July 23, 2020

To Whom It May Concern:

We are providing a letter of participation on behalf of Michael Nath, a student who attended the Pre-College Program in Computational Biology (PCCB) in July 2020. PCCB (http://www.cbd.cmu.edu/education/pre-college-program-in-computational-biology/) is a three-week, full-time educational summer experience for rising high school juniors and seniors held in Carnegie Mellon University's School of Computer Science. The selection process for PCCB is very competitive; Michael was one of 50 students who attended the program from a pool of over 200 very strong applicants from around the United States. The program was held remotely this year due to the COVID-19 pandemic.

Much of modern biological research is inherently computational, and our PCCB students obtained an educational experience in computational biology that is not available to high school students anywhere else in the world. After collecting a sample and processing it experimentally, we must apply cutting-edge algorithmic approaches to analyze the data obtained. In our program, students wrote code to implement a plethora of computational approaches for analysis of genomic and metagenomic data. We then showed how we could apply students' code to analyze the microbial system present in Pittsburgh's Three Rivers as well as analyze the spread of the coronavirus around the world and determine how to detect novel mutations arising as it spreads.

Specifically, our students focused on building algorithms for analysis of biological data across six main areas.

1. Microbiome analysis using multiple diversity metrics.
2. Sequence alignment for identification of bacterial species in a given sample of microbial DNA.
3. Genome assembly from multiple short sequencing reads to reconstruct the coronavirus genome.
4. Identification and annotation of coronavirus and bacterial genes.
5. Evolutionary tree generation and inference of mutations in coronaviruses lurking within the evolutionary tree.

Our work with PCCB students was heavily communal, with students working in groups online to solve a few dozen different coding challenges of increasing complexity. As such, it makes it difficult to assess our students' individual strengths and weaknesses. However, every student group was able to work together to solve some very difficult coding challenges that we would typically expect from undergraduates. These groups amazed us with their ability to progress rapidly through the material, despite the fact that every group had at least one student who had not programmed before our program.

Furthermore, PCCB students' engagement in their work was superlative, with students committing for four hours, five days per week, to the program. We had fifty wonderful students who arrived on time every day and had perfect attendance. At the end of our program, they still found the energy to produce recorded final presentations explaining a complex topic in computational biology to their peers.

In short, Michael was part of a very enjoyable and talented group of students this year. We would love to see these students become the next generation of computational biologists, but we are sure that that their future is bright in whatever they set their minds to.

Regards,

Phillip Compeau, Ph.D.
Co-Director, PCCB
Computational Biology Department
School of Computer Science
Carnegie Mellon University

Joshua Kangas, Ph.D.
Co-Director, PCCB
Computational Biology Department
School of Computer Science
Carnegie Mellon University