

**Delaware EPSCoR RII Track 2**

**Final Report – August 31, 2011**

**Reporting Period – September 1, 2011 – August 31, 2012**

**Report Sections**

**Project Summary**

The *Northeast Cyberinfrastructure Consortium* (NECC - <http://www.necyberconsortium.org/>) is comprised of five EPSCoR states – New Hampshire, Maine, Rhode Island, Vermont and Delaware, and completed its third and final year of Track-2 funding. The goal of NECC is to develop a regional cyber-based collaborative network of scientists that can benefit from core centers and expertise available at many of the participating institutions. Two pilot projects have been launched under NECC - the genome sequencing of the Little Skate, *Leucoraja erinacea*, and a study of metagenomics of algal blooms. These pilot projects are highly de-centralized and distributed among the NECC partner states and take significant advantage of the cyberinfrastructure upgrades. Delaware is leading the NECC sequencing effort through its *Sequencing and Genotyping Center* (SGC) and is providing one of two regional database centers at UD's *Center for Bioinformatics and Computational Biology* (CBCB - <http://bioinformatics.udel.edu/>), under the leadership of Dr. Cathy Wu, Edward G. Jefferson Endowed Professor, Department of Computer & Information Sciences. The *Northeast Bioinformatics Collaborative* (NEBC) is comprised of key bioinformatics experts in the NECC.



*Table 1: Members of the Northeast Bioinformatics Collaborative (NEBC)*

<b>Name</b>	<b>Organization</b>	<b>NECC State</b>
Cecilia Arighi	Delaware Biotechnology Institute	Delaware
Clinton Chichester	University of Rhode Island	Rhode Island
Benjamin King	Mount Desert Island Biological Laboratory	Maine
Carolyn Mattingly	Mount Desert Island Biological Laboratory	Maine
Jason Moore	Dartmouth University	New Hampshire
Daniel Udvary	University of Rhode Island	Rhode Island
James Vincent	University of Vermont	Vermont
Cathy Wu	Center for Bioinformatics and Computation Biology	Delaware

For our NECC colleagues in the New England states, the predominant need was to install significantly fiber upgrades to support broad bandwidth Internet access. In Delaware the cyberinfrastructure goals have been different. The University of Delaware (UD) and Delaware State University (DSU) have been connected to Internet-2 via MAGPI in Philadelphia for several years with partial support provided by NIH-IDeA through the Delaware INBRE program. UD is linked to Internet2 through a 10 Gb MAGPI node. In 2006, DSU connected to MAGPI via UD at 45 Mb and has been upgrading research and teaching facilities across the campus to higher bandwidth. In addition, all of the partner institutions were looking into significant on-campus cyber upgrades, including enhancement of intercampus redundancy, wireless access for students, and bandwidth upgrades within and between campuses.

One of the key features of Delaware's partnership in the NECC consortium is the ability to create a virtual partnership with colleagues in other EPSCoR States. To that end, Delaware has taken the lead in the genome sequencing of the Little Skate and the metagenomics sequencing project for algal blooms. Delaware also has provided leadership in the annual sequence annotation workshops and serves as one of the two regionally distributed datacenters. Finally, the related Track-2 C2 funding awarded by NSF-EPSCoR has catalyzed the establishment of a statewide "Delaware Environmental Monitoring and Analysis Center (DEMAC)," hiring an environmental informatics specialist who coordinates the collection and organization of environmental sensor data from a diverse set of sensor sources.

During the third Annual NECC Meeting, held on March 16, 2012 at the University of Delaware in Newark, DE, we reviewed progress on the installation of fiber for the regional network, cyber-enabled research, workforce development and diversity, external engagement and sustainability. Dr. Sian Mooney represented the NSF EPSCoR Office, and Dr. Steven Meacham attended as a second NSF observer. The result of the workshop was that we have made significant progress as an unprecedented consortium among five EPSCoR states, and have laid the foundation for strong, sustainable research collaborations among our five states.

#### Goals for the Track-2 Program

- Provide cyberinfrastructure for research and education in the northeast EPSCoR region;
- Develop human infrastructure to create a virtual organization for distributed bioinformatics and data analysis;
- Launch pilot projects to develop the regional expertise to analyze genomics data, especially deep sequencing data, in a distributed manner; and to foster collaborations across the region;
- Establish collaborations on cyber-education, cyber-knowledgeable workforce development and diversity and outreach;
- Leverage resources to accomplish our goals;

#### Progress to Date

- The fiber installation and upgrade projects on the Delaware campuses have been completed;
- Pilot projects in metagenomics of toxic algal blooms in the region and the Little Skate Genome have helped to build a virtual organization that allows us to analyze data remotely in a distributed and highly collaborative manner – learning and training opportunities;
- Pilot projects have inspired novel collaborations on water and cyber-enabled research in the NECC region;
- Collaborations on the Watershed Project leveraged the water education outreach programs in the region for workforce development and diversity; and
- Novel collaborations were fostered that would not have happened without NECC and its expanded cyber-infrastructure.

#### Novel Aspects across the NECC Network

Cyber-enabled research collaborations that require NECC resources.

- Open access shared database resource for eukaryotic metagenetics research (NH, VT)
- Virtual Organization for bioinformatics and data analysis (NEBC, all 5 states)
- EOS Webster – GIS and positional information overlaid on field sample data for water research (DE, NH and VT)

- Five-state metagenome project to better understand algal blooms that impact the state economies and that require the coordinated efforts of bioinformaticians working remotely – work flow, data storage, data management
- Watershed watch and watershed projects collaborate in RI and NECC at large

### **Intellectual Merit**

#### **The Little Skate Genome**

The genome of *Leucoraja erinacea*, the Little Skate, has been sequenced in a collaborative effort among NECC members, and with support by the *Northeast Bioinformatics Collaborative* (NEBC). The Little Skate is a chondrichthyan fish that evolved approximately 450 million years ago. It is one of the most primitive jawed vertebrates with paired limbs. As such it is used as a model organism for the study of the origin and evolution of developmental processes, as well as human physiology, immunology, toxicology and other fields.

Research progress has been slowed by lack of detailed genomic information. The only reported chondrichthyan genome sequence is a very low coverage draft of the non-elasmobranch elephant shark. Completion of the Little Skate Genome will provide an important tool for researchers.

