National Center for Genome Analysis Support (NCGAS)

supported by
National Science Foundation
Division of Biological Infrastructure
Award #1458641

Annual User Assessment – 2019 Summary Report

prepared for

Thomas G. Doak, Ph.D.
Sheri Sanders, Ph.D.
Bhavya Papudeshi
National Center for Genome Analysis Support
Research Technologies
Indiana University Pervasive Technology Institute

prepared by

Julie A. Wernert
Program Evaluation and Assessment
Indiana University Pervasive Technology Institute

Harmony Jankowski
Community Engagement and Interoperability
Research Technologies
Indiana University Pervasive Technology Institute

August 2019
Version 1.1

Table of Contents

A. Document History

Entire document	.9	8/8/19	Created baseline document from template	H. Jankowski
Entire document	.9	8/21/19	Added data tables and narrative	J. Wernert
Entire document	1.0	8/28/19	Added discussion section	Doak, et al
Entire document	1.0	8/30/19	Edited and reformatted document	H. Jankowski

B. Document Scope, Findings, and Recommendations

This document summarizes the results of the National Center for Genome Analysis Support (NCGAS) Annual User Assessment conducted among 240 users from June 3 through July 8, 2019. This survey is conducted yearly or semi-yearly to determine the quality and appropriateness of NCGAS services, and to predict future trends and needs. This year, the content of the survey and the population surveyed has changed somewhat. In previous years, only users with NCGAS allocations were queried, but this year users of Trinity Galaxy and GenePattern were included. The NCGAS allocations are primarily those served through the NSF mandate, and we know from former surveys that the majority of these researchers have NSF funding. The Trinity Galaxy service is funded by an NIH ITCR grant, and is available to anyone world-wide; GenePattern is the same. Trinity and GenePattern are specifically aimed at cancer researchers; we know that many general biologists also use Trinity, but we know less about GenePattern users. We expect that the addition of GenePattern-users has increased the number of NIH-funded respondents.

B.1 Findings

Findings from the NCGAS Annual User Assessment include the following:

- A vast majority of respondents (88%) are "satisfied" or "extremely satisfied" with their overall experience using NCGAS resources.
- The most highly rated NCGAS services are its Carbonate cluster (4.45/5.0) and its Genome Analysis and Bioinformatics consulting services (4.32/5.0); while still rated very highly, its Blogposts (4.22/5.0) and Talks and Guest Lectures (4.22/5.0) are the lowest rated among the items evaluated.
- The most used NCGAS services are the Galaxy at Indiana University and NCGAS Trinity Galaxy services, and the Carbonate cluster.
- Over 40% of respondents report that NCGAS resources are essential in carrying out their research programs; another 49% describe NCGAS resources as helpful in conducting their research programs.
- Over 60% of respondents are affiliated with doctoral-granting institutions; just over four percent described their institution as a Minority-Serving Institution (MSI) and less than six percent indicated their institution was located in an EPSCoR state.
- Twenty-eight percent of respondents were from outside the United States, including nearly 16% from developing countries of the Global South.
- Over 38% of respondents report having first encountered NCGAS resources as a graduate student; 9% of respondents indicate first using NCGAS resources as an undergraduate or as a student at a two-year institution.
- Nearly 14% of all respondents are affiliated with Indiana University, NCGAS' host institution.
- Most users (77%) report having learned about NCGAS and its resources from colleagues and/or word-of-mouth interactions.
- Considering current services offered by NCGAS, users report the greatest areas of need for increased support are genome assemblers and annotation and long read and scaffolding technologies (e.g., PacBIO, 10x, Nanopore, Hi-C, Bionano, etc.)

- Nearly one-third (30.8%) of respondents' research involves RNA-Seq approaches, with another 16% citing genome assembly as a component of their research program.
- Over half of respondents reporting are funded, at least in part, by two agencies: the National Science Foundation (32.4%) and/or the National Institutes of Health (21.4%). Another significant portion of users who indicated a funding source (12.6%) are funded by international agencies and/or government entities, all of whom are Trinity and GenePattern users.
- Nearly a quarter of respondents indicate that NCGAS should increase its current levels of support for genome assemblers and annotation, with 19% of users citing long-read and scaffolding technologies for additional support.
- No clear consensus was established when respondents were asked to identify which services NCGAS should consider adding to its service portfolio in the future, but slightly greater needs were indicated for the installation of published bioinformatics applications, installed/curated genomics applications, and Bioinformaticians On-Call.
- Users cite memory during computations, followed by CPU speed/number of availably cores, among their greatest general research needs.

B.2 Recommendations

- Building on increased participation by users from non-research-focused institutions and those in underserved areas, continue to increase outreach efforts to Minority-Serving Institutions (MSIs) and teaching-focused institutions (e.g., Associate's, Bachelor's, and Master's degree-granting institutions), as well as those located in EPSCoR states.
- Continue ongoing outreach and training efforts targeting undergraduate and graduate students.
- NCGAS should consider increasing its support for genome assemblers and annotation, and long-read scaffolding technologies.
- NCGAS should consider adding resources to its service portfolio that would support the installation of published bioinformatics applications, installed/curated genomics applications, and/or Bioinformaticians On-Call.
- As there was no clear consensus about the existing or new services and resources for which NCGAS should add or increase its support, NCGAS should endeavor to examine users' needs and research trends carefully to ensure the greatest use and impact. To inform these decision-making processes, a more granular, focused survey instrument, user interviews, and/or focus group sessions should be considered in the data collection phase of these processes.
- Qualitative and quantitative data indicate that users are extremely satisfied with their overall NCGAS experience, especially in the area of user support. NCGAS should continue its emphasis on user support functions and user interactions to maintain its high-quality user experience.
- New services and delivery models should be deployed within the context of users' overall research environments, which are chronically plagued with less-than-optimal CPU speeds, too few available cores, and lagging computational memory capacity.
- With regard to the specifics of the annual user survey deployment, care should be taken to ensure that populations are clearly defined and differentiated so as to make the best use of data, and so that findings can be targeted to specific populations. Further, if possible, ensure that populations members are active users who have interacted with NCGAS resources within the 12-month period preceding the data collection effort.
- Regarding the data collection instrument, in general, reduce the number of questions to increase participation; in particular, reduce the number of questions that allow respondents to "select all that apply," perhaps opting to limit the number of responses to such questions to three or fewer. This would likely result in more conclusive, actionable data. Further, to the extent practical, attempt to maintain the data collection instrument as deployed for at least two subsequent data collection cycles to allow for longitudinal trend analysis.

C. Survey Methodology and Results

C.1 Collection Methodology

The NCGAS Annual User Assessment was conducted from June 3, through July 8, 2019, and was available to all current users, without regard to user status or institutional affiliation (the data collection instrument is included in Section E.1). NCGAS users were not required to participate and were able to opt out of the survey. Those who chose to participate were only eligible to do so once during the deployment period; the survey was accessed via a custom link specific to each member of the population and after completing the survey (or opting out), the survey was no longer accessible.

The total number of NCGAS users in the population was 1601, with 240 choosing to participate. Respondents were not required to answer any questions and were free leave the survey at any time or to skip questions they did not wish to answer and, therefore, the N varies from question to question. After recoding some responses—where the respondent agreed to participate but did not provide any further responses—as implicit refusals, the effective rate of response was 13.2%. In 2018, there were 1076 in the population and there were 115 responses, with a 10.6% rate of response. The increase in overall population is due to the addition of Trinity Galaxy and GenePattern users, who, in many cases ,are unaware they are using NCGAS resources; the addition of these users to the population may have affected the results reported here, but we are not able to differentiate between the user populations confidently and exclusively.

