



Evaluation Summary Report
Tutorial: *Computing in the Cloud: What Every
Computational Life Scientist Should Know*

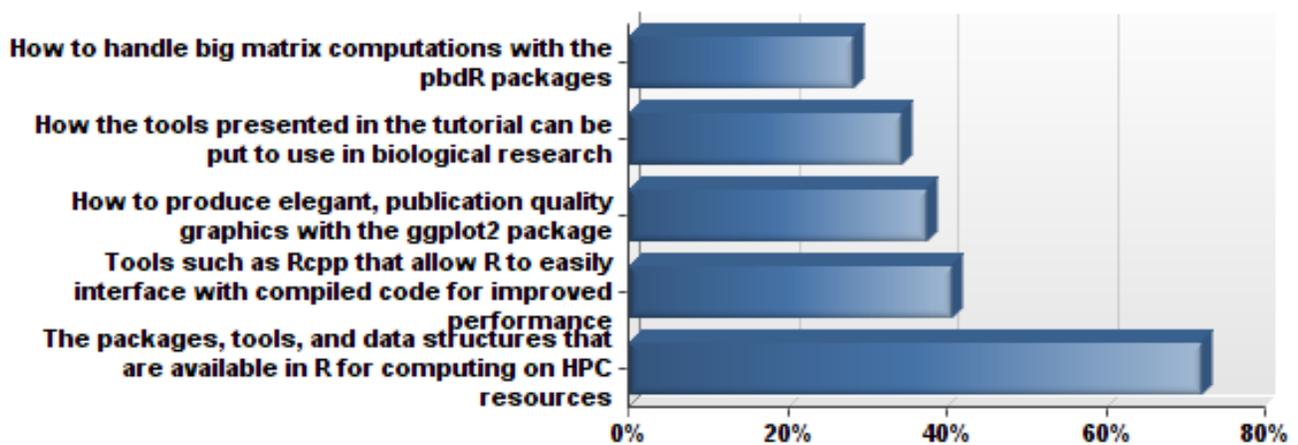
April 6-8, 2014

Pamela Bishop
Evaluation Manager
National Institute for Mathematical and Biological Synthesis
May, 2014

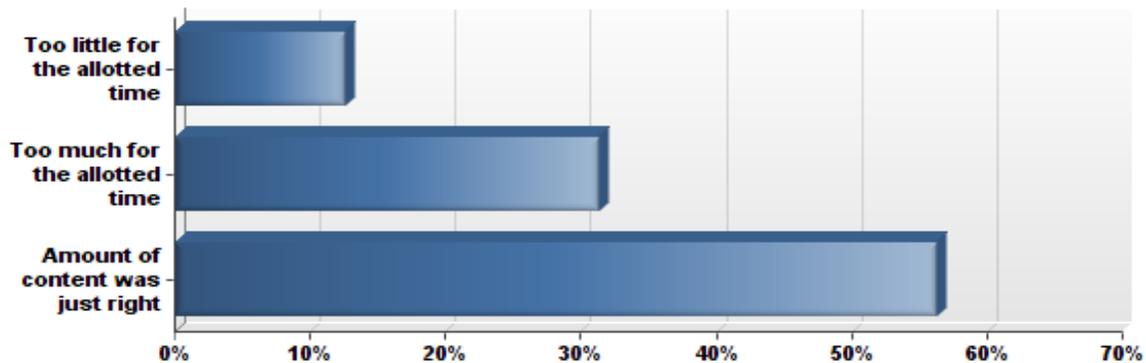
1. Please indicate your level of agreement with the following statements about this tutorial:



2. As a result of participating in this tutorial, I have a better understanding of:



3. How do you feel about the amount of content offered during the tutorial?



4. What topics would you have liked to have covered in this tutorial if given more time?

More examples of implementing biological calculations using pbdR (how different types of workflows can be distributed). I also would have appreciated more organized introductions among tutorial participants and discussion of research topics (to find people with similar biological questions)

More hands-on exercises and discussion of what people in the room are using, what their scientific workflows are, and some examples of how to choose which parallel resources are best for which kinds of examples.

