To: Dr. Ashley Shade
From: STEM E.D., LLC (Libarkin)
Re: Edamame Workshop Evaluation
Date: October 20, 2014. FINAL.

A pre-workshop evaluation of the Edamame Workshop 2015 was conducted on June 21 & 22, with a post-workshop evaluation occurring June 30, 2015. Observations were also conducted at the start and end of the workshop. In all, 32 participants completed the pre-survey and 31 completed the post-survey; data from 30 participants could be matched between the pre- and post-tests.

Demographics

Demographic data are reported for pre-workshop respondents. Remaining participants ranged in age from 22-57, with an average age of 32.0 ± 8.0 years. Participants were 72% female (n=23), with the remainder male (n=9) and no transgendered participants. Participants were also 69% exclusively Caucasian (n=22). Remaining participants were: Asian/Asian American (n=4); Latino(a) or Latino(a) and Caucasian (n=5); and African/African American/Black (n=1). Finally, the academic status of participants included graduate students (n=21), M.S.-holding professionals (n=1), and Ph.D.-holding professionals (n=10).

EXECUTIVE SUMMARY OF RESULTS

We summarize evaluation results below. Following this summary, we provide: 1) A description of scales used to measure participant characteristics, including statistical analyses. This includes scale analysis of one new scale; and 2) results for qualitative and close-ended questions related to participant impressions of the workshop. In summary, we found that:

1. Observations suggest that *the structure of the EDAMAME workshop and interaction between leaders and students were excellent.*
2. *Participants perceived greater computational ability after engagement in the workshop.*
3. *Participants perceived greater computational understanding after engagement in the workshop.*
4. *Participants perceived greater coding ability after engagement in the workshop.*
5. *Participants were more comfortable with workshop content after engagement in the workshop.*
6. *Participants were generally very satisfied with the workshop*
7. *Participants felt the workshop met their needs and would overwhelmingly recommend it to others.*
8. Participants were generally positive about the workshop in their open-ended comments. Suggestions for improvement include: *encouraging some speakers to provide more context for lectures* and *content that might require more coverage.* See end of report for additional suggestions.

WORKSHOP EVALUATION AND RESULTS

The workshop was evaluated through observation, Likert-type scales, and open-ended prompts. Perceptions scales were established and validated in a previous workshop evaluation (see 2014 report); a scale related to comfort with computational tasks was validated. Participants completed identical pre- and post-workshop surveys containing four scales: *Perceptions of Computational*

Ability, *Computational Understanding*, and *Coding Ability*, and *Comfort With Computational Tasks*. Respondents also completed several demographic questions and responded to two open-ended questions and one box-and-arrow prompt covering workshop content. The surveys also contained questions about expectations (pre) and overall perceptions (post) of the workshop.

Table 1. Pre and Post Workshop Scale Averages

Scale	**Pre**	**Post**
Perception of Computational Ability	1.92 ± 0.53 (No to Low Ability)	3.24 ± 0.40 (Intermediate to High Ability)
Computational Understanding	2.45 ± 0.61 (Disagree to Agree)	3.57 ± 0.38 (Agree to Strongly Agree)
Coding Ability	2.27 ± 0.52 (Disagree to Agree)	3.38 ± 0.37 (Agree to Strongly Agree)
Comfort with Computational Tasks	1.59 ± 0.68 (Very Uncomfortable to Uncomfortable)	3.21 ± 0.39 (Comfortable to Very Comfortable)

Perception of Computational Ability

Participants rated their ability in each as No Ability, Low Ability, Intermediate Ability, or High Ability on five questions related to computational ability. A score of 1 implies No Ability, while a score of 4 implies High Ability. An average across these five items is used to calculate a score for Perception of Computational Ability. See 2014 evaluation results for scale analysis.

Results. Scores on the Perception of Computational Ability scale were calculated for the pre- and post-workshop survey. Overall scale scores were calculated as averages across all five items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=30). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different ($Z = -4.794$, $p \leq 0.001$). Statistical significance and mean results (Table 1) indicate that perceptions of computational ability increased as a result of instruction, from Low/No Ability to Intermediate/High Ability.

