SCTR's Pune Institute of Computer Technology
Dhankawadi, Pune

AN INTERNSHIP REPORT ON

Implementation of new features and improving the functionality of PretextView.

SUBMITTED BY
Name: Atharv Udayraj Fakatkar
Class: TE 2
Roll no: 31221

Under the guidance of
Prof. Kopal Gangrade

DEPARTMENT OF COMPUTER ENGINEERING
ACADEMIC YEAR 2024-25

DEPARTMENT OF COMPUTER ENGINEERING

SCTR's Pune Institute of Computer Technology
Dhankawadi, Pune
Maharashtra 411043

CERTIFICATE

This is to certify that the SPPU Curriculum-based internship report entitled
"Implementation of new features and improving the functionality of PretextView."

Submitted by
Atharv Udayraj Fakatkar
(Exam No. 31221)

has satisfactorily completed the curriculum-based internship under the guidance of *Prof. Kopal Gangrade* towards the partial fulfillment of third year Computer Engineering Semester VI, Academic Year 2024-25 of Savitribai Phule Pune University.

Prof. Kopal Gangrade
Internship Guide
PICT, Pune

Dr. G. V. Kale
Head
Department of Computer Engineering
PICT, Pune

Place:
Date:

Acknowledgement

It gives me great pleasure in presenting the internship report on "Implementation of new features and improving the functionality of PretextView.".

First of all I would like to take this opportunity to thank my internship guide Prof. Kopal Gangrade for giving me all the help and guidance needed. I am really grateful for her/his kind support and valuable suggestions that proved to be beneficial in the overall completion of this internship.

I am thankful to our Head of Computer Engineering Department, Dr. G.V.Kale, for her indispensable support and suggestions throughout the internship work.

I would also genuinely like to express my gratitude to the Department Internship Coordinator, Prof.P.P.Joshi, for her constant guidance and support and for the timely resolution of the doubts related to the internship process.

Finally, I would like to thank my mentor, Jonathan Wood for his constant support during the overall internship process. It would not have been possible without his help.

Contents

List of Figures

1 Title

Implementation of new features and improving the functionality of PretextView.

2 Introduction

This report details my internship experience working on the PretextView project, where I was responsible for implementing new features and improving its functionality. PretextView is a visualization tool designed for the manual curation of Hi-C maps, originally developed by Ed Harry under the supervision of Zemin Ning from the High Performance Assembly Group, LIMS and Informatics Team at the Wellcome Sanger Institute, UK.

As part of the **Future Innovators Mentorship Program**, I had the opportunity to collaborate with the development team and end users to enhance the tool's usability and performance. My contributions were guided by mentor Jonathan Wood, who provided a wishlist of requested features, helping to prioritize key improvements. Through requirement gathering, discussions with end users of the Genome Reference Informatics Team (GRIT), and iterative development, I worked on optimizing PretextView to better serve its intended purpose.

This report outlines the objectives of the internship, the methodologies followed, the challenges faced, and the solutions implemented. It also highlights the impact of the enhancements on the tool's functionality and future directions for further development.

3 Problem Statement

PretextView is a desktop application tailored for visualizing and editing Hi-C contact maps, offering real-time functionalities such as breaking and joining scaffolds, as well as re-orienting genomic segments. However, there were existing gaps in the application's feature set and areas where enhancements and quality of life improvements were needed. This project aimed to address these shortcomings through collaborative efforts with end users to implement new features, thereby advancing the capabilities of the desktop application.

4 Objectives and Scope

- **Bug Fixes**
  - Stability improvements – fix random crashing.
- **Feature Enhancements**
  - Toggle extensions (tracks) with keyboard shortcuts.
    - Currently available in the submenu, but adding a top-level menu shortcut would improve accessibility.
  - Update default metadata tags.
    - Add: HAP1, HAP2, Target, Contaminant, X1, X2, Y1, Y2, Z1, Z2, W1, W2, Roman numerals (I-V), B1, B2, B3, U.
    - Allow users to change the order.
  - Add horizontal waypoint markers.
  - Enable jumping to the main diagonal with a hotkey (e.g., pressing "J" when hovering over shrapnel).
  - Improve scaffold navigation when using the "Input Sequences" list.
    - Instead of placing the scaffold in the middle of the screen, highlight or zoom to the relevant region.
  - Add user profile settings.
    - Change default track color for each track.
    - Resolve metadata tag/map color clashes.
    - Modify app background color.
  - Enable sorting/undo sorting by metadata tag type.
  - Improve menu usability.
    - Adjust menu fonts and enable text resizing when switching monitors.
    - Implement a searchable menu to find input scaffolds and search directory contents for maps.
  - Support regional keyboard layouts (e.g., AZERTY).
  - Enable deletion of embedded tracks via CLI.
  - Implement a log file for each map to track time spent on curation.
  - Allow opening the directory from where a map was loaded when saving or loading a file.
  - Enable greying out of scaffolds tagged as haplotigs (should be switchable).
- **Keyboard Shortcuts & UI Improvements**
  - Restore the grid toggle (perhaps mapped to 'L').
  - Bind scaffold paint toggling to a keyboard key (currently only available in the top-level menu as "Scaffolds Always Visible").
  - Add background color options: Black, White, and Grey.
  - Enable tag-specific sorting by metadata tag.

