Facilities, Equipments, & other Resources

Collaborations: Through our collaboration with the DNAmark mark project (**letter: Gilbert**), we will have access to genome-skims of close to a thousand species that the Gilbert lab will sequence in the upcoming years (Prof. Gilbert's sequencing project is already funded through Danish industrial partnerships). The Gilbert lab also has access to several (as of now) unpublished genomic datasets that they will share with us for testing our methods. For example, a large dataset of wolves and dogs is created in his lab, and the data will be useful for our Aim 1.

Facilities

The Bafna Lab is in the UC San Diego Computer Science and Engineering (CSE) Department, located within the Computer Sciences building, a 135,000 sq. ft. high technology facility within the Jacob School of Engineering campus. Vineet Bafna's Bioinformatics laboratory encompasses approximately 1600 sq. ft. of dry lab space within the CSE building, and is adjacent to other labs focusing also on bioinformatics. There are 4 rooms, each about 40x10 sq feet, containing interactive 6x6 ft. cubicles, space for interacting and resting, with enough space for about 30 researchers. Irwin Jacobs, the founder of Qualcomm, has longstanding relations with our department, having been a professor at UCSD for many years. UCSDs CSE Department embodies the University's tradition of excellence as a world-class leader in computer science and engineering education and research. CSE is in a period of exciting growth and opportunity. Ranked in handful of top programs in the country, the CSE is dedicated to research, education and overall excellence. The Bioinformatics Group within CSE has a focus on Genomics/Genetics, studying problems relating to genome rearrangements, structural variation, deep-sequencing, and population genetics, with a focus on novel analyses of data from high throughput technologies including sequencing, genotyping and mass-spectrometry. CSE Bioinformatics is an integral part of a vibrant Bioinformatics PhD program on campus with faculty from Biochemistry, Bioengineering, Biology, Mathematics, Medicine, Pharmacy, and Cancer Biology.

The Mirarab lab is located in the Electrical and Computer Engineering (ECE) department of UC San Diego located at Jacobs Hall. It's a large dry lab that hosts eight graduate students. PI Mirarab has an office in the same building.

Equipment

The proposed work requires computational platforms for studying the performance of the methods that will be developed. Both PIs have access to a variety of computational resources to conduct the work, as described below.

SDSC We are linked with the rest of campus via a 10 Gb/sec fiberoptic network, which provides access to the San Diego Supercomputer Center (SDSC), a unique national facility, with a variety of vector, multithreaded, and parallel supercomputers as well as a state-of-the-art high-performance visualization tools. PI Mirarab currently has close to 1,000,000 hours of allocations on the SDSC

cluster through NSF's XSEDE initiative. PI Mirarab will write renewal XSEDE proposals for continuing his work on phylogenetics. If this proposal is granted, we will request further allocations from XSEDE for conducting this project and analyzing real biological datasets that would accompany this project.

TSCC. PI Mirarab has access to the Triton Shared Computing Cluster (TSCC) cluster (`http://idi.ucsd.edu/computing/`) at San Diego Computing Center (SDSC). The PI has purchased an equivalent of 1.6 million cpu hours on the cluster that will last for the next two years, and the structure of TSCC will enable his lab to use at least those many hours of computing time. More specifically, the PI has purchased two nodes, each with 24 cores (Intel Xeon processor @ 2.5GHz) and 128GB of memory. The cluster is maintained by SDSC, with subsidized maintenance costs, and access to extra computational resources when the work load permits.

Mirarab's big data cluster. PI Mirarab has created a cluster of 70 server nodes, each with an Intel(R) Xeon(R) CPU iwth 7 cores. The members of the lab have full access to this cluster, which the PI shares with 4 other faculty in the ECE department.

Vineet Bafna's Bioinformatics laboratory. Each cubicle contains a customized computer terminal, running an Intel Core i7 processor (with 8 processing threads @ 3Ghz), with 16 GB RAM and 2 TB of storage per computer. The Lab also has a Central Computer Linux Cluster, with 5 nodes, each with 24 cores and 128 GB RAM, and 128 TB storage space in total for data-intensive projects.

Other

All of the members of the both Mirarab and Bafna lab are proficient in the following computer languages: C++ (used to generate data filtering pipelines), PERL (used to interact with online databases), Python, Java, as well as Matlab and R for statistical analysis. Each person is familiar with algorithm development, data structure classes. Almost all of the proposed research uses free software, but additionally, PIs have access to simulation tools such as MATLAB and SIMULINK, if needed.