Computational Understanding

Participants were asked to indicate the extent to which they understood specific concepts, with ratings of Strongly Disagree, Disagree, Agree, and Strongly Agree on 13 items. A score of 1 implies low understanding (Strong Disagreement); a score of 4 implies high understanding (Strong Agreement). Items were modified significantly from 2014 items based on feedback from the workshop facilitator, prompting reanalysis of scale validity and reliability.

Factor Analysis. Exploratory factor analysis of responses to pre-workshop Computational Understanding items was undertaken to evaluate the presence of one or more scales. Four factors emerged based on eigenvalue analysis, while one very strong dominant factor was suggested by scree plot analysis. Based on this, confirmatory analysis for one scale was conducted. This analysis had an adequate Kaiser-Meyer-Olkin measure of sampling adequacy of 0.66, just above the 0.6 value recommended for factor analysis. The data set met other minimum conditions necessary for factor analysis. First, the majority of items correlated with at least one other item at a level over 0.3, indicating that a factor structure could be expected to emerge. The Bartlett's Test of sphericity was significant (χ^2 (78) = 215.6, p<0.001). Communalities for most items were above 0.3, indicating shared variance with other items. Given these data, confirmatory factor analysis was performed on all 13 items. Subsequent removal of two items due to poor loading yielded one satisfactory scale containing 11 items. This single scale of explains 40.7% of the variance in the data and all items yielded factor loadings over 0.32, again suggesting the presence of a single scale (Table 2).

Table 2. Factor loadings and communalities for Computational Understanding Scale. *Items modified from 2014 are in italics.*

Items	Factor Loadings	Communalities
I know how to assess the quality of Illumina data	.493	.244
I know the computational commands to summarize sequencing data	.463	.214
I know how to run R	.646	.418
I know the main differences in analyses offered by QIIME and mot	.662	.439
I am familiar with biom formatted files	.559	.313
I can name at least two different microbial metagenomic databases	.562	.316
I know what an R package is	.670	.448
I understand the structure of an OTU table	.825	.680
I know what a kmer is	.426	.182
I know the difference between within sample and comparative diversity	.757	.573
I know how to use metadata to guide community analyses	.810	.656

Reliability analysis. Cronbach's alpha was calculated to establish internal consistency of items. The scale has high reliability, with a calculated alpha for the sample of 0.873. Despite the small sample, this is above the recommended minimum of 0.70 for analysis of individual scores.

Results. Scores on the Computational Understanding scale were calculated for both the pre- and post-workshop surveys. Overall scale scores were calculated as averages across all 11 items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=30). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different ($Z = -4.785$, $p \leq 0.001$). Statistical significance and mean results (Table 1) indicate that perceived computational understanding increased as a result of instruction.

Coding Ability

Participants were asked to indicate the extent to which they agreed with statements about their coding ability, with ratings of Strongly Disagree, Disagree, Agree, and Strongly Agree on 13 items. A score of 1 implies low ability (Strong Disagreement), while a score of 4 implies high ability (Strong Agreement). Inversely worded items, those containing the word CANNOT, were reverse coded before reliability analysis. Minor modifications prompted reanalysis of scale containing 13 items; four items used in the 2014 evaluation were deleted. Scale metrics were similar to 2014 evaluation results.

Results. Scores on the Coding Ability scale were calculated for both the pre- and post-workshop surveys. Overall scale scores were calculated as averages across all 13 items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data (n=30). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different ($Z = -4.707$, $p \leq 0.001$). Statistical significance and mean results (Table 1) indicate that perceived coding ability increased as a result of instruction.

Comfort with Computational Tasks

Participants were asked to indicate the extent to which they felt comfortable performing specific computational tasks, with ratings of Very Comfortable, Comfortable, Uncomfortable, and Very Uncomfortable offered on eleven items. A score of 1 implies low comfort (Very Uncomfortable); a score of 4 implies high comfort (Very Comfortable). Participants were also allowed to respond "I do not know what this is"; this response was given a score of 0.