- **Technical Improvements**
  - Increase scaffold grid limit (current limit: 10,000).
  - Increase resolution and zoom layers.
  - Enable opening multiple maps within the same instance of the program.
  - Dynamically display assembly statistics (e.g., L90, N90) to track how much more needs to be placed to reach 90% assignment.
  - Allow mirroring of edits between different map versions.
    - Example: Apply edits made in a high-resolution map to a low-resolution version for seamless switching during curation.
  - Enable scaling of Hi-C read mapping quality within the map.
    - Adjust between multi-mapping reads and unique reads.
  - Allow selection of multiple non-adjacent scaffolds.

5 Methodological Details

The development and enhancement of PretextView followed an iterative and structured approach based on the Agile methodology. At the beginning of the internship, I underwent a training session on Agile practices, which provided a strong foundation for managing tasks efficiently and adapting to evolving project requirements.

5.1 Agile Workflow

The Agile methodology was implemented throughout the internship, ensuring continuous development, feedback, and improvements. The project followed a two-week sprint cycle, with clearly defined objectives and deliverables. Each sprint concluded with a report meeting held every fortnight on Friday, where progress was reviewed, blockers were discussed, and the plan for the next sprint was outlined.

5.2 Sprint Structure and Training Sessions

Within each sprint, multiple sessions were conducted to enhance both technical and professional skills. These sessions covered a range of topics, including:

- **High-Performance Computing (HPC)** – Understanding parallel computing and optimizing code for large-scale data processing.
- **Software Development Best Practices** – Writing maintainable and efficient code, version control strategies, and debugging techniques.
- **Presentation and Communication Skills** – Structuring effective presentations, engaging technical audiences, and communicating project outcomes clearly.

5.3 Development Cycle

Each sprint cycle involved the following stages:

1. **Requirement Gathering and Planning** – Discussing feature requests, prioritizing tasks from the wishlist, and defining sprint goals.
2. **Implementation and Iterative Development** – Developing features, fixing bugs, and refining the software based on user feedback.
3. **Testing and Validation** – Conducting functional tests, reviewing code, and ensuring software stability and usability.
4. **Sprint Review and Retrospective** – Presenting progress in the sprint review meetings, gathering feedback, and improving processes for the next sprint.

By following this structured methodology, I was able to systematically implement new features and improve the functionality of PretextView while gaining valuable experience in Agile software development, HPC, and scientific communication.

6 Modern engineering tools used

The development and enhancement of PretextView involved the use of several modern tools and technologies to streamline collaboration, improve efficiency, and ensure software stability. These tools spanned version control, project management, development, testing, data visualization, and documentation.

6.1 Version Control & Collaboration

- **Git & GitHub** – Used for version control, collaborative development, and issue tracking. Code changes were managed through branches and pull requests.

6.2 Project Coordination

- **Slack & Outlook** – Facilitated team communication, discussions, and sprint coordination.
- **Zoom** – For sprint review meetings.

6.3 Development & Programming

- **C++ & OpenGL** – Core technologies used to develop and enhance PretextView, ensuring efficient rendering and interactivity.
- **CMake** – A build system generator to manage the compilation and linking of the project across different platforms.

7 Outcome/ results of internship work

A new enhanced version of PretextView was made available on the GitHub repository for open-source use, with the wish-list of features implemented.

7.1 implemented features

The implementation of the user profile feature allows users to customize default track colors, background colors, and resolve metadata tag color clashes. Figure 1 illustrates the user profile interface.

