

ABI Development: National Center for Genome Analysis Support

*Craig A. Stewart
William K. Barnett
Matthew W. Hahn
Michael R. Lynch*

Indiana University

PTI Technical Report PTI-TR15-009

December 15, 2015

Citation:

Stewart, S.A., Barnett, W.K., Hahn, M.W., and Lynch, M.R. "ABI development: National center for genome analysis support", Indiana University, Bloomington, IN. PTI Technical Report PTI-TR15-009, Dec 2015.



**PERVASIVE TECHNOLOGY
INSTITUTE**

INDIANA UNIVERSITY



**RESEARCH
TECHNOLOGIES**

INDIANA UNIVERSITY

University Information Technology Services
Pervasive Technology Institute

Table of Contents

ABI Development: National Center for Genome Analysis Support	i
1. Accomplishments	3
1.1. What are the major goals of the project?	3
1.2. What was accomplished under these goals?	3
1.2.1. Major Activities	3
1.2.2. Specific Objectives	5
1.2.3. Significant results	5
1.2.4. Key outcomes or other achievements	6
1.3. What opportunities for training and professional development has the project provided?	6
1.4. How have the results been disseminated to communities of interest?	7
1.5. What do you plan to do during the next reporting period to accomplish the goals?	7
2. Products	7
2.1. Products resulting from this project during the specified reporting period	8
2.1.1. (Peer-reviewed) Journal Articles	8
2.1.2. Conference Papers and Presentations	8
2.1.3. Other Publications	9
3. Participants	9
3.1. Individuals	9
3.1.1. Full details of individuals who have worked on the project	10
3.2. Partner organizations	11
3.2.1. Full details of partner organizations	12
3.3. Have other collaborators or contacts been involved?	14
4. Impact	14
4.1. What is the impact on the development of the principal discipline(s) of the project?	14
4.2. What is the impact on other disciplines?	14
4.3. What is the impact on the development of human resources?	15
4.4. What is the impact on physical resources that form infrastructure?	15
4.5. What is the impact on institutional resources that form infrastructure?	15
4.6. What is the impact on information resources that form infrastructure?	15
4.7. What is the impact on technology transfer?	15

4.8.	<i>What is the impact on society beyond science and technology?</i>	15
5.	<i>Changes/ Problems</i>	16
5.1.	<i>Changes in approach and reasons for change</i>	16
5.2.	<i>Actual or Anticipated problems or delays and actions or plans to resolve them</i>	16
5.3.	<i>Changes that have significant impact on expenditures</i>	16
5.4.	<i>Significant changes in use or care of human subjects</i>	16
5.5.	<i>Significant changes in the use or care of vertebrate animals</i>	16
5.6.	<i>Significant changes in the use or care of biohazards</i>	16

1. Accomplishments

1.1. What are the major goals of the project?

The major goals of the NSF ABI Award to support the National Center for Genome Analysis during the 48-month funding period of this award were as follows:

- Support biological and bioinformatics research related to genome analysis by the US open research community, focused on researchers doing biological research of relevance to the NSF mission
- Support software and computational resources relevant to genome analysis
- Maintain, harden, tune, enhance, and distribute software relevant to genome analysis
- Provide online help, consulting, and tutorials related to genome analysis

In all of these areas the focus of NCGAS' activities was on "whole-genome" and metagenomics research. Emphasis was placed on genome and transcriptome assembly at the technically challenging end of the spectrum of current bioinformatics activities - for example de novo genome assembly and transcriptome assembly.

1.2. What was accomplished under these goals?

1.2.1. Major Activities

NCGAS is one of several cyberinfrastructure and service providers that are part of the Indiana University Pervasive Technology Institute and as such has significant facilities, human resources, and administrative support from IUPTI and IU generally. NCGAS has had significant accomplishments using NSF funding, along with funding and facilities from IU, to aid discovery and innovation in the biological sciences in the US. NCGAS has aided in discoveries that range from better understanding of basic biological processes to discoveries that will aid the management of economically important animals and plants.

1.2.1.1. Support biological and bioinformatics research related to genome analysis by the US open research community, focused on researchers doing biological research of relevance to the NSF mission.