Factor Analysis. Exploratory factor analysis of responses to pre-workshop Comfort with Computational Tasks items was undertaken to evaluate the presence of one or more scales. Four factors emerged based on eigenvalue analysis, while one very strong dominant factor was suggested by scree plot analysis. Based on this, confirmatory analysis for one scale was conducted. This analysis had an adequate Kaiser-Meyer-Olkin measure of sampling adequacy of 0.745, above the 0.6 value recommended for factor analysis. The data set met other minimum conditions necessary for factor analysis. First, the majority of items correlated with at least one other item at a level over 0.3, indicating that a factor structure could be expected to emerge. The Bartlett's Test of sphericity was significant ($\chi^2(45) = 127.4$, $p<0.001$). Communalities for most items were above 0.3, indicating shared variance with other items. Given these data, confirmatory factor analysis was performed on all eleven items. Subsequent removal of one item due to poor loading yielded one satisfactory scale of ten items. This single scale explains 37.9% of the variance in the data and all items yielded factor loadings over 0.32, again suggesting the presence of a single scale (Table 3).

Table 3. Factor loadings and communalities for Comfort with Computational Tasks scale.

Items	Factor Loadings	Communalities
Navigating around a Unix file system to view and manipulate files	.626	.392
Making and using a "key" for secure EC2 connection	.383	.147
Installing axillary software on an Amazon EC2 instance	.725	.526
Using FastQC to assess the overall quality of raw sequencing data	.590	.348

Obtain summary information about sequence files (fasta, fna, fastq)	.706	.499
Trimming or filtering out errors from sequences by identifying low coverage kmers in high coverage areas	.642	.412
Assembling a metagenome with MEGAHit	.641	.410
Differentiating data modes in R, including numeric, complex, logical, character modes	.672	.452
Testing hypotheses about differences in community structure across samples or sites	.651	.424
Identifying features of public sequence databases that are useful to your research	.426	.181

Reliability analysis. Cronbach's alpha was calculated to establish internal consistency of items. The scale has high reliability, with a calculated alpha for the sample of 0.914. Despite the small sample, this is above the recommended minimum of 0.70 for analysis of individual scores.

Results. Scores on the Comfort with Computational Tasks scale were calculated for both the pre- and post-workshop surveys. Overall scale scores were calculated as averages across all ten items for all respondents. Data were inappropriate for parametric tests, so related samples nonparametric tests were run on matched data ($n=30$). Results from the Wilcoxon Signed Ranks Test indicate that pre- and post-workshop results are statistically different ($Z = -4.787$, $p \leq 0.001$). Statistical significance and mean results (Table 1) indicate that participants moved from being uncomfortable to comfortable with the material after instruction.

Open-Ended Questions and Box-and-Arrow Diagram

Two open-ended questions and one box-and-arrow diagram related to the concepts covered in the workshop were completed pre- and post-workshop (Table 4).

Open-ended questions were analyzed for evidence of increased understanding of workshop concepts, specifically as reflected in the complexity of participant responses. Response rates were higher post-workshop for individual questions, suggesting greater ability to respond to questions; as one participant put it pre-workshop, "*This is fairly over my head to be honest. I wouldn't even know where to start with this.*" In general, complexity of responses increased from pre- to post-workshop, with much more discussion of what the participants did not know on pre-workshop surveys. Further analysis of open-ended responses for accuracy is possible with collaboration between evaluators and workshop facilitators.

The box-and arrow diagram asked participants to "use the Illumina platform to evaluate both the phylogenetic diversity and functional potential" of a new microbial community. An example box-and-arrow model was provided to facilitate participant ability to respond. In general, post-workshop diagrams were much more complex and detailed than pre-workshop diagrams. Further analysis of box-and-arrow diagrams should occur in collaboration with the workshop facilitators.

Table 4. Open-Ended Response Rates

Open-Ended	Pre Response Rate*	Post Response Rate
Suppose that you are doing a 16S marker analysis of the gut microbial community from a non-model worm. This worm has a unique digestive tract evolved for the breakdown of a specialized diet. Discuss a sequencing strategy, a number of challenges you expect to face, and what kinds of additional resources are needed to help explore your data set.	68%	84%
Suppose that you are using the Illumina platform to sequence total DNA from soil samples. For this particular study, the soil samples come evenly distributed from two plot types on a single farm - one growing corn, the other dairy cow pasture. Discuss one or more approaches you would take to analyze the data, as well as your expected sensitivity and specificity to taxonomic and functional sequence identification. Include in your discussion the technical limitations of your analysis.	68%	84%

*Blank,"no idea", or similar responses are treated as non-responsive.

