Warren J. Baker Endowment
for Excellence in Project-Based Learning
Robert D. Koob Endowment *for Student Success*

I. Project Title
Unveiling mechanisms of regeneration in a tunicate model using genome and transcriptome sequencing

II. Abstract
In this multi-disciplinary genomics study, we propose to use affordable industry-standard genomic sequencing technologies to assemble the first genome of *Botrylloides violaceus,* an invertebrate chordate that can regenerate its entire body from only its external vasculature. We will use informatic analysis of the *B. violaceus* genome to shed light on its evolutionary history and figure out genes involved in whole body regeneration. Using the same technologies, we will sequence RNA molecules to generate transcriptomes at distinct steps of regeneration. Transcriptome sequencing will allow us to observe how certain genes are turned on or off during regeneration and will add a layer of functionality to the genome sequence. Applications of this study will allow future researchers to exploit *B. violaceus* as a model organism in regenerative medicine. The techniques described in this proposal can be applied to any organism, especially those that lack genome sequencing data, and will provide a new resource to the University's researchers.

III. Objective(s)
Introduction

The first human genome was released in 2001 and took 15 years and three billion dollars. Today the human genome can be sequenced in one week with $1,000[1]. Traditional methods of nucleic acid sequencing have been replaced with high-throughput Next Generation Sequencing (NGS) technologies that allow millions of DNA strands to be sequenced simultaneously[2]. While wildly influential in the biological sciences, traditional methods are to NGS as the Guttenberg Press is to an iPad. With the advent of NGS, genomics research is in an era of "big data" and requires complex computational analysis. Here we propose an experimental and computational pipeline for cost effective whole genome (DNA) and transcriptome (RNA) sequencing of the invertebrate chordate, *Botrylloides violaceus*. A genome is an organism's collection of genes and other genetic material, and a transcriptome is a snapshot of all genes that are expressed, or turned on, in an organism at given point in time. Hybrid assembly, a cutting-edge genomics technique, combines long read (10,000bp) and short read sequencers (75bp) to assemble a near complete genome at a lower sequencing coverage than either sequencing method can independently[3]. We will use Cal Poly's new industry quality short read sequencer (Illumina MiniSeq) and purchase a compact long read sequencer (Nanopore MinIon) to sequence the first *B. violac*eus genome. With this data, we will collaborate with Computer Science and Mathematics majors to optimize algorithms for allowing this to be a tool to serve a broad range. **Optimizing this pipeline will open affordable genomic and bioinformatic approaches to Cal Poly's students and faculty and their broad range of research interests. Additionally, data generated in this study will be adapted as a series of exercises for students in Cal Poly's BIO 441 class, Bioinformatics Applications, so that any enrolled students can apply bioinformatics methods to raw industry standard data.**

B. violaceus is a colonial ascidian capable of whole-body regeneration (WBR)[4]. Colonial ascidians are the closest relatives to vertebrates that retain WBR, making them an interesting model organism. By sequencing the *B. violaceus* genome, we hope to shed light on its evolutionary history and gain insight into why colonial ascidians can undergo WBR but vertebrates cannot[5]. We will adapt publicly available algorithms to compare various genomes to *B. violaceus*, focusing on its evolution, structure and annotation. In parallel, we will sequence the transcriptome of *B. violaceus* at distinct

stages of the regenerative process. We will use these data to explore how regeneration is regulated through changes in gene expression**. This research will be in academic collaboration with Dr. Elena Keeling, who studies colonial ascidian regeneration.**

Formal objectives include the following:

(1) Optimize DNA extraction protocol for both Nanopore and Illumina sequencing (NGS technologies).
(2) Generate all sequencing libraries for analysis, with detailed metadata organization.
(3) Complete sequencing for genome and transcriptomes
(4) Interpret and synthesize results using bioinformatic pipelines

IV. Methodology

(1) Optimize DNA extraction protocol for both Nanopore and Illumina sequencing

DNA and RNA will be extracted from *B. violaceus* with standard commercially available kits (Qiagen DNeasy Blood & Tissue Kit & Zymo Quick-RNA MiniPrep Plus Kit) previously purchased by Dr. Elena Keeling. DNA quality, purity and quantity will be assessed using two instruments (Nanodrop Spectrophotometer and Qubit Fluorometer) located in the Center for Applications in Biotechnology. Extraction procedures will be optimized to meet literature standards for Nanopore and Illumina sequencing. One student (Jack Sumner) is also a member of Dr. Keeling's lab and is experienced with *B. violaceus* RNA and DNA extraction protocols. Several DNA and RNA samples have already been prepared.

(2) Generate all sequencing libraries for analysis, with detailed metadata organization.

Using extracted nucleic acids, we will prepare libraries (input material for sequencers) for the Illumina MiniSeq Sequencer (bought by Cal Poly in early 2018) and for a Nanopore MinIon Sequencer (will be purchased with endowment funds). We will use two standard commercial kits to prepare sequencer-specific libraries (Illumina Nextera DNA Library Preparation Kit and Nanopore Ligation Sequencing Kit). Quantity and size of libraries will be assessed using available instruments (Qubit Fluorometer and gel electrophoresis) and library preparation protocol will be optimized as needed. Dr. Jean Davidson, a sequencing and bioinformatics expert, was recently hired from the biotechnology industry and will oversee library preparation and sequencing. Additionally, one student (Jack Sumner) recently completed an internship at a Silicon Valley biotech company developing novel NGS library preparation methods and has used both Illumina and MinIon platforms.

