



Office of the Registrar  
901 W. Illinois Street, Suite 140  
Urbana IL 61801  
(217) 333-6383

## One-Click Degree Verification

Varenya Jain

Institution: University of Illinois Urbana-Champaign  
Degree: Bachelor of Science in Liberal Arts and Sciences  
Graduation Date: December 23, 2024



A handwritten signature in blue ink that reads "Meghan E. M. Hazen".

Meghan E. M. Hazen, University Registrar

This eDiploma was digitally Signed and Certified on Feb 5, 2025.

**Michael Sutter**  
DOCUMENT CERTIFICATION SERVICES

### eDiploma - Signed and Certified PDF

To validate the Signature and Certification of this eDiploma, you must open this PDF with Adobe Reader or Adobe Acrobat. You will see a Blue Ribbon at the top of the document validating the digital Signature used to Certify this eDiploma.





*By authority of the Board of Trustees of the*

# **UNIVERSITY OF ILLINOIS**

*and upon recommendation of the Senate  
at Urbana - Champaign*

**Varenya Jain**

*has been admitted to the Degree of  
Bachelor of Science in Liberal Arts and Sciences*

*and is entitled to all rights and honors thereto appertaining  
Witness the Seal of the University and the Signatures of its Officers  
this twenty-third day of December, two thousand and twenty-four.*



  
Chair of the Board of Trustees

  
Secretary of the Board of Trustees

  
Timothy L. Killeen  
President of the University of Illinois

  
Timothy L. Killeen  
Chancellor, University of Illinois Urbana-Champaign  
Vice President, University of Illinois



**COMPUTATIONAL SCIENCE & ENGINEERING**

*certifies that*

**Varenya Jain**

*while enrolled in the undergraduate program*

**Integrative Biology**

*has satisfied the requirements for a*

**Certificate of Undergraduate Specialization in  
Computational Science and Engineering**

*Luke Olson*

---

Luke Olson, Director  
Computational Science & Engineering

3/6/2025

Date

**COLLEGE OF LIBERAL ARTS & SCIENCES**

Student Academic Affairs Office  
2002 Lincoln Hall, MC-446  
702 S. Wright St.  
Urbana, IL 61801

January 6, 2025

TO WHOM IT MAY CONCERN:

This is to certify that Varenya Jain (DOB: 05/21/02) has completed all requirements for a degree of Bachelor of Science in Liberal Arts and Sciences in the College of Liberal Arts and Sciences at the University of Illinois at Urbana-Champaign with a major in Integrative Biology. The degree conferral date is December 23, 2024.

Sincerely,

A handwritten signature in black ink that reads "Jeanine A. Meyer".

Jeanine Meyer

Admissions/Records Officer

College of Liberal Arts and Sciences

University of Illinois at Urbana-Champaign

[jammeyer@illinois.edu](mailto:jammeyer@illinois.edu)



**COLLEGE OF LIBERAL ARTS & SCIENCES**

Student Academic Affairs Office  
2002 Lincoln Hall, MC-446  
702 S. Wright St.  
Urbana, IL 61801

November 19, 2024

TO WHOM IT MAY CONCERN:

This is to verify that Varenya Jain (DOB: 05/21/2002) upon successful completion of courses in progress for the Fall 2024 semester, will complete the requirements for the degree of Bachelor of Science in Liberal Arts and Sciences at the University of Illinois at Urbana-Champaign with a major in Integrative Biology. The degree conferral date is December 23, 2024.

Sincerely,

A handwritten signature in black ink that reads "Jeanine A. Meyer".

Jeanine Meyer

Admissions/Records Officer

College of Liberal Arts and Sciences

University of Illinois at Urbana-Champaign

**College of Liberal Arts & Sciences**

Student Academic Affairs Office  
2002 Lincoln Hall, MC-446  
702 S. Wright St.  
Urbana, IL 61801

July, 2024

Dear Varenya :

**Congratulations!** On behalf of the College of Liberal Arts & Sciences, I am very pleased to inform you that your academic performance has qualified you for Spring 2024 Dean's List recognition. Achieving this honor distinguishes you among the excellent student body at Illinois. Your grade point average places you within the top 20 percent of our college.

In recognition of your accomplishment, the honorary designation of "Dean's List" for Spring 2024 will be noted on your official University of Illinois transcript. In addition, your name will be posted on the University of Illinois News Bureau's website later this spring: [news.illinois.edu](http://news.illinois.edu).

The criteria for Dean's List for the academic year 2023-2024 are as follows:

The Dean's List is prepared each semester to honor all full-time students whose grade-point average (GPA) for that semester ranks in the upper 20 percent of their college. The minimum GPA establishing eligibility for the LAS Dean's List in 2023-2024 is 3.93. To be eligible for Dean's List recognition, students must complete at least 12 academic semester hours taken for a letter grade (A through F). This excludes courses graded credit/no credit, satisfactory/unsatisfactory, and test-based credit that is graded pass/fail. No consideration is given for the Dean's List until all final traditional grades are posted.

Students who are registered with the Center for Wounded Veterans (CWV) or with Disability Resources and Educational Services (DRES) who are enrolled less than 12 but a minimum of nine graded semester hours who are in the top 20 percent of their college are also eligible.

Learn more about College Distinctions online: [las.illinois.edu/students/honors/distinctions](http://las.illinois.edu/students/honors/distinctions).

As a result of your academic success, you are invited and encouraged to work with the campus' National and International Scholarships Program to consider applying for various awards. These scholarships are particularly suited for students interested in eventually pursuing a graduate degree. You may visit [topscholars.illinois.edu](http://topscholars.illinois.edu) and contact the office directly to discuss scholarship opportunities that are appropriate for you.

The College of LAS is proud of your accomplishments! We applaud your efforts and wish you continued success in the attainment of your educational goals.

Sincerely,

A handwritten signature in black ink that reads "Venetria K. Patton".

**Venetria K. Patton**  
Harry E. Preble Dean  
College of Liberal Arts & Sciences  
VKP:ab

# Varenya Jain

646-306-1672 | [varenya.jj@gmail.com](mailto:varenya.jj@gmail.com) | [Linkedin](#) | [Github](#) | [Website](#)

## EDUCATION

### University of Illinois Urbana-Champaign

Class of 2024

Bachelor of Science in Integrative Biology, Minor in Computational Science and Engineering

- Dean's List (Spring 2024)
- Biology: Physiology, Anatomy, Bioinspiration, Evolution, Genetics, Behavioral Genetics, Genomics and Human Health, Biostatistics, Analysis of Biological Data in R, Programming for Genomics
- Electrical Engineering: Analog Signal Processing, Computer Systems and Programming, Digital Signal Processing, Probability with Engineering Applications, Applied Parallel Programming

## EXPERIENCE

### Contract Bioinformatician

December 2022 - Present

Advanced Genomics Institute and Laboratory Medicine

Paramus, NJ

- Configured and optimized the Illumina ICA platform and BaseSpace Sequence Hub to enhance genomic data processing and bioinformatics capabilities.
- Established FTP-based data transfer workflows, ensuring seamless real-time collaboration and efficient management of large genomic datasets.
- Spearheaded DRAGEN Enrichment workflows, improving variant calling pipelines for enhanced precision in genomic analyses.
- Led the Copy Number Variation (CNV) analysis, setting up DRAGEN Baselines for structural variant detection and integrating results into the Franklin by Genoxx platform for downstream interpretation.

