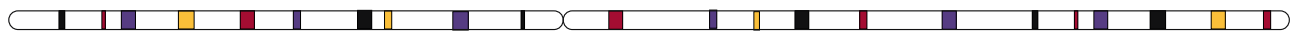# Summary of the National Center for Genome Analysis Support (NCGAS) 2018 *de Novo* Transcriptome Workflow and Workshop

Sheri A. Sanders
Carrie L. Ganote
Bhavya Papudeshi
Craig A. Stewart
Thomas G. Doak

Indiana University
PTI Technical Report

June 7, 2018

**PERVASIVE
TECHNOLOGY INSTITUTE**

The National Center for Genome Analysis Support is a Pervasive Technology Institute Center at Indiana University in Bloomington

<u>Contents</u>

Introduction

The National Center for Genome Analysis Support (NCGAS) held a workshop entitled "*de novo* Assembly of Transcriptomes using HPC Resources" on April 30th, 2018 through May 1, 2018. This workshop was in serving NCGAS's mission of enabling the biological research community to analyze, understand, and make use of the genomic information now available by packaging our now seven years of experience assisting with *de novo* transcriptome assemblies and running High Performance Computing (HPC) resources into a documented, easily approachable workflow for our users. The workshop covered common questions and problems that our users have had in HC (such as job handling, resource availability, data management, and troubleshooting) and in the construction of transcriptomes (such as software choices, combination of assemblies, and downstream analyses). The two-day workshop also highlighted the available resources for US scientists, concentrating heavily on available XSEDE resources for analyses, visualization, and archiving of data.

While the primary goal of the workshop was to provide advanced training on a common methodology, NCGAS also sought to implement an internal workflow for converting knowledge and experience gained from working on a breadth of projects with our users into easily transferable products (workflow/workshop).

Methods

1) Compile of our previous presentations, scripts, and tickets on the topics

NCGAS has spent years presenting on transcriptome assembly and analyses at numerous national conferences (i.e. Galaxy, Plant and Animal Genome) and guest lectures (i.e. Bethune-Cookman, Mount Desert Island Biological Laboratories Environmental Genomics Workshop). While independently desperate, the library of presentations, demos, and lectures compiled over the last seven years of NCGAS covered most topics in *de novo* assembly. NCGAS's work with Keithanne Mockaitis has resulted in an established workflow for assembly. As such, we had scripts premade for running almost all of the pertinent software, and hundreds of tickets covering the most common problems in working in HPC. This pool of resources served as the material to create the workflow and workshop without much additional developmental effort.

2) Clean up, annotate, and test to create best practices workflow

We took our scripts, made them as generic and readable as possible for the two main machines NCGAS clients utilize - IU's Carbonate and PSC/XSEDE's Bridges. All steps to the multi-software assembly were organized and linked to form the workflow. Two major considerations were made here. First, that this was not meant

to be "push button". We want users to get familiar with the job scripts and commands while taking away the stress of building them from scratch. Common methodology was made the default, but links and documentation for other situations (polyploidy, stranded sequencing, etc.) is listed in documentation. Second, despite lowering the learning curve to get started running the software, we wanted to preserve best practices - specifically that multiple parameters and multiple assemblers should be used to account for individual software biases. As a result, four assemblers (SOAP *de novo*, Trinity, TransABySS, and Velvet/Oases) were included with several kmers for each assembler, resulting in 19 individual assemblies. These were then set up to be merged by concordance via the software EviGenes.

This step resulted in an easy to run, easy to read, easy to modify workflow that followed current best practices.

3) Test workflow on our own work with clients

During the first three months after development, this workflow was used in house to test for bugs, usability, etc. on client projects NCGAS was contracted to complete. This step resulted in minor adjustments of the workflow.

Evaluation:
The workflow was developed by one team member and handed off to the other team members to work with. Direct feedback was handled internally but allowed for testing on multiple projects.

4) Test workflow with independently working clients

After guest lecturing at Mount Desert Island Biological Laboratory's Environmental Genomics course, NCGAS received several requests for assistance with assembly of *de novo* transcriptomes using Illumina data. We used this opportunity to beta test our workflow with students at least semi-familiar with command line. These individuals got support, experience, and training, while NCGAS got feedback on the workflow.

Evaluation:
Direct feedback was solicited from clients in the form of tickets and informal survey. Comments on design and general organization of the workflow were solicited during presentation of the workflow at a demo at Plant and Animal Genomes 2018.