It would have been nice to have more time for hands on applications with some of the material or working groups to troubleshoot using HPC in our own work.

I would have liked more lab sessions. Some things I feel were over my head and would have been better understood in a lab session, but I understand that that would take SIGNIFICANTLY more time, and this was a 3-day workshop. All-in-all I thought it was an excellent introduction for me, a beginning R user, just starting to play with big data, while at the same time useful for those experienced big data R users.

I think that more of an introduction to HPC at the very beginning would have been extremely helpful. This was covered at the end when students were getting hardware concepts confused.

More of the hands on -- actually *how* to do some of the stuff (e.g., writing a bash script to run a job on Amazon -- I've never written a bash script in my life and right now couldn't do that if my life depended on it); or actually going through code of doing a simple job in parallel.

An expanded demonstration of Rcpp applications for improved performance

More hands-on examples and practice of what we learned in small groups

R for genomic and metagenomic data

Advanced visualization in R

I would have liked the audience to have had more time to present their particular problem/ work and get some feedback from colleagues and the instructors on how to handle it. This happened a little on the final day, but could have been much more. I would have also liked to have direct demonstration of sending commands to local clusters rather than the Amazon cluster.

I would have liked a short primer on multi-threading.

What to do if my data is too big to read into R. What exactly am I doing when setting up an Amazon EC2 instance? More worked examples. More practice

Go through an example that requires HPC through ALL the steps

I would have liked a period of time to discuss and present on what the tutorial participants were working on and the problems that they encounter with their data. We could then brainstorm solutions.

More information/tips on better programming practices within serial R.

I would have liked to have more time for hands on, and a somewhat slower and interactive exposition of the very technical matters.

The HPC seminars led by Schmidt were the most useful for my purposes. Overviews and sample scripts for HPC. While the science talks by phylogeneticists and biogeographers were interesting, they didn't have any practical HPC benefits w.r.t. my research.

Some of my problems in moving to scale stem from inability to handle errors in an elegant way. So, some instruction in exception handling and testing frameworks would be useful for me. However, there was only so much time, and I really value what I learned!

We could have done more on BASH and access university clusters from Linux. Less focus on amazon

and more focus on clusters, since it is a more common (and free, at least for us, resource).

More details about the presented toolboxes to provide insights in how the parallelization is achieved. Synchronization of processes, threads and nodes was unfortunately not discussed either.

While I gained a tremendous amount, I was expecting there to be a hands on session where we actually set up a batch job in an HPC environment running a say a model written in C. We set up an interactive session on the Amazon cloud service with R but it was a complicated process and given the username and passwords required which were set up temporarily just for the class I'm not sure if I'll be able to use it again. I suppose if there were more 'computer practical' sessions the course this would have been useful.

I would have liked more tailored instruction for our specific problems

5. What do you feel was the most useful aspect of the tutorial?

The explanation of hardware architectures, pbdR, Rcpp, ggplot2

HPC resources in R, ggplot, benchmarking code, R Markdown

Just getting info on the terminology for hardware and software, and naming the available resources was really helpful. Seeing examples of bad vs good code, and discussion of how to problem-solve slow or un-runnable code (because of computing time and/or data size) was really helpful. The scientific talks were great too. And the tutorial was a great place to network and meet others taking similar approaches.

Great exposure to a variety of tools and examples. Also, I got a lot from interactions with other participants.

For me personally, I enjoyed the sessions on how to use Markdown and ggplot2. Those were the places I was most lacking. I also enjoyed seeing the more advanced topics even if I couldn't quite grasp them or understand how I would apply them to my work just yet. It's always good to stretch your rubber band.

The most useful aspect of this tutorial was to have the opportunity to ask questions to a diverse group of scientists. There were many people participating with overlapping expertise. This allowed for excellent group discussion because participants could translate concepts between disciplines by coming up with new analogies.

The inclusion of ggplot2, Rcpp, teaching with R, and HPC all in one tutorial, as participants likely came for only one of these topics but received good exposure to all these topics which are potentially powerful tools for researchers and educators.