### Student Researcher

Jan 2024 - December 2024

Illinois Natural History Survey - Tan Lab of Biodiversity Genomics

Urbana, IL

- Engaged in a phylogenomic study focusing on homology and ultraconserved elements within zebrafish genomes, enhancing understanding of evolutionary relationships.
- Managed and optimized bioinformatics workflows for high-throughput sequencing data analysis, involving quality control with FastQC, data trimming with BBduk, and sequence assembly using SPAdes, Phyluce, Scipio, and MEGAHIT to enhance phylogenomic reconstruction accuracy using the CAPTUS toolkit.
- Contributed to developing and refining a streamlined protocol for phylogenomic data assembly, extraction, and analysis, while actively engaging in lab meetings to address technical challenges and strategize research methodologies.

### Student Researcher

Aug 2023 - Dec 2023

Illinois Natural History Survey - Tan Lab of Biodiversity Genomics

Urbana, IL

- Conducted comprehensive research on morphological diversity in Sisoridae and Amphiliidae catfish adaptation to fast water environments.
- Utilized Geometric Morphometrics in R Studio for precise quantification of body shape differences. Applied PCA and GPA statistical methods to explore correlations and variations.
- Presented comprehensive results, highlighting significant differences in shape disparity in relation to adhesive organ presence. Provided insights into the evolutionary constraints on body shape in fastwater-adapted catfishes.

### Student Researcher | Metabolomics & Proteomics Core Facilities

June 2023 – Aug. 2023

Roy J. Carver Biotechnology Center

Urbana, IL

- Developed an untargeted metabolomics pipeline for post-processing, Quality Assurance/Quality Control, and data analysis.
- Provide support for data processing on LC-MS Untargeted Metabolite Profiling.
- Refine untargeted analyses to support various biomedical studies at the Carver Metabolomics Core.

### SLC Conference Planning Committee | IEEE

Jan 2022 - Jan 2023

The Institute of Electrical and Electronics Engineers

Chicago, IL

- Assisted in organizing and managing conference logistics, scheduling, and coordinating with stakeholders.
- Worked with Midwest Region 4 Committee members to develop and implement strategies for a successful conference.
- Proactively identified and addressed potential challenges during the planning process.

### SPIN Research Intern | The NEAT Project v4.0

Aug. 2021 – Sep. 2022

National Center for Supercomputing Applications

Urbana, IL

- Designed NGS toolkit for HAL Cluster, increasing Parallel Processing speed by 7%, and streamlined Empirical Mutation and Sequencing Models. Revised bash scripts and input flags.
- Managed genomic pipeline (FASTA, FASTQ, SAM, BAM, VCF) and implemented relevant Bioinformatics algorithms: Smith-Waterman, BLAST, Localized String Alignment.
- Presented HPC for Computational Genomics findings at the NCSA Exhibition during UIUC's Engineering Open House and the REU FoDOMMaT/SPIN Showcases.

<b>Outreach Committee Lead   Pulse 2022</b>	Aug 2021 - Aug 2022
<i>Department of Electrical and Computer Engineering</i>	<i>Urbana, IL</i>
<ul style="list-style-type: none"> <li>• Coordinated materials and event location logistics with the ECE department during COVID-19.</li> <li>• Develop Software to exhibit Computer Engineering principles: C++ Data Structures, Polymorphism, Command-Line Interface, Stack/Heap Memory Management, Address Space, etc.</li> <li>• Design Hardware activities to guide freshmen through simulated Electrical Engineering projects: series vs parallel circuits, Pulse Width Modulation motor control, Thermistor and LDR implementation</li> </ul>	

<b>Phys 211 Experienced Learning Assistant</b>	Jan 2021 - Dec 2021
<i>Loomis Laboratory Of Physics</i>	<i>Urbana, IL</i>
<ul style="list-style-type: none"> <li>• Continued the study of Physics pedagogy by instructing PHYS 211 labs, providing guidance to 30+ students.</li> <li>• Answered student questions, clarified critical lecture material, and built/corrected IOLab setups.</li> <li>• Developed advanced teaching, communication, and leadership skills through interactions with students and seminar presentations alongside lab staff.</li> </ul>	

<b>IOT Research Lab Assistant   Caesar Lab</b>	Jan. 2021 – Jun. 2021
<i>Coordinated Science Laboratory</i>	<i>Urbana, IL</i>
<ul style="list-style-type: none"> <li>• Remodeled a Reinforcement Learning System intended for UAV-Assisted Emergency Response.</li> <li>• Upgraded the fully-distributed communication environment for USAF usage.</li> <li>• Implemented communication trees via Python Message Passing Interface (MPI) standard in under 6 months.</li> </ul>	

<b>Research Intern   Drs. Spitalnik, Hod, La Carpia</b>	June 2018 - Aug 2018
<i>Columbia University Medical Center - Lab of Transfusion Biology</i>	<i>New York, NY</i>
<ul style="list-style-type: none"> <li>• Investigated the effects of transfusional iron overload on gut microbiota due to intravenous infusion.</li> <li>• Conducted initial studies with a mouse model to retrieve data on iron-deficient erythropoiesis in blood donors and red blood cell recovery after transfusion of hematopoietic red blood transplant.</li> <li>• Utilized basic Spearman Correlation meta-analysis of bacteria communities, performed blood analysis tests, and used Flow Cytometry to collect sample cell data.</li> </ul>	

## HARDWARE PROJECTS

---

<b>BadgeDev   SIGPWNY</b>	Aug. 2024 – Dec 2024
<ul style="list-style-type: none"> <li>• Aid in software development for FallCTF participant, sponsor, and staff badges</li> <li>• Assemble and perform QC on Pico soldering and test firmware before distribution</li> <li>• Fix and replace badge firmware/hardware during event</li> </ul>	