NCGAS' most significant accomplishments related to support for biological and bioinformatics research were in aiding the completion of the following RNA and DNA transcriptome or genome assemblies:

RNA transcriptome assemblies

- NCGAS provided consulting and computational resources that aided in the complete sequencing of the RNA transcriptome of the fruit fly, *Drosophila melanogaster*. The fruit fly is the largest and most complex organism for which a complete RNA transcriptome has been assembled. This is utterly critical in understanding the biological functioning of the genetic information of an organism since it is the DNA that is transcribed and expressed in the cell that determines and regulates these processes.

DNA genome or RNAseq assemblies. NCGAS aided researchers in completing de novo assemblies of the following organisms:

- Cacao plant
- Coffee

- Loblolly Pine
- Mango
- Peanut
- Litchi
- Daphnia
- North Atlantic copepod
- Pacific Neocalanus copepod
- Filarial Nematode and Wolbachia endosymbiont
- Paramecium species
- Rotifer Species
- Barred tiger salamander
- European green crab
- Diatoms
- Heliconius butterflies
- Stalk-eyed Flies

In addition, NCGAS supported a total of 259 biologists doing research in the general area of genome analysis (103 named allocations), and contributed to the publication of a total of 55 peer-reviewed scientific publications.

1.2.1.2. Support software and computational resources relevant to genome analysis

NCGAS established a new model for support of software on major NSF-funded cyberinfrastructure. In particular, NCGAS established new models for management and support of software on XSEDE (the eXtreme Science and Engineering Discovery Environment). XSEDE is, according to its own web site (xsede.org), "... supports 16 supercomputers and high-end visualization and data analysis resources across the country. "XSEDE is composed of multiple partner institutions known as Service Providers or SPs, each of which contributes one or more allocatable services." In the past, the model for such NSF-funded cyberinfrastructure (XSEDE, TeraGrid, and NSF supercomputer centers before that) was that the entities that proposed to host hardware were in essence awarded funding to manage support as a side effect of winning awards for hardware hosting. NCGAS has changed that model, proposing to support genome analysis software (and users) using the resources supported by XSEDE and funded by the NSF. This support is provided in cooperation with XSEDE and the SPs that deliver resources to the national research community via XSEDE - but formally, and in terms of grants from the NSF, NCGAS is structurally independent of XSEDE. This is precedent setting. NCGAS set standards for XSEDE and XSEDE-supported SPs for what genome analysis software should be available on XSEDE-supported systems, down to specifying version numbers. While each SP was responsible for their own decisions whether to comply with these recommendations or not, the publication of this list increased consistency of software available to the national research community.

NCGAS notably helped operate BLAST as a distributed computing application on the Open Science Grid. While sequence similarity searching in general was not central focus of NCGAS, we were presented with an interesting opportunity to implement BLAST on the high throughput computing system operated by the Open Science Grid. As a result of creating a BLAST portal, millions of hours of CPU time have been provided to the research community to run BLAST jobs, all running on the largely volunteer OST. NCGAS is notable in that it is one of few entities in the US that supports significant suites of software running both on XSEDE and the OSG.

In addition, NCGAS and the IU Pervasive Technology Institute made the large-memory cluster Mason available for use by the national research community. Mason was commissioned in 2011, and was one of

a handful of nationally-available computing systems that broke the mold of attentiveness to the Top500 list. Rather than being configured for maximizing the rate of floating point operations per second on a dense matrix algebra task, as are many supercomputers today, Mason was configured with 16 nodes each with half a terabyte of RAM (Random Access Memory), specifically to support the DNA genome assembly software most used and most widely trusted by practicing biologists. NCGAS supported more than 1,000 users who received allocations of computer time on Mason via the XSEDE-supported national allocation process over the four years of NSF funding that are the subject of this report.

1.2.1.3. Maintain, harden, tune, enhance, and distribute software relevant to genome analysis

NCGAS' most significant accomplishment in enhancing genome analysis software was NCGAS contribution to the widely used RNA-seq (RNA sequence assembly) software Trinity. IU's enhancement to the software sped up analysis by a factor of 8, and these software enhancement have been included in the definitive software distribution of Trinity.

NCGAS has also supported the creation of the "XSEDE National Integration Toolkit" (XNIT). XNIT is a suite of software available for download and installation on computational clusters. XNIT was developed by XSEDE to address the campus bridging problem of software on campus clusters being very different from software on XSEDE resources. XNIT is distinct in its ease of use in that it is based on tools for at-will or automated download and installation of software so that it is very easy for campus cluster administrators to install any of the software tools included in XNIT. Further, campus cluster administrators have complete control of what software within the overall XNIT suite is downloaded and installed. At the request of the research community NCGAS has added the whole suits of bioinformatics software supported by NCGAS on XSEDE. This means that every computational cluster administrator in the US has the ability to easily download and install locally all of the software supported by NCGAS.