Learning about the available resources (packages/tools) for High Performance Computing in R as well as hearing what problems others are solving with HPC.

learning about new tools available in R that I previously had not heard about

The knowledge of new resources

Brian O'Meara's talk and also the ggplot session and also the Rstudio demonstrations

The hands-on exercise was the best (setting up an AWS machine).

Knowledge of the different packages out there. Knowing how to interface with R using Amazon EC2

Hands-on exercises

The mix of subjects. It was a good introduction to many topics.

Networking with fellow participants, exposure to the tools available for HPC in R, ggplot2 tutorial

Teaching with R, Rcpp, parallel, general information on potential options

Seeing someone setup and use the HPC facilities. I can't quite do it on my own, but at least I know that it can be done. That's a big help.

The HPC seminars led by Schmidt.

The chance to get together with a bunch of other biologists who use R, including some experts, talk to everyone, and see the experts in action.

The informal discussions were more helpful than the formal lectures. The PBD packages were especially interesting.

Hands on exercises.

Presentation of HPC tools for R and how to put them to use.

Where to begin? Everything was incredible. From the nice broad overviews of different parallel paradigms to the extraordinary presentations on phylogenetic inference and phylogeography. Drew was amazing. All the lectures were fantastic. Getting to see everyone's workflows was super. I really liked the ggplot lecture (that was never very accessible, previously), and I thought Danny Kaplan's demonstration of Rstudio and Rmarkdown and mosaic were brilliant.

Getting more of a general feel for what HPC services are available to me and what R resources for HPC are accessible.

Knowledge of the options that are available for interfacing with the cloud, and great instruction about R presentation and teaching tools

6. What would you change about the tutorial?

I did not find the portion about teaching with R particularly useful, or the setting up of an Amazon AWS account (too many glitches that were unnecessary). Instead of those modules, I would have preferred more structured time to interact with other tutorial participants and possibly split into working groups based on common research themes (though this may not have been feasible). If not by research topic, then simply having smaller groups to work through more concrete examples of pbdR and multicore (for parallelization) would have been nice.

Instead of using AWS, perhaps using University resources

More time for hands-on exercises. I need lots of practice in order to really feel comfortable with the packages. We only briefly looked at AWS - it might have been helpful to revisit the second day (although I did get a small charge,

More breakout sessions for to experiment or implement the tools

I thought it was excellent. 90% of the presenters were not only experts but great expositors, which made even the most advanced (and currently outside of my understanding) topics interesting.

It is always extremely hard to teach to a diverse group, but more background on HPC in general would have been helpful. See comments above. Although all of the content was great, it seemed to be chopped up. Perhaps this was because of scheduling issues. For example, teaching R to undergraduates could have been left for the end of the day and done in one slightly longer session.

I didn't find the "Teaching statistics with R" useful -- or, really, in context. Though some of the stuff would be helpful

I wouldn't change anything about the way this tutorial was set up

I, as well as many other participants I spoke with, would have liked to interact more with the other tutorial

participants in a formal setting. Firstly, I would have liked for all of the participants to do an introduction (name, where from, what sort of project, what experience with HPC/R, etc.) at the beginning of the tutorial. This would help us talk to others who may be working on similar problems. Secondly, I would have liked to work in small groups with other participants to solve a toy problem and/or real problem using the tools we learned about in the sessions. Groups could be divided by familiarity with R.

less content or more time

More days, and an equal amount of "frontal lessons" and hands on activities.

I think the R markdown was not within the topic of the tutorial, though it was very interesting and good resource for teaching- it was a side topic to the computing in the cloud. Also I think several presentations (Matzke, Gilcrest, Wei Chen), while very interesting and well done, did not cover much at all about the computing that was involved. The presenters spent almost all the time on their biological/ molecular problems but didn't well explain the computational challenges or how they used computer clusters. A small amount of time on the biology and then direct demonstration of how computation was used would have been much better. I also think the presentations were all out of order and jumbled. I think some intro material was not presented till day 2 or 3. A careful scheduling could have made sure intro material (such as the basic structure of nodes, some coding and Rcpp) would have been best on day 1. The presenters were also not well coordinated.

