Proposal All Reviews: 1711984

Agency Name:	National Science Foundation
Agency Tracking Number:	**1711984**
Organization:	
NSF Program:	Broadening Participation of Groups Underrepresented in Biology
PI/PD:	Pierce, Nuri
Application Title:	NSF Postdoctoral Fellowship in Biology FY 2017: Improving RNA-Seq Analysis through Graph-based Analysis and Computational Indexing

Review 1

Rating:

Multiple Rating: (Excellent/Very Good)

Review:

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

The successful completion of the postdoctoral fellowship research project proposed will improve RNA-Seq analysis, particularly for organisms without a reference genome. The two focus areas that will guide this research are 1) development of new debrujin graph-based annotation and expression analysis approaches for RNA-Seq analysis. 2) Improve indexing of publicly available RNA-Seq data to enable efficient searching and comparative analysis for 2a) sequence or functional category similarity and 2b) expression level similarity. These approaches will be developed and tested using the Marine Microbial Eukaryotic Transcriptome Sequencing Project data. The PI will serve as a role model and mentor continuing with several outreach programs she previously participated in as a graduate student. She will develop a lecture module that combines hands on marine biology experience with coding and analysis of experimental data.
Research Strengths
The research outlined here, particularly in Aim 2 addresses an important problem in analysis of biological data. The
Research Weaknesses
While the PI does an excellent job explaining the validation data set and comparisons between the newly developed approaches and existing analysis, interpretation of the validation is lacking. How will the PI discern and quantify if the newly developed approach provides and improvement over existing approaches. Will it be based on gain in speed alone, percent mapped, functional enrichment? It is not clear if there is a contradiction between the results from the new approaches and previous analysis how this will be evaluated.
A specific example of how aim 3 might be biologically relevant would be beneficial.

In the context of the five review elements, please evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

Improving RNA-seq analysis will have large potential effects across a range of research areas. Establishing a method for indexing and searching available transcriptome sequencing data will enhance the value of all researchers performing RNA-Seq analysis if these approaches are applied to multiple data sets.

Please evaluate the strengths and weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if applicable

Mentoring Plan
The statement by the sponsoring scientists indicates that there is no overlap between the research proposed by the PI and existing projects in the PIs lab. The statement provides thoughtful and detailed evaluation of the opportunities for the PI to continue the research after the fellowship period. The sponsoring scientist provides no discussion concerning the commitment of this project to broadening participation and the time required.
Broadening Participation

The PI will serve as a role model continuing to participate as a mentor through organizations such as Expanding Your Horizons, where she also worked as a graduate student. In addition, she will develop a lecture module that combines hands on ûactivities related to microbial ecology and computational approaches for analysis.

Data Management Plan

No new experimental data will be generated. Details are provided for sharing of scripts and code generated in this project and shared by GitHub. Publication of research through preprint servers and open source manuscripts is described..

Potential for Career Goals/Development

The PI has chosen a laboratory and research institution that is well-aligned with her career goals. There is some concern that the proposed research is very similar to the PI's graduate work and may not offer the breadth of training necessary to meet the PI's goals of being prepared for a successful academic position.

Summary Statement

The PI is a strong candidate likely to complete the outlined research activities relating to bioinformatics. The experiments are well-justified and validation approaches are outlined. Interpretation of the validation results is not presented in the project description. The PI has a demonstrated commitment to outreach activities and will continue to serve as a role model and participate in outreach activities including developing a marine ecology focused lecture module employing computational approaches with hands-on activities.

Review 2

Rating:

Excellent

Review:

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

This is a timely and well-written proposal that seeks to improve and standardize RNA-seq analysis pipelines. These improvements are badly needed. The focus on improving the comparative utility sequencing data that are currently in databases (and will be added in the future) is particularly compelling. Overall, the proposal is well-reasoned, the PI has a strong record for her career stage and the choice of mentor and training institution is excellent. The mentoring plan is very strong and so are the letters of support.

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

The direct broader impacts of the work are also impressive. I fully agree with the PI's description of the current state of RNA-seq analyses in non-traditional models species being a patchwork of quality and approaches. Standardization and increased rigor and transparency are badly needed. The development of these tools alone are very strong broader impacts. Added to this tool development, are the PI's outreach efforts in broadening participation though data analysis and coding training. These training efforts seem as though the will be effective and can be realistically implemented in the timeframe of the fellowship. In sum, the broader impacts of the research, like its intellectual merits, are excellent.

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if
applicable

Summary Statement

Overall this is an excellent and timely proposal with strong intellectual merit. The developed pipeline are badly needed by the community, and as a result, the broader impacts of the research are also excellent.

Review 3

Rating:

Excellent

Review:

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to intellectual merit.

The applicant will complete their PhD in 2017 at the Scripps Institute of Oceanography where they investigated the effect of O2 conditions on squid embryos. They gained extensive experience in gene expression analysis and would be joining the Brown Lab at UC Davis to improve RNA-seq data analysis tools. Specifically, the goal for Aim 1 is to improve analysis methods for individual transcriptomes using graph-based approaches. These are expected to improve read assembly and reduce computational resource use due to read error correction. The results obtained using the new analysis method would be compared to previous analysis done using standard methods. With Aims 2 and 3, the goal is to increase the accessibility and utility of available data such as the SRA archives. The approach is to use computational indexing to facilitate query of sequencing and expression data. The usefulness of these tools will be tested on data found in the Marine Microbial Eukaryotic Transcriptome Sequencing Project (MMETSP).
Overall the project is concerned with building computational tools that would improve sequence data assembly, annotation, and sharing. In embarking on such a project, the applicant would extend their training as a computational biologist with skills in developing, modifying, and applying analysis and indexing algorithms to large biological data. Their ultimate goal is to obtain a faculty position in marine bioinformatics. The Brown lab at UC Davies is well suited for the proposed work and a strong mentoring plan is provided.

In the context of the five review elements, please
evaluate the strengths and weaknesses of the proposal with respect to broader impacts.

Aside from developing open-source tools that will serve the community of users, the applicant will continue their activity in data science training and outreach by teaching basic coding that is context relevant. As Hispanic woman and first generation American, they appreciate the value of providing opportunities and support to a wide array of young students.

Please evaluate the strengths and
weaknesses of the proposal with respect to any additional solicitation-specific review criteria, if
applicable

Summary Statement

This is a strong proposal that could provide the applicant the training and environment to successfully launch their career as a faculty in Marine Bioinformatics.