The data collection instrument was submitted under protocol #1407478943A004/exempt to the Indiana University Institutional Review Board, and was granted "exempt" status. As such, the resulting data may be published externally in reports, presentations, and other documents, so long as data are presented in aggregate form and no identifying information is divulged. While all identifying information is redacted for public consumption, please note that responses were initially associated with the respondent's email address and cannot be considered anonymous; in some cases, this may have a bearing on the responses provided and should be considered when reviewing and analyzing the data.

C.2 Results

In this survey, NCGAS users were asked to rate their level of satisfaction, ranging from "extremely dissatisfied" to "extremely satisfied," with a number of NCGAS services, and with their overall experience interacting with NCGAS. Overall, 88% of all respondents report they are either "somewhat satisfied" to "extremely satisfied" in their overall experience using NCGAS services. Applying a standard Likert scale to the responses provided, with "1" being "extremely dissatisfied" and "5" being "extremely satisfied," the mean satisfaction with all services evaluated is well above 4.0 on a 5.0 scale. The Use of the Carbonate Clusters (4.45) and Genome Analysis and Bioinformatics Consulting (4.32) are the most highly rated services. Notably, mean satisfaction with users' overall experience with NCGAS is 4.49 on a 5.0 scale. The satisfaction means and rating distributions are presented in Figure 1.

On a 1-5 scale, with "1" being "very dissatisfied" and "5" being "very satisfied", please rate your satisfaction with the following NCGAS services. If you have no basis for evaluation, please select not applicable.								
	Number of valid responses	Mean satisfaction	Very dissatisfied	Dissatisfied	Neither satisfied or dissatisfied	Satisfied	Very satisfied	Not applicable/ never used
Genome analysis/bioinformatics consulting	95	4.32	4%	6%	9%	14%	66%	117
Use of software workflow systems deployed by NCGAS	120	4.29	3%	6%	9%	25%	58%	92
Use of the Carbonate cluster	69	4.45	1%	3%	14%	12%	70%	143
Educational materials	87	4.23	2%	5%	11%	31%	51%	125
Talks and guest lectures	49	4.22	4%	4%	18%	12%	61%	163
Response time	102	4.27	4%	4%	13%	20%	60%	110
Blogposts	49	4.22	4%	2%	16%	22%	55%	163
NCGAS RNAseq pipleline	80	4.23	6%	4%	9%	24%	58%	132
Overall satisfaction with NCGAS	138	4.49	2%	3%	7%	20%	68%	74

Figure 1: Distribution and mean satisfaction with NCGAS services

Further, in addition to being the resource with which users are most satisfied, the Carbonate Cluster, along with the Galaxy at Indiana University and NCGAS Trinity Galaxy services (part of NCGAS's larger collection of resources and services), is among the most highly used resources. Figure 2 represents use of NCGAS services and resources by respondents.

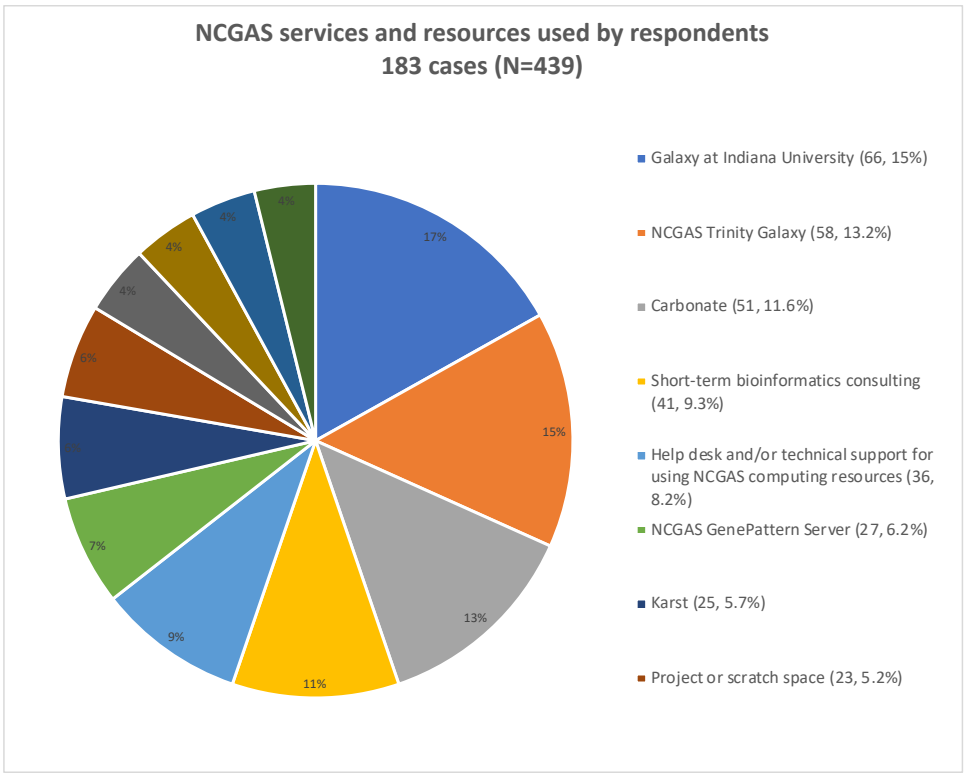

Figure 2: NCGAS resources and services used by respondents

In addition to very high levels of satisfaction, 90% of respondents report that NCGAS resources are at minimum "helpful" in carrying out their research programs. Of these, 41% rate NCGAS resources as essential, indicating their research "could not be conducted" without use of these resources. Figure 3 represents all users' responses regarding the importance of NCGAS resources and services.

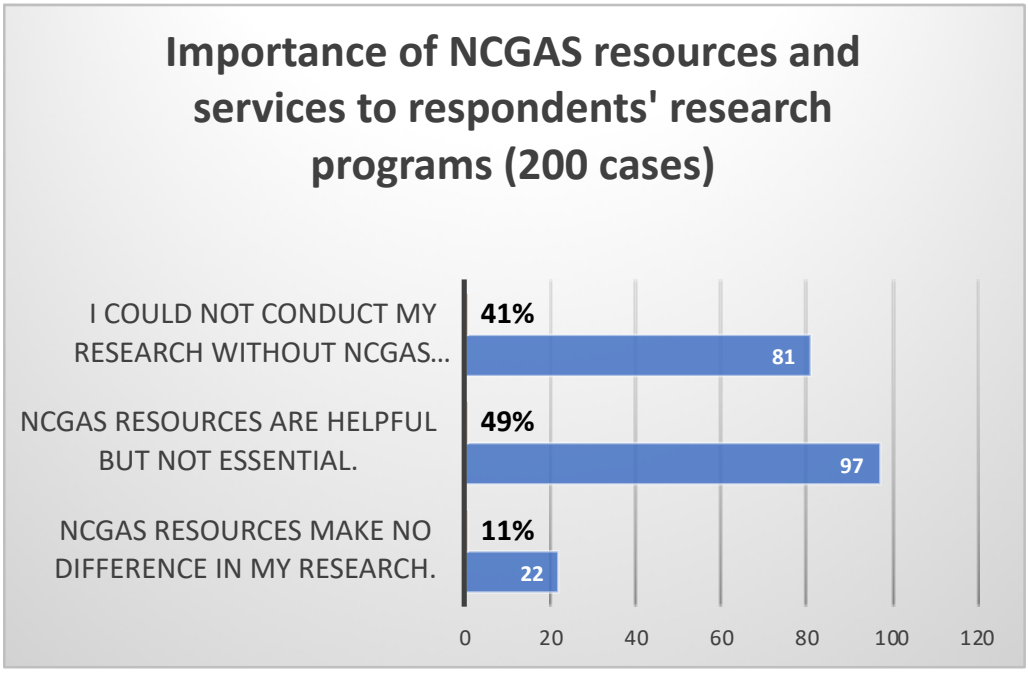

Figure 3: Importance of NCGAS resources and services

Qualitative data fully supports and underscores NCGAS users' high levels of usefulness, satisfaction, and criticality found in the quantitative data:

- I am a professor at a small college with limited resources. The resources, assistance, and collaboration at NCGAS has made it possible for me to tackle an RNAseq project that will be the ground work for my future research. The college does not have the computing hardware or bioinformaticians to support such a project. Through the use of NCGAS hardware, collaboration with the NCGAS staff, and attending an NCGAS workshop, I am confident I will be able to complete de novo assembly of transcripts. That is a great feeling. NCGAS empowers people at small colleges with the same tools that are found at large universities. Thank you so much for that!
- NCGAS resources - the computers and the help from staff - make possible all of our bioinformatics research. IU is a leader in this and our research program would be much weaker without NCGAS.
- NCGAS staff and computing resources have been vital and instrumental to the bioinformatic analyses for my projects. I have learnt basic and advanced bioinformatic skills from NCGAS staff. These skills and IU/NCGAS computing resources have made my research possible.
- The personal assistance provided by NCGAS in dealing with all of the bioinformatics tools available makes it possible for me to do my work. Left on my own I would never have attempted many of my RNA-Seq and model organism database creation projects.
- We work with non-model species requiring de novo assemblies or RNA-Seq data. NCGAS has provided infrastructure and advice. We have also been exploring genome sequencing, and regularly engage in conversations with NCGAS scientists on this topic and others like pros and cons of sequencing platforms, new technologies, new software tools.
- I simply don't have access to any other easy-to-use computational resources as powerful as those at NCGAS; my institution is too small to support our own cluster, and I don't have a budget to pay for something like AWS.
- NCGAS provides access to installed and configured software pipelines to automate complex analyses, including tools not readily accessible elsewhere.
- What I really like is [NCGAS'] computing resources. This is because I come from a country [where] high-end computing resources is very expensive. Moreover, I am originally from a marine science background. NCGAS resources (blogs, FB posts, tutorials) really did help me in increasing my understanding regarding advanced computational biology.
- NCGAS resources are intelligent, agile, and quick to respond.
- My project required nearly 10,000 hours of computational time that was provided by IU via Carbonate over a 6 month period. Without access to Carbonate, these projects would never finish using internal server resources.

All qualitative comments are included in Section E.2-4.

Respondents were also asked a series of questions related to current and future resource needs, in particular regarding the types of applications for which NCGAS might consider increasing its user support. Twenty-three percent of users report that NCGAS should increase its current levels of support for Genome assemblers and annotation, while 19% cite long-read scaffolding technologies for additional support. Figure 4 presents the full range of services for which NCGAS may consider increasing its current levels support.

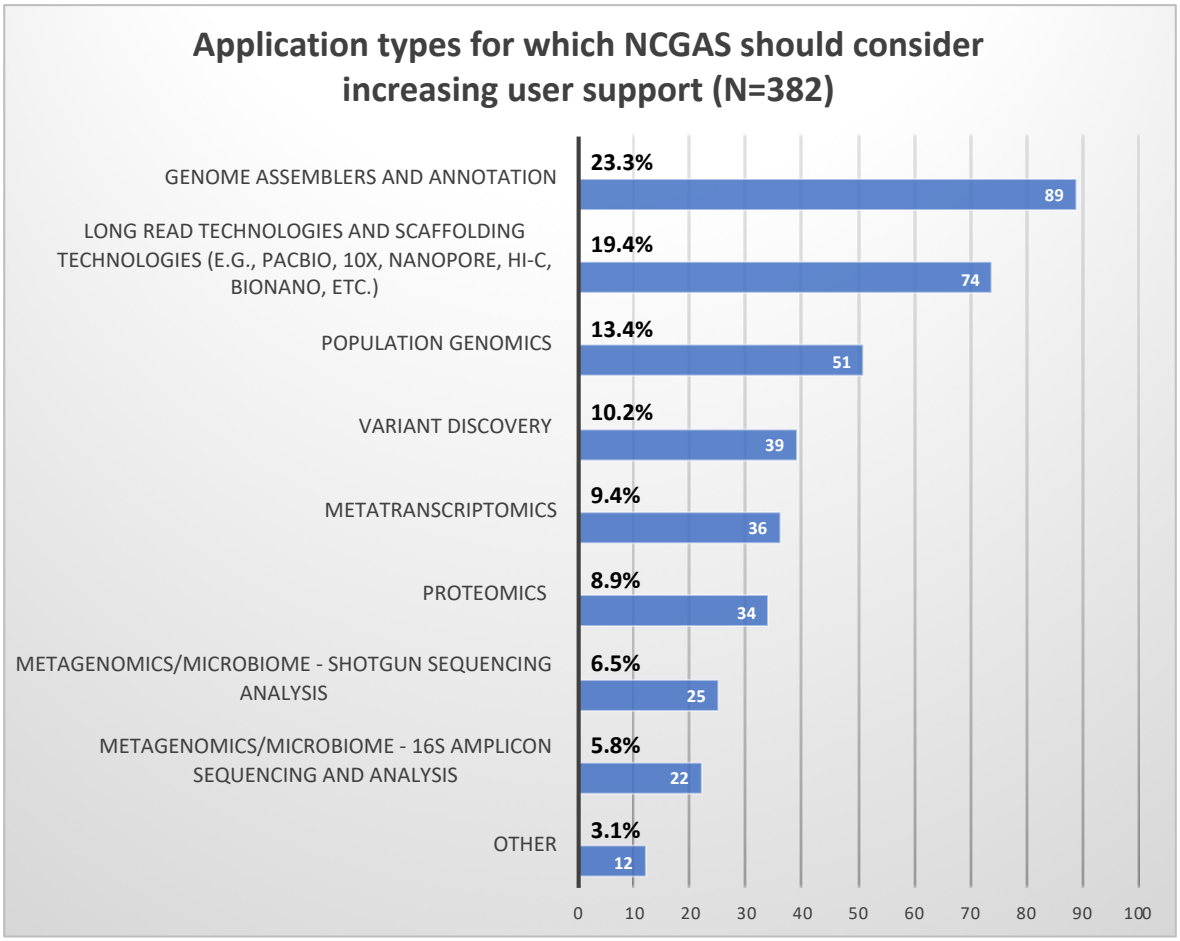

Figure 4: Application types for which NCGAS should consider increasing user support

When asked about resources NCGAS should consider adding to its service portfolio in the future, responses varied, establishing no clear consensus, but a slightly greater need was indicated for the installation of published bioinformatics applications (12.8%), installed/curated genomics applications (12.6%), and Bioinformaticians On-Call (12.6%). Figure 5 details the full range of resources users would like to see added to the NCGAS service portfolio.