There was too much lecture and very little hands-on activity. I can get lectures easily over the Internet, but the value of an in-person tutorial is having a person being able to see what you're doing incorrectly.

The instructors assumed a very high level of understanding that a lot of participants did not have. There was a lot of jargon and comments like "you're all aware of this", which was not true.

Less direct-lecturing and more hands-on tutorial time. Also, don't assume a group of biologists is familiar with the minutiae of computer architecture.

I would build in more time for participants to present very briefly on the work they do and the problems they have to facilitate brainstorming and networking. I would start off the tutorial with a broad overview of HPC in R - parallel, cloud computing, share memory vs distributed, etc.. Defining key terms so that everyone is on the same page from the beginning. I would spend less time going over the code for solving mixed models and setting up the Amazon AMI. I would have the demonstrations focus more heavily on professional workflows instead of pure demonstration... e.g. not using the inline package for Rcpp and showing us how it would actually work if we were to do it ourselves

Some presentations were incomprehensible and not useful: too much background, no real information on the programming challenges and solutions which should have been the main point. The "hands-on activities" was a complete waste of time. All it did was force us to sign up to a paid service, without any actual practical application. Much better would have been a short tutorial in taking one or more simple problems and implementing them in parallel code, etc.

Slower pace for most sessions. Breaking up sessions across days so there's some time to communicate with the session leader and to work out glitches.

More HPC overviews and samples starting from basic (e.g., how to benchmark code, parallel computing) to advanced (e.g., other languages, compiling, distributed computing). Yes, I know you had these, but taking time to flesh out these topics and intermediate steps. Reducing the amount of time spent on non-methods science talks. Having a progression from basic to advanced so that users can decide what timeframe is most useful for their purposes; e.g., allowing advanced users to tune out for basic talks early in the workshop if needed and noobs to tune out when supercomputing topics are brought up.

I would add an extra day to allow a bit more hands-on activity and workshopping. Some of the hands-on activities were a little bit rushed, such that if I fell behind I couldn't catch up easily. It would also be very

useful to have a list of expectations of installed packages and libraries beforehand, to keep me from spending tutorial time on that sort of thing.

I would start with an introduction of each person, why they are there, what are the computing challenges and add more break out groups. A lot of the lunches were the most helpful time, and that could have been structured into groups interested in advanced statistical methods, local parallelization, cloud platforms and different R packages. These are all different interests, and could have refined some of the more general conversations.

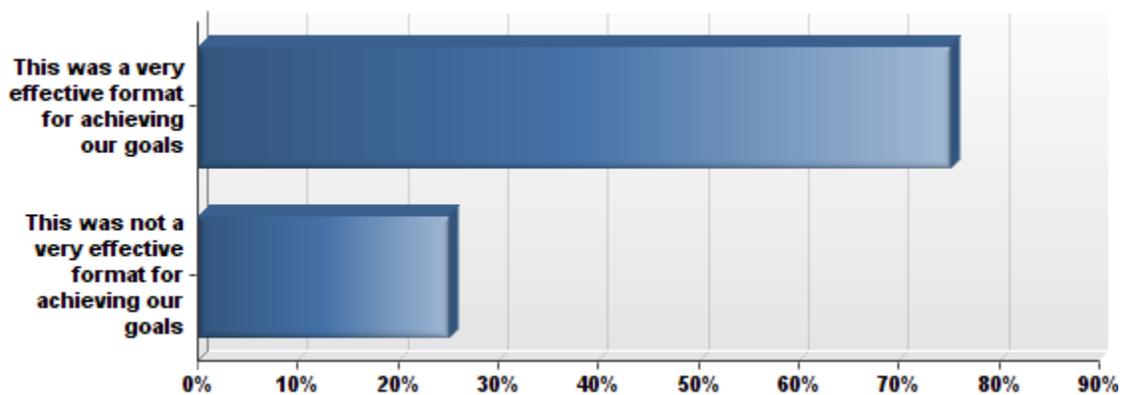
More hands-on exercises, perhaps working in a group with other participants. Introductions so we could learn names and a bit about what other participants were doing.