(3) Complete sequencing for genome and transcriptomes

Libraries will be sequenced on their respective sequencers. Nanopore sequencing is completed within several hours and Illumina sequencing within one day.

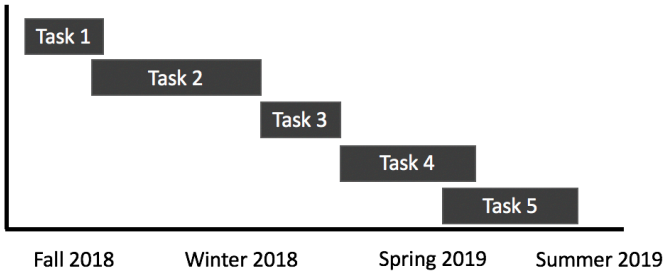
(4) Interpret and synthesize results using bioinformatic pipelines

We will use publicly available software to analyze our data sets. Several algorithms will be used and adapted with the help of two collaborating students with majors in Mathematics and Statistics both with minors in Data Science (Charlie Liou and Jenna Landy). Some programs that we will adapt are the Burrows Wheeler Algorithm (bwa), MaScURA assembler, Pilon assembler, QUAST, and Picard Tools. Many of these programs have an open source code that can be changed as necessary. Dr. Davidson is also an expert in bioinformatics and one student (Jack Sumner) has experience working with the above programs.

V. Timeline

Task 1: Optimize DNA extraction protocol for both Nanopore and Illumina sequencing.
Task 2: Generate all sequencing libraries for analysis, with detailed metadata organization.
Task 3: Complete sequencing for genome and transcriptomes
Task 4: Interpret and synthesize results for poster and manuscript.

Task 5: Disseminate findings through conference (2019 Stem Cell Conference, Asilomar CA), journal, and collaboration outreach.



VI. Final Products and Dissemination

Our dissemination goals are three-fold. First, we will attend the Bay Area Stem Cell Conference in Asilomar, May 8-11th 2019. We will apply to present both posters and oral presentations to discuss progress and findings with leaders in the regenerative medicine field. This will provide interaction with potential employers and graduate school programs, as well as an opportunity to troubleshoot any issues that arise with scientific colleagues. Second, a major goal will be to publish these findings, as we actively participate in the writing, editing and submission of this manuscript. Potential journals targeted include Journal of Regenerative Medicine, G3 (Genes, Genomes, Genetics), and PLoS Genetics. Last, establishing this protocol of hybrid genome assembly and sequencing can be broadly utilized by a diversity of research agendas across the campus. Many Cal Poly researchers rely on non-traditional model organisms which do not have a well-established reference genome, thereby limiting genomics and bioinformatics applications. This proposal will establish a pipeline to expand genomics resources to a wide diversity of model systems and will be fully implemented by undergraduate and graduate researchers.

VII. Budget Justification

Travel:

Bay Area Stem Cell Conference, in Asilomar, CA - May 8-11th 2019, travel and lodging: $500
Poster Printing for Asilomar Conference: $200

Operating Expenses:

Nanopore Minion Sequencer (all reagents and sequencer): $1000 Starter Kit
Illumina MiniSeq Reagents (library reagents and sequencing chip): $1000
Cloud compute analysis tools (AWS and BaseSpace): $800
Bay Area Stem Cell Conference Registration: $500
Journal Publication costs (Genome Informatics, Genes, Genomes, Genetics): $1000

References

[1] Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) Accessed [10/10/18]

[2] Moorthie S, Mattocks CJ, Wright CF. Review of massively parallel DNA sequencing technologies. Hugo J. 2011;5(1-4):1-12.

[3] Mun Hua Tan *et al.;* Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (Amphiprion ocellaris) genome assembly, GigaScience, Volume 7, Issue 3, 1 March 2018, gix137

[4] Rinkevich B, Shlemberg Z, Fishelson L. Whole-body protochordate regeneration from totipotent blood cells. Proc Natl Acad Sci U S A. 1995;92(17):7695-9.

[5] Zondag LE, Rutherford K, Gemmell NJ, Wilson MJ. Uncovering the pathways underlying whole body regeneration in a chordate model, Botrylloides leachi using de novo transcriptome analysis. BMC Genomics. 2016;17:114. Published 2016 Feb 16. doi:10.1186/s12864-016-2435-6

Warren J. Baker Endowment
for Excellence in Project-Based Learning
Robert D. Koob Endowment *for Student Success*

CAL POLY

PROPOSAL BUDGET

Student Applicant(s): Jack Sumner, Charlie Liou, Jenna Landy, Evan Dagy	
Faculty Advisor: Jean Davidson	
Project Title: Unveiling mechanisms of regeneration in a tunicate model using genome and transcriptome sequencing	**Requested Endowment Funding**
Travel *subtotal*	**$500**
Travel: In-state	**$500**
Travel: Out-of-state	**$**
Travel: International	**$**
Operating Expenses *subtotal*	**$ 4500**
Non-computer Supplies & Materials	**$2,000**
Computer Supplies & Materials	**$800**
Software/Software Licenses	**$**
Printing/Duplication	**$200**
Postage/Shipping	**$**
Registration	**$500**
Membership Dues & Subscriptions	**$**
Multimedia Services	**$**
Advertising	**$**
Journal Publication Costs	**$1,000**
Contractual Services *subtotal*	**$**
Contracted Services	**$**
Equipment Rental/Lease Agreements	**$**
Service/Maintenance Agreements	**$**
TOTAL	**$5000**