5) Design workshop around presentations and workflow

Once we had the workflow solidified, we built a workshop around the use of it. We ordered previous talks and demos into a logical order to cover the material and common problems from our library of tickets. We used publicly available data as a test set for using the assembly and talked about our experience with downstream analyses of the transcriptomes we have generated from this workflow.

This step resulted in a scaling up of knowledge transfer, further documentation of the workflow and surrounding topics, and conversion of the information into digital resources that we can point new clients to and low effort development of an in-person workshop.

Evaluation:
Pre and post surveys for the workshop were designed to measure comfort levels with included topics, comment on the successful and superfluous aspects of the workshop and suggest future topics for workshops.

Participants

NCGAS contacted all current and previous clients and all participants of previous workshops NCGAS via our mailing list. Information was posted via the NCGAS twitter and Facebook page, Indiana University's IT News and Events page as well as the Evolution Directory List Serve.
All applicants were directed to a survey (Appendix A) allowing them to provide information on their background, projects, and status (faculty, staff, Ph.D. student, etc.). All applicants were independently ranked by all NCGAS staff on the following criteria:
  • Appropriateness of project
  • Appropriateness of learning goals in relations to this workshop
  • EPSCoR status

The ranks were averaged across and the 40 top-scoring applicants were offered space in the workshop. Several had to decline due to scheduling conflict and will be reconsidered in the second iteration of the workshop in the fall.

NSF funding for participant travel to workshops was used to support 26 student/staff travel vouchers. The workshop itself was free of charge, as NCGAS is already supported by NSF funds and we did not want participants to use their NSF grant money to pay for time already covered by the NSF.

Results

The primary products of the work were the workflow and the workshop. The workflow allows us to streamline work to publication and posters, as well as presentations and websites on the tool. The workshop provides contact with our clients, generates more clients, and transfers knowledge in a quantifiable way.

1) Publications - Internal use of workflow

Yang, T., L. Fang, S. Sanders, S. Jayathi, G. Rajan, R. Ppdicheti, S.K. Thallapuranam, K. Mockaitis, F. Medina-Bolica. (2017). Stillbenoid prenyltransferases define key steps in the diversification of peanut phytoalexins. Journal of

Biochemistry. Retrieved from
http://www.jbc.org/content/early/2017/11/17/jbc.RA117.000564

Several others will result from work with beta testers.

2) Presentations using workflow

Ganote, C., S. Sanders, L. Wu, T. Doak, and K. Mockaitis. (2018). Solving the
challenges of complex genome analysis collaborations on-line using XSEDE
resources. In Plant and Animal Genomics 2018, San Diego, CA. Retrieved from
http://hdl.handle.net/2022/21903

Sanders, S., R. Podicheti, T. Yang, L. Fang, S. Jayanthi, G. Rajan, T.K.S. Kumar, F.
Medina-Bolivar, and K. Mockaitis. (2018). Stilbenoid prenylation pathway
discovery in peanut using targeted transcriptomics. In Plant and Animal
Genomics 2018, San Diego, CA. Retrieved from
http://hdl.handle.net/2022/21902

Sanders, S., and M. Pfrender. (2017). *de novo* assembly and annotation of
Ambystoma laterale and *Ambystoma jeffersonianum* transcriptomes: the first
steps toward investigating polyploid salamander expression. In Evolution
Conference. Retrieved from
https://scholarworks.iu.edu/dspace/handle/2022/21599

3) Presentations on workflow

Sanders, S., C. Ganote, and T. Doak. (2017). *de novo* Transcriptome Assembly. In
Mount Desert Island Biological Laboratory. Retrieved from
https://scholarworks.iu.edu/dspace/handle/2022/21645

Sanders, S., C. Ganote, B. Papudeshi, K. Mockaitis, and T. Doak. (2018). NCGAS
makes robust transcriptome analysis easier with a readily usable workflow
following *de novo* assembly best practices. In Plant and Animal Genomics
2018, San Diego, CA. Retrieved from http://hdl.handle.net/2022/21904

4) Websites for available data

GitHub: https://github.com/NCGAS/Transcriptome-assembly-workshop-2018

Workflow Documentation Page: http://ncgas.org/WelcomeBasket_Pipeline.php

5) Beta test commentary excerpts - full comments in Appendix B

"The use of this pipeline has saved me tons of time from having to figure out the
script for each assembly program and it is VERY easy to use, especially for a person
like myself who barely understands Linux!"