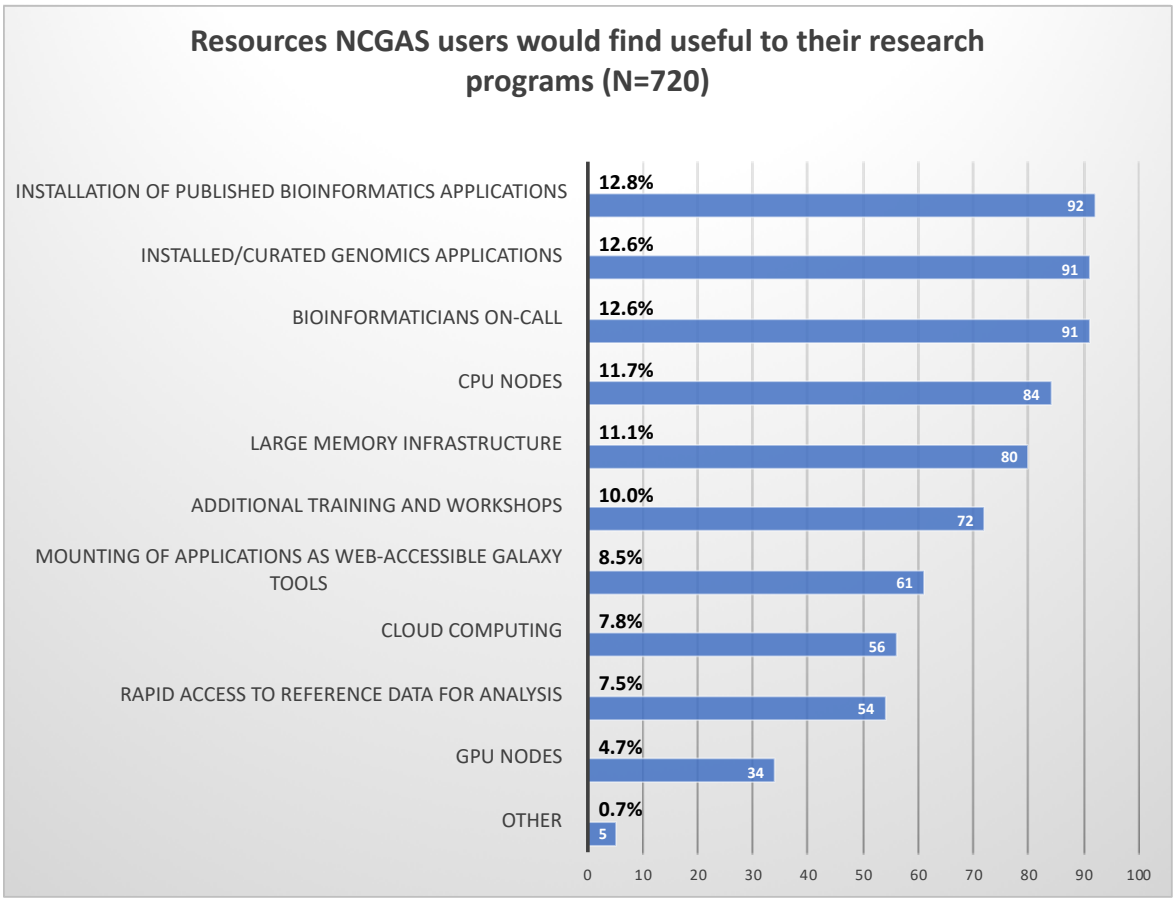

Figure 5: Resources NCGAS users would find useful to their research programs.

To provide context of users' overall research environments, users were also asked to indicate their greatest general research needs. Not surprisingly, nearly sixteen (15.9) percent of respondents cited memory during computations, followed by some 15% citing CPU speed/number of availably cores among their greatest research needs. Figure 6 details the full spectrum of NCGAS users' greatest research needs.

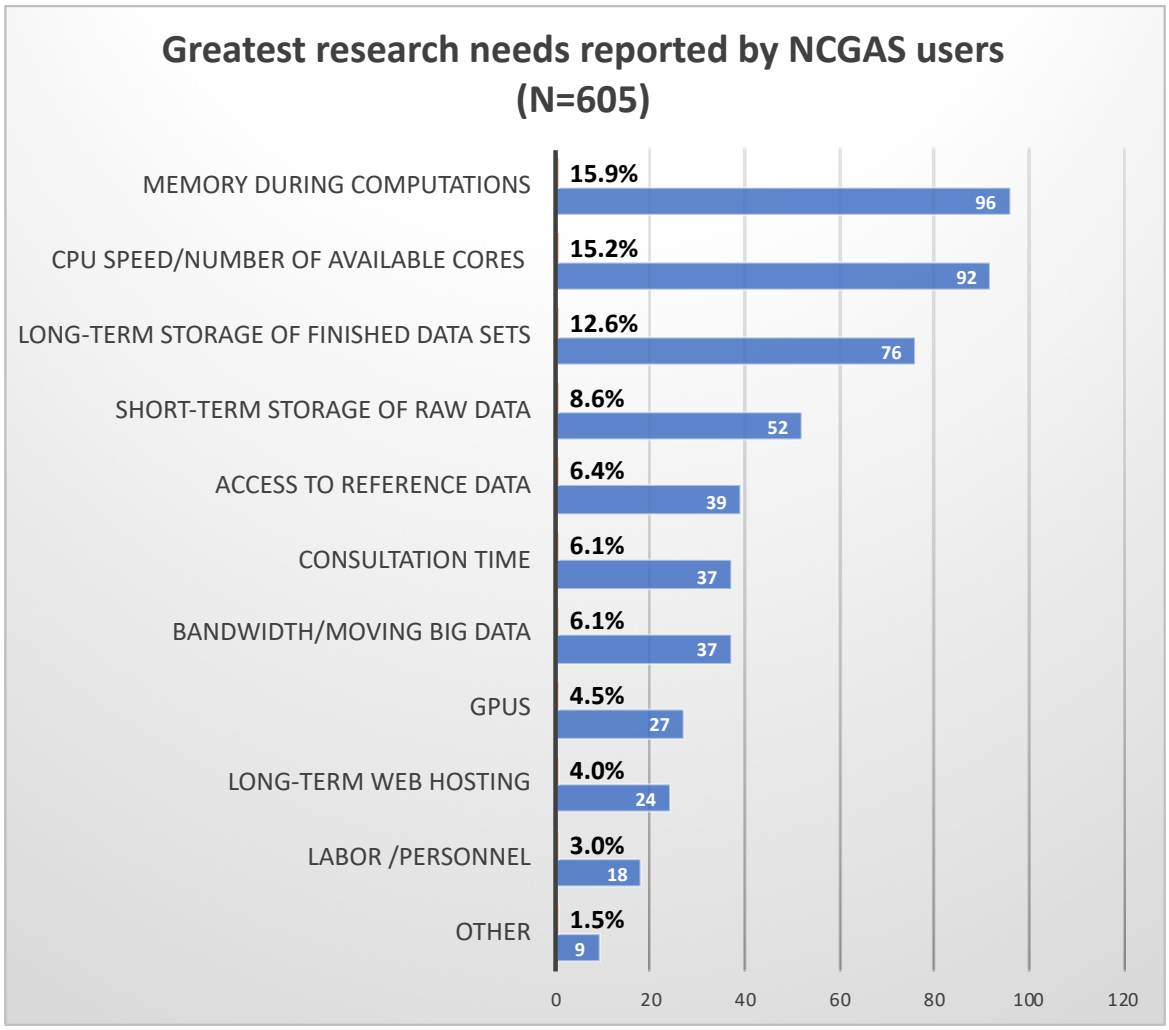

Figure 6: Greatest research needs reported by NCGAS users.

D. Discussion

When compared to previous years' survey results (https://scholarworks.iu.edu/dspace/handle/2022/21622; https://scholarworks.iu.edu/dspace/handle/2022/21614; https://scholarworks.iu.edu/dspace/handle/2022/22009), there are only small changes in results, though Trinity Galaxy users are included this year and the number of respondents has doubled.

Observations about the results include the following:

- Other surveys have established that NCGAS serves researchers in all states and Puerto Rico, including all EPSCoR states. Given the user population profile, the near-six percent of users from EPSCoR states responding to this survey aligns with expectations. The number of users we per state is well-correlated with the population of a state; given that EPSCoR states typically have lower populations when comparted to non-EPSCoR states, it is expected that NCGAS have only a few users per EPSCoR state.
- Additional outreach to EPSCoR states and Minority-Serving Institutions (MSI) is an NCGAS goal; it is, however, personnel-intensive. Workshops both at home and at meetings is adding to the center's outreach, although it does not currently have travel money for participants.