<b>AM Radio   ECE 210</b>	Jan. 2022 – May 2022
<ul style="list-style-type: none"> <li>• Construct a functioning AM Radio using a Superheterodyne Receiver; Convert digital .wav audio input to analog 3.5mm speaker output</li> <li>• Utilize Fourier Transforms to convert Time domain signals to Frequency domain responses</li> </ul>	

<b>Virtual Gloves   ECE 120 Honors</b>	Jan. 2021 – May 2021
<ul style="list-style-type: none"> <li>• Develop a “virtual keyboard” by moving fingers attached to flex sensors and provide haptic feedback</li> <li>• Collect data from flex resistors in Arduino Studio and use C++ to send output confirmations to LEDs</li> </ul>	

<b>LoopKit   Personal</b>	July. 2020 – Sep. 2021
<ul style="list-style-type: none"> <li>• Adapt the LoopKit source code to customize a program for personal monitoring functionalities</li> <li>• Modify Swift code to circumvent authentication requisites for bolus delivery within the Loop system</li> <li>• Implement precision adjustments to default carb absorption parameters, optimizing glycemic control dynamics</li> </ul>	

**Student Organization Membership:** IEEE@UIUC, iRobotics, MRDC, Vex Robotics, ACM, SIGPWNY

## LEADERSHIP

---

### Conferences

- NCSA Letter of Outstanding Student Leadership Recognition (2023-2024)
- NCSA 2nd Annual Student Research Conference - Planning Committee, Panel Moderator, and Industry Chair (2024)
- ACMG Annual Clinical Genetics Meeting (Metro Toronto Convention Center) - Attendee, 2024
- ISPD 28th International Conference on Prenatal Diagnosis and Therapy (Westin Copley Place, Boston) - Attendee, 2024
- NCSA Student Research Conference - Planning Committee and Panel Moderator (2023)
- IEEE Nexus Region 4 Student Leadership Conference - Conference Planning Lead (2022-2023)

### Presentations

- University of Illinois - Undergraduate Research Symposium (2024)
- National Center for Supercomputing Applications - SPIN Lightning Talks (2021-2022)
- National Center for Supercomputing Applications - Engineering Open House (2021)

### Teaching

- ECE PULSE - Outreach Committee Lead (2022)
- Loomis Laboratory Of Physics - “Expert Learning Assistant” for Mechanics (2021)

### Volunteering

- UIUC ACM SIGPWNY - Contributed to solving cybersecurity “Capture the Flag” competitions and developed badge software for FallCTF
- De Dilse Charitable Inc. - Co-Founder and Lead Volunteer (2019-2022)
- Kaplen Jewish Community Center on the Palisades - Senior Science Counselor (2019)
- American Wheat Mission - Volunteer (2018)

## TECHNICAL SKILLS

---

**Lab Equipment:** Flow Cytometer, Clinical Benchtop Centrifuge, ThermoFisher Q-Exactive Mass Spectrometer

**Bioinformatics Tools:** BLAST+, Smith-Waterman Algorithm, Rstudio, Biopython, Chromosome Analysis Suite (ChAS), OMIM, UCSC Genome Browser, Ensembl, Illumina Connected Analytics, BaseSpace SequenceHub (DRAGEN)

**Analytics Tools:** Google Cloud (SDK, Analytics, BigQuery)

**Developer Tools:** Git, GitHub, GitLab, Docker, VirtualBox, Kernel-based Virtual Machines (KVMs), Digital Ocean Droplets, VSCode, IntelliJ

**Programming Experience:** Unix Command Line Interface, Bash shell scripting, C/C++, CUDA, Python MPI, R

**Development Hardware and Electronics:** Breadboards, Raspberry Pi (Zero/3b/4), Arduino UNO, Intel MAX 10 FPGA

**Electrical Engineering Fundamentals:** Oscilloscope, Network Analyzer, RLC circuits, Operational Amplifiers, Band-Pass Filters, Quartus Prime utility, ModelSim utility, SystemVerilog

**GNU/Linux:** Debian, Fedora, CentOS 8.5, AlmaLinux 8.7, Arch Linux, Ubuntu Server 22.04, Manjaro, EndeavorOS



## NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

1010 NCSA Building, MC-257  
1205 W. Clark St.  
Urbana, IL 61801

Offer Letter

August 6, 2021

Dear Mr. Varenya Jain,

On behalf of Dr. Christina Fliege, Mr. Joshua Allen, and the National Center for Supercomputing Applications (NCSA), it is my pleasure to offer you a position in the Academic Year 2021-22 Students Pushing Innovation (SPIN) Internship Program. You will be working on the *Development of NEAT Advanced Features* project. The program will begin on August 23, 2021 and conclude on December 8, 2021. Through the program, you will work 5 hours a week and will be paid \$12.50 per hour as an Undergraduate Hourly employee.

One of the SPIN program requirements is that you need to be a full-time student enrolled in at least 12 credit hours per term. As a SPIN Intern, you are required to submit a research plan and research report on the dates indicated by Olena Kindratenko, NCSA Senior Research Coordinator, give one SPIN lightning talk and participate in the next year's Engineering Open House. You are required to attend all scheduled SPIN Lightning Talks. You are expected to meet with your SPIN mentor regularly. You should also plan on participating in various NCSA and SPIN activities, including a one-on-one meeting with Olena Kindratenko and other professional development and/or social events.

It is your responsibility to regularly report your work hours in the NCSA Biweekly Timesheets reporting tool. If you cannot continue your work on the project for two consecutive weeks without notifying your SPIN mentor or the NCSA Senior Research Coordinator, the SPIN program may choose to terminate your internship.

The SPIN interns are required to follow NCSA's COVID-19 policies and be prepared to work in person if NCSA adopts such policies for its staff. More updates will follow.

You will share a workspace in the NCSA building and will be provided with the appropriate computer system, if needed. It is the policy of the University of Illinois and NCSA that reasonable accommodations be made for persons with disabilities. Should you need accommodations, please notify NCSA Human Resources.

Please let us know if you accept or decline this offer by August 10, 2021.

If you have any questions, please email us at [SPIN@ncsa.illinois.edu](mailto:SPIN@ncsa.illinois.edu). We look forward to hearing from you with your decision.

Sincerely,  
Olena Kindratenko  
Senior Research Coordinator  
National Center for Supercomputing Applications

I accept this offer: \_\_\_\_\_ Date: \_\_\_\_\_

I decline this offer: \_\_\_\_\_ Date: \_\_\_\_\_



## NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

1010 NCSA Building, MC-257  
1205 W. Clark St.  
Urbana, IL 61801

Offer Letter

August 6, 2021

Dear Mr. Varenya Jain,

On behalf of Dr. Christina Fliege, Mr. Joshua Allen, and the National Center for Supercomputing Applications (NCSA), it is my pleasure to offer you a position in the Academic Year 2021-22 Students Pushing Innovation (SPIN) Internship Program. You will be working on the *Development of NEAT Advanced Features* project. The program will begin on August 23, 2021 and conclude on December 8, 2021. Through the program, you will work 5 hours a week and will be paid \$12.50 per hour as an Undergraduate Hourly employee.