"If this pipeline was not available, I would have most likely used only package and at one kmer size for my assembly, and it would have probably taken me just as long to figure out and run."
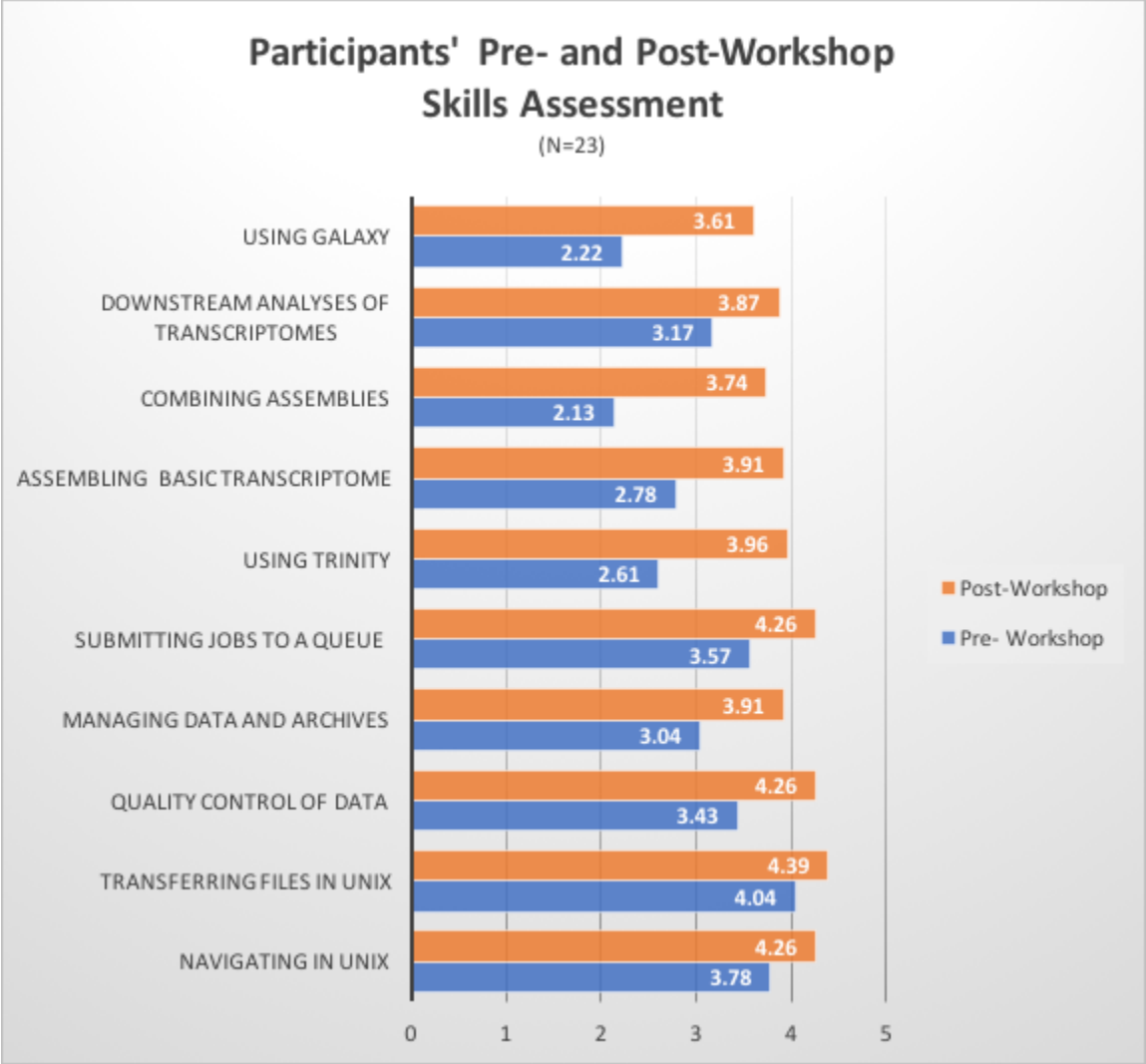
6) Workshop results

● Interest Level – We had 69 applicants for the 30 seats we had planned for the workshop. These applicants came from 42 institutions in 19 states/1 territory (6 EPSCoR states/1 territory), and three foreign countries.