Observations

The workshop was observed twice, once at the beginning and once at the end. The EDAMAME workshops were very structured and designed to encourage interaction among the students and instructor – structure was even clearer than in 2014 workshops. Students appeared attentive and interested in the topics covered in the workshop. Visuals used by speakers had improved visual style, including better use of font size for visibility – this may be in response to 2014 evaluation recommendations. Overall, the interaction between leaders and students seemed adequate and productive for student learning.

Overall Workshop Impressions

Participants were asked to respond to close-ended questions related to overall impressions of the workshop (Table 5). Participants also responded to open-ended questions about their expectations for the workshop (pre-survey) and their perceptions of the workshop (post-survey).

Table 5. Overall Workshop Impressions. Plus (+) or minus (-) sign indicates potentially meaningful increase (+) or decrease (-) in score from 2014 workshop.

Workshop Components	Average Score (ideal score)	% Yes
Concept: Within Sample Diversity	4.45 (5)	NA
Concept: Comparative Diversity	4.45 (5)	NA
Amplicon analyses in QIIME	4.45 (5)	NA
Amazon EC2	4.71 (5) +	NA
Amplicon analyses in mothur	3.45 (5) -	NA
Shotgun metagenomics	3.90 (5) +	NA
Ecological Analyses in R	4.03 (5)	NA

Overall Workshop Rating	4.81 (5)	NA
Overall Invited Speaker Rating	4.61 (5)	NA
Meet Needs?	3.87 (4)	NA
Learn What Hoped to Learn?	NA	93%
Computational Understanding Change?	NA	97%
Recommend Workshop?	NA	97%

Close-Ended Questions. Twelve questions asked participants to rate components of the workshop, as well as the overall workshop and speakers, on a 5-point Likert scale of Very Poor-Poor-Adequate-Good-Very Good. A 1 corresponds to a Very Poor rating and a 5 corresponds to a Very Good Rating. Participants were also asked if the workshop met their needs on a 4-point scale (4=very well), if they learned as expected from the workshop, if understanding of computational science changed, and if they would recommend the workshop.

Results. Post-workshop responses to questions about the efficacy of the workshop indicate that participants were generally very satisfied with the workshop (Table 6). In general, participants rated the workshop components as Good-Very Good. Two exceptions (amplicon analysis in mother; shotgun metagenomics) were rated Adequate-Good. The shotgun metagenomics evaluation increased almost one full point from 2014; the amplicon analysis experienced a decrease. Workshop leaders may want to reconsider how this content is taught.

Participants generally felt the workshop met their needs and would overwhelmingly recommend it to others. Participants gave the workshop a rating of 4.61 out of 5, indicating the workshop met their needs well to very well. Ninety-three percent (n=29) of participants felt they had learned what they hoped to learn, while 97% (n=30) felt their computational understanding changed and that they would recommend the workshop to others.

Qualitative Data: Participant Expectations and Perceptions. On the pre-survey, participants responded to a prompt: "*Please provide any additional comments about your expectations for the workshop below*." Participants also responded to a similar post-survey question: "*Please provide any additional suggestions or comments about the workshop below*."

Results. In all, 19 participants provided comments about expectations on the pre-survey (Table 6). These comments related a desire to become comfortable with computation, fill knowledge gaps, and being able to perform better statistical analyses. Given their brevity, pre-workshop expectations are provided verbatim:

- *To become a computational master. I want to understand the workflow, but more importantly, familiarize myself with the semantics of bioinformatics so I can intuitively understand what particular parameters are doing and how I can optimize parameters for a particular data set. I also want to be able to know what multivariate analysis is appropriate for a particular hypothesis I want to test.*
- *Be able to answer ′very comfortable ′ to the all asked questions in this pre-course surveys.*
- *I have some knowledge of the topics presented in this workshop. I expect that this course will help to fill in some of the gaps in what I already know.*
- *I have a lot to learn, and many folks at my institution who hope that what I learn can be*

applied to several projects in my department.