NCGAS has also provided access to bioinformatics software through three online web portals:

- Galaxy web portal - providing access to the widely used Galaxy workflow system on Mason and other XSEDE-supported resources
- Trinity RNA-Seq portal, running on IU's Karst supercomputer.
- Assistance provided to the Texas Advanced Computing Center for the iPlant project and its online web portals.

1.2.2. *Specific Objectives*

The software supported by NCGAS as of the end of the first four years of NSF funding includes 47 packages described in detail in the attached file on significant software activities.

1.2.3. *Significant results*

Provide online help, consulting, and tutorials related to genome analysis.

Key highlights of NCGAS support include:

- Publication of 45 Knowledge Base documents about bioinformatics software. The IU-developed Knowledge Base (KB) provides a 7 x 24 x 365 self-help tool. Researchers can ask a question of the KB via the NCGAS web pages and XSEDE web portal, and receive an answer in the form of a well written document any time of day. If the written documentation is not sufficient to answer a user's questions then there is a web form for additional follow up and contact with a NCGAS consultant.
- Consulting. NCGAS has during its 48 months of funding from the NSF completed a total of 741 short term consulting engagements (those taking less than 4 hours of staff time to resolve) and 235 long term consulting engagements (taking more than 4 hours of staff time to resolve). Many of the long term consultations were research collaborations that went on for months or even years. In such research collaborations NCGAS staff became partners in research endeavors and played a critical role in some of the most important scientific discoveries accomplished by scientists who received help from NCGAS.
- NCGAS completed tutorials and training and outreach activities attended by thousands of attendees during the course of four years of funding by the NSF.

Consultant services were provided by telephone, email, and in-person consultations. Consulting hours are typically 8 am to 5 pm weekdays, but when there is time pressure on a researcher support activities often extend beyond local business hours.

During the course of the four years of NSF support, NCGAS carried on a total of 49 extended consultations, described in the attached file on consulting and significant results.

1.2.4. Key outcomes or other achievements

The key outcome of NCGAS was the establishment of an effective consulting service that focused specifically on accelerating the research of biologists and bioinformaticians, and in so doing accelerated biological discoveries in the US.

1.3. What opportunities for training and professional development has the project provided?

The most important training and professional development activities have been tutorials provided at national and international conferences. A summary of these tutorials is provided below:

- 900 participants in tutorials (2254 contacts at events). In the last year, some of the events attended were:
 - Galaxy Community Conference 2014/ Baltimore MD;
 - Genomics in July, IU - Cyberinfrastructure Building, IU;
 - Genomics in July / IUB - Cyberinfrastructure Building, IU;
 - NEON / Boulder, CO;
 - IU Bioinformatics Clinic 2014 IU;
 - Indiana Statewide IT Conference;
 - Wonderlab Museum Bloomington, IN;

- The International Conference for High Performance Computing, Networking, Storage and Analysis / New Orleans, LA.

Through a Research Experiences for Undergraduates (REU) supplemental award, NCGAS trained two students in systems administration skills for high performance computing systems important for genomics research. Both of these students were offered jobs in this specialty within 6 months of completing their internships.

Dr. Thomas Doak was, at the beginning of the funding period, a postdoctoral fellow. Dr. Doak was promoted to the rank of Assistant Scientist at Indiana University, and is now the PI of a NSF award to continue the services of NCGAS.

Staff member Carrie Ganote has begun a graduate degree program in bioinformatics while in the employ of NCGAS.

1.4. How have the results been disseminated to communities of interest?

Results have been disseminated to communities of interest in a variety of ways, including:

- Publications in scientific journals
- Presentations
- Birds of a feather sessions at technical conferences
- Displays and booths at national and international technical conferences
- Articles in the lay press, most notably in Science Node (formerly International Science Grid this week - now at <https://sciencenode.org>)
- NCGAS web site at ncgas.org
- In-person contacts
- Email list distribution
- Newsletters

1.5. What do you plan to do during the next reporting period to accomplish the goals?

Nothing to report.