● Workshop Attendance – We had 30 people attend the workshop from 26 institutions (4 Minority Serving Institutions) in 13 states/1 territory (4 EPSCoR states/1 territory). The sex ratio was about even at 13 male and 10 female participants (of the 23 reported on the survey). There were 13 white, 6 Asian, and 2 African American participants (of the 21 reported on the survey).

● Efficacy – In general, the format for the workshop was well received, with 22 of the 23 respondents stating they preferred the workshop be in person. One suggested having live webinars monthly. **Overall, the comfort level of the participants increased 31% in general transcriptome assembly.** When queried about ten individual skills before the workshop, four were in the "no hands-on experience" range, four were in the "can run limited examples" range, and one was in the "ability to run realistic examples" range. The average skill level was 3.07 out of 5. After the two-day workshop, six of these skills were on average in the "ability to run limited examples", and four were in the "ability to run realistic examples" range, with an average skill level of 4.02 out of 5 – a full 20% more confident in their skill set and on average able to run real data.

Comfort in overall transcriptome assembly - 3.74 to 5.94 on 7-point scale (increase of 31% in comfort level)

Comfort level in each skill – on a 5-point Likert scale as defined below:
1 – No previous experience of knowledge
2 – Knowledge of its function, but no hands-on experience
3 – Ability to run very limited examples, such as small data sets and tutorials
4 – Ability to run more realistic examples, such as real data
5 – Ability to troubleshoot tasks for myself and others

Participants' Pre- and Post-Workshop Skills Assessment

(N=23)

Skill	Pre-Workshop	Post-Workshop
USING GALAXY	2.22	3.61
DOWNSTREAM ANALYSES OF TRANSCRIPTOMES	3.17	3.87
COMBINING ASSEMBLIES	2.13	3.74
ASSEMBLING BASIC TRANSCRIPTOME	2.78	3.91
USING TRINITY	2.61	3.96
SUBMITTING JOBS TO A QUEUE	3.57	4.26
MANAGING DATA AND ARCHIVES	3.04	3.91
QUALITY CONTROL OF DATA	3.43	4.26
TRANSFERRING FILES IN UNIX	4.04	4.39
NAVIGATING IN UNIX	3.78	4.26

● Select Comments on Efficacy of workshop – In free commentary, it appears our workshop was well received and perceived as a useful two days in boosting confidence, knowledge, and size. Many participants were surprised to learn about XSEDE (one had heard of it prior to the course) and all the resources available to them through NSF. Many were surprised (and happy) to hear about the breadth of services NCGAS provided as well - we've received six new allocations, in addition to the five that had allocations before the workshop.

Responses to "Favorite Part":
"This was the first workshop of its kind that I have attended and one of the most useful workshops in my doctoral degree. So, I received everything and think that everything is very important and directly applicable to my research."

"Fantastic workshop. I learned how to do in 2 days what I have been trying to do on my own for more than 5 months (make a Trinity pipeline -> annotation -> EdgeR analysis)"

"Successfully completing this exercise gave me the confidence to tackle and use published bioinformatics pipelines that I would otherwise have been too intimidated to try and run."

"You guys really know how to handle this type of workshop and I think the teacher to student ratio was spot on. I've been to some of these things where we get through less than this in much more time. Of course that was all on command line and not Galaxy, but you guys are seriously working magic! Galaxy is soooooooooooo nice!"

"Being able to run through the entire pipeline myself in a streamlined fashion - I have attended workshops where there was no flow to what we were learning and everything seemed very disjointed, and in those instances, it has always been very hard to follow what we are actually supposed to be doing, but that was not the case here."

Responses to "Most surprising thing you learned":
"The resources that are available to NSF researchers. I've been struggling on my own with my limited coding knowledge for quite some time. NCGAS is an absolutely amazing resource and cuts down on wasted time for me. It all makes analyzing transcriptomes and genomes much easier, faster and more efficient, which contributes to the scientific community overall."

"The number of NSF-funded super-computing resources available to researchers; the presenters did a great job of publicizing such resources"

"you being available for anyone when they need some help to move to the next step and availability to answer the questions related our independent work as well"

- Internal comments on workshop structure efficacy - Having five instructors in the room at all times was critical. We had the NCGAS main staff present - Tom Doak (Manager), Carrie Ganote (Bioinformatic Analyst), Bhavya Papudeshi (Bioinformatic Analyst), Sheri Sanders (Bioinformatic Analyst) - as well as Keithanne Mockaitis (long-term collaborator). Having one instructor presenting and the rest moving around to prevent any participants from falling behind kept everyone on track and made sure everyone was able to complete the exercises. Building in obvious outputs to some activities (graphical output to screens for Galaxy and Data Transfer) facilitated this further, as staff could walk around the room and easily survey when everyone had hit a checkpoint. We plan to add another instructor to help this go more smoothly in the future.

