

## Blue Waters Education Allocations Final Report

As a component of your application for a Blue Waters education allocation, you agreed to provide a report at the end of your project. Now that your project is complete, we would appreciate your submission of the following information within two weeks. Please send this report, and any supporting documents, to Scott Lathrop, [lathrop@illinois.edu](mailto:lathrop@illinois.edu).

This information will be shared with the Blue Waters team and the National Science Foundation. Portions of the report (we will omit names of participants) will be posted on the Blue Waters portal for public access.

### Project Information

Project Name	<a href="#"><u>Improving the calability of Mothur for large metagenomic studies</u></a>
Names of project staff (instructors, TAs, etc) and their department and institutions	<a href="#"><u>Charles Peck, Computer Science, Earlham College (instructor),</u></a> <a href="#"><u>Kristin Muterspaw, Computer Science, Earlham College</u></a> <a href="#"><u>(student), Tara Urner, Physics, Earlham College (student)</u></a>
URL for the project	<a href="#"><u>NA (to-date)</u></a>
Provide links to or attach materials made available to participants (e.g. slides, articles, exercises, etc.) that may be made publicly available	<a href="#"><u>NA (to-date)</u></a>
Provide links to or attach any photos (with captions to describe activities)	<a href="#"><u>NA (to-date)</u></a>
Start date	<a href="#"><u>June, 2014 (proposal approved December, 2014)</u></a>
Completion date	<a href="#"><u>May, 2015 (May, 2016 proposed)</u></a>

### Information about the Participants

# Participants	# Faculty or staff	# students	# other (e.g. industry)	# under-represented (e.g. women, minorities)	# institutions represented by participants
<a href="#"><u>4</u></a>	<a href="#"><u>1</u></a>	<a href="#"><u>3</u></a>	<a href="#"><u>0</u></a>	<a href="#"><u>3</u></a>	<a href="#"><u>1</u></a>

Please describe the scope and purpose of this project. Also, please indicate if there were any changes implemented from the original proposed plan, and briefly describe why they were made.

This project seeks to document and improve the strong and weak scaling of Mothur. Mothur is popular open-source bioinformatics software package that is used for analyzing 16S rRNA gene sequences. Through our experiments, we have discovered that the software has many technical issues, especially those related to distributed and shared memory implementations of the code. Very early on we learned that the Blue Waters runtime environment does not support code that uses fork/exec, the technique used by much of Mothur to implement parallelism. Just this past week we learned from Mobeen Ludin at the Petascale Institute that Darshan, a runtime performance analysis tool, also has the ability to wrap fork/exec calls in such a way as to make it possible to run on Blue Waters. This will greatly enhance our ability to make progress since we won't have to remove all the fork/exec calls (over 100) at once.

Please describe the learning outcomes of the participants. How did this project enhance the learning of the participants? What did the participants learn as a result of the use of Blue Waters system that they could not have learned using other systems?

The sheer scale of the resource is making it possible for us to do combined analysis on more data sets than has been possible on any other resource. This makes us especially anxious to implement parallelism to see how much speed-up can be accomplished with improved approaches.

Please describe lessons learned from the project. What would you do differently next time?

The allocation process was more complex than we had experienced before, but once we learned the details it was straightforward. Next time we would start earlier. The fork/exec restriction was also unknown to us until relatively late in the process, it appears as though we now have a work-around for that.

What would you recommend that the Blue Waters team do to enhance the success of education projects in the future?

Nothing so far, the support has been very good.

Please provide a summary of any surveys or evaluations you conducted of the participants. Feel free to attach any related documents.

NA (to-date)

Please provide any anecdotal stories we may share with NSF and the public.

NA (to-date)

How would you rank the overall experience?

	Excellent	Very Good	Good	Fair	Poor	N/A
Education allocations process		<u>X</u>				
Blue Waters support		<u>X</u>				
Blue Waters computing system	<u>X</u>					
Blue Waters documentation			<u>X</u>			
Blue Waters training		<u>X</u>				

Do you plan to request an education allocation for other future events that will use BW?

Please describe the plans for future events, including the frequency (each semester, yearly, etc.).

Yes, due to the complexities described above we have only used 231 hours of the 48K we were allocated. Since the project itself has been approved for a second year (to accomplish what we set-out to do last year) we would like to extend our current allocation for Tara Urner to pick-up and work with. Kristin Muterspaw (2014 BW Intern) and another student will also be working with Tara.

Please provide any other comments or suggestions.

Thank-you very much for the opportunity, it has benefited my students and I greatly to work with Blue Waters.