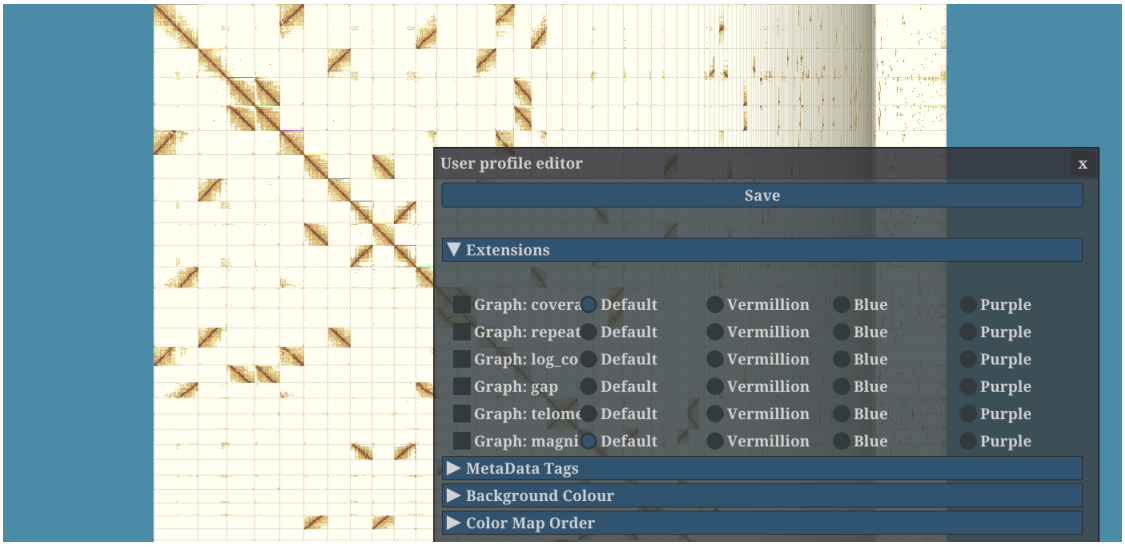


Figure 1: User profile implementation.

Additionally, the metadata tag system was updated to include new default tags such as `HAP1`, `HAP2`, `Target`, and `Contaminant`, with support for additional labels like `X1, X2, Y1, Y2, Z1, Z2`. The updated metadata tagging system is shown in Figure 2.

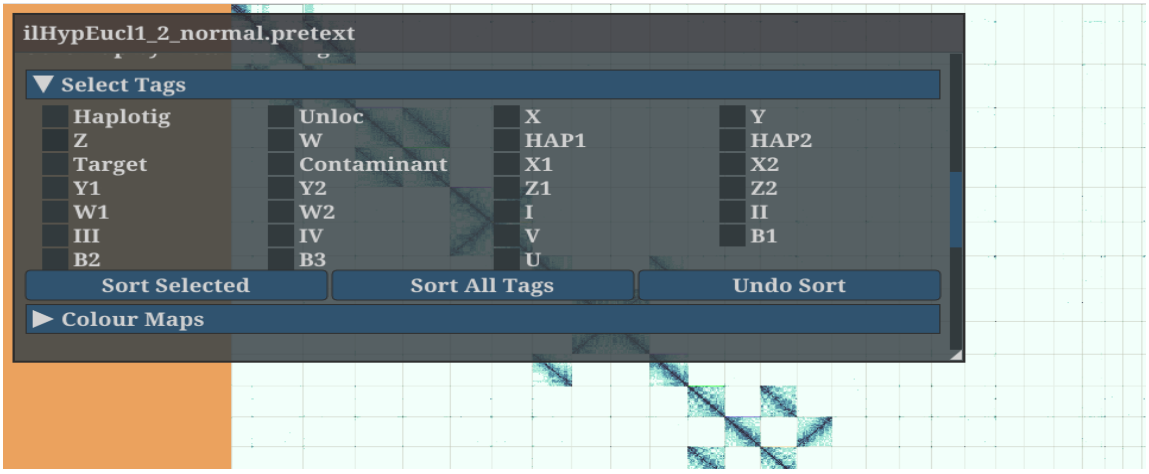


Figure 2: Metadata tags implementation.