The genome of the Little Skate is slightly larger than the human genome. While much technical progress has been made in the ability to generate large amounts of sequencing data from DNA with massively parallel sequencing instruments, the assembly of this data into a draft genome is a very labor intensive process requiring a broad range of expertise.

The NECC has generated a substantial set of skate genome sequence data, where each base of the 3.4 Gbp genome has been sequenced 59 times on average. These data consist of three batches of data generated from two different sequencing libraries that were prepared from a single stage 32 skate embryo. Two batches of data are from a paired-end library with a 500bp insert and the other is a mate-pair library with a 3.5kbp insert size. The paired-end library data consist of 770 million 125bp paired-end reads generated using an Illumina Genome Analyzer IIx, and 1 billion 100bp paired-end reads recently generated using an Illumina HiSeq2000. The mate-pair library data consist of 96 million 130bp mate-pair reads generated using an Illumina Genome Analyzer IIx. An initial *de novo* genome assembly made using the initial set of 770 million paired-end reads has over 3.1 million contiguous sequences (contigs) that together represent 1.6 Gbp of genomic sequence. The large number of contigs is due to repetitive elements in the genome and that the sequencing library was made from short (500bp) DNA fragments.

#### **Metagenomes of Cyanobacterial Blooms**

Lakes, estuaries and bays in NH, ME, RI and VT are plagued by algae blooms composed of cyanobacteria species and, frequently, the toxins they produce. These blooms range from irritating nuisances to health threats to animals and humans. These health threats also have a negative impact on tourist-based economies that are crucial in the Northeast region.

Metagenomics is the determination of genome sequences from a community by shotgun sequencing directly from an environmental sample, such as a water sample. Metagenomics gives a composite snapshot of the population, and provides insights into both species identification and genetic functional potential. This method allows the identification of microorganisms that are not easily cultured, without prior hypotheses of which ones are present.

The conditions that favor cyanobacterial blooms and subsequent production of toxins are not well understood. The data generated by the NECC metagenome pilot project will provide a baseline survey of algal blooms over time from five water bodies in the Northeast. This new knowledge will provide insight into the diversity of populations present over the time course of blooms, as well as the genetic potential of the species present.

The goal of the Northeast Cyberinfrastructure Consortium (NECC-[www.necyberconsortium.org/](http://www.necyberconsortium.org/)) is to systemically improve inter-institutional collaborations in the region among academic and biomedical research institutions by significantly enhancing connectivity and communications. The fundamental infrastructure enhancements completed have proven extremely productive and impactful; enabling researchers across our states access to resources, which enable them to perform data-intensive, cutting-edge life science research. This collaboration continues to drive infrastructural, educational, and cooperative development of bioinformatics capabilities among institutions across the states. It has both utilized new fiber connecting the member institutions to transfer the very large genome sequencing files produced by the project, but has also produced novel resources, such as a shared data center to further enable the sharing of data across the partners.

The NECC Little Skate Genome project serves multiple roles: promoting scientific goals of characterizing a potential biomedical model, demonstrating and further building our Delaware and NECC cyberinfrastructure, and acting as a platform for training cyber-knowledgeable scientists in our states.

- The Little Skate Genome project has developed significant resources related to the genome and its annotation. Initial sequencing and assembly of the genome produced a rough draft sequence, totaling 2,962,365 contigs and 1,555,444,314 bp of skate genomic data. The Little Skate Genome assembly and completed mitochondria annotation have been deposited to the GenBank public repository. This has contributed to several high-impact scientific papers published in the *Proceedings of National Academy of Sciences* [1] and a recent publication, led by Dr. Benjamin King at Mount Desert Island Biological Lab in Maine, “A natural deletion of the HoxC Cluster in elasmobranch fishes” in the journal *Science* in December 2011 [2], and a Research Highlight, “Case of the missing cluster,” in the journal *Nature*, also in December 2011.
- In support of the mission to create a cyber-knowledgeable life science workforce, the Little Skate Genome project has successfully completed three intensive weeklong workshops to train students in bioinformatics for the purpose of annotating the Little Skate Genome. The workshops were hosted by the *Center for Bioinformatics and Computational Biology* (CBCB) at UD in May 2010, at Mount Desert Island Biological Laboratory (MDIBL) in October 2010, and UD-CBCB in May 2011 (See Figure 2), respectively (<http://skatebase.org/workshops>), provided valuable training and workforce development across all five states. The annotation workshops provided a minimum of thirty-two hours of training for 56 participants, with ten instructors and fourteen guest lecturers from the NECC states. The Genome Annotation workshop is intended to inspire a new generation of bioinformaticians. The May 2010 workshop



*Fig 2: The May 2011 NECC Annotation Workshop at UD attracted over 100 people*

attracted 35 attendees to UD – 16 males and 19 females. The May 2011 workshop held at UD attracted 41 attendees with 29 males and 12 females.