2. Products

2.1. Products resulting from this project during the specified reporting period

2.1.1. (Peer-reviewed) Journal Articles

- Couger, Brian M; Pipes, Lenore; Squina, Fabio; Prade, Rolf; Siepel, Adam; Palermo, Robert; Katze, Michael G; Mason, Christopher E; Blood, Philip D (2014). Enabling large scale next-generation sequence assembly with Blacklight. *Concurrency and Computation: Practice and Experience*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: doi: 10.1002/cpe.3231
- Farlow, A., H. Long, S. Arnoux, W. Sung, T.G. Doak M. Nordborg and M. Lynch. (2015). The Spontaneous Mutation Rate in the Fission Yeast *Schizosaccharomyces pombe*.. *Genetics*. 201:737 . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://www.genetics.org/content/201/2/737.full>
- Ghaffari, N.; Sanchez-Flores, A.; Ryan, D.; Garcia-Orozco, K.D.; Chen, P. L.; Ochoa-Leyva, A.; Lopez-Zavala, A. A.; Carrasco, J. S.; Hong, C.; Briebe, L. G.; Rudino-Pinera, E.; Blood, P. D.; Sawyer, J. A.; Johnson, C. D.; Dindot, S. V.; Sotelo-Mundo, R.R.; Criscitiello, M. F. (2014). Improved transcriptome of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture. *Nature Scientific Reports*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = No
- Long, H., S Kucukyildirim, W. Sung, E. Williams, H. Lee, M.S. Ackerman, T. G. Doak, H. Tang, M. Lynch. (2015). Background mutational features of the radiation-resistant bacterium *Deinococcus radiodurans*.. *Molecular Biology and Evolution*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://mbe.oxfordjournals.org/content/early/2015/06/17/molbev.msv119.abstract>
- Long, H., W. Sung, S. F. Miller, M. S. Ackerman, T. G. Doak, M. Lynch. (2015). Mutation rate, spectrum, topology and context-dependency in the DNA mismatch repair (MMR) deficient *Pseudomonas fluorescens* Migula ATCC948.. *Genome Biology and Evolution*. 23:262-71. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://gbe.oxfordjournals.org/content/7/1/262.full.pdf+html?sid=e5570fd3-bf72-4c6c-a744-975adcdccdda>
- M Wang, TG Doak, Y Ye. (2015). Subtractive assembly for comparative metagenomics, and its application to type 2 diabetes metagenomes.. *Genome Biology*. (16:243), . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://www.genomebiology.com/2015/16/1/243>
- Petersen, I. T. and Bates, J. E. and Dodge, K. A. and Lansford, J. E. and Pettit, G. S. (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood.. *Development and Psychopathology*. (PubMed), <http://www.ncbi.nlm..> Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; DOI: doi: 10.1017/S0954579414000789.
- Pipes, Lenore; Li, Sheng; Bozinovski, Marjan; Palermo, Robert; Peng, Xinxia; Blood, Phillip; Kelly, Sara; Weiss, Jeffrey M; Thierry-Mieg, Jean; Thierry-Mieg, Danielle (2013). The non-human primate reference transcriptome resource (NHPRT) for comparative functional genomics. *Nucleic acids research*. PubMed 41:D906-D914.. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://www.ncbi.nlm.nih.gov/pubmed/23203872>
- Uy, Karen; LeDuc, Richard; Ganote, Carrie; Price, Donald. (2014). Physiological effects of heat stress on Hawaiian picture-wing *Drosophila*: genome-wide expression patterns and stress-related traits.. *Conversation Physiology*. doi: 10.1093/conphys. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://conphys.oxfordjournals.org/content/3/1/cou062.full>

2.1.2. Conference Papers and Presentations

- Blood, Philip D; Marcus, Shoshana; Schatz, Michael C (2014). Large-scale Sequencing and Assembly of Cereal Genomes Using Blacklight. *XSEDE 14 Conference Proceedings*. . Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <http://dl.acm.org/citation.cfm?id=2616502&CFID=737967943&CFTOKEN=61560666>

