Board 427: US-Sweden Bioinformatics IRES Year 2: Improving Student Preparedness

Prof. Mark A Chapman, University of San Diego

Mark Chapman is an assistant professor at the University of San Diego in the Department of Integrated Engineering. His interests lie in the fields of skeletal muscle mechanics, muscle disease, exercise physiology, international education and engineering education. He earned his MS and PhD in bioengineering from the University of California, San Diego and a B.S. in biomedical engineering from the University of Minnesota.

US-Sweden Bioinformatics IRES Year 2: Improving Student Preparedness

Abstract

This National Science Foundation (NSF) International Research Experience for Students (IRES) site centers around a common theme of bioinformatics. Students selected for this program spend 10 weeks at the Science for Life Laboratory (SciLifeLab) in Stockholm, Sweden using and developing bioinformatics tools to analyze '-omics' (i.e. genomic, transcriptomic, and proteomic) data generated from human subjects. Since the program's inception, a significant amount of effort has been made developing programming to prepare students for their research projects. Specifically, technical training sessions focused on coding languages and professional development sessions focused on scientific communication have been developed and provided to the IRES students during the first two years of the program. While both technical and non-technical trainings have been developed for this IRES site, the focus of this paper is to highlight the technical training sessions since these materials were identified as an area for improvement in the external evaluation following the program's first year.

Summary of the Program

Following the open application and student selection phases, this program is composed of 3 different parts: (1). Virtual Training Series: from March-May the PI meets with the selected students weekly over Zoom to discuss program logistics, acquaint students with their projects and to facilitate group bonding, (2). Pre-departure Symposium: during the first week of June, the PI hosts a 4-day in-person symposium aimed at providing technical and professional development training for the IRES students. Additionally, this symposium serves to further encourage team bonding amongst the students and PI, and (3). 10-week research experience in Stockholm: From June-August the students work in different host labs at SciLifeLab on a diverse set of bioinformatics-related projects. During the last week of the program, the PI hosts a research symposium where all IRES students present their research findings to their mentors and peers. Lastly, all IRES students with accepted abstracts are given the opportunity to present at the annual biomedical engineering society meeting in the October following the summer program. A full description of the pre-training program as well as a qualitative investigation of student experiences in this program can be found in previously published works [1–3].

Year 1 Initial Lessons Learned

A formal external evaluation was performed in the fall of 2021 following the first year of the program. Overall, the program's evaluation was positive where the evaluator described the IRES site as a, "well-organized program offering students valuable research experiences abroad during a challenging time for global travel [due to the COVID-19 pandemic]". Furthermore, the evaluation highlighted that, "the program was particularly successful at inspiring participants to pursue graduate programs and/or research careers in bioinformatics and at providing authentic research experiences that led to publications or presentations". Although the evaluation of the program was mostly positive, there were some key areas for improvement identified. Specifically, the main area for improvement was related to the pre-training activities primarily

centered around building the proper technical coding skills. The IRES students stated that they would like to have additional experience with coding, exercises tailored to their specific projects, and perhaps more time to learn some of the background of their projects.

Efforts to Improve Student Preparedness

In the lead-up to the first year of this IRES program, the PI created a custom course in DataCamp (https://www.datacamp.com/) for students to independently learn coding skills that would be useful during their summer project. DataCamp is an online coding boot camp platform that allows students to work on exercises to improve their coding skills in various languages. Based on input from host laboratories, the PI included exercises in Python (including: *Introduction to Python, Intermediate Python, Introduction to Importing Data in Python*) and R (including: *Introduction to R, Intermediate R,* and various courses on the Bioconductor R package as well as data visualization exercises). On their own time from April-May, the students were instructed to work on specific exercises individualized to their project. Additionally, students were given 2 hours/day (from 3-5pm) to work on these exercises during the 4-day pre-departure symposium. As stated above in the 'initial lessons learned' section, students did not think that the coding exercises were relevant enough to their summer projects. As a result of this feedback, for the 2022 program, the PI worked with each host lab to determine optimal coding tutorials that would develop the coding skills important for student success. Specifically, this included the host labs identifying DataCamp tutorials, R/Python packages as well as gene browser databases to become familiar with prior to starting in the lab.