- The request for more RNAseq support is somewhat surprising, given that it is NCGAS's greatest area of focus. The center has developed a pipeline and offers a 2.5 day workshop, for de novo RNAseq assembly that may further address this need. A similar product for genome assembly is in development. NCGAS currently supports various long-read and scaffolding methodologies, and is expanding these—this is a field that is changing very quickly!
- It is gratifying for the NCGAS team to know that it has maintained a high level of user support and satisfaction. It will request supplemental funding to extend this support to more workshops and online trainings, coupled with more travel funding to disseminate this training.
- The NCGAS team represents has seen a drop in its approval—although it's still good—that can be attributed to GenePattern users. While NCGAS provides the GenePattern server, it does not offer support for users—perhaps reflected in a lower satisfaction with the service.

Recommendations from this survey relevant to next year's assessment, include the following:

- Continue following long-read and scaffolding use, given how heterogeneous the field now is.
- Continue to monitor the effect of outreach and training—who has participated, and who is aware it exists.
- NCGAS is developing training in high performance computing for biologists; a question relevant to this effort should be developed.
- NCGAS is expanding support of epigenetic methods; develop a question that will express more about what chemistries are being used would be useful.
- Assess the use of container/packages users prefer to use, and if they have a way to use them.
- The NCGAS team believes that novel user-groups are picking up epigenetic approaches to their research questions (e.g., ecology, experimental evolution, etc.); develop questions that will document this important trend.

E. Appendix

E.1 Informed Consent and Questionnaire

National Center for Genome Analysis Support User Survey

INFORMED CONSENT: You are invited to participate in a research study conducted by the principal investigator of the National Science Foundation-funded National Center for Genome Analysis Support (NCGAS) project. We ask that you read this statement and ask any questions you may have before agreeing to take part in the survey. This study is administered on behalf of the NCGAS project by the Indiana University Pervasive Technology Institute and is funded, in part, by the National Science Foundation.

STUDY PURPOSE: The purpose of the NCGAS User Survey is to assess current and future computational needs of the biological research community, as well as additional services we can provide to help your research/analysis. Survey information will be used to guide NCGAS personnel in (1) assessing and addressing current and future needs, (2) focusing outreach and training efforts, (3) making decisions related to resource provisioning, and (4) informing plans for future expansion of resources and services. Further, this study inquiries about training methods, research tools, institutional characteristics, users' funding agencies, etc., and informs research publications, grant and supplemental solicitations, agency reporting requirements, best practice theories, and conference presentations.

PROCEDURES FOR THE STUDY: If you agree to take part in the study, you will complete an online survey in which you will not be ***required*** to provide any identifying information. You will be asked if members of the research team may contact you for additional follow-up (i.e., to gain further insights about your experiences or concerns, to invite your participation in case studies, etc.) but you may decline to consent. If you consent to further contact, you will be asked to provide your name and contact information. You will also be given the opportunity to associate your comments with your contact information for the purpose of facilitating further contact. The survey will remain confidential, and survey responses will not be associated with any identifying information, unless you specifically grant permission to do so.

You will receive via email an initial letter of invitation, followed by up to three (3) reminder messages. After the initial letter of invitation, only those who have not responded will receive subsequent messages. You will have the opportunity to opt out of all future communications upon receipt of the initial letter of invitation. The survey should not take more than 10 minutes to complete, with an average time for completion in the seven- to eight-minute range.

CONFIDENTIALITY: Efforts will be made to keep any personal information that you might inadvertently disclose confidential. We cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Your identity will be held in confidence in reports in which the survey results may be published and/or databases in which results may be stored.

Organizations that may inspect and/or copy survey records for quality assurance and data analysis include groups such as the study investigator and his/her research associates, the Indiana University Institutional Review Board or its designees, the study sponsor, the National Science Foundation, and (as allowed by law) state or federal agencies, specifically the Office for Human Research Protections (OHRP).

CONTACTS FOR QUESTIONS OR PROBLEMS: For questions about the study, contact Indiana University Information Manager Julie Wernert at (812) 856-5517 or jwernert@iu.edu. For questions about your rights as a participant or to discuss problems, complaints, or concerns about a research study, to obtain information, or to offer input, please contact the IU Human Subjects Office at (812) 856-4242 or by email at irb@iu.edu.

VOLUNTARY NATURE OF STUDY: Taking part in this study is voluntary. You may choose not to take part or may leave the survey at any time. Leaving the survey will not result in any penalty. Your decision whether or not to participate in this survey will not affect your current or future relations with the Indiana University Pervasive Technology Institute, the NCGAS project, or the National Science Foundation.

This study was approved by the Indiana University Institutional Review Board on May 21, 2019. Please reference study #1407478943A004/exempt when inquiring.

Do you agree to participate?

I agree (1)
I disagree (2)

Skip To: If National Center for Genome Analysis Support User Survey INFORMED CONSENT: You are invited to par... = I agree
Skip To: End of Block If National Center for Genome Analysis Support User Survey INFORMED CONSENT: You are invited to par... = I disagree

Please tell us which of the following your research involves. *Select all that apply*.

Genome assembly
RNA-Seq approaches
Microbiome and metagenomics assembly and analysis
Whole genome structural analysis, i.e., nucleosome mapping, high-resolution DNase sensitivity, transposon insertion sequencing, etc.
Protein-DNA interactions with ChIP-seq
Methylation studies
Proteomics and mass spectrometry
Genome annotation and ortholog discovery
SNPs and variant analysis
Other: _____

Which NCGAS services and/or resources have you (or a member of your group) used? *Select all that apply*.

- Short-term bioinformatics consulting
- Long-term bioinformatics consulting
- Help desk and/or technical support for using NCGAS computing resources
- Assistance with education and/or training
- Carbonate
- Bridges
- Galaxy at Indiana University
- Project or scratch space
- Scholarly Data Archive
- NCGAS Trinity Galaxy
- NCGAS GenePattern Server
- NCGAS genome browsers
- Jetstream (IU/TACC)
- Big Red II
- Wrangler
- Karst
- Not sure
- None of the above/not applicable

Display This Question:
If Which NCGAS services and/or resources have you (or a member of your group) used? Select all that... = Jetstream (IU/TACC)

Do you use a Jetstream allocation? *Select all that apply.*

- I use my own
- I use NCGAS's
- Unsure
- No

End of Block: Default Question Block

Start of Block: Block 1

How did you learn about NCGAS? *Select all that apply.*

- Conference or workshop
- Academic journal, paper or website
- Course lecture or class presentation
- Press coverage
- Newsletter, email solicitation, or survey
- From a colleague (word of mouth)
- Social media
- NCGAS website and blog
- Other: _____

What percentage of your bioinformatics computation is done using NCGAS resources?

	0 10 20 30 40 50 60 70 80 90 100
Percentage of bioinformatics computation using NCGAS resources ()	

How essential are NCGAS resources to the execution of your research project(s)?

- I could not conduct my research without NCGAS resources.
- NCGAS resources are helpful but not essential.
- NCGAS resources make no difference in my research.

Skip To: If How essential are NCGAS resources to the execution of your research project(s)? = NCGAS resources make no difference in my research.
Skip To: If How essential are NCGAS resources to the execution of your research project(s)? = NCGAS resources are helpful but not essential.
Skip To: If How essential are NCGAS resources to the execution of your research project(s)? = I could not conduct my research without NCGAS resources.