- An additional series of virtually distributed Annotation Jamborees involved 29 trainees to annotate the mitochondrial genome were held in each state in September 2011. These Jamborees marked the complete assembly of the mitochondrial genome from the Little Skate, as well as the complete annotation of the various features of the mitochondrial genome, culminating in a *Database* journal paper entitled “Community annotation and bioinformatics workforce development in concert – Little Skate genome annotation workshops and jamborees,” [3] co-authored by all Jamboree participants, as well as a presentation at the April 2012 *International Biocuration Conference* in Washington, DC.
- A 30-minute special for Vermont Public Television, entitled “*Regenerating our Future*” was produced during 2011 and first aired November 9, 2011. The program outlines the cyberinfrastructure enhancements put in place across the NECC partner states, and describes the scientific impact of the collaborative research project and its potential for regenerative medicine. Delaware’s contributions are prominently featured in the program. A web version of the telecast is available at <http://www.vpt.org/show/16403/403>.
- The NECC Shared Data Center (SDC) was developed and deployed to enable sharing of the large data sets produced in the project. The SDC co-developed by Delaware and Vermont and hosted by UD-CBCB with a mirror site at the University of Maine is critical to the collaborative research of the NECC. Based on a redundant set of storage servers that are geographically distributed and custom software for replication and failover, SDC has allowed sharing of large data sets. In particular, the SDC web site allows for a user-friendly method for transferring and sharing the very large data files such a project produces.
- A dedicated genome portal, SkateBase (<http://skatebase.org>), and underlying infrastructure were developed to support ongoing annotation and genome work. Hosted by UD-CBCB, the SkateBase portal serves as a central hub for the Little Skate Genome project—both as a *curation portal* for collaborative annotation and as a *public portal* for project dissemination. The online resource provides project information and results to the larger scientific community, while also providing the needed infrastructure to drive the project’s annotation and training activities. It currently provides a number of tools for file exchange, sequence analysis, genome visualization, and curation interface to support the NECC collaborative annotation and training. The SkateBase also provides public dissemination of research data, including the Little Skate Genome assembly and mitochondria annotation deposited to the GenBank.
- The expansion of the Sequencing and Genotyping Core would not have been possible without a world-class group of faculty dedicated to genomics research. In Fall 2011, three UD faculty members, Blake Meyers (PI for the NSF MRI grant that funded the new PacBio sequencer), Janine Sherrier, Professor in the department of Plant and Soil Sciences, and Pam Green, Crawford H. Greenewalt Chair, Professor of Plant and Soil Sciences and



*Fig 3: A 2011 Vermont Public TV Special “Regenerating our Future” that focuses on NECC and features Delaware researchers.*

Professor of Marine Studies, published their findings in the journal Nature, “The Medicago Genome Provides Insight into the Evolution of Rhizobial Symbiosis,” They teamed up with the international Medicago Genome Consortium to sequence, assemble and annotate the genome of *Medicago truncatula*. The sequenced Medicago species is often used as a model, since it is similar to other important legume crops such as soybeans, peas and peanuts. Legumes are important crop plants because they interact with microbes to get nutrition from the environment, revitalizing nutrient-depleted soils, providing proteins and minerals.

### **Programmatic Terms and Conditions**

#### **Broadening Participation**

Delaware keeps close track of participants in Track-2 activities. Table 2 provides a breakdown of the participants by category, percent of women and percent of underrepresented minorities.

*Table 2: Demographics of the State of Delaware and RII Academic Partners*

<b>Delaware</b>	<b>Total Number</b>	<b>% Female</b>	<b>% African-American, Hispanic and Native American</b>
State of Delaware, 2011	907,135	52%	31%
<b>STEM** Student Demographics by Institution</b>			
	<b>Total</b>	<b>Female</b>	<b>% African-American, Hispanic and Native American</b>
University of Delaware (STEM grad. students)	2,100	47%	10% (210)
Delaware State Univ. (STEM grad. students)	111	68%	38% (42)
University of Delaware (STEM undergraduates)	7,871	57%	8% (630)
Delaware State Univ. (STEM undergraduates)	1048	58%	81% (845)
Wesley College (STEM undergraduates)	138	60%	56% (77)
Delaware Tech (STEM undergraduates)	1,844	24%	29% (535)
<b>Total STEM Graduate Students</b>	<b>2,211</b>	<b>48%</b>	<b>11% (252)</b>
<b>Total STEM Undergraduates</b>	<b>10,901</b>	<b>52%</b>	<b>25% (2,732)</b>
<b>Faculty Demographics</b>			
	<b>Total</b>	<b>Female</b>	<b>% African-American, Hispanic and Native American</b>
University of Delaware (Faculty)	<b>1,172</b>	<b>40%</b>	<b>19% (222)</b>
Delaware State University (Faculty)	<b>211</b>	<b>40%</b>	<b>36% (76)</b>
Wesley College (Faculty)	<b>155</b>	<b>50%</b>	<b>10% (15)</b>
Delaware Tech (Faculty)	<b>1,149</b>	<b>63%</b>	<b>11% (128)</b>
<b>Total Delaware Faculty</b>	<b>2,687</b>	<b>50%</b>	<b>16% (441)</b>

\*\* STEM populations here include some human health-related fields.

#### **Increasing Institutional Engagement**

The Delaware EPSCoR program continuously seeks to increase engagement and collaborations among institutions of higher education, K-12, community colleges, government and non-profits. Table 3 lists the institutions participating in these programs.

Table 3: Institutions participating in the NECC Track-2 programs:

Higher Education Institutions	Non-Profit(s) Organizations
<ul style="list-style-type: none"> <li>Delaware State University</li> <li>University of Vermont</li> <li>University of Maine (Machias &amp; Orono)</li> <li>University of Rhode Island</li> <li>University of New Hampshire</li> <li>Dartmouth Medical School</li> <li>Johnson State College</li> <li>University of Southern Maine</li> <li>Johns Hopkins University</li> <li>Franklin and Marshall</li> <li>Millersville University</li> <li>Wesley College</li> </ul>	<ul style="list-style-type: none"> <li>Stroud Water Research Center</li> <li>Delaware Nature Society</li> <li>Partnership for the Delaware Estuary</li> <li>Delaware 4H</li> <li>Newark Center for Creative Learning</li> </ul>
Government Organizations	Others
<ul style="list-style-type: none"> <li>Delaware Department of Natural Resources and Environmental Control</li> <li>Delaware Economic Development Office</li> <li>National Science Foundation</li> <li>Delaware Department of Education</li> </ul>	<ul style="list-style-type: none"> <li>Delaware Geological Survey</li> <li>American Geological Institute</li> <li>Philadelphia Global Water Initiative</li> </ul>
High Schools	Community College(s)
Charter School of Wilmington	Delaware Technical and Community College

**Progress on Program Elements**

**Status**

The fiber network upgrades are fully installed at each of the Delaware institutions. In Maine, the network has reached Canada and is now connected to *Canada’s Advanced Research and Innovation Network (CANARIE)*. The Vermont fiber is complete to Albany and is completed to the New Hampshire border and linked into the New Hampshire network. In New Hampshire, the fiber build-out is complete from UNH to Maine/Boston. The Rhode Island fiber network is completed. Details of these networks are provided in the individual state progress reports and an overview schematic is shown in Figure 4.