One of the SPIN program requirements is that you need to be a full-time student enrolled in at least 12 credit hours per term. As a SPIN Intern, you are required to submit a research plan and research report on the dates indicated by Olena Kindratenko, NCSA Senior Research Coordinator, give one SPIN lightning talk and participate in the next year's Engineering Open House. You are required to attend all scheduled SPIN Lightning Talks. You are expected to meet with your SPIN mentor regularly. You should also plan on participating in various NCSA and SPIN activities, including a one-on-one meeting with Olena Kindratenko and other professional development and/or social events.

It is your responsibility to regularly report your work hours in the NCSA Biweekly Timesheets reporting tool. If you cannot continue your work on the project for two consecutive weeks without notifying your SPIN mentor or the NCSA Senior Research Coordinator, the SPIN program may choose to terminate your internship.

The SPIN interns are required to follow NCSA's COVID-19 policies and be prepared to work in person if NCSA adopts such policies for its staff. More updates will follow.

You will share a workspace in the NCSA building and will be provided with the appropriate computer system, if needed. It is the policy of the University of Illinois and NCSA that reasonable accommodations be made for persons with disabilities. Should you need accommodations, please notify NCSA Human Resources.

Please let us know if you accept or decline this offer by August 10, 2021.

If you have any questions, please email us at [SPIN@ncsa.illinois.edu](mailto:SPIN@ncsa.illinois.edu). We look forward to hearing from you with your decision.

Sincerely,  
Olena Kindratenko  
Senior Research Coordinator  
National Center for Supercomputing Applications

I accept this offer: Varenya Jain

Date: 08/06/2021

I decline this offer: \_\_\_\_\_ Date: \_\_\_\_\_



## NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

1010 NCSA Building, MC-257  
1205 W. Clark St.  
Urbana, IL 61801

Reappointment Letter  
May 13, 2022

Dear Varenya Jain,

On behalf of Mr. Joshua Allen, Dr. Christina Fliege, and the National Center for Supercomputing Applications (NCSA), it is my pleasure to offer you a position in the Summer 2022 Students Pushing Innovation (SPIN) Internship Program. You will be working on the *Development of NEAT Advanced Features* project. The program will begin on June 6 and conclude on July 29, 2022. Through the program, you will work 20 hours a week and will be paid \$15.00 per hour as an Undergraduate Hourly employee.

As a SPIN Intern, you are required to submit a research plan and a research report on the dates indicated by the NCSA Senior Research Coordinator and give one SPIN lightning talk. You are required to attend all scheduled SPIN Lightning Talks. You are expected to meet with your SPIN mentor regularly. You should also plan on participating in various NCSA and SPIN activities, including one-on-one meeting with Olena Kindratenko, the Summer 2022 NCSA Undergraduate Research Poster Session, and other professional development and/or social events.

It is your responsibility to regularly report your work hours in the NCSA Biweekly Timesheets reporting tool. If you cannot continue your work on the project for two consecutive weeks without notifying your SPIN mentors or the NCSA Senior Research Coordinator, the SPIN program may choose to terminate your internship. If you are not responding to your SPIN mentors or the NCSA Senior Research Coordinator emails for three consecutive weeks, your internship may be terminated as well.

You will share a workspace in the NCSA building and will be provided with the appropriate computer system if needed. It is the policy of the University of Illinois and NCSA that reasonable accommodations be made for persons with disabilities. Should you need accommodations, please notify NCSA Human Resources.

Please let us know if you accept or decline this offer by May 17, 2022.

If you have any questions, please email us at [SPIN@ncsa.illinois.edu](mailto:SPIN@ncsa.illinois.edu). We look forward to hearing from you with your decision.

Sincerely,  
Olena Kindratenko  
Senior Research Coordinator  
National Center for Supercomputing Applications

I accept this offer: \_\_\_\_\_ Date: \_\_\_\_\_

I decline this offer: \_\_\_\_\_ Date: \_\_\_\_\_



## NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

1010 NCSA Building, MC-257  
1205 W. Clark St.  
Urbana, IL 61801

Reappointment Letter  
May 13, 2022

Dear Varenya Jain,

On behalf of Mr. Joshua Allen, Dr. Christina Fliege, and the National Center for Supercomputing Applications (NCSA), it is my pleasure to offer you a position in the Summer 2022 Students Pushing Innovation (SPIN) Internship Program. You will be working on the *Development of NEAT Advanced Features* project. The program will begin on June 6 and conclude on July 29, 2022. Through the program, you will work 20 hours a week and will be paid \$15.00 per hour as an Undergraduate Hourly employee.

As a SPIN Intern, you are required to submit a research plan and a research report on the dates indicated by the NCSA Senior Research Coordinator and give one SPIN lightning talk. You are required to attend all scheduled SPIN Lightning Talks. You are expected to meet with your SPIN mentor regularly. You should also plan on participating in various NCSA and SPIN activities, including one-on-one meeting with Olena Kindratenko, the Summer 2022 NCSA Undergraduate Research Poster Session, and other professional development and/or social events.

It is your responsibility to regularly report your work hours in the NCSA Biweekly Timesheets reporting tool. If you cannot continue your work on the project for two consecutive weeks without notifying your SPIN mentors or the NCSA Senior Research Coordinator, the SPIN program may choose to terminate your internship. If you are not responding to your SPIN mentors or the NCSA Senior Research Coordinator emails for three consecutive weeks, your internship may be terminated as well.

You will share a workspace in the NCSA building and will be provided with the appropriate computer system if needed. It is the policy of the University of Illinois and NCSA that reasonable accommodations be made for persons with disabilities. Should you need accommodations, please notify NCSA Human Resources.

Please let us know if you accept or decline this offer by May 17, 2022.

If you have any questions, please email us at [SPIN@ncsa.illinois.edu](mailto:SPIN@ncsa.illinois.edu). We look forward to hearing from you with your decision.

Sincerely,  
Olena Kindratenko  
Senior Research Coordinator  
National Center for Supercomputing Applications

I accept this offer: Varenya Jain Date: 05/17/2022

I decline this offer: \_\_\_\_\_ Date: \_\_\_\_\_

# The NEAT Project v3.2, The NExt-generation sequencing Analysis Toolkit

Varenya Jain, Joshua Allen, Christina Fliege

National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign

## INTRODUCTION

The goal of the project is to further develop a genomic read simulator. We are working on improving, debugging, and optimizing the main utilities, which give access to simulations of statistical properties of genetic data, visualization tools for viewing and troubleshooting parameters, and data sets to help shape the simulations.

