENCE & ENGINEERING
ARTIFICIAL INTELLIGENCE HELPING STORIES RESEARCH

Schmidt Sciences has selected **Kai Wang** as one of 19 researchers to receive this year's AI2050 Early Career Fellowship. In doing so, Wang becomes the first AI2050 fellow to represent Georgia Tech.

SSE Mission and Vision Statement

Mission Statement: The Center for Scientific Software Engineering (CSSE) is focused on the development and dissemination of software engineering best practices to accelerate both the quality and pace of scientific discoveries at Georgia Tech and throughout the scientific community.

Vision: The Georgia Tech Center for Scientific Software and Engineering aims to engage both Georgia Tech and external scientific researchers in the advancement of scientific computing through the application of software engineering tools and best practices.

CSSE Research Direction

We aim to communicate key software engineering principles to academic researchers to enable cohesive and comprehensive scientific software solutions, as demonstrated by efforts like the DoE's E4S software stack.

- Currently, research code development is focused on getting to publication without providing paths for reproducibility and sustainability

To do this, we will build a body of "use cases" that a) demonstrate software engineering (SE) best practices as applied to promising scientific software and b) contribute to and generate a body of knowledge that allows other researchers to easily integrate software engineering best practices into their own codes.

Techniques used include:

- i. working on codes with specific SE solutions (build system improvements, testing, commit standards, acceleration, packaging)
- ii. developing curriculum for our undergrads and online masters (OMSCS) program focused on improving SE literacy for scientists
- iii. participating in outreach to build a sustainable group of research software engineers (RSEs) at Georgia Tech and across the USA.

If we are successful, CSSE will contribute to a more sustainable vision of research software engineering at GT and other partners, and the participating user codes will see more effective adoption and external contributions by the general user community.



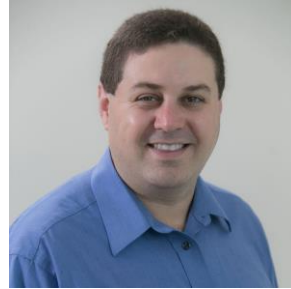
Team Composition



Team Members



Alex Orso
Co-Director, Co-PI



Jeff Young
Co-Director



Irfan Essa
Co-PI,
Chair of Advisory Board



Dave Brownell
Head of Engineering



Kate Rachwal
Principal Research Engineer



Robin Fievet
Senior Research Engineer



Ketan Bhardwaj
Research Engineer II



Varun Narayan
Research Engineer



Nirvana Edwards
Program and Operations
Manager





Team Background



Combined 61 years of industry experience



Strong background in consulting and entrepreneurship



Passion for academia



Expertise in UX Design, Frontend development, Backend development, HPC, AI & Machine Learning



Love of learning



Community Outreach





Community Outreach

March 2023 SIAM workshop with local scientific software groups

New Open Source Program Office (OSPO) bridges the efforts of the SSE Center with institute research computing (PACE) and GT's Library

- Standardize GT's open source practices
- Training and summer internships
- Archival techniques to preserve open source software
- *Provides GT's SSE center with another avenue to reach and work with campus community*

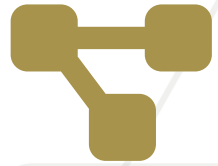
External leadership at events like ADSA/RSE workshop, SIAM CSE 2023, US-RSE, and Supercomputing (2023-2024)