- *Career-saving opportunity to get up to speed in a rapidly changing area of analysis that is fundamental to my work*
- *I am looking forward to this workshop for a few reasons. Primarily, to get a better understanding on using bioinformatic tools in the amazon cloud space, and secondarily, to become more familiar with the techincal aspects of all the tools (Mothur and R, definitely!)*

  And since I'm the only research that uses QIIME at my office and I'm self taught, it would be nice to see how other users work with the program.
- *Looking forward to meeting inspiring people with great ideas :)*
- *Hope to learn more about quality control for 16S metagenomics, R programs, and shotgun metagenomcis.*
- *Learn lots and have fun doing it!*
- *I am really excited to learn metagenomic analysis and assembly.*
- *My expectations for this course are to*
  1. Have fun
  2. Learn what databases are available for me to use for comparative studies of microbial communities and the contexts in which they are appropriate to use.
  3. Gain insight into the gold standard statistical program that is R.
  4. Gain insight into the technical limitations and assumptions made within the various platforms.
  5. Increase comfortability in working with large data sets (how to format the data such that programs can be written to execute tasks).
  6. Have fun :-)
- *I feel that I have some familiarity with several of the concepts in this survey, but lack sufficient practice and experience to claim confidence or comfort therewith. In addition, my experience is distinctly biased towards only a few of the programs and tools mentioned, where I have zero practice with the others. I expect to gain more confidence and skill with both my current knowledge and those things I have yet to attempt.*
- *I am just looking forward to gaining some experience with all the platforms and programs mentioned in order to do better statistical analysis of my metagenome.*
- *I hope I can handle the bioinformatics. I'm nervous about it, but excited to give it a try.*
- *I hope I will be able to circle more "YES" when I finish this workshop if I have the chance to take the same survey again.*
- *I'd like to be able to feel good when evaluating generated sequence data from beginning to end and knowing I chose the right programme to do it.*
- *Thanks!*
- *To learn more about the processes in working with metagenomics and working with computational analysis.*
- *It will be great, all information from your schedule are useful*

In all, 23 post-survey responses expressed either satisfaction with the workshop or suggestions for improvement. In general, comments were positive. Participants appreciated the the access to tutorials, the variety of talks, and the overall workshop experience. Examples are provided below:

- *The EDAMAME worshop is just amazing and awesome, all the informations that got from the tutorials, the talks and different discussion groups are so precious and I'll be able to quickly apply theme in my own data. LONG LIFE TO EDAMAME !*
- *…the Instructors did an outstanding job in helping us work with this huge amount of data*
- *This was truly a fabulous workshop.*

- *Ashley is a fantastic instructor/researcher with an amazing presence in the room. Her personality and wit have made this course a pleasure. Her openness to answer questions and provide open-source resources for the community cannot be recognized enough. This course has impacted our microbial ecology community immensely in a very positive way. The TAs were fantastic and always there to help when I was lost - they have been an amazing resource for me and the entire class.*

Suggestions for the workshop included:

1. Move evening speaker to earlier in the evening
2. Provide a formal "common mistake" file generated highlighting some of the pitfalls we ran into during this course
3. Provide a tutorial for windows users specifically
4. Encourage speakers to prepare an outline and tutorials before lecture, to provide goals at start of lecture, and/or to provide access to speaker slides before or after lecture (x2)
5. Provide more instruction on: shotgun metagenomic sequencing (x3), MG-RAST (x2), Xander (x2), mothur (x2), RDP, methods to map reads to contigs, command syntax, R in general
6. Consider having fewer speakers to allow more time for participants to practice
7. Have participants download all materials before workshop

Table 6. Perception and Expectations Response Rates

Prompt/Question	**N responses**
Please provide any additional comments about your expectations for the workshop. (pre)	19
Please provide any additional suggestions or comments about the workshop. (post)	23

END REPORT.