The tutorial about teaching R was unexpected did not fit in the announced content. I think that time could have been used better.

More 'computer practicals as I mentioned before for different settings, e.g. (1) setting up an interactive session (like we did); (2) setting up a batch job with a model written in C; (3) setting up a batch job which consists of a mixture of R and C (e.g. model written in C but an MCMC algorithm carried out in R).

I would have preferred more time working in small groups on specific problems

7. How do you feel about the format of the tutorial?



8. The tutorial format would have been more effective if:

There was more preparation for some of the hands on sessions- there were many small problems and also confusion about whether Amazon would be a pay session. We spent a lot of time setting this up and waiting. I think a direct login to a local cluster may have been much more efficient and also better teaching. There could have also been more "break" time- not for breaks as such but for pairing up in small groups to discuss particular problems and to network. Lastly, some time for presentations from the audience would have been good- allow to present our own particular needs and problems (maybe 15 minutes each for 5 or 6 people) to get feedback from the audience and instructors. I know several audience members (not me, haha) had a lot of great expertise to share.

I would have liked much more hands-on activity versus lectures. Also, some of the lecture topics were not pertinent to my skill level or interests (e.g., Rccp, teaching R). I think these might have been valuable for some people, but it might have been better if there had been a few concurrent sessions among which participants could choose.

Needed more hands-on, especially in using Amazon EC2 and with our own data. We set it up but didn't do anything with it. Mostly we listened and watched without having enough time to follow what the

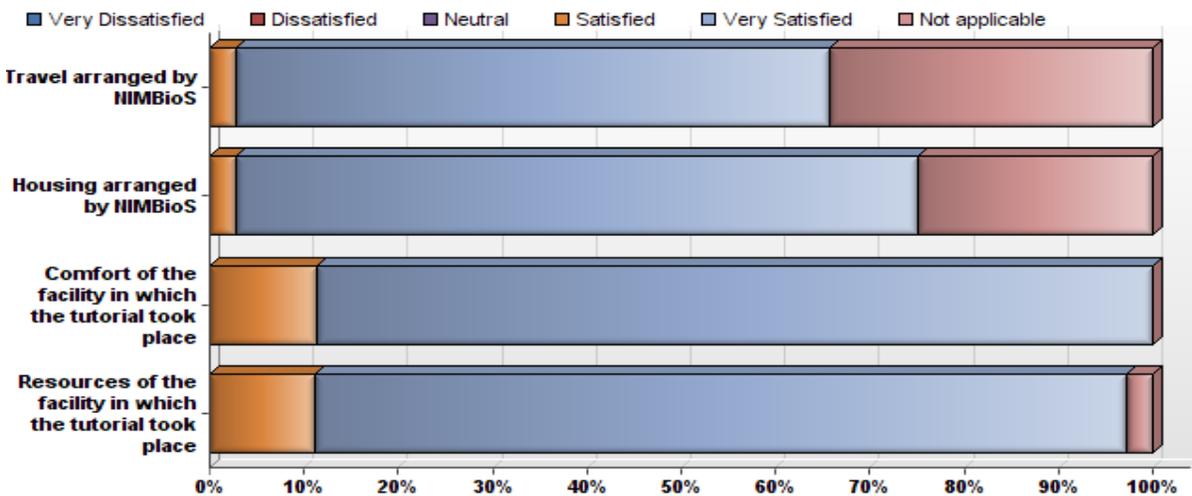
instructors were doing or understand the lines of code. A primary concern with the tutorial is that it was advertised as what "every life scientist needs to know about computing in the cloud". As a participant, I expected to learn how to compute in the cloud, yet this is the one thing we didn't really learn. What we learned was high performance computation. Great, but I think a big picture start would have helped. For instance, the first talk could have started with, "So you need to compute in the cloud. Why? Your analysis is computationally intensive and you need to free up your computer to do other things? Your data is too big to fit on a single thread? Great. We'll show you how to run your analysis in the cloud but we'll also show you how to speed up your analysis through better coding and running in parallel." It's not enough to show me what you are doing; I need to know why and when I should use package "x".