My goal as an intern is to take advantage of the opportunities to learn Python, data visualization, and software production. My end goal is to make a meaningful contribution by designing and testing packages while keeping my changes well documented.

## Computational Genomics Using A Fine-Grained Read Simulator

### Background

Previous research by Stephens, Zachary, et al with the NEAT toolkit revealed that there is a great lack of "ground truth" in sequencing models, meaning our processes of validating and benchmarking methods are limited. Additionally, there is a great struggle with donor privacy: we need more test datasets to improve our analysis and understanding of genomic data, yet there is no collection of free and publicly available human genome datasets.

While read simulators can fulfill this requirement, their limited flexibility and limited resemblance to true data poses a roadblock. Therefore, the NEAT toolkit v1.0 was developed to provide researchers with an "easy-to-use simulator" along with a variety of other tools.

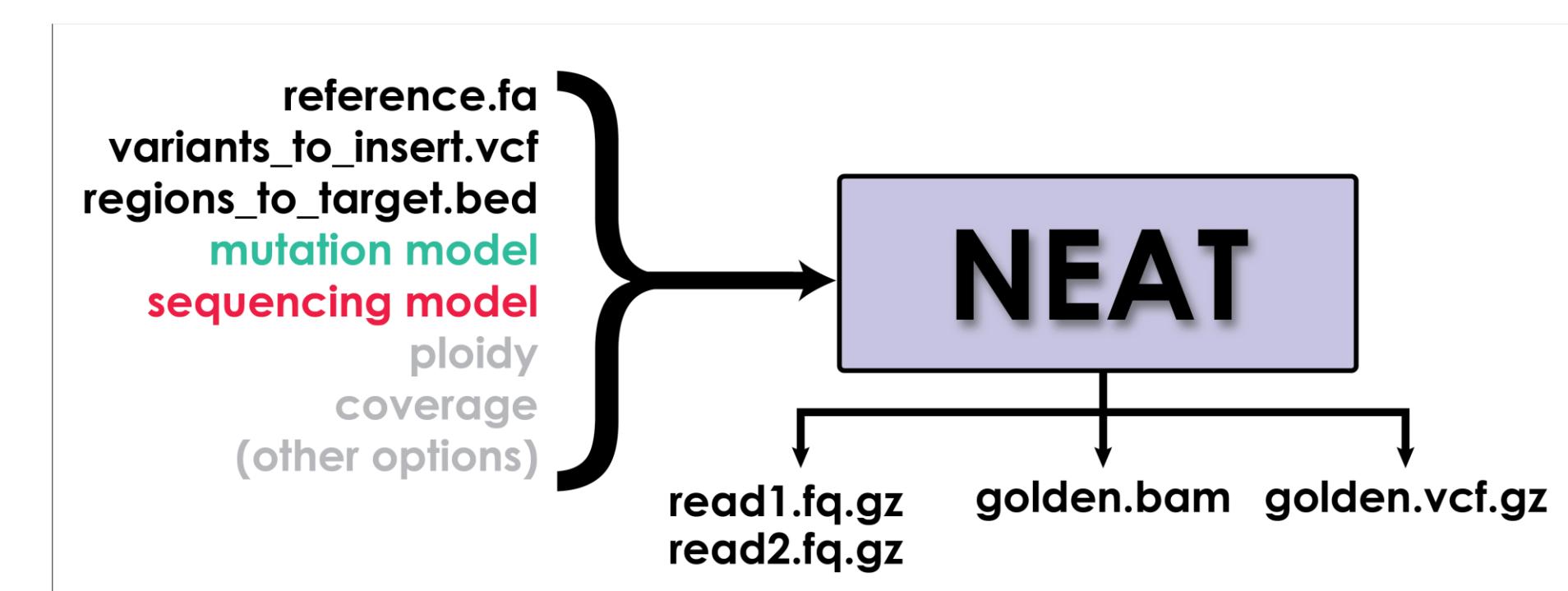
NEAT is a fine-grained read simulator. It simulates real looking data using models learned from specific datasets. There are several supporting utilities for generating models used for simulation and for comparing the outputs of alignment and variant callers to the golden BAM and golden VCF produced by NEAT. NEAT is one of the few simulators to generate a BAM file, which is a file that gives information on where each read matches up with the reference.

NEAT's core functionality is invoked using the `gen_reads.py` command. The simplest command can produce a single ended fastq file with Read Length, Ploidy, and Coverage specified by 3 input parameters. The real strength of NEAT lies in the ability for the user to customize many sequencing parameters, produce 'golden,' true positive datasets.