**Leveraging NSF Programs**

The first face-to-face meeting among INBRE PIs and EPSCoR PDs was held in Vermont in 2008, leading to the establishment of NECC and subsequently to a successful, collaborative EPSCoR Track-2 award in January 2009 and simultaneous, complementary supplement requests to NIH-IDEa in Spring 2009. In addition, NSF-EPSCoR RII Track-2 C2 awards were made to Delaware and Rhode Island in September 2010. To date, NSF-EPSCoR has funded \$8.3M in cyberinfrastructure improvements to NECC states. Several other cyberinfrastructure

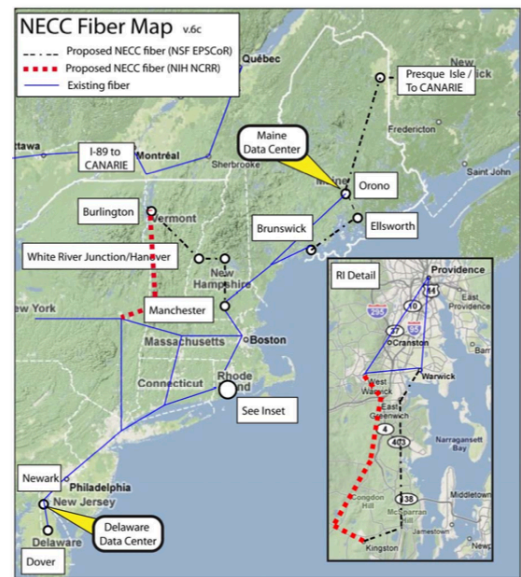


Fig 4: Fiber Upgrade Implementation Plan for NECC shows links across New England and into Canada.

grants have been secured by the partner states to leverage this significant NSF investment, as shown in Table 4.

Table 4: Leveraging of EPSCoR Track-2 Initiatives with other Funded Programs

Leveraging and Synergies with ARRA funding (in Millions)						
Agency	Program	Delaware	Maine	New Hampshire	Rhode Island	Vermont
NTIA	BTOP	10.9	25.4	65.9	21.7	47.1
USDA	RUS		1.3			116
Other Sources					12.3	

The NECC initiative is highly leveraged. Together, the NECC states have secured \$8.3 Million in funds from the NIH-IDEA program and \$8.4 Million in funds from NSF EPSCoR to develop the fiber network, carry out cyber-enabled research and education, and conduct other activities for broader impact and sustainability. Beyond these two funding sources, the other projects awarded to the NECC states have added value in particular to the fiber projects. These projects are described below.

In addition to the EPSCoR RII (5 years, \$15 Million from NSF plus \$5 Million from the State of Delaware, September 2008 to August 2013) and Track-2 funding (\$1,050,079) described in this report, Delaware has also received a \$1,176,470 award from NSF EPSCoR under the Track-2 C2 award mechanism (September 2009 to August 2012), overlapping with the performance dates for this Track-2 grant. This grant supports cyberinfrastructure upgrades at each of the four Delaware EPSCoR partner institutions – UD, DSU, Delaware Technical Community College and Wesley College. The PI of the Track-2 award, Dr. Karl V. Steiner, Senior Associate Provost for Research Development and Professor of Electrical and Computer Engineering, also serves as Co-PI on the EPSCoR RII grant and as Co-PI on the Track-2 C2 award, thus assuring close coordination among these EPSCoR cyber-related awards.

A related grant for the NECC initiatives in Delaware is supported by the NIH-IDEA program through the Delaware INBRE in the amount of \$979,810, with Dr. Karl Steiner as PI. This three-year award also started August 2009 and ends in August 2012. The NIH-INBRE supplement supports much of the sequencing of the Little Skate Genome, providing funds for supplies and consumables in UD’s *Sequencing and Genotyping Center (SGC)*, and its director, Bruce Kingham, to conduct the actual sequencing – see [www.dbi.udel.edu/core/dnasequencing.html](http://www.dbi.udel.edu/core/dnasequencing.html)).

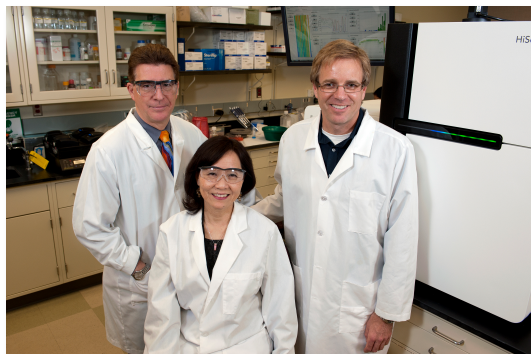
In the summer of 2010, Delaware received \$2.8 Million through the *Broadband Technologies Opportunities Program (BTOP)* program for the Delaware Job/Learning Lab Network. The funds include a \$750k grant by the Bill and Melinda Gates Foundation. The Foundation also provides \$150k to help Delaware secure additional federal E-rate funding to sustain broadband connection costs in the future. The project serves residents by providing public access computers, assistive technology, videoconferencing capability and wireless networking at Delaware’s libraries.

In the Fall of 2010, a UD team led by Dr. Blake Meyers, Edward F. and Elizabeth Goodman Rosenberg Professor and chair of the Department of Plant and Soil Sciences, was awarded an NSF *Major Research Instrumentation (MRI)* award to purchase a Pacific Biosystems (PacBio) RS DNA Sequencer. Funding in the amount of \$1.062 Million is provided by the NSF-MRI (\$745k) plus UD cost share. The PacBio allows for high-throughput long-read length sequencing of single DNA molecules and supports building *de-novo* genomic scaffolds of



genomes, resolving repeats and large genomic variations. The PacBio platform was installed in September 2011 and is now operational and heavily utilized by researchers in Delaware and beyond.

Over the past year, UD has provided funds to upgrade another genome sequencer - the Illumina GAllx platform - to the next generation HiSeq2000 platform. Funding in the amount of \$600,000 for this upgrade was provided by the EPSCoR RII and through UD cost share. With the upgrade, completed in Spring 2011, the HiSeq2000 system is now able to generate up to four times the number of reads at no extra cost per run.