Please elaborate on how NCGAS resources are useful to your work? *Please do not include any identifying information in your comments.*

_____
_____

Please indicate other national resource providers you use in your work. *Select all that apply.*

- CyVerse
- XSEDE resources
- NCBI
- Galaxy Main
- The BROAD Institute
- Center for Computational Biology (CCB)
- DIAG Computing
- Open Science Grid
- CIPRES
- Your local resource
- Amazon Web Services (AWS)
- Google Cloud
- Other: _____

For which additional application types should NCGAS consider providing greater levels of support? *Select all that apply.*

- Genome assemblers and annotation
- Long read technologies and scaffolding technologies (e.g., PacBIO, 10x, Nanopore, Hi-C, Bionano, etc.)
- Metagenomics/Microbiome - 16s amplicon sequencing and analysis
- Metagenomics/Microbiome - Shotgun sequencing analysis
- Metatranscriptomics
- Proteomics
- Variant discovery
- Population genomics
- Other: _____

Which of the following needs/resources would be of the **greatest** benefit to your research? *Select all that apply.*

- Short-term storage of raw data
- Long-term storage of finished data sets
- Memory during computations
- CPU speed/number of available cores
- GPU
- Labor/personnel
- Bandwidth/moving big data
- Long-term web hosting
- Access to reference data
- Assistance with education and/or training
- Assistance with education and/or training
- Consultation time
- Other: _____

Which genome sequencing platforms do you work with now, or will you work within the next year? *Select all that apply.*

- Sanger
- 454
- Illumina
- IonTorrent
- Oxford Nanopore
- PacBio SMRT sequencing
- 10X Chromium
- Bionano Saphyr
- Hi-C and related
- Other: _____

Which of the following resources would be helpful to your research program? *Select all that apply.*

- Bioinformaticians on-call
- Large memory infrastructure
- CPU nodes
- GPU nodes
- Installed/curated genomics applications
- Installation of published bioinformatic applications
- Mounting of applications as web-accessible Galaxy tools
- Rapid access to reference data for analysis
- Additional training and workshops
- Cloud Computing
- Other: _____

Do you have access to a genome browser for your research/organism?

- Yes
- No

Would support for genome browsers be helpful to your research program and/or community?

- Yes
- No

If you have a preferred genome browser application (e.g., GBrowse, JBrowse, etc.), please list it below:

_____

On a 1-5 scale, with "1" being "very dissatisfied" and "5" being "very satisfied", please rate your satisfaction with the following NCGAS services. *If you have no basis for evaluation, please select not applicable.*

	Very dissatisfied	Dissatisfied	Neither satisfied or dissatisfied	Satisfied	Very satisfied	Not applicable/never used
Genome analysis/bioinformatics consulting	○	○	○	○	○	○
Use of software workflow systems deployed by NCGAS (e.g., Galaxy, GenePattern, JBrowse, etc.)	○	○	○	○	○	○
Use of the Carbonate clusters	○	○	○	○	○	○
Educational materials	○	○	○	○	○	○
Talks and guest lectures	○	○	○	○	○	○
Response time	○	○	○	○	○	○
Overall satisfaction with NCGAS	○	○	○	○	○	○
Blogposts	○	○	○	○	○	○
NCGAS RNAseq pipeline	○	○	○	○	○	○

Which of our workshops/trainings have you attended? *Select all that apply.*

- R for biologists
- De novo transcriptome assembly
- MDIBL Environmental genomics
- PAG workshops
- Galaxy workshops
- Bioinformatics2go
- Galaxy tutorials

We are developing the following workshops; which would you be interested in attending? *Select all that apply.*

- HPC for Biologists
- Microbiome/Metagenomic analysis
- Genome assembly and annotation
- Python for Biologists
- Data management and collaboration in HPC
- Epigenetics (methylomics and other methods)

What other workshops would you be interested in NCGAS offering?

_____

Which of the following NCGAS social media outlets do you follow? *Select all that apply.*

- Twitter (@ncgas)
- Facebook (@ncgasiu)
- Youtube (IU PTI)
- NCGAS (ncgas.org/NCGAS_Blog.php)
- None of the above

What do you like about NCGAS resources? *Please do not include any identifying information in your comments.*

_____
_____

How can we improve our support of your research needs? *Please do not include any identifying information in your comments. You are always free to contact us at help@ncgas.org for project specific comments and requests.*

_____
_____

End of Block: Block 1

Start of Block: Block 2

To help us better understand your responses please answer the following demographic questions. To ensure your survey responses remains confidential, please do not include any identifying information in your comments. If you have questions or concerns and would like to be contacted by NCGAS staff, please contact us at **help@ncgas.org**.

Have you used NCGAS services as an undergraduate and/or a two-year student?

- Yes
- No

Have you used NCGAS services as a graduate and/or PhD student?

- Yes
- No

Are you a member of the Indiana University community?

- Yes
- No

What is your *primary* research field or field of study?

▼ Arts and Humanities (1) ... Not applicable (22)

From which of the following agencies do you have funding? *Select all that apply.*

- National Science Foundation
- National Institutes of Health
- Department of Energy (3)
- US Department of Agriculture
- Department of Defense
- Environmental Protection Agency
- Other: _____

Have you been funded as a principal investigator or co-principal investigator on an NSF grant award?

- Yes
- No

Are you funded by the NSF for projects that use NCGAS resources?

- Yes
- No

Are the activities for which you have used NCGAS resources in an area of research funded by the National Science Foundation, even if you're not funded by it currently?

- Yes
- No

Please best describe your status as a user of NCGAS services and resources.

- I am the owner of an NCGAS allocation
- I am on someone else's allocation
- I am not associated with an NCGAS allocation

Does your institution offer workshops/courses on using high performance clusters (like Carbonate) or other computer science courses targeting to biologists?

- Yes
- No
- Not sure

Please describe the size of your academic institution:

- Small (fewer than 3,000 degree-seeking students)
- Medium (3,000 - 10,000 degree-seeking students)
- Large (more than 10,000 degree-seeking students)
- Not applicable

Please describe your institution/organization. *Select all that apply*.

- Institution located in an EPSCoR state
- Minority-serving Institution (MSI)
- Associate's College (all degrees are at the Associate's level)
- Baccalaureate College/University
- Master's College/University
- Doctorate-granting University
- Teaching-focused Institution
- Research-focused Institution
- Government Lab or Center
- High performance computing resource provider (e.g., NCSA, TACC, etc.)
- Non-profit Organization (non-academic)
- Corporate/Industrial Organization

Considering your primary institutional affiliation, please indicate the country in which your institution is located.

▼ United States (1) ... Other (253)

Please select your primary role.

- Executive leadership (e.g., director, CIO, etc.)
- University faculty or equivalent
- University/Center research staff or equivalent (non-postdoctoral)
- University/Center non-research support staff (or equivalent)
- Postdoctoral fellow
- Graduate student
- Undergraduate student
- Other (please specify):_____

What is your race? *Select all that apply.*

- Alaskan Native
- American Indian
- Asian
- Black or African-American
- Native Hawaiian or Pacific Islander
- White
- Other: _____
- Prefer not to disclose

What is your gender?

- Female
- Male
- Non-cisgender
- Other: _____
- Prefer not to disclose

End of Block: Block 2

Start of Block: Block 3

If you have published any work that used or acknowledges NCGAS services or resources, please send us a separate email to **help@ncgas.org** with information about your publication. This is the best way to document our contributions, and is extremely important to our continued funding.

May we contact you regarding your responses?

- Yes
- No

Display This Question:
If May we contact you regarding your responses? = Yes You have granted consent for additional contact. Please provide your name and preferred contact information.

- Name: _____
- Contact information: _____

E.2 Open Text Responses: Please elaborate on how NCGAS resources are useful to your work.

