Statement of Purpose: UC San Diego Department of Computer Science and Engineering
Marcus William Fedarko

I want to pursue a doctoral degree in computer science, with the goal of following an academic research career. My research interests are centered on bioinformatics, and include metagenomics and sequence assembly. I am also interested in applying research in visualization and human-computer interaction to these fields in order to produce useful tools for bioinformatic analysis.

My primary motivation for pursuing this goal is twofold. During my research experience as an undergraduate I've come to enjoy solving computational problems and producing visualizations of large-scale biological data, and I think a career in computer science and bioinformatics would be an engaging way to apply my education in computer science thus far. More personally, though, I've seen firsthand the effects of chronic medical conditions; cancer and autism, among other conditions, have impacted close members of my family for all of my life. I believe that technological advances have the potential to mitigate or even prevent the health problems faced by many individuals: that insights gleaned from research in bioinformatics and computational biology can result in medical advances that improve people's quality of life. I want to devote my career to contributing to this onward march of progress.

Since summer 2016, I have worked in the University of Maryland's Center for Bioinformatics and Computational Biology on the visualization of metagenomic assembly graphs. Knowledge of the structural details of these often-complex graphs can yield significant insights into the underlying biological data's characteristics. However, state-of-the-art work on visualizing assembly graphs tended to produce broad visualizations of the overall graph structure, which made analysis of small-scale patterns in the graph difficult. In order to rectify this I designed and developed MetagenomeScope, a web application for interactively visualizing metagenomic assembly graphs, under the supervision of Professor Mihai Pop and Dr. Todd Treangen. The project started with a simple idea from Professor Pop of visualizing assembly graphs semilinearly, as opposed to extant force-directed layouts used in assembly graph visualizations. We expanded its scope to include more features to support exploratory analysis as I continued to work on the tool and discuss its possibilities with Professor Pop, Dr. Treangen, and other bioinformaticians inside and outside of the Pop Lab. With the support of the Rita Colwell Travel Fellowship, I was recently able to present a poster on MetagenomeScope at the 25th International Symposium on Graph Drawing and Network Visualization in Boston during September 2017 [1], and we are currently preparing a manuscript on this work for journal submission.

My work in MetagenomeScope has also spurred advances in other projects in the Pop Lab. One of MetagenomeScope's novel features, visualizing scaffolds of contigs as overlaid features on the underlying assembly graph, was used to identify an error in MetaCarvel, our lab's under-development assembly scaffolder [2]. Once the error in question was fixed, the regenerated scaffolds could be visualized in the same context to show that that instance of

the error was rectified. Professor Pop has also used MetagenomeScope to display assembly graphs as part of the 2017 "Strategies and Techniques for Analyzing Microbial Population Structure" research training course held at the Woods Hole Marine Biological Laboratory. It is encouraging to see my work used in these ways, and I hope to continue working on research that has a positive impact.

I also look forward to gaining further teaching experience during my graduate studies. During my time working with Dr. Anwar Mamat as a teaching assistant for CMSC 330, an over 400-student course on programming languages and related theory at Maryland, I was responsible for holding twice-weekly office hours and designing and grading exam questions and quizzes. I hope to continue building my teaching skills in the future by serving as a teaching assistant during some of graduate school.

I am very interested in the UCSD faculty's ongoing research in computer science and bioinformatics, for both its academic merit and its practical potential. In particular, **Prof. Pavel Pevzner**'s research on sequence assembly and **Prof. Rob Knight**'s research on microbial communities both seem compelling to me. I would like to continue contributing to research in assembly and metagenomics during my graduate studies, and I believe working with these professors and their respective labs would afford many excellent opportunities to do so. For these reasons, I believe that studying computer science at the University of California, San Diego would be an incredible opportunity.

References

[1] **Marcus Fedarko**, Jay Ghurye, Todd Treangen, and Mihai Pop. MetagenomeScope. In *Proceedings of the 25th International Symposium on Graph Drawing and Network Visualization (poster abstract, to appear)*. Springer, 2017.

[2] Jay Ghurye, Todd Treangen, Sergey Koren, **Marcus Fedarko**, W. Judson Hervey IV, and Mihai Pop. Graph-based metagenomic analysis uncovers genome dynamics in Human-associated microbiota. In preparation.