*Fig 5: Karl Steiner, Cathy Wu and Bruce Kingham (from left) in front of the new Pacific Biosciences Sequencer in UD's Sequencing & Genotyping Center.*

With the successful installation of both the PacBio and the HiSeq2000 platforms, which provide critical complementary technologies, UD can now provide state-of-the-art genome sequencing and genotyping services to faculty across the state and to outside collaborators across the region. Only a limited number of laboratories across the Nation currently have access to both platforms, putting Delaware in elite company. The sequencing instruments will be used in unique ways, branching out into areas such as environmental and agricultural studies. Several recent large sequencing grants have both contributed and benefitted from these upgrades that require a strong cyberinfrastructure component.

Under the NECC, Kingham and the SGC serve as the core center for the sequencing of the Little Skate. This full-genome sequencing effort of the Little Skate was the first project of this scale conducted and led by the SGC and served as an important technology leader and proof-of-concept effort to further establish the center.

#### Complementary Grants

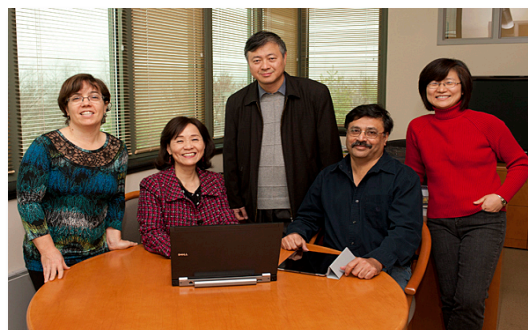
In Spring 2012, Dr. Eric Wommack, professor in UD's Department of Plant and Soil Sciences, was awarded a \$200k NSF EAGER Collaborative Research grant focusing on exploratory application of single-molecule real time (SMRT) DNA sequencing in microbial ecology research. Wommack will focus on three areas of research: 1) Single-cell genomics, in which he and his collaborator at the Bigelow Marine Laboratory in Maine will pull a single bacterial cell out of an environmental sample and sequence its genome; 2) Characterizing the composition of a microbial community using a single gene; and 3) Shotgun meta-genomics of viral communities.

Dr. Cathy Wu, Director of CBCB, is the PI of two new research grants awarded in late 2011, and aimed at improving computerized databases to store, organize and index bioinformatics data and to create specialized tools to view and analyze the data. The UD research team, including Drs. Vijay Shanker, Hongzhan Huang and Cecilia Arighi (all UD Computer Information Sciences) and Dr. Jung-Youn Lee (Plant and Soil Sciences), will develop a new bioinformatics research infrastructure to increase understanding of proteins. While genomes provide the genetic blue prints of organisms, the proteomes of all proteins serve as the key mechanical components of life's processes.

- Under a \$1.6 Million grant from the *Advances in Biological Informatics* program at NSF, the UD research team will establish a centralized resource the plant research community can use to answer important questions about post-translational modifications (PTMs) – the biochemical modifications of proteins, how they are brought about and what their impacts

are to functions of the plant cells and organisms.

- A separate \$3.2 Million renewal grant from NIH targets the development of biomedical ontologies, virtual libraries that organize biological knowledge using a universal language. Ontologies are increasingly important in systems biology research, where complex data need to be integrated and scientific data needs to be represented accurately.



*Fig 6: Cecilia Arighi, Cathy Wu, Hongzhan Huang, Vijay Shanker and Jung-Youn Lee (from left) team up for several major bioinformatics grants.*

**Cyberinfrastructure-enabled Science and Engineering Projects**

“The Metagenomes of Cyanobacterial Blooms” is a pilot project of the NECC designed to develop a baseline survey of species and genetic functional potential in toxic algal blooms that occur in fresh water lakes and ponds throughout the northeast. As part of the project, NECC members have developed the expertise, infrastructure and collaborations necessary to carry out other projects of this scope. The distributed nature and scale of the project requires the efforts of several groups of people across the NECC for sample collection, sample preparation, sequencing, data management and analysis. These regional collaborations would have been unlikely without existing NECC partnerships. NECC members currently participating in the project are given in Table 5.

*Table 5: Metagenome Project Collaborators*

<b>Water Sample Collection</b>		
Seabasticook Lake, ME Yawgoo & Trustom Ponds, RI Lake Winnepesaukee, NH Highgate Springs, VT	Kristin Ditzler, Jasmine Saros Linda Green Jeff Schloss, Jeff Haney Pat Pearson	University of Maine University of Rhode Island University of New Hampshire University of Vermont
<b>Laboratory Processing</b>		
	Tim Hunter James Vincent Scott Tighe Pat Pearson	University of Vermont University of Vermont University of Vermont University of Vermont
<b>DNA Sequencing</b>		
	Bruce Kingham	University of Delaware
<b>Bioinformatics and Data Management</b>		
	James Vincent Marc Farnum Rendino Chelsea Mitchell Colin Delaney Shawn Polson	University of Vermont University of Vermont & UD University of Vermont Saint Michael’s College, VT University of Delaware

Samples from five water bodies in four states were collected over a fourteen-week period in the summer and fall of 2010. Over 170 one-liter samples were collected and processed at the DNA Sequencing Facility at the University of Vermont (UVM). Processing included filtration of bulk

water and extraction of algae, algae counting by three methods and preparation for DNA sequencing at UD.

Workforce development progress includes a new full time position created in the UVM DNA Sequencing Facility and two full time summer internships in bioinformatics at the University of Vermont.

The *NECC Shared Data Center* (SDC) is established and has been developed primarily by Marc Farnum Rendino, who worked with engineers at both UD and the University of Maine to establish a redundant, live failover data center for sharing large data sets among NECC members and between NECC members and external collaborators. The SDC has been used to share and manage sequencing data from the Skate Genome project. It will play a key role in the management and sharing of DNA sequencing data from the metagenome project.

### **Cyber-enabled Research Awards**

*Regional Awards for Cyber-Enabled Research* (RACER) are a pilot-award mechanism to initiate collaborative research on a small scale. Projects parallel to the metagenomics pilot project, related to water-related environmental research, bioinformatics or use of shared data facilities are given priority. A pre-proposal stage is used to allow investigators to identify potential collaborations; proposals are required to involve a partnership between at least two jurisdictions. Merit review of proposals is conducted by a subset of the Water Research committee, with one representative from each jurisdiction. Vermont EPSCoR administers the proposal evaluation process.