- It makes bioinformatics analysis easier for those who have less experience or programming knowledge.
- it is very useful since some resources do not need deep knowledge in bioinformatics so it is useful for me as agronomist.
- Galaxy has provided computing capacity and easy to use programs.
- It is the only chance I have to run a transcriptomic assembly.
- The Galaxy implementation has sufficient storage capacity for my work. Other public implementations do not.
- I used Trinity de novo assembly in the beginning at Galaxy. After a few weeks, I decided to run it locally.
- Convenient, fast, and useful pipelines.
- The Trinity instance is useful because that amount of computing would be prohibitively expensive to us (a smaller genomics lab at a small public university).
- We just use ESPPredictor although it appears to be currently unavailable.
- Computational resources at my university are expensive and wait times are long.
- I use NCGAS to process all my data and ran jobs on Carbonate.
- The transcriptome assembly pipeline has been phenomenal in helping me progress with my own transcriptome assembly and differential gene expression quests. Also, before my institution had a decent cluster set up that could handle the computations I needed to do, I was doing all of my work on Carbonate.
- High computation resources are necessary in my research; awesome tech support of NCGAS.
- To carry out my bioinformatics work and problems were solved by Bhavya.
- I am non-bioinformatician student as compared to other tools that are very tricky its very useful and easy to handle. I am able to find out fusion genes in my samples easily and it takes very less time also.
- 40vCPU 400Gb and 5T
- Workflow already done and easy to follow
- I simply don't have access to any other easy-to-use computational resources as powerful as those at NCGAS; my institution is too small to support our own cluster, and I don't have a budget to pay for something like AWS.
- I am using this service for assembly of seq data
- High capacity computing with easy and functional access
- If the platform or server for TRINITY is not available, NCGAS is very necessary because TRINITY requires a lot of RAMs and CPU cores which I cannot execute TRINITY from my laptop and the output from TRINITY is very important for RNA-seq analysis.
- Already installed software with complex installation is great! And servers.
- NCGAS provides access to installed and configured software pipelines to automate complex analyses, including tools not readily accessible elsewhere
- The other computing resources I have access to often have long wait times due to heavy usage, and have lower computing power. NCGAS resources help me to increase my productivity by reducing wait time for jobs.

- assembly and downstream gene expression analyses done using the NCGAS servers frees up a lot of the computational load in our institute and allows us to do other tasks.
- Has state of the art software that we can't run locally.
- I have fairly large RNA-seq files which could not be analyzed at my home institution, or at least they do not have the pipeline set up with all the software.
- Consulting, help with code and software.
- Provides a reliable platform for comparison
- Assembly pipeline tools are really helpful in my transcriptome assembly and analysis as a beginner.
- Provide the access to Carbonate and all the relevant bioinformatics support from experienced people.
- Carbonate/Bridges are much more accessible to our research than our university's current setup. It is easy to access, install, and manage modules. NCGAS and Carbonate reduce our work time by almost 50%
- Lots of help for people who aren't familiar with bioinformatic processes.
- Saving time in script errors.
- I can get a hint of my genes of interest in big datasets.
- I have used NCGAS resources primarily for transcriptome assembly, but I also used them to create a novel core ortholog set for HaMStR, with much help from NCGAS staff.
- At present I have comparable hardware at my own institution and have not needed additional bioinformatics support. That said, I used NCGAS previously before I had sufficient RAM for eukaryotic genome assemblies.
- I found the gene sequence for my protein of interest in a transcriptome assembled here.
- The initial consultation was critical for adopting new approaches.
- The specific tools implemented in NCGAS fully cover a main stage of my research, but other tools from other sources are helpful in other analyses.
- Carbonate allows for manipulation of large data files.
- It's the best alternative for people who do not have access to a personal server.
- NCGAS allows analysis with high precision.
- Was useful to use another Galaxy Server.
- My project required nearly 10,000 hours of computational time that was provided by IU via Carbonate over a 6 month period. Without access to Carbonate, these projects would never finish using internal server resources.
- Analysis
- We work with non-model species requiring de novo assemblies or RNA-Seq data. NCGAS has provided infrastructure and advice. We have also been exploring genome sequencing, and regularly engage in conversations with NCGAS scientists on this topic and others like pros and cons of sequencing platforms, new technologies, new software tools.
- Used Jetstream resources for teaching a workshop.
- For my PhD, the NCGAS resources played a significant role. Besides the computational power, the space was another significant feature. The scratch directory allowed me to process the large scale genomic data sets necessary for my dissertation.
- I used NCGAS resource for de novo based analysis
- Blogs, guidance, resource access

- I use the galaxy for its specialization on RNAseq
- NCGAS resources have enabled me to quickly run analyses, for which I do not have any command line scripts already written. That is it is a great way to get a quick look at the data to better guide downstream analyses.
- I did not use the NCGAS resources today, but I could in the future, usually I do my own workflows.
- Currently we have not used this resource however are interested in the services.
- I use bioinformatics in my research and teaching, so my students and I utilize a lot of different resources. This allows me to help students with their own projects and help them learn to be flexible and innovative in their approaches.
- I am a professor at a small college with limited resources. The resources, assistance and collaboration at NCGAS has made it possible for me to tackle an RNAseq project that will be the ground work for my future research. The college does not have the computing hardware or bioinformaticians to support such a project. Through the use of NCGAS hardware, collaboration with the NCGAS staff, and attending an NCGAS workshop I am confident I will be able to complete de novo assembly of transcripts. That is a great feeling. NCGAS empowers people at small colleges with the same tools that are found at large universities. Thank you so much for that!
- Without NCGAS resources, I don't have access to enough computing resources for de novo transcriptome assembly.
- RNAseq data analysis, fusion gene finding
- We would not be able to do RNAseq without NCGAS
- NCGAS resources make it possible for me to do my own computational analysis, rather than relying on collaborators (which I also do).
- Our local HPC resources are limited and sometimes we experience some down time.
- Mainly it helped me to reconstruct non-model organisms' transcriptomes
- I don't need bigger computer
- Supervised and unsupervised clustering analysis
- We are using the SDA for archiving purposes
- The ability to free myself from data storage, software upkeep and large horsepower concerns
- I am running my allocations on several HPCs to capacity and I have my own server, but I really need the allocation through NCGAS for RNA-Seq analysis.
- It provided the training in bioinformatics that has been crucial to my research
- NCGAS resources - the computers and the help from staff - make possible all of our bioinformatics research. IU is a leader in this and our research program would be much weaker without NCGAS.
- Assembly of very large numbers of sequencing reads.
- The personal assistance provided by NCGAS in dealing with all of the bioinformatics tools available makes it possible for me to do my work. Left on my own I would never have attempted many of my RNA-Seq and model organism database creation projects.
- I have used for assembly for RNAseq data.
- NCGAS has the most extensive tool sets, which enables high efficiency in data analysis.
- The Trinity instance on the Galaxy server allowed me to perform de novo assembly of a transcriptome when I did not have an otherwise readily available solution.
- I rely on NCGAS for bioinformatics.

- Support on complex bioinformatics analysis.
- Wide range of installed bioinformatics software, can work with a lot of data at a time, jobs are completed quickly.
- I will be using some NCGAS programs in the future that are essential; I just have not used them yet.
- In training for genomic data analysis, I need to show users where to find the best training and computational resources for labs who can't access the heavy compute architecture for larger-scale genomic data analysis.
- HPC and long-term data storage.
- Quick running of programs.
- Submitting large jobs that don't require most/all of my local computing power.
- The availability of bioinformatics program packages that difficult to install locally and are mostly used for nonmodel organism research like genome annotators is really helpful.
- Galaxy.
- NCGAS staff and computing resources have been vital and instrumental to the bioinformatic analyses for my projects. I have learnt basic and advanced bioinformatic skills from NCGAS staff. These skills and IU/NCGAS computing resources have made my research possible.
- All of my genome analysis is aided by NCGAS resources.
- Carbonate and Karst are invaluable for running computationally intensive genomics workflows.
- I have used NCGAS resources like Carbonate, participated in RNA-seq workshop and performed consultations with NCGAS personnel regarding project planning and troubleshooting problems that I encountered in my transcriptome analysis.
- It helps with clustering analysis.
- It helped me to start with some complicated analyses quickly. Later I learned different approaches.
- It's the only way we could assemble so many transcriptomes.
- I don't have the resources to perform bioinformatic analyses without the infrastructure and facilitation that NCGAS resources offer to researchers.
- We have not yet engaged in using NCGAS, though we anticipate doing so in the near future.
- NCGAS offers some software that I can't easily install on local resources.
- We use Carbonate, Karst, and Big Red systems frequently in our research projects.
- NCGAS was essential, but now we have a cluster on campus.
- It was extremely helpful when we did not have our own bioinformatics set up. It is a great resource for labs that don't have expensive computational infrastructure. I highly recommend the continuation of support to this excellent bioinformatics resource.
- Bioinformatics.