- Fischer, J., R. Knepper, E. Coulter, C. Peck, C. A. Stewart (2015). XCBC and XNIT - Tools for Cluster Implementation and Management in Research and Training,. *IEEE Cluster Computing Conference*. pp.857-864. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: [http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7307692&filter=AND\(p_Publication_Number:7307300\)](http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7307692&filter=AND(p_Publication_Number:7307300))
- Kuhn, D. N, Dillon, N. L., Innes, D. J., Wu, L. S., Mockaitis, K (2014). Development of Single Nucleotide Polymorphism (SNP) Markers from the Mango (*Mangifera indica*) Transcriptome for Mapping and Estimation of Genetic Diversity. *Plant and Animal Genome XXII Conference*. <https://pag.confex.c>. Status = PUBLISHED; Acknowledgment of Federal Support = Yes ; Peer Reviewed = Yes ; OTHER: <https://pag.confex.com/pag/xxii/webprogram/Paper11012.html>
- Seetharam, Arun; Gomez, Antonio; Purcell, Catherine M; Hyde, John R; Blood, Philip D; Severin, Andrew J (2015). NCBI-BLAST programs optimization on XSEDE resources for sustainable aquaculture. *XSEDE 2015 Conference Proceedings*. Status = PUBLISHED; Acknowledgment of Federal Support = Yes; Peer Reviewed = Yes ; OTHER: <http://dl.acm.org/citation.cfm?id=2792749>
- Wu L.S., C. L. Ganote, T.G. Doak, W. Barnett, K. Mockaitis, C.A. Stewart. (2015). *Cyberinfrastructure resources enabling creation of the loblolly pine reference transcriptome*.. XSEDE 15. St. Louis, MO. Status = PUBLISHED; Acknowledgement of Federal Support = Yes
- Hayashi, Soichi and Gesing, Sandra and Quick, Rob and Teige, Scott and Ganote, Carrie and Wu, Le-shin and Prout, Elizabeth (2014). *{Galaxy Based BLAST Submission to Open Science Grid Resources.}*. International Symposium on Grids and Clouds 2014. <http://pos.sissa.it/archive/conferences/210/025/IS>. Status = PUBLISHED; Acknowledgement of Federal Support = Yes
-

2.1.3. Other Publications

- C. L. McGrath, J. F. Gout, P. Johri, T. G. Doak, and M. Lynch. (2014). *Differential retention and divergent resolution of duplicate genes following whole-genome duplication*.. *Genome Research*. Status = PUBLISHED; Acknowledgement of Federal Support = Ye

3. Participants

3.1. Individuals

Table 1. Individuals that have worked on the project

Name	Most Senior Project Role	Nearest Person Month Worked
Stewart, Craig	PD/PI	1
Barnett, William	Co PD/PI	1
Hahn, Matthew	Co PD/PI	1
Lynch, Michael	Co PD/PI	1
Henschel, Robert	Other Professional	1
Miller, Therese	Other Professional	1

Quick, Robert	Other Professional	1
Wu, Leshin	Other Professional	6
Doak, Thomas	Staff Scientist (doctoral level)	8
Gannote, Carrie	Staff Scientist (doctoral level)	6

3.1.1. Full details of individuals who have worked on the project

Craig A Stewart

Email: stewart@iu.edu

Most Senior Project Role: PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Project oversight as P.I.

Funding Support: Indiana University

International Collaboration: No

International Travel: Yes, germany - 0 years, 0 months, 7 days

William K Barnett

Email: barnettw@indiana.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Co-PI responsible for outreach to new groups to generate users/projects using the NCGAS service.

Funding Support: NSF/Indiana University

International Collaboration: No

International Travel: No

Matthew W Hahn

Email: mwh@indiana.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Scientific leadership of NCGAS activities

Funding Support: Indiana University

International Collaboration: No

International Travel: No

Michael R Lynch

Email: milynych@indiana.edu

Most Senior Project Role: Co PD/PI

Nearest Person Month Worked: 1

Contribution to the Project: Co-PI doing genome research.

Funding Support: Indiana University.

International Collaboration: No

International Travel: No

Robert Henschel

Email: henschel@iu.edu

Most Senior Project Role: Other Professional

Nearest Person Month Worked: 1

Contribution to the Project: software optimization

Funding Support: Indiana University

International Collaboration: No
International Travel: No

Therese Miller

Email: millertm@iu.edu

Most Senior Project Role: Other Professional

Nearest Person Month Worked: 1

Contribution to the Project: Financial and reporting management

Funding Support: IU

International Collaboration: No

International Travel: No

Robert Quick

Email: rquick@iu.edu

Most Senior Project Role: Other Professional

Nearest Person Month Worked: 1

Contribution to the Project: OSG software implementation

Funding Support: IU

International Collaboration: No

International Travel: No

Leshin Wu

Email: Lewu@iu.edu

Most Senior Project Role: Other Professional

Nearest Person Month Worked: 6

Contribution to the Project: bioinformatician consultant / programmer

Funding Support: NSF, IU

International Collaboration: No

International Travel: No

Thomas Doak

Email: tdoak@iu.edu

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 8

Contribution to the Project: NCGAS operational management

Funding Support: NSF, NIH, IU

International Collaboration: No

International Travel: No

Carrie Gannote

Email: cgannot@iu.edu

Most Senior Project Role: Staff Scientist (doctoral level)