Software Engineering Minisymposium at SIAM CSE 2023



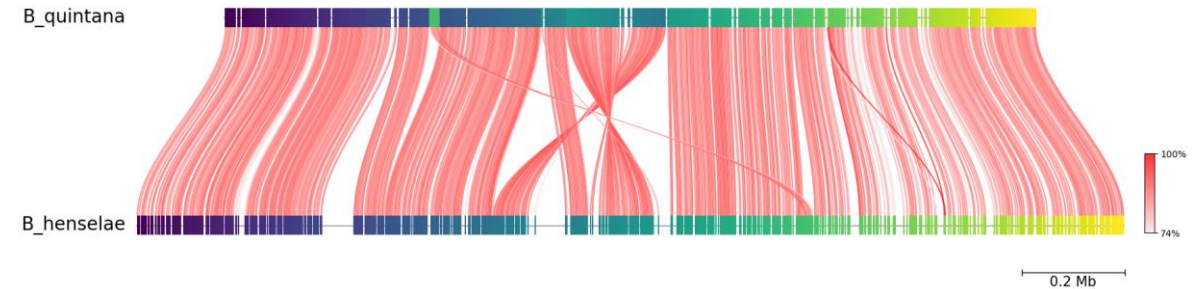
Project Results (Selected)



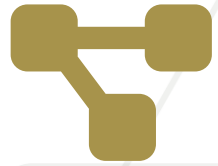
FastANI (S. Aluru, GT)

FastANI is an HPC-oriented Python package to perform whole-genome Average Nucleotide Identity (ANI).

FastANI allows for quickly comparing two genomes to find matches in their genetic structure.



51K downloads on BioConda over 4 years



FastANI: Goals and Results



Develop tests and CI/CD to allow for future updates

- ✓ • Required developing a heuristic to determine "correct" result since ANI techniques typically are not deterministic
- ✓ • Fixed an OpenMP multithreading bug during the process
- ✓ • Added Github Actions workflow after discussion with SSEC team on pros/cons of different CI/CD solutions



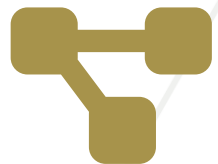
Update Build System and Repository Clean Up

- ✓ • Added CMake build support and CTest testing to make cross-platform usage easier
- ✓ • Cleaned up stale issues, pull requests, and merged in old bug fixes



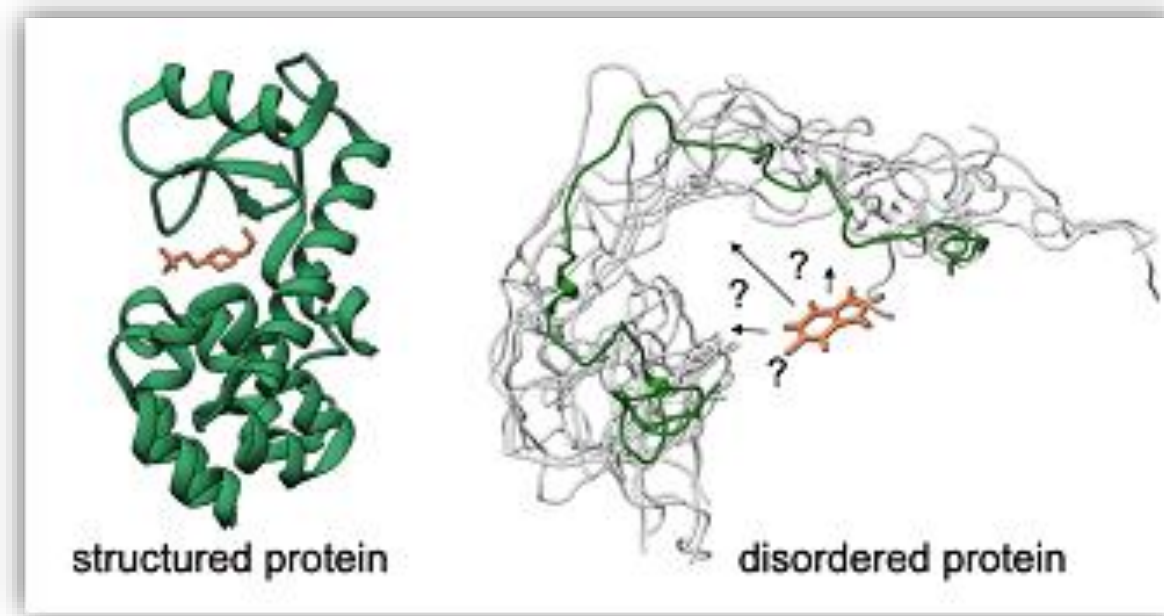
Release a new version of FastANI

- ✓ • New BioConda release has 1200 downloads in the first month alone
- ✓ • 2 new issues raised by users and bugfix applied to this release based on user testing



Heller / Lohr

How to target Intrinsically Disordered Proteins (IDPs) with therapeutic small molecules.