E.3 Open Text Responses: What do you like about NCGAS resources?

- Provide online service which requires lot of CPU and memory spaces
- The personnel are outstanding.
- They are easy to use
- Galaxy provides opportunities for non-computational savvy biologists to analyze sequence data.
- Availability, quality. Big computational resources, Gratuity
- The web interface is well organized and tools are readily accessible
- Very resourceful. Great space for working, many applications installed already. And technical support is always helpful.
- Taking very less time and easy to handle
- Easily accessible
- The ability to use Carbonate without worrying about allocation limits is extremely helpful. It allows me to text out different types of analyses without the risk of using up my allocation.
- Web-based interface and RNAseq assembly.
- NCGAS is very friendly for the beginner users and easy to execute the software without any knowledge on Linux, UNIX, or R. It saves time when [we] use NCGAS because the users don't need to install their software by themselves.
- It is available to the entire US research community
- Ease of use and well-adapted software accessible within the servers.
- They are free and open source!
- The RNA-seq pipeline is relatively easy and well-explained (but could use some updating)
- Personalized to my project.
- Free of charge support.
- Curated course material and pipeline.
- Best consultation for beginners.
- Everything.
- They allow me to do assemblies quickly, in parallel. Thus far, I have only done transcriptome-based phylogenomic work, which requires me to do de novo Trinity assemblies of SRA data for many taxa. Each one is usually fairly quick, but I may have to do dozens. NCGAS makes that possible over days or weeks, rather than the months it might take if I only used my local resources.
- Usability.
- Easy interface.
- Large memory multiple cpu
- Simplicity, valuable infrastructure
- Mainly, more FREE space to upload and analyze NGS files, RNAseq, and WES.
- Focus on genomics.
- Everyone is extremely helpful and courteous. I was extremely impressed with response times and the ease of access as well as the solutions provided.
- Easy access to computing power and software tools. WE have had to run certain analyses multiple times for a variety of reasons – and the ability to go back and try different settings has been extremely helpful.
- Availability to users.

- The amount of space (i.e., scratch) and helpful staff.
- Good resource.
- Very informative, quick response times.
- Accessibility, policy.
- Rapid response times, ease of use.
- Ease to use lots of applications.
- I do not use NCGAS resources right now, maybe I will check again and start to use it, I have to say that I almost forgot about NCGAS resources, I sued Galaxy one time during my PhD, then another time last year with galaxy Trinity, in order to see if it would give me a better transcriptome assembly, but it gave me a very similar result that I got in the local server that I work, so I left it and kept working locally in our server.
- Possibility to run some analyses and test some new pipelines.
- Availability to anyone.
- It's easy to handle and required less time to run a file.
- The people employed by NCGAS do their best to make computational resources accessible, user-friendly, and functional for my research.
- Their availability for free.
- What I really like is its computing resources. This is because I come from a country that high-end computing resources is very expensive. Moreover, I am originally from a marine science background. NCGAS resources (blogs, FB posts, tutorials) really did help me in increasing my understanding regarding advanced computational biology.
- That is a community effort to fix bottlenecks in big data biology.
- That I have the ability to run their RNAseq pipeline. That they are available for consultation if I need it.
- NCGAS resources are intelligent, agile, and quick to respond.
- Excellent.
- Personal service, availability of hosting and work space, tutorials.
- It is a very thoughtful design and collection of tools every biologist or computational biologist needs.
- Uniquely available Trinity use on Galaxy.
- Yes.
- Ease of access. Technical support. Training available. Very friendly staff.
- Easy to access.
- I cannot help because I ended up not using the service, sorry.
- Tom Doak is always responsive and thorough in solving problems; data transfer is easy.
- Easy to run jobs and good organization of loadable modules.
- Freely available and are remote cluster-based.
- Accessibility, quick response times and comprehensive knowledge of NCGAS staff.
- I haven't used them but the GUIs seem well-designed.
- Knowledgeable staff when I have questions about Carbonate and Karst. Also, the fact that Carbonate and Karst are free for use.
- Availability is great.
- I've been meaning to check out your resources but haven't had the time.

- Access through my NSF grant.
- Access to Carbonate, Karst, and Big Red like servers with quick access to frequently used bioinformatics software.

E.4 Open Text Responses: How can NCGAS improve its support of your research needs?

- Better explanations of the format and parameters to enter to the analysis. Not referring to the original work to find what I need to enter my data.
- Online training would be useful.
- Galaxy often has problems that are not explained well or the system is down for extended times.
- Don't know.
- Add newer tools for analyzing noncoding RNAs (lnc and circular RNAs)
- Better response time to errors. It's been months for one error now and it still isn't resolved or responded to.
- It would be great to see ESPPredictor back online.
- Sometimes we faced problem during running a file by using different tools, because of lack of knowledge like which type of file it accepts, so if we get the whole knowledge about each and every tool so it will be beneficial for us.
- Better diagnostic of problems during runs.
- Please fix your Trinity de novo assembler
- Increasing the CPU cores and memory
- Additional training offerings, and more Galaxy modules.
- Greater bandwidth for big datasets.
- Cannot think of much other than updating some instructions and adding a few more programs for transcriptome comparison and consolidation.
- Larger computation volumes for big data sets in RNAseq experiments.
- You are doing a great job already.
- Aside from the usual (I wish I didn't have to wait in queue as long), nothing.
- More and faster CPUs and more memory (though I rarely need more memory)! But isn't that always the answer? MORE MORE MORE!
- I'm OK so far.
- More stable instances on Jetstream for temporary accounts.
- Easier access to clusters.
- I really wasn't aware of all that is offered – perhaps more dissemination?
- Access to additional tools, and tutorials on their use. We ran into trouble on analyzing/reducing PacBio reads.
- Updating of datasets; lesser time for queue of job order; posting of more bioinformatics analyses/software that can be used.
- Idk.
- Again, I have to say that I was surprised to be called to this assay, because I am not a real user of NCGAS resources, but now I am going to see if you can help me with my transcriptome issue. I need help with enrichment analysis, most specifically the visualization of enrichment analysis, generating plots. I did the complete workflow of Trinity - Trinotate pipelines, but I am having difficulty to show the results visually. This is something that I would really appreciate. I suggest you to include metabarcoding in your set of workflows, since this approach is rising every day, lots of teams working in several different habitats, huge amount of biodiversity information and a lot of different pipelines, like OBItools, VSEARCH and even Blast-MEGAN more traditional workflow.
- Continue!

- Put new tools on Galaxy as quickly as possible, and explain what is replacing what and why, because applications change so fast.
- Nothing to add.
- I think everything is great for now. Thank you so so much for all your hard work in providing such valuable resources.
- I have not used NCGAS resources for at least 5 years, so I clicked through the survey so that I'd stop getting the 'please take the survey' emails.
- Short term support (accommodation) to collaborate/attend courses. Some of the elements of the now defunct NESCENT model.
- Just keep doing what you're doing.
- I haven't run into issues that need more support than I've received.
- Make the workshop or talks accessible online.
- Availability of latest versions of software, e.g. Trinity, on the web.
- A little more than 60 days on the data capacitor – can be hard to complete an analysis if one does not have focused time every day.
- Larger offer of online bioinformatic tools.
- Support for request of specific programs.
- It would be nice to decrease the response time for bug reports, but I can live with the current response time.
- Increase staffing at NCGAS and add genomic analysis capabilities and workshops.
- n/a
- More information and resources for scRNA-seq.
- By not having to renew annually and suddenly cutting people off.
- Providing students from non-IUB campuses equal opportunity to attend the workshops.