NCGAS has developed a close relationship with the rest of the HPC center at IU over the years, which added an additional twist to this workshop that is often missing from similar workshops. HPC topics learned from working with the HPC center were integrated into the analysis, emphasizing the importance of clusters as useful tools that can be used more efficiency with more training/knowledge. Novice job handling and data management can add unnecessary time to analyses and burden on systems, but these topics are seldom taught explicitly to biological users. Covering this material was appreciated by participants, as several users sighted these topics as the most useful.

The close association with HPC staff was also helpful in allowing NCGAS to pull in system administration staff to quickly fix a system-side issue immediately (while staying on schedule) and talk to the participants when technical questions arose. The HPC staff also provided us with knowledgeable guides for tours of the IU Data Center (a much appreciated activity) and helpful comments during the evening chats over dinner. Lightly including the HPC personnel in the workshops helped expose biologists to conversing with and humanizing the computer scientists.

Conclusions and Further Directions

We were able to synthesize and package our previous knowledge and work into a workflow for *de novo* transcriptome assembly and a workshop based around it. The two-day workshop was well attended and received. We were able to elevate the skill set across a diversity of HPC and bioinformatic skills for a diverse set of students, staff, and professors across the country. We introduced almost all of them to XSEDE resources to power their new skills, toured the data center while talking about what it takes to run and manage these machines, and introduced them to some HPC staff to help provide the biologists with a better comfort level in working with HPC systems/staff.

At present, one publication has used the pipeline, but this will likely increase quickly as three of our own posters have featured the pipeline and we are working with several early adopters on publications. Two presentations and two websites have also come out of the workflow. Due to the significant interest in the workshop, it is our intent to have a second session in the fall. We will incorporate comments and suggestions to improve the workshop and will recruit an additional instructor to help answer questions and troubleshoot as we go through demos and exercises.

Additionally, it is our hope (pending continuing funding of NCGAS) to follow similar protocol for transfer of knowledge for our other work on genome annotation, microbiome and metagenomics work, and basic skills (R, Python, Unix). The development of similar workshops on these topics will require limited development time, as we can again utilize our large volume of presentations, experience, scripts, and tickets from the last seven years. We already have this process started for a four-session R workshop, to be given late summer semester of 2018.
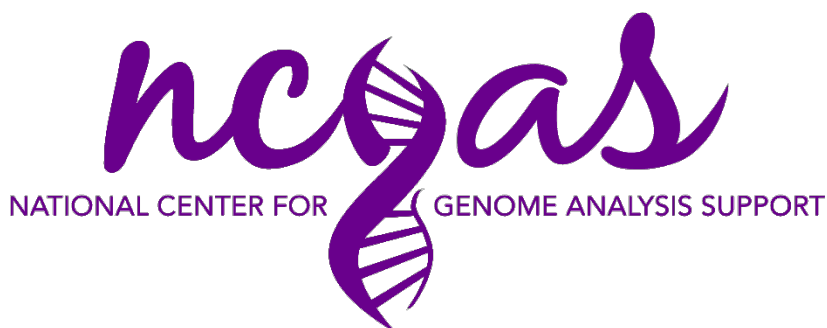
Spring 2018 Workshop Registration Form

NCGAS is preparing a 2-day workshop in Spring 2018 to provide training in basic bioinformatics concepts, with hands-on tutorials and walkthroughs. Our staff will share our collective expertise to help you become more familiar with High Performance Computing (HPC) environments, available lab technologies to answer your experimental questions, web-based visual bioinformatics tools such as Galaxy, data management tips and troubleshooting approaches. We have room for ~30 applicants so this form will help us make tough choices should we get too many responses!

NOTE: This workshop is now at capacity. If you would like to be added to the waiting list, please feel free to fill out the form!