In 2010, the first RACER award was presented to Dr. Thomas Kelly at the University of New Hampshire for his work on “An open-access, shared database resource for eukaryotic metagenetics research.”

In 2011, three full RACER proposals were submitted; two awards were made to a collaborative effort between investigators. Julia Daly’s, University of Maine, project is looking at developing a cyber-enabled Northeastern Monitored Lake Temperature (MeLT) Network and Benjamin King’s, Mount Desert Island Biological Laboratory, project is focused on discovery of microRNAs expressed in response to hypoxia in the estuarine fish, *Fundulus heteroclitus*. The projects use the NECC network and data centers to serve the needs of the research community to create a data repository and bioinformatics tools resource that addresses the unique challenges of metagenetics analysis.

### **Data Sharing**

Due to the high-level of interest in the skate genome and metagenomics sequence data and the nature of the NECC collaboration, Mr. King and Dr. Mattingly of MDIBL developed a data access policy to clarify how these data can be shared and made public. The policy was reviewed and approved by the NECC Executive Committee. The spirit of the policy is to enhance research opportunities of investigators within the five NECC states during the sequencing, assembly and annotation phases of the projects.

### **Diversity Plan**

Year 2 of the Track-2 award marked the launch of the Watershed Project, a collaborative effort among the five EPSCoR jurisdictions, to leverage our outreach and workforce development programs. The Watershed Project is modeled after the Vermont EPSCoR Streams Project, which engages high school teams and undergraduates in watershed research. High school teams from Delaware participated for a weeklong training at Saint Michael’s College in

Colchester, Vermont in 2010 and 2011. These teams then returned to their home states existing watershed education programs, where they integrated their training experience and monitored local streams. A total of seven undergraduate interns from Delaware participated in a weeklong training in the summers of 2010 and 2011 in Vermont. All undergraduate students then conducted an internship (summer of academic year) within their jurisdictions. Both the high school teams and the undergraduates joined students from VT, RI, PR, ME and NY, during their training weeks. The Watershed Project Spring Symposium, held in April 2012, marked the culmination of the program, where participants presented the results of their research through nine oral presentations and 30 poster presentations.

### ***Dissemination and Communication***

The communication and dissemination plan for the NECC in Years 2 and 3 included the following efforts:

#### ***Public Dissemination***

- Local print and web coverage extended beyond the NECC member states in Year 2 to coverage in the NIH-NCRR Reporter “Leveraged funds maximize the impact of biomedical and translational research at institutions across the country”.  
[http://www.ncrr.nih.gov/publications/ncrr\\_reporter/fall2010/recovery\\_act.asp?p=all](http://www.ncrr.nih.gov/publications/ncrr_reporter/fall2010/recovery_act.asp?p=all)
  - The NECC was one of three awards featured in the article with interviews from Karl Steiner (DE) and Judith Van Houten (VT)
- VT hosted a press conference on October 28, 2010 with U.S. Senator Patrick Leahy. Coverage followed in print and broadcast media including the Associated Press, The Dartmouth (College) Pager, The Burlington Free Press, MSNBC, local television affiliates for NBC, ABC, CBS and Fox
  - Timing of press conference coincided with the completion of one leg of the fiber project in Vermont
- All media coverage is available on the VT NECC web site at  
<http://www.uvm.edu/~epscor/necc/>

NECC Leadership gave numerous public presentations about the plans for and progress of the programs. These events include:

- October 4, 2010: Karl Steiner and Judith Van Houten presented progress on the network and Ben King presented the progress on the little skate genome at the NIH INBRE annual meeting in Bethesda, MD
- October 8, 2010: Karl Steiner, Judith Van Houten and Jeff Letourneau presented the fiber progress and cyber-enabled NECC research at the NSF EPSCoR Cyberinfrastructure Workshop in Arlington Virginia
- December 7, 2010: Judith Van Houten and Karl Steiner presented an overview of the NECC project to the EPSCoR/IDeA Foundation and Coalition meeting in Washington, DC
- February 9, 2011: Kelvin Chu and Judith Van Houten presented an overview of NECC to representatives of Sandia National Lab;
- August 12, 2011, NECC members organized a half-day workshop as part of the third Northeast Regional IDeA Conference, held at Salve Regina University in Newport, RI.

#### ***NECC Website***

The Rhode Island EPSCoR office assumed responsibility for the NECC website in January of 2011. The site is now hosted on a server within the Rhode Island EPSCoR office. The redesign included taking the structure of the former website and implementing it more efficiently, a logo

upgrade, and creating a page explaining how to use the wiki. The original site was archived and taken down in 2011.

The new partner's page includes logos from all the INBRE and EPSCoR partners in the NECC that links to their individual websites. The other pages are all individual wiki pages for news, information, committees, projects, meetings and solicitations. These pages are readable by anyone but only those with credentials can login to the edit the pages. Anyone can also comment on the posts on the website. Login access can be requested through email from Kate Wilson at [kate@riepsc.org](mailto:kate@riepsc.org). The NECC website is available under [necyberconsortium.org/](http://necyberconsortium.org/).

There are varying ways the site has been used. News articles have been posted involving the NECC such as the Burlington Press report about the most recent Water Symposium. Maine EPSCoR used the site to link to their coverage of the Second Annual Meeting. Delaware listed the presentations from the Third NECC Workshop held at UD on in March 2012. When selected, the Maine EPSCoR site is embedded within the NECC site recognizing the credit of both organizations. Steve Exler used the solicitations page to link the UVM hosted application. This page also is embedded with the NECC site as opposed to taking the visitor from the site.

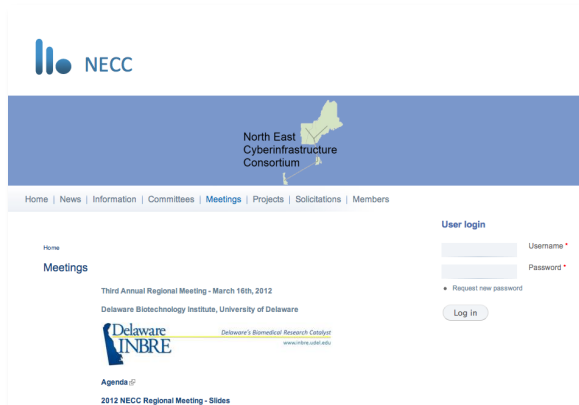


Fig 7: The NECC Web Page provides access to presentations at prior workshops.