Nearest Person Month Worked: 6

Contribution to the Project: bioinformatics consultant / programmer

Funding Support: NSF

International Collaboration: No

International Travel: No

3.2. Partner organizations

Table 2. Partner organizations

Name	Type of Partner Organization	Location
Pittsburgh Supercomputing Center, Carnegie Mellon University	Academic Institution	Pittsburgh, PA
San Diego Supercomputer Center, University of California San	Academic Institution	San Diego, CA
Technische Universitaet Dresden	Academic Institution	Dresden, Germany
Texas Advanced Computing Center, University of Texas	Academic Institution	Austin, TX
XSEDE	Other Nonprofits	United States

3.2.1. *Full details of partner organizations*

3.2.1.1. [Pittsburgh Supercomputing Center, Carnegie Mellon University](#)

Partner's Contribution to the Project

- In-Kind Support
- Facilities
- Collaborative Research
- Personnel Exchanges

More Detail on Partner and Contribution: PSC has provided facilities, computer time, and storage space on the Blacklight supercomputer in support of NCGAS activities and in support of biological researchers who have used NCGAS services. Staff of this institution have also engaged use of NCGAS staff and facilities, and have made available resources at their site to NCGAS staff. Some of the support provided by this institution has been provided in-kind, and this institution has engaged in collaborative research on genome analysis software, particularly as regards use of Galaxy and software that requires the large shared memory architecture of Blacklight. This institution has also participated in education, outreach, and dissemination efforts of NCGAS.

3.2.1.2. [San Diego Supercomputer Center, University of California San](#)

Partner's Contribution to the Project

- In-Kind Support
- Facilities
- Collaborative Research
- Personnel Exchanges

More Detail on Partner and Contribution: SDSC has provided facilities, computer time, and storage space on the Gordon supercomputer in support of NCGAS activities and in support of biological researchers who have used NCGAS services. Staff of this institution have also engaged use of NCGAS staff and facilities, and have made available resources at their site to NCGAS staff. Some of the support

provided by this institution has been provided in-kind, and this institution has engaged in collaborative research on genome analysis software. This institution has also participated in education, outreach, and dissemination efforts of NCGAS.

3.2.1.3. Technische Universitaet Dresden

Partner's Contribution to the Project

- Financial support
- In-Kind Support
- Facilities
- Collaborative Research
- Personnel Exchanges

More Detail on Partner and Contribution: Staff of TU-D have engaged use of NCGAS staff and facilities, and have made available resources at their site to NCGAS staff. Some of the support provided by this institution has been provided in-kind, and this institution has engaged in collaborative research on genome analysis, particularly as regards movement of genome analysis over international high-speed networks. TU-D has provided financial support, paying the cost of PI Stewart's collaborative visit to TU-D in June of 2013. TU-D is the most active of the NCGAS partner institutions in personnel exchanges, with IU and TU-D staff regularly visiting each other's main campus.

3.2.1.4. Texas Advanced Computing Center, University of Texas

Partner's Contribution to the Project

- In-Kind Support
- Facilities
- Collaborative Research
- Personnel Exchanges

More Detail on Partner and Contribution: TACC has a subcontract from NCGAS. In addition, it has provided facilities, computer time, and storage space on the Stampede supercomputer in support of NCGAS activities and in support of biological researchers who have used NCGAS services. Staff of this institution have also engaged use of NCGAS staff and facilities, and have made available resources at their site to NCGAS staff. Some of the support provided by this institution has been provided in-kind, and this institution has engaged in collaborative research on genome analysis software. This institution has also participated in education, outreach, and dissemination efforts of NCGAS.