1) **First Name**

2) **Last Name**

3) **Best Contact Email**

4) **Institution/University/Organization Name**

5) **State or province of your institution**

6) **What is your current status?**

Student - Undergrad

Student - Masters

Student - PhD

Post-doc

Faculty

Staff

Other

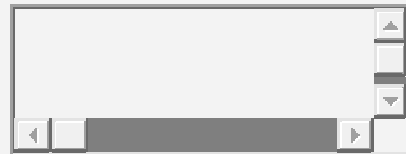**7) List any grants with which you're involved (investigator, researcher, consultant, etc):**

Expand

8) Are you (or your institution) able to cover the full cost of lodging + transportation + meals during the workshop?

Yes

No
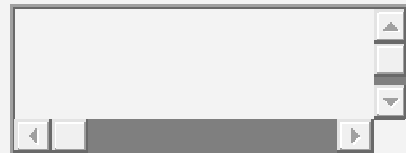
reset

9) What is the biggest challenge you've faced so far when dealing with bioinformatics?
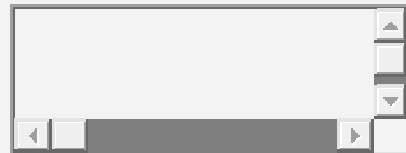
Expand

10) Describe in 1-2 paragraphs what you hope to learn during this workshop.

Expand

11) Describe your current work in 1-2 paragraphs - are you part of a lab project, do you lead your own work, or are you studying something particularly compelling?

Expand

APPENDIX B: Beta Testing Commentary

REQUEST:
Hey all!

The three of you were using my pipeline, and I was wondering if you had a second to respond to a couple questions:

1) Has this made it easier to run the software packages?

2) Got a one sentence review you want to leave?

3) Suggested Improvements?

RESPONSE:
FIRST:
Hi Sheri,

Sure thing! To be honest, I haven't finished running through the pipeline yet. We had some delays getting data, and the Thomas fire really threw us for a loop here in Santa Barbara. But now I'm finally catching back up on things! 😬

1) Definitely. If this pipeline was not available, I would have most likely used only package and at one kmer size for my assembly, and it would have probably taken me just as long to figure out and run.

2) The pipeline has compiled and organized everything I needed, saving me time and frustration on a process that otherwise would have taken me months to disentangle.

3) The biggest problem I've had is running into virtual memory and wall time limits, as it takes some time for before it's actually hit those limits and I've been alerted the job has failed, then I adjust the parameters and try again. I've tried setting vmem to 500, but I've still run into limit problems. For some of the runs, I've actually tried breaking up the code into separate jobs (i.e., running 34, 45, and 55 separately). I'm not sure if that's the best way to go about things?
One suggestion might be to include in the READMEs, what it should look like when steps are done running (i.e., what files you should expect to have to know everything has run okay).

Thanks again for all of your help, Sheri!
Best,

Juliet

SECOND:
1) Has this made it easier to run the software packages?
Extremely! All I have to do is tell it my file names and run each program. This is so much easier than figuring out the script for each program.

2) Got a one sentence review you want to leave?

The use of this pipeline has saved me tons of time from having to figure out the script for each assembly program and it is VERY easy to use, especially for a person like myself who barely understands Linux!

3) Suggested Improvements?

So far I have none.

Heather Walsh

APPENDIX C: Workshop Schedule
NCGAS SPRING WORKSHOP 2018 Schedule
Bloomington, IN

Day 1	Activity		Lead
8:30 am	Registration	All	
9:00 am	Introduction to NCGAS and staff		Tom
9:30 am	Introduction to Clusters and Other Resources		Sheri
11:00 am	BREAK		
11:15 am	Optimizing Jobs		Carrie
12:15 pm	Data Center Tours/Lunch	Staff	
2:00 pm	Data Management and Movement		Bhavya
3:30 pm	BREAK		
3:45 pm	Introduction to Assembly and Pipeline		Sheri
5:00 pm	Publicly Available Resources		All
6:30 pm	Dinner at Nick's		

Day 2	Activity		Lead
9:00 am	Using Galaxy		Carrie
10:00 am	Hands on Pipeline Use		All
11:15 am	BREAK		
11:30 am	Getting Help in HPC		Sheri/All
12:30 pm	Lunch		
1:30 pm	Hands on Pipeline Use Cont'd		Sheri/All
3:00 pm	BREAK		
3:15 pm	DE Analysis	Sheri	
4:45 pm	Downstream Analyses Discussion		Sheri
5:45 pm	Drinks at Upstairs		