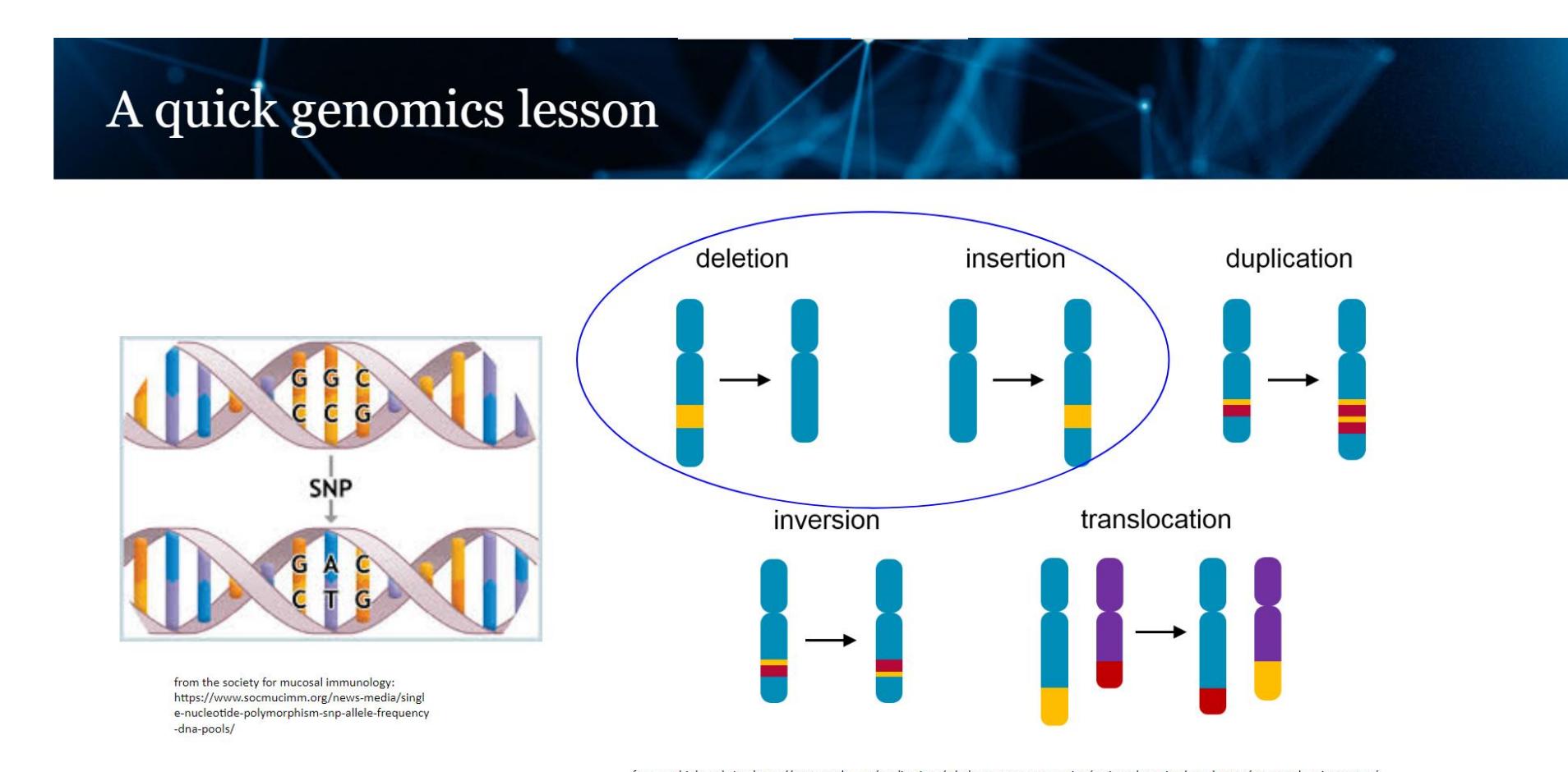
## Pipeline

In general, NEAT takes multiple arguments, including but not limited to:

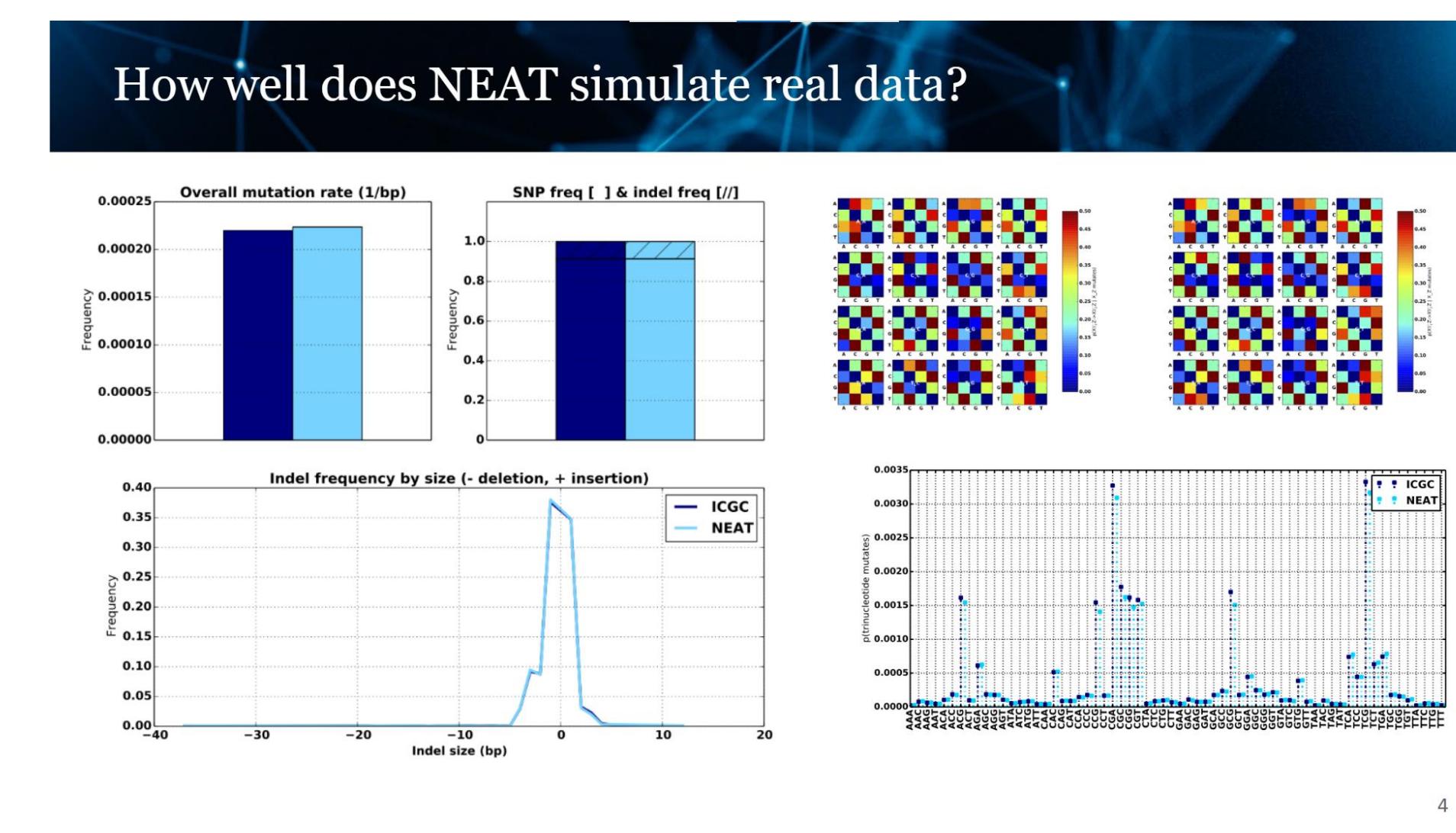
- A "reference file" of nucleotide sequences in the FASTA format.
- A VCF file storing gene sequence variations
- Sequencing Read Length - number of base pairs sequenced from a DNA fragment
- Sample Ploidy (the number of complete sets of chromosomes in a cell)



Some common mutations shown below



In the graphics below, you can see that NEAT performs well simulating Single Nucleotide Polymorphism (SNP) and Insertion/Deletion Polymorphism (INDEL) variations. Heat maps of real data vs NEAT simulated data show minor differences.