3.2.1.5. XSEDE

Partner's Contribution to the Project

- In-Kind Support
- Facilities
- Collaborative Research
- Personnel Exchanges

More Detail on Partner and Contribution: Staff of the NSF-funded XSEDE project have engaged use of NCGAS staff and facilities, and have made available resources at their site to NCGAS staff. Some of the support provided by XSEDE has been provided in-kind, and this institution has engaged in collaborative research on genome analysis software. XSEDE has particularly played a strong role in education, outreach, and dissemination efforts of NCGAS.

3.3. *Have other collaborators or contacts been involved?*

No

4. Impact

4.1. *What is the impact on the development of the principal discipline(s) of the project?*

Results have been disseminated to communities of interest in a variety of ways, including:

- Publications in scientific journals
- Presentations
- Birds of a feather sessions at technical conferences
- Displays and booths at national and international technical conferences
- Articles in the lay press, most notably in Science Node (formerly International Science Grid this week - now at <https://sciencenode.org>)
- NCGAS web site at ncgas.org
- In-person contacts
- Email list distribution
- Newsletter

4.2. *What is the impact on other disciplines?*

The primary other discipline on which NCGAS has had an impact is computational science and cyberinfrastructure. The largest impact that NCGAS has had in computational science has been to establish a model of a “scientific service center” focused on a particular subdiscipline, independent of federally-funded cyberinfrastructure computational resources. That is, we have decoupled federal funding for supercomputers and federal funding for supercomputer application support. This is tremendously important as it ensures that a community that is relatively newer to use of supercomputers - biology for example - has a support resource funded by the BIO directorate of NSF and attuned to the needs of the current most important research.

We have also established new models for distribution of software relevant to biological research which improves the nation's ability to use its aggregate cyberinfrastructure resources.

4.3. *What is the impact on the development of human resources?*

The most important training and professional development activities have been tutorials provided at national and international conferences. This includes more than 900 participants in tutorials (2254 contacts at events). In the last year, some of the events attended were:

- Galaxy Community Conference 2014/ Baltimore MD;
- Genomics in July, IU - Cyberinfrastructure Building, IU;
- Genomics in July / IUB - Cyberinfrastructure Building, IU;
- NEON / Boulder, CO;
- IU Bioinformatics Clinic 2014 IU;
- Indiana Statewide IT Conference;
- Wonderlab Museum Bloomington, IN;
- The International Conference for High Performance Computing, Networking, Storage and Analysis / New Orleans, LA.

Through a Research Experiences for Undergraduates (REU) supplemental award, NCGAS trained two students in systems administration skills for high performance computing systems important for genomics research. Both of these students were offered jobs in this specialty within 6 months of completing their internships.

Dr. Thomas Doak was, at the beginning of the funding period, a postdoctoral fellow. Dr. Doak was promoted to the rank of Assistant Scientist at Indiana University, and is now the PI of a NSF award to continue the services of NCGAS.

Staff member Carrie Ganote has begun a graduate degree program in bioinformatics while in the employ of NCGAS.

4.4. *What is the impact on physical resources that form infrastructure?*

Nothing to report.

4.5. *What is the impact on institutional resources that form infrastructure?*

The software distributed by NCGAS has improved the effectiveness and ease of use of cyberinfrastructure resources throughout the nation.

4.6. *What is the impact on information resources that form infrastructure?*

NCGAS has facilitated the publication of several data sets important to basic biological research and to management of important plant and animal stocks.

4.7. *What is the impact on technology transfer?*

The primary impact of NCGAS on technology transfer is in the addition of major upgrades to the open source Trinity sequence software.

4.8. *What is the impact on society beyond science and technology?*

The impact on society beyond science and technology is hard to gauge now, but could be tremendous over the course of coming decades. What is the impact of understanding the genome of the pine tree,

cacao, and mango so that these important crop plants can properly be managed over coming decades? The potential impact of science supported by NCGAS on society through better management of food supplies and better understanding of how organisms adapt to global climate change could be of fundamental importance to US and global populations.

5. Changes/ Problems

5.1. Changes in approach and reasons for change

Nothing to report.

5.2. Actual or Anticipated problems or delays and actions or plans to resolve them

Nothing to report.

5.3. Changes that have significant impact on expenditures

Nothing to report.

5.4. Significant changes in use or care of human subjects

Nothing to report.

5.5. Significant changes in the use or care of vertebrate animals

Nothing to report.

5.6. Significant changes in the use or care of biohazards

Nothing to report.