NCGAS Pre-Workshop Survey

Start of Block: Informed Consent

Q10 Thank you for registering to attend the NSF-funded National Center for Genome Analysis Support (NCGAS) Spring Workshop. To inform workshop content, we ask that you participate in this short survey for the purpose of gauging the needs and experience levels of attendees. Data will also be used in evaluating workshop effectiveness and future workshop content.   Your participation, as well as all survey questions, are optional and your responses are confidential. Data will be reported in the aggregate without any identifying information that you choose to, or inadvertently, disclose.   Your decision to participate will not in any way affect your relationship with the NSF, the NCGAS project, the Pervasive Technology Institute, or Indiana University.
 For questions about this survey, including problems with accessing the survey, please contact cesg@iu.edu and reference protocol #1804120218/exempt, approved on April 26, 2018, by the Indiana University Institutional Review Board.

 Do you agree to participate?
1. Yes (1)
2. No (2)

Skip To: End of Block If Thank you for registering to attend the NSF-funded National Center for Genome Analysis Support (N... = Yes
Skip To: End of Survey If Thank you for registering to attend the NSF-funded National Center for Genome Analysis Support (N... = No

End of Block: Informed Consent

Start of Block: Default Question Block

Q1 Have you ever used command line before?
3. Yes (1)
4. No (2)
5. Not sure (3)

Q2 How comfortable are you using the computer via command line?

6. Extremely uncomfortable (1)
7. Moderately uncomfortable (2)
8. Slightly uncomfortable (3)
9. Neither comfortable nor uncomfortable (4)
10. Slightly comfortable (5)
11. Moderately comfortable (6)
12. Extremely comfortable (7)

_____

Q3 Have you previously worked with Unix?

13. Yes (1)
14. No (2)
15. Not sure (3)

_____

Q4 Are you or are you currently working with any bioinformaticians?
16. Yes (1)
17. No (2)

_____

Q5 Which, if any, bioinformatics tools have you used?

_____

_____

_____

_____

_____

_____

Q6 Do you have existing data that you plan on assembling?

18. Yes (1)
19. No (2)

End of Block: Default Question Block

_____

Q7 Please provide a brief description of the data you plan on assembling.

_____

_____

_____

_____

_____

- -

Q8 Are there any specific issues you have encountered with your data to-date?

_____

_____

_____

_____

_____

Q9 What is your skill level for each of the following:

	No previous experience or knowledge (1)	Knowledge of its function, but no hands-on experience (2)	Ability to run very limited examples, such as small data sets and tutorials (3)	Ability to run more realistic examples, such as real data (4)	Ability to troubleshoot tasks for myself and others (5)
Navigating in Unix (1)	20.	21.	22.	23.	24.
Transferring files in Unix (2)	25.	26.	27.	28.	29.
Quality Control of Data (3)	30.	31.	32.	33.	34.
Managing Data and Archives (4)	35.	36.	37.	38.	39.
Submitting Jobs to a Queue (5)	40.	41.	42.	43.	44.
Using Trinity (6)	45.	46.	47.	48.	49.
Assembling a Basic Transcriptome (7)	50.	51.	52.	53.	54.
Combining Assemblies (8)	55.	56.	57.	58.	59.
Downstream Analyses of Transcriptomes (9)	60.	61.	62.	63.	64.
Using Galaxy (10)	65.	66.	67.	68.	69.

Q17 How comfortable are you assembling a transcriptome?

70.Extremely comfortable (1)
71.Moderately comfortable (2)
72.Slightly comfortable (3)
73.Neither comfortable nor uncomfortable (4)
74.Slightly uncomfortable (5)
75.Moderately uncomfortable (6)
76.Extremely uncomfortable (7)

End of Block: Skill Level

Start of Block: Demographics

Q11 Please describe your institution/organization: *Please select all that apply.*
1. Institution located in an EPSCoR state (AL, AK, AR, DE, HI, IA, KS, KY, LA, ME, MS, MT, NE, NV, NH, NM, ND, OK, RI, SC, SD, VT, WV, WY) (1)
2. Minority-Serving Institution (MSI) (2)
3. Associate's College (all degrees are at the associate's level) (3)
4. Baccalaureate College/University (4)
5. Master's College/University (5)
6. Doctorate-Granting University (6)
7. Teaching-Focused Institution (7)
8. Research-Focused Institution (8)
9. Government Lab or Center (9)
10.High performance computing resource provider (e.g. NCSA, TACC, etc.) (10)
11.Non-Profit Organization (non-academic) (11)
12.Corporate/Industrial Organization (12)