Future plans for the site include stimulating the collaboration among the NECC members through using the wikis. An example would be to post committees' resources such a minutes and reports. The wiki can be enabled to require a login to see certain pages if everything should not be public.

### **Evaluation and Assessment Plan**

Because funding for the NECC project is an admixture of funds from NIH-IDEA and NSF EPSCoR funds, assessment and evaluation for the five NECC jurisdictions is conducted by a combination of INBRE and EPSCoR offices. This flexibility allows NECC states to increase efficiency, avoids duplication of efforts and allows for efficient use of existing federal funds.

- Delaware: external reviews of the cyberinfrastructure efforts have been included as part of the regular INBRE and RII programs. On the INBRE side, the most recent External Advisory Board evaluation occurred April 20, 2012 and the INBRE AAAS review in February 8, 2012. On the EPSCoR side, the RII EAB review occurred March 22, 2012.
- Maine: evaluation via the INBRE External Advisory Committee occurred in August 2011 and by the AAAS evaluation in Fall 2011. The EPSCoR Track-2 evaluation occurred as part of Maine's Track-1 process in May 2011.
- New Hampshire: assessment of the fiber transmission for the southern route occurred in May 2011.
- Rhode Island: the initial INBRE supplement assessment occurred in 2010; the next review occurred in 2012.
- Vermont: the AAAS panel in June 2011 provided formative assessment of the entire NECC program as part of the Vermont INBRE. In addition, Vermont EPSCoR worked with Dr. Joy Livingston, Flint Springs Associates, to perform formative assessment for participants in the Annotation Jamborees.

### **Sustainability Plan**

At the Third Annual NECC Meeting at UD, we discussed the future of NECC and ways to sustain it as an organization. Grant opportunities through a future Track-2 and DOE proposals were discussed at some length. The NECC executive committee has continued this conversation and planning over the summer of 2012. Unfortunately, we have learned that neither NSF-EPSCoR nor NIH-IDEA will be able to provide supplemental funding opportunities for NECC projects in FY 2012. The discussion with EPSCoR and IDEA leadership continues for FY 2013.

Part of our approach to sustainability is to foster cyber-enabled research through collaborative research project funded through pilot funds, call RACERs, as described in more detail above.

### **Management and Coordination Plan**

The NECC Executive Committee continues to meet through videoconferencing on a regular basis. The Executive Committee has held twelve videoconferences over the past two years. These conferences helped us update each other on progress, address issues and welcome new members to the committee. We took recommendations from the Water Research Committee on RACER awards and the selection of RACER awardees; approved Year 2 RFP. The January 6 videoconference included Dr. Jennifer Schopf, who discussed the Annual Progress Report for the NSF Track-2 grant.

The other standing committees meet regularly as well, including the Technical Committee, Water Research Committee and NE Bioinformatics Collaborative (NEBC). The Technical Committee has been considering the need for Shibboleth for the NECC in order to share data files with expediency and security. The Water Research Committee reviewed and recommended a RACER award in Year 2 and two awards in Year 3.

### **Unobligated Funds**

As of the end of August 2012, all funds related to the Track-2 project have been spent.

**References:**

1. Schneider, I., I. Aneas, A.R. Gehrke, R.D. Dahn, M.A. Nobrega, and N.H. Shubin, Appendage expression driven by the Hoxd Global Control Region is an ancient gnathostome feature. *Proceedings of National Academy of Sciences, USA*, 2011. 108(31): p. 12782-6. PMID:PMC3150877
2. King, B.L., J.A. Gillis, H.R. Carlisle, and R.D. Dahn, A natural deletion of the HoxC cluster in elasmobranch fishes. *Science*, 2011. 334(6062): p. 1517. PMID:PMC3264428
3. Wang, Q., C.N. Arighi, B.L. King, S.W. Polson, J. Vincent, C. Chen, H. Huang, B.F. Kingham, S.T. Page, M.F. Rendino, W.K. Thomas, D.W. Udway, C.H. Wu, and North East Bioinformatics Collaborative Curation Team. Community annotation and bioinformatics workforce development in concert - Little Skate Genome Annotation Workshops and Jamborees. *Database (Oxford)*, 2012. 2012: p. bar064. PMID:PMC3308154

**APPENDIX 1:**

**NECC Year 3 Annual Meeting**

**Hosted by Delaware EPSCoR**

**March 16, 2012**

**Delaware Biotechnology Institute**

**University of Delaware, Newark, DE**



**2012 Annual NECC Meeting  
Friday, March 16, 2012**



**Delaware Biotechnology Institute, University of Delaware**

**7:30 am Registration and Continental Breakfast**

**8:00 am Welcome and Introductions**

*Karl Steiner, DE INBRE PI and EPSCoR Co-PI  
Judith Van Houten, VT EPSCoR PD and INBRE PI  
Siân Mooney, Program Director, NSF EPSCoR*

**8:15 am NECC Cyber Projects: Fiber and Data Centers  
(Panel Session – 5-minute presentations each)**

*New Hampshire Fiber Progress: Scott Valcourt, NH  
Maine Fiber Progress and Data Center: Jeff Letourneau, ME  
Rhode Island Fiber Progress and NECC Website: David Porter, RI  
Vermont Fiber Progress: Kelvin Chu, VT  
Delaware Cyber Progress and Data Center: Karl Steiner, DE*

**8:45 am NECC Cyber-enabled Collaborative Research Programs  
(10-minute presentations with 5 minutes Q&A)**

*NECC Little Skate Genome Project: Ben King, ME  
NECC Metagenome Project: James Vincent, VT  
NECC Sequencing Resources: Bruce Kingham, DE  
NECC RACER Awards: Ben King, ME and Sudarshan Chawathe, ME  
Cyber-enabled Environmental Sensor Research: Jennifer Specker, RI  
Delaware Environmental Monitoring and Analysis Center: Dan Leathers, DE*