Presentations focused on practical problems and solutions to them using the techniques to be discussed. Some presentations did this, but many did not.

There were more informal, small discussion groups, with more hands on time.

I would have preferred that it focus on a few relevant topics, with more emphasis on actually doing an analysis in the cloud. There was quite a bit of time devoted to pbdR tools, including a beta program that we were instructed not to use. I felt that time would have been better spent doing hands-on work.

9. Please indicate your level of satisfaction with the tutorial accommodations:



10. Comments about accommodations:

Great!

The locked doors, warnings about security badges, etc. make it forbidding for visitors. The lack of a real sign for NIMBioS on the stairway going down other than a taped sheet of paper is also a bad signal.

I liked that the hotel was walking distance to the workshop.

These were some of the nicest meeting accommodations I have experienced.

Accommodations were wonderful and the NIMBioS facility was perfect.

The presentation room would get unreasonably hot when full of people.

Excellent, no-sweat arrangements--thank you!

Amazing.

A very nice hotel. A pleasant walk to lots of small restaurants.

Everything was fantastic. I'm excited about coming to NIMBioS again in the future.

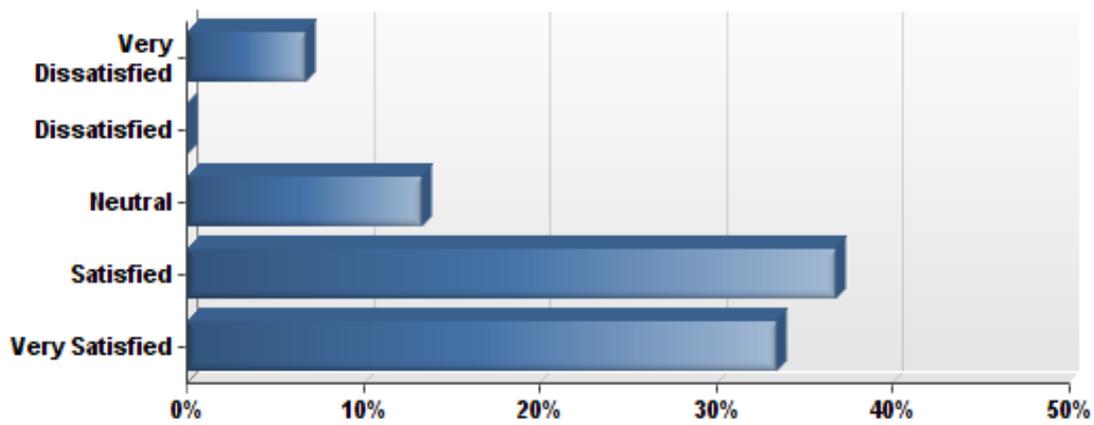
This is a comment about the travel arrangements. I put "somewhat satisfied" because I had three flights on the way in, and a total of eight hours waiting in airports on the way back. However, I live in a place where travel is not easy to arrange...

I appreciate the professionalism by all NIMBioS members

Outstanding!

Impeccable, couldn't have asked for better accommodations and a nicer group of people to interact with.

11. How satisfied were you with the opportunities provided during tutorial presentations and discussions to ask questions and/or make comments?



12. Please indicate any suggestions you have for facilitating communication among participants during the tutorial:

I think there should be a structured session on the first day in which each participant introduces him/herself and describes his/her field of research and particular need for HPC. This would allow people to know which of the participants to track down with similar interests/needs. Additionally, if a few of the modules were removed, the extra time could be used to allow people with similar interest to join in a working group and discuss how what was learned could be applied to their research.

More discussion time would be great. So many participants had different experiences and questions, and I think the one time that we did stop and take 30 mins for discussion was really helpful for me. I learned a lot!