Page
Break

Q12 What is your gender? *Select all that apply.*
13.Female (1)
14.Male (2)
15.Non-Cisgender (3)
16.Other (4)
17.Prefer not to disclose (5)

Q13 What is your ethnicity?
77.Hispanic or Latino (1)
78.Not Hispanic or Latino (2)
79.Prefer not to disclose (3)

Q14 What is your race? *Please select all that apply.*

80. American Indian (1)
81. Alaska Native (2)
82. Asian (3)
83. Black or African-American (4)
84. Native Hawaiian or Other Pacific Islander (5)
85. White (6)
86. Other: (7) _____
87. Prefer not to disclose (8)

_____

Q15 Do you consider yourself to be a person living with a disability?
88. Yes (1)
89. No (2)
90. Prefer not to disclose (3)

End of Block: Demographics

NCGAS Post-Workshop Survey

Start of Block: Informed Consent

Q13 Thank you for attending to attend the NSF-funded National Center for Genome Analysis Support (NCGAS) Spring Workshop. We ask that you participate in this short survey for the purpose of gauging your workshop experience, as well as the efficacy of the content and delivery format of the workshop of attendees. Data will also be used in evaluating content and formats of future workshop.   Your participation, as well as all survey questions, are optional and your responses are confidential. Data will be reported in the aggregate without any identifying information that you choose to, or inadvertently, disclose.   Your decision to participate will not in any way affect your relationship with the NSF, the NCGAS project, the Pervasive Technology Institute, or Indiana University.
 For questions about this survey, including problems with accessing the survey, please contact cesg@iu.edu and reference protocol #1804120218/exempt/exempt, approved on April x, 2018, by the Indiana University Institutional Review Board.

 Do you agree to participate?
91. Yes (1)
92. No (2)

Skip To: End of Block If Thank you for attending to attend the NSF-funded National Center for Genome Analysis Support (NCG... = Yes
Skip To: End of Survey If Thank you for attending to attend the NSF-funded National Center for Genome Analysis Support (NCG... = No

End of Block: Informed Consent

Start of Block: Block 1

Q12 Given your participation in the workshop, what is your skill level for each of the following:

	No previous experience or knowledge (1)	Knowledge of its function, but no hands-on experience (2)	Ability to run very limited examples, such as small data sets and tutorials (3)	Ability to run more realistic examples, such as real data (4)	Ability to troubleshoot tasks for myself and others (5)
Navigating in Unix (1)	93.	94.	95.	96.	97.
Transferring files in Unix (2)	98.	99.	100.	101.	102.
Quality Control of Data (3)	103.	104.	105.	106.	107.
Managing Data and Archives (4)	108.	109.	110.	111.	112.
Submitting Jobs to a Queue (5)	113.	114.	115.	116.	117.
Using Trinity (6)	118.	119.	120.	121.	122.
Assembling a Basic Transcriptome (7)	123.	124.	125.	126.	127.
Combining Assemblies (8)	128.	129.	130.	131.	132.
Downstream Analyses of Transcriptomes (9)	133.	134.	135.	136.	137.
Using Galaxy (10)	138.	139.	140.	141.	142.

Q3 How comfortable are you now assembling a transcriptome?

    143.   Extremely comfortable (1)
    144.   Moderately comfortable (2)
    145.   Slightly comfortable (3)
    146.   Neither comfortable nor uncomfortable (4)
    147.   Slightly uncomfortable (5)
    148.   Moderately uncomfortable (6)
    149.   Extremely uncomfortable (7)

Q2 What was your favorite part of the NCGAS Spring Workshop?

_____

_____

_____

_____

_____

Q4 What is the most surprising or interesting thing you learned?

_____

_____

_____

_____

_____

Q5 Which aspect(s) of the workshop did you find most useful?

_____

Q6 If you were to add one thing to the workshop, what would it be?

_____

_____

_____

_____

_____

Q7 What aspect(s) of the workshop did you find least useful?

_____

_____

_____

_____

_____

Q8 If you feel you would benefit from a more extensive workshop, what you would add?

_____

_____

_____

_____

_____

Q9 Considering the content of the workshop, which delivery format do you believe would be *most* effective?

    150.   In-person workshop (1)
    151.   Live webinar (2)
    152.   Recorded webinar (3)
    153.   Other: (4) _____

Q10 What other workshops would you like to see us offer in the future?

_____

_____

_____

_____

_____

End of Block: Block 1