## PROGRESSION

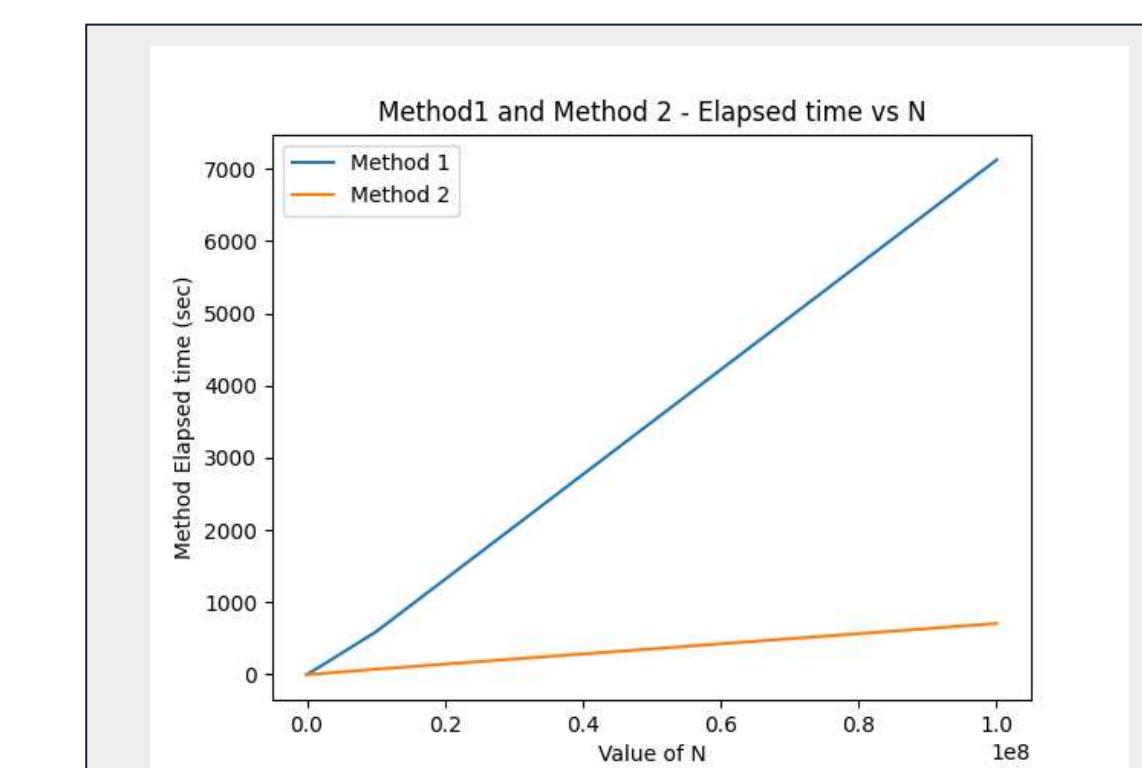
### NEAT Assignments

In order to further develop NEAT, I have taken different routes to improve its capabilities. Some modifications I have completed include:

1. Testing Utilities
2. Parallelization Runtime
3. Local BLAST Development
4. BLAST/Smith-Waterman Integration
5. Continuous Integration Updates

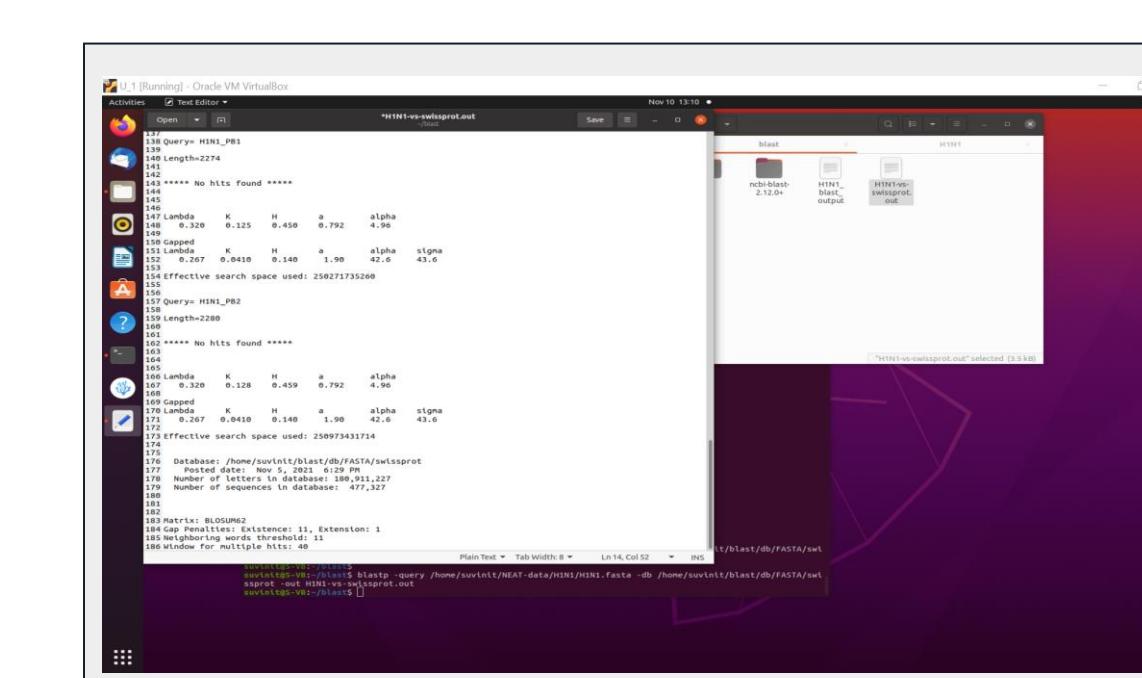
```
p(GAA --> GGA | GAA mutates) = 0.9651162790697765
p(GAA --> GTA | GAA mutates) = 0.023255813953488372
p(GAC --> GCC | GAC mutates) = 0.021739130434782608
p(GAC --> GGC | GAC mutates) = 0.8478260869565217
p(GAG --> GGG | GAG mutates) = 0.130434782608695655
p(GAG --> GTG | GAG mutates) = 0.9333333333333333
p(GAT --> GCT | GAT mutates) = 0.066666666666666667
p(GAT --> GCT | GAT mutates) = 0.01639344262295082
p(GAT --> GTT | GAT mutates) = 0.9016393442622951
p(GAT --> GTT | GAT mutates) = 0.08196721311475409
p(GCA --> GAA | GCA mutates) = 0.13157894736842105
p(GCA --> GGA | GCA mutates) = 0.13157894736842105
p(GCA --> GTA | GCA mutates) = 0.7368421052631579
```

A small look at the output of testing `gen_mut_model.py`. The actual output can be a couple hundred lines which can tell a researcher the overall mutation rate and total variants processed.