Based on the host lab suggestions, the PI provided each IRES student with specific exercises in DataCamp to focus on. Although a majority of the resources provided were on DataCamp, some students were also directed to review a specific package's documentation (e.g. Seurat single cell genomics toolkit in R: https://satijalab.org/seurat/, DESeq2 documentation: https://bioconductor.org/packages/release/bioc/html/DESeq2.html, etc…). Compared to the feedback from students in the 2021 cohort, there was only a modest improvement in how useful students viewed the coding tutorials. Specifically, in the external evaluation, students stated, "I think the DataCamp is helpful but did not make a huge difference" and, "I think the coding sessions in DataCamp were a bit less useful [compared with other non-coding preparatory activities] because they often came at the end of the day, and we didn't really know what we were actually going to be using in our project, but I think it was a cool resource to have, especially when we started our actual projects". These statements show that the students did find the coding exercises somewhat useful, but since they were unaware of how these exercises related to their upcoming project, they found them less useful in the leadup to the program. Interestingly, this student commented about how they found the DataCamp tutorials to be more useful to revisit while actually working on their project. Additionally, timing of the exercises during the pre-departure symposium proved to be difficult since coding sessions were assigned from 3-5pm following a full day of other preparatory activities.

Challenges to Overcome in the Future

Redesigning the preparatory coding sessions between 2021 and 2022 showed mixed results. Although student perceptions of the coding sessions improved in 2022, there are several areas to

improve upon for the 2023 cohort. These improvements can be classified into a 'curricular' and a 'logistical' category. Curricular improvements relate to the coding activities themselves – specifically, a clear relevance of the coding exercises to the student's project needs to be presented to the students when the exercises are assigned. In the past, when designing the coding exercises, the PI and host labs focused on providing students with a list of 'typical' bioinformatics packages used by the lab instead of targeted ones based on the planned project. Moving forward, the PI will attempt to better tailor the coding exercises to each student's project as well as clearly explain how the exercises relate to their future project. The PI hypothesizes that this will improve student motivation to complete the exercises and will allow students to identify the relevance of these exercises earlier in program. However, the PI acknowledges some potential pitfalls with this approach. Specifically, it is not always possible to know exactly which project students will be working on due to uncertainty in experimental progress (i.e. will the necessary data be available by the time the student arrives in June?) or the availability of lab personnel (i.e. who will be around in the summer to mentor the student?). Given that these issues are a natural part of scientific research and cannot be avoided in some situations, the PI will make the IRES students aware of the fact that their project may change before arriving in Stockholm. Logistical improvements in the coding exercises relate to when students work on the coding exercises during the 4-day pre-departure symposium. The external evaluation revealed that students found it difficult to work on coding exercises at the end of the day (3-5pm). Given this, the PI aims to distribute the coding activities more evenly throughout the pre-departure symposium. Additionally, students will have access to their coding exercises in April, so the PI will encourage them to be working on these exercises before coming to San Diego for the pre-departure symposium. Lastly, in addition to providing these coding exercises, the PI will give students the tools needed to learn how/where to access help if they run into roadblocks in their projects.

Although students have expressed a desire to feel 'more prepared' for their coding work in the external evaluation, the PI is also cognizant of the fact the students have significant commitments during the semester (school work, sporting teams, jobs…etc.). These other obligations need to be met, so the PI does not expect that all students will be able to work on coding exercises during the spring semester. If students are not able to work on their coding activities during the semester, it is not realistic to expect them to complete all necessary coding exercises during the 4-day pre-departure symposium. Thus, the PI believes that it is acceptable if there is some on-the-job training during the summer. This is a reality for most research/industry positions and no one expects new hires to be 100% ready on day 1 to tackle all challenges. The PI will make this clear to the incoming 2023 cohort, which may put them at ease in knowing that they do not need to experts in the necessary coding language in order to have a successful summer.

Conclusions

Overall the first two years of this bioinformatics IRES site have been a success. However, a major challenge that persists is properly preparing students for the computational aspects of their research projects. Based on student feedback collected through the external evaluation of this program, the PI has developed a strategy to better prepare students for the summer. This new approach involves defining summer projects early, developing targeted coding training materials,

restructuring the pre-departure symposium and training students in problem-solving techniques that will allow them to overcome challenges they may encounter throughout the summer.

References

1. Chapman M. US-Sweden Bioinformatics NSF-IRES Year 1: Program Development and Initial Lessons Learned. In: ASEE Annual Conference & Exposition. 2022. p. 1–5.
2. Chapman M. WIP: Development of a training program for undergraduate students participating in an immersive bioinformatics summer research internship. In: ASEE Virtual Annual Conference Content Access, Virtual Conference. 2021. p. 1–5.
3. Chapman M, Forbes M. US-Sweden Bioinformatics IRES: Investigating Engineering Students' Attitudes and Perspectives Throughout a 10-week International Research Program. In: ASEE Annual Conference & Exposition. 2022. p. 1–15.