More allotted time for breakout sessions, though this would have been difficult without lengthening the number of days for the tutorial

When there is a diverse group no one wants to admit that they do not know something. I think that the tutorial organizers could have started discussion by asking obvious questions to the speakers to get the ball rolling. This opens the floor a little more. This was done for some sessions. Also, a large part of a tutorial like this is the social aspect. As people got more comfortable with each other questions were being asked more openly. Perhaps having a reception the night before the tutorial starts could ease the social anxieties of questions.

More time to have audience present their own problems and get feedback, also some small group time. The sessions often went over time with not enough question time.

There is no such thing as a stupid question. A problem may be "embarrassingly parallel" but that does not mean that we know how to run it in parallel. Instructors should be aware that we are missing something if a question gets asked multiple times.

Block out time for Q & A and discussion sections as a group. Make us give a very brief (1 - 2 min) presentation on what we work on and what we're hoping to get out of the tutorial. At the very least, have introductions

Have the participants work in small groups to perform examples of the HPC calculations. Have a simple case study that guides users through the entire process, e.g. from reading in data, to formatting it, to performing the calculations, to displaying the output.

Very open and interactive presentations.

Have people pair up and work through some stuff.

More small group discussion time would have been helpful, so we could share about our computational problems with the hope of developing future collaborations.

I think you should just keep doing what you are doing! Having breakfast and lunch together was GREAT! I met so many people and learned so much from them.

It was a friendly group with lots of opportunities to ask questions. I didn't find the chance to ask questions simply because most of the material was outside of my area of expertise and I didn't know what to ask.

13. Additional comments:

I really, really enjoyed the tutorial! I feel like I learned a lot and have (at the very least) the foundation (familiarity with lingo) required to learn more in-depth in the future. I also liked that the meals and evenings were so free to socialize with other participants. I had the chance to make new friends and explore Knoxville. Overall I had a wonderful experience, and hope such opportunities will continue to exist for others.

Live streaming was a great idea, as was live-tweeting.

Overall, it was a great experience and I learned a lot over a short period of time. I hope to attend other tutorials or have the opportunity to return to NIMBioS in some capacity. Many thanks!

I thought this was an excellent workshop. Well-organized and executed. I hope to bring some of the speakers out to my institution to give talks.

Thank you for a great tutorial!

Very well -organized tutorial. Great set of speakers. Facilities were appropriate for the event. Nice and helpful staff caring for the participants

Great workshop :)

I know it's difficult to teach computing since everyone comes in at different levels. This was my first experience coming in at a level below what most of the lectures assumed I knew. I don't have a general solution to this sort of problem, but making the lectures into hands-on activities may have ameliorated part of (my) problem.

I learned a lot from this tutorial and felt the presenters and organizers did a great job. I just feel that the actually workshop did not quite match the advertisement, that the organizers assumed everyone was at a certain level, that a "big-picture" presentation would have helped me understand how and when I could

use all the information provided - which would have helped me get more out of the tutorial. I definitely needed more "hands-on" time.

Thank you!! The tutorial was a really great experience. I was really impressed by almost every aspect.

It was a very nice tutorial. The sessions ought to have been more interactive. (The AWS session tried to be, but there was too much confusion from the changes to the interface. If that had been smoothed out beforehand, it would have been good.)

This was such a wonderful opportunity, and I think it will provide a great foundation for the next stage in my career. Your administrative staff deserves a special shout-out for their excellent organization of this complex event, and the ease of interacting with them.

While i tried to provide constructive criticism, i don't want to give the impression that the course was not helpful. It was very good and i thoroughly appreciate the help.

Thank you all at NIMBioS for this opportunity. It was really the most exciting three days I have spent learning about computational tools, ever! I was very impressed!

It was a great workshop with a very engaged audience. I am already applying some of the things I learned in my research!

Thanks to everyone involved in organizing and instructing this workshop! I really enjoyed it. If there were one recommendation I could leave you with, it would be to target a less broad audience. I am an ecologist and need to use high performance computing, but my background with computation and programming was not sufficient to keep up with the material being presented. It seemed better targeted to computational biologists or programmers rather than ecologists who dabble out of a necessity to understand and use these resources.