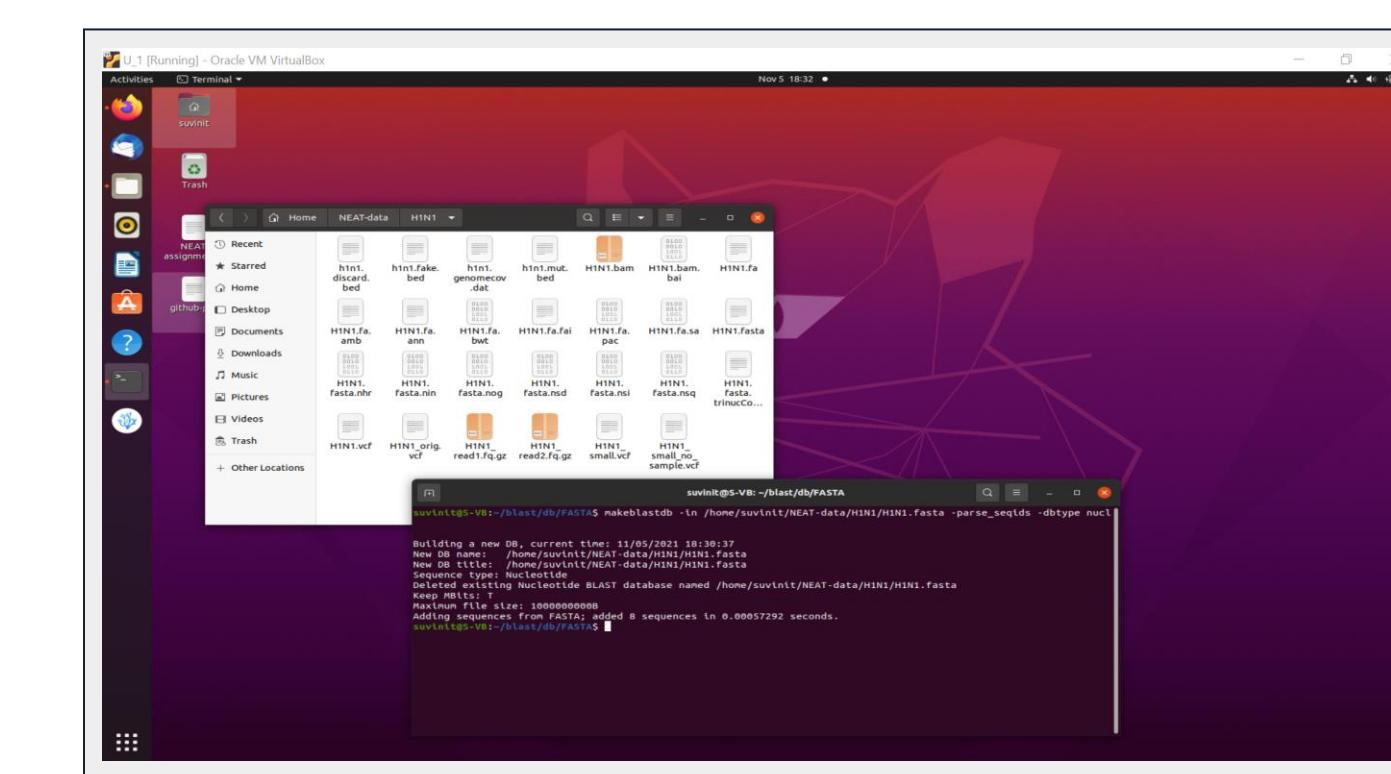
A Comparison of Parallelization Runtime Methodologies

High-Performance Biological Computing requires lots of processing power and time to run datasets. Therefore, we can save time by running many tasks at once.



An example output of the BLAST algorithm

This method produces a quick alignment and comparison file



A command-line view of making databases out for FASTA files for use in BLAST comparisons

## RESULTS

The final training / testing accuracies of the current model are not yet available. The goal of this project is not only to obtain the highest testing performance, rather it is to also to gauge the efficiency of NEAT when processing a myriad of datasets and parameters.

As of its current version (3.2), some of the new features of NEAT allows the user to:

- customize the read length for analysis
- introduce random mutations at will, pre-determined mutations from a VCF file
- emulate multi-ploidy heterozygosity
- simulate targeting sequencing via BED input specifications
- introduce sequencing substitution errors

More features are coming soon (parallelization, lower memory utilization, etc.) as we continue to update and upgrade our current model.

## CONCLUSIONS AND ACKNOWLEDGEMENTS

There is still much work to be done, especially with parallelization, polyploidy, and tumor simulation

Colleagues: Keshav Gandhi, Raghid Alhazmy, Yash Wasnik

This work was completed as part of the NCSA SPIN program

- [1] Stephens, Zachary D., Matthew E. Hudson, Liudmila S. Mainzer, Morgan Taschuk, Matthew R. Weber, and Ravishankar K. Iyer. "Simulating next-generation sequencing datasets from empirical mutation and sequencing models." *PloS one* 11, no. 11 (2016): e0167047.
- [2] <https://github.com/ncaa/NEAT>
- [3] <https://wiki.ncsa.illinois.edu/pages/viewpage.action?spaceKey=LH&title=varenya3++NCSA+SPIN+-+NEAT+Assignments>



## NATIONAL CENTER FOR SUPERCOMPUTING APPLICATIONS (NCSA)

1010 NCSA Building, MC-257  
1205 W. Clark St.  
Urbana, IL 61801

Recognition Letter

May 23, 2024

Dear Varenya Jain,

On behalf of the National Center for Supercomputing Applications (NCSA), it is my pleasure to present you with this recognition letter to acknowledge your accomplishments during the Academic Year 2023-24 Students Pushing Innovation (SPIN) internship program. You have been nominated by Ms. Olena Kindratenko, NCSA Senior Research Coordinator, for your outstanding student leadership.

For the past two years, you have excelled as co-chair of the NCSA Student Research Conference and skillfully moderated the conference mixed panel that comprised industry representatives, students, and faculty. You have demonstrated exceptional communication skills, attention to detail, and ability to appreciate diverse perspectives that have significantly enriched our academic community. Your outstanding work was also noticed by Alaina Kanfer from the NCSA Strategic Partnerships program, as well as the industry representatives. Your boundless enthusiasm, inquisitive nature, and unwavering work ethic make you a true asset to the SPIN program and NCSA.

Thank you for your dedication and contributions that you have made in the course of your internship. I wish you the best in your future endeavors.

Sincerely,

A handwritten signature in blue ink that reads "William Gropp".

William Gropp  
Director, NCSA  
Grainger Distinguished Chair in Engineering  
Department of Computer Science  
University of Illinois Urbana-Champaign