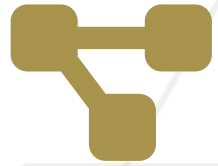


FAST. FLEXIBLE. FREE.
GROMACS

PLUMED



Georgia Tech College of Computing
Center for Scientific
Software Engineering



Heller / Lohr: Goals and Results



Improve efficiency of Metainference Simulations with PLUMED

- ✓ • 68% improvement over CPU baseline / 53% improvement over GPU baseline
 - Improvements based on:
 - Command line flags: STRANDS_CUTOFF eliminates calculations for data unlikely to impact the local scope
 - Compiler flags: OPTIMAL-FAST uses algorithms that are more efficient than those described in the most often cited PLUMED whitepaper



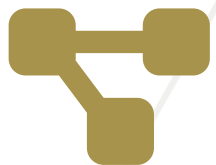
~~Add GPU to PLUMED to improve efficiency of Metainference Simulations~~

- ✓ • Profiled multiple simulations, revealing PLUMED is not a significant bottleneck in these simulations (with flags above)
- ✓ • Added documentation so that others can generate profile data



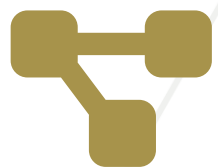
Ensure backwards compatibility for existing PLUMED simulations

- ✓ • Recommended conventions to make flags above more discoverable



Heller / Lohr: Final PLUMED Meeting

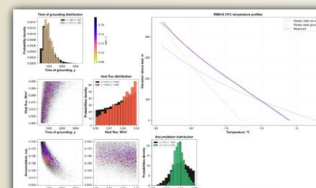




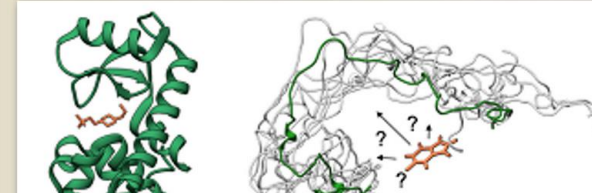
Project Showcase



A numerical heat flow model that infers long-term climate and glacial history using Bayesian inversion of in-situ borehole temperature measurements previously taken in Antarctica



How to target Intrinsically Disordered Proteins (IDPs) with therapeutic small molecules



Check out the highlights of recent projects by CSSE at
<https://gt-sse-center.github.io/project-showcase/>



Engaging with CSSE



SSE Skill Profiles

Software Engineer

- Write code and apply best-of-industry practices to solve problems and automate processes and repetitive activities.

Software Architect

- Design software, infrastructure, and technology stacks that help propel a team forward.

Product Management

- Apply [lean software principles](#) to validate hypotheses and introduce minimum viable products (MVPs).

Program Management

- Ensure successful outcomes by introducing practices that detect issues before they derail progress.

UI/UX

- Create user interfaces and experiences that are efficient, engaging, easy to use, and require little to no training.

Technical Writer (Long term)

- Introduce prose that communicates complicated software ideas and product details to broad audiences.



SSE Opportunities for Engagement

1. Prototyping

- Create solutions that demonstrate system capabilities

2. Partnering (dependent upon work defined within the engagement)

- Augment a team's existing capabilities by working side-by-side for a predetermined set of time

3. Guiding

- Periodic interaction with existing teams to increase the likelihood of a successful project

4. Mentoring

- Periodic interaction with team members to help them grow in their role

5. Consulting

- One-time engagement to address specific concerns about a project

6. Profiling

- Scoped engagement to identify and propose resolutions to performance bottlenecks

7. Curriculum Development and Training

- Develop and present curriculum that is broadly applicable to industry and academia

Questions?