8 Achievements

During the course of this internship, I was able to make significant contributions to the development and enhancement of PretextView, which were well recognized by the community. The key achievements from this internship are as follows:

- **Presentation at the Genome Informatics Conference:** I was given the prestigious opportunity to present my work on PretextView at the Genome Informatics Conference held at the Wellcome Sanger Institute, UK. This event, which took place in November, allowed me to showcase the features I had developed and interact with leading researchers in the field of bioinformatics. The trip was fully sponsored by the organization, and I spent 15 days at the Wellcome Genome Campus, engaging in discussions and gaining valuable insights from experts in genome assembly and visualization.
- **Recognition and Further Offer:** Following my contributions to the project, I was offered an opportunity to continue working on PretextView for an additional two years with an agreed-upon stipend. This offer was a testament to the impact of my work and the trust placed in my capabilities. However, I made the decision to decline this opportunity, as I did not want to narrow down my career scope into bioinformatics at such an early stage. Instead, I chose to keep my career path open to exploring a broader range of topics in system programming, artificial intelligence, and software development.
- **Technical Growth and Industry Exposure:** Throughout the internship, I enhanced my skills in software development, debugging, and optimization while working with large-scale genomic data. I also gained exposure to high-performance computing (HPC), Agile development practices, and effective team collaboration, all of which have strengthened my ability to contribute to complex software projects in the future.
- **International Networking and Collaborations:** One of the most valuable takeaways from this experience was the opportunity to build an international professional network. Through interactions with researchers, developers, and scientists from various institutions, I established connections that could open doors for future collaborations in interdisciplinary research and software development. Engaging with experts from the Wellcome Sanger Institute, EMBL-EBI, and other global research organizations provided me with fresh perspectives on genome informatics and software engineering best practices.

This internship provided me with an invaluable experience, allowing me to contribute meaningfully to an active research project, gain industry exposure, build a global professional network, and refine my software development skills while making an informed decision about my future career direction.

9 Internship Completion

12 March 2025

To Whom It May Concern,

Subject: Confirmation of Internship Completion – Atharv Fakatkar

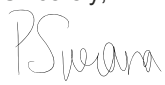
This letter is to confirm that Atharv Fakatkar (Pune Institute of Computer Technology, India) successfully completed an internship at Wellcome Sanger Institute from July to November 2024.

During this period, Atharv worked as part of the Biodata Developers (BioDev) Network's Future Innovators Mentorship scheme, contributing approximately 10 hours per week. His responsibilities included code contributions to the PretextView software for the Genome Reference Informatics Team in the Tree of Life programme at Sanger.

We appreciate Atharv's dedication and contributions during their time with us and wish them success in their future endeavours.

Please feel free to contact us if you require any further information.

Sincerely,



Priyanka Surana, PhD
Programme Lead, The Biodata Developers (BioDev) Network
Senior Scientific Manager, Informatics and Digital Solutions
Wellcome Sanger Institute, Hinxton, UK

Wellcome Sanger Institute
Wellcome Genome Campus
Hinxton, Cambridge
CB10 1SA

Genome Research Limited
Registered Office:
Wellcome Sanger Institute
Wellcome Genome Campus
Hinxton, Cambridge
CB10 1SA

Wellcome Sanger Institute
Human Resources
T: +44 (0)1223 494788
W: www.sanger.ac.uk
E: hrservices@sanger.ac.uk

11/06/2024

PRIVATE AND CONFIDENTIAL
Atharv Fakatkar

Dear Atharv ,

Virtual Contingent Worker Agreement

We are pleased to confirm our agreement to host you as a Collaborator - Offsite, Contingent Worker at the Wellcome Sanger Institute ["the Institute or GRL"].

During this term of engagement commencing on 17/06/2024 until 30/05/2025 you will be accountable to Priyanka Surana who will be your host manager.

During your period of engagement with the Institute, it is expected that your pattern of work will be on a Part time basis.

We understand that during your time as a remote contingent worker, you will remain employed by your substantive employer. The Institute will endeavour to provide you with the same facilities as one of its contracted employees when working remotely. Employees of the Institute adhere to certain terms and conditions and as a Collaborator - Offsite, contingent worker; you are required to accept some of these terms during your period of engagement. Please find enclosed one copy of our Standard Terms for Virtual Contingent Workers. Please acknowledge your acceptance of these terms in Workday as soon as possible.

The Wellcome Sanger Institute is part of the Wellcome Genome Campus, which has a comprehensive Health, Safety & Wellbeing Policy, a copy of which can be provided to your home organisation upon request and is available to you via our Intranet. Only those provisions that relate to working at home or off-site will be relevant to you as a virtual contingent worker. If you need to visit the Campus, you will be treated as a visitor and subject to the same requirements as any other visitor.

Please review and where applicable complete the documents shared via Workday as soon as possible, prior to your start date.

If you have any queries, please do not hesitate to HR Services or your host manager.

Yours sincerely,

Wellcome Sanger Institute
Human Resources
T: +44 (0)1223 494788
W: www.sanger.ac.uk
E: hrservices@sanger.ac.uk

Harry Bell
HR Services On- boarding Administrator
hrservices@sanger.ac.uk