**10:15 am Break and Poster Session**

**10:30 am Workforce Development and Diversity  
(10-minute presentations each)**

*NECC Little Skate Annotation Jamborees: Cathy Wu, DE  
NECC Watershed Project: Miranda Lescaze, VT; Amy Broadhurst, DE;  
Michelle Gregoire, NH; Vicky Nemeth, ME*

**10:50 am NECC Committee Meetings (Breakout Session # 1)**

- Northeast Bioinformatics Consortium (NEBC)
- Cyberinfrastructure Upgrades - Generation 2
- Environmental Sensing and Analysis
- Other Research & Education Collaboration Opportunities

**12:00 Lunch, Networking, Poster Session**

Poster Session in DBI Lobby

**12:45 pm Tour of Delaware Biotechnology Institute Facilities**

Center for Bioinformatics and Computational Biology, *Shawn Polson, DE*

Sequencing and Genotyping Center, *Bruce Kingham, DE*

Bioimaging Center, *Kirk Czymmek, DE*

**1:15 pm Feedback from Breakout Groups**

**1:45pm NECC Committee Meetings (Breakout Session #2)**

- Northeast Bioinformatics Consortium (NEBC)
- Cyberinfrastructure Upgrades - Generation 2
- Environmental Sensing and Analysis
- Other Research & Education Collaboration Opportunities

**2:45pm Break and Poster Session**

**3:00 pm NECC Evaluation and Assessment:**

Executive Committee Report: *Kelvin Chu, VT*

Delaware NECC Evaluation Plans: *Leslie Cooksy, DE*

**3:30 pm Planning Ahead – The Next Steps**

Assessment Discussion

Key Accomplishments under NECC

Plans for Additional Collaborative Research Projects

**4:30 pm Meeting Adjourned**

**2012 Annual NECC Meeting Attendees  
March 16, 2012**

Last Name	First Name	Title/Position	Institution/Organization
Broadhurst	Amy	Assistant Director/Delaware EPSCoR PA	University of Delaware, Delaware Environmental Institute
Callahan	Christina	Environmental Informaticist, DEOS	University of Delaware
Chawathe	Sudarshan	Associate Professor, Computer Sciences	University of Maine
Chu	Kelvin	VT EPSCoR Associate Project Director	University of Vermont
Cousins	Stephen	Supercomputer Engineer/Administrator	University of Maine
Gamache	Lillian	VT EPSCoR Project Coordinator	University of Vermont
Gregoire	Michelle	Program Manager	New Hampshire EPSCoR
Grim	Daniel	Executive Director, IT, Networking, Systems Services	University of Delaware
Hand	Patricia	PI, Maine INBRE	Mount Desert Island Biological Laboratory
Jacobson	Carl	Vice President, Information Technologies	University of Delaware
King	Benjamin	Staff Scientist	Mount Desert Island Biological Laboratory
Kingham	Bruce	Director, Sequencing and Genotyping Core	University of Delaware
Leathers	Daniel	Director, Delaware Earth Observing System	University of Delaware
Lescaze	Miranda	VT EPSCoR CWDD and Streams Project Director	University of Vermont
Letourneau	Jeff	Executive Director, Networkmaine	University of Maine System
Meacham	Steven	Program Director	National Science Foundation
Mooney	Sian	Program Director	National Science Foundation
Nemeth	Vicki	Director of Research Administration & Maine EPSCoR	University of Maine
Polson	Shawn	Assistant Professor/Bioinformatics Core Coord.	University of Delaware
Porter	David	Director, Media & Technology	University of Rhode Island

		Services	
Sacher	Richard	Associate Director, IT-Client Support & Services	University of Delaware
Sine	Patricia	Director, IT-Client Support & Services	University of Delaware
Segee	Bruce	Technical Director, Maine Supercomputer	University of Maine
Specker	Jennifer	Associate Project Director, Rhode Island NSF	University of Rhode Island
Steiner	Karl	Senior Associate Provost for Research Development	University of Delaware
Thomas	Kelley	Hubbard Professor in Genomics and Director Hubbard Center for Genomics Studies	University of New Hampshire
Todd	David	Associate Vice President/CIO	University of Vermont
Valcourt	Scott	Director, Project Management & Consulting Services	University of New Hampshire
Van Houten	Judith	VT EPSCoR Project Director	University of Vermont
Vincent	James	Director, Bioinformatics Core, Vermont Genetics Network	University of Vermont
Wilson	Kate	Information Technologist	University of Rhode Island
Wu	Cathy	Director, Center for Bioinformatics and Computational Biology and	University of Delaware

# NECC Annual Meeting

University of Delaware

Delaware Biotechnology Institute

March 15/16, 2012

**North  
East  
Cyberinfrastructure  
Consortium**



<b>Evaluation Sheet (35 Total)</b>	<b>Very Satisfied</b>	<b>Satisfied</b>	<b>Dissatisfied</b>
Please rate your overall satisfaction with the NECC Meeting	<b>16</b>	<b>8</b>	<b>0</b>
Please rate your overall satisfaction with the format of the NECC Meeting (i.e.- presentations and break-out sessions, tours)	<b>16</b>	<b>8</b>	<b>0</b>
Please rate your satisfaction with the Session on Cyber Projects: Fiber and Data Centers	<b>15</b>	<b>9</b>	<b>0</b>
Please rate your satisfaction with the Session on Collaborative Research Programs	<b>20</b>	<b>4</b>	<b>0</b>
Please rate your satisfaction with the Session on Workforce Development and Diversity	<b>19</b>	<b>5</b>	<b>0</b>
Please rate your satisfaction with the Tour of the Delaware Biotechnology Institute	<b>22</b>	<b>2</b>	<b>0</b>
Please rate your satisfaction with the Session on Evaluation and Assessment	<b>14</b>	<b>9</b>	<b>1</b>
Please rate your satisfaction with the Session on Next Steps	<b>10</b>	<b>12</b>	<b>2</b>
Please rate your overall satisfaction with the location of the NECC Meeting (UD - Delaware Biotechnology Institute)	<b>18</b>	<b>6</b>	<b>0</b>
Please rate your overall satisfaction with the Accommodations for the NECC Meeting	<b>14</b>	<b>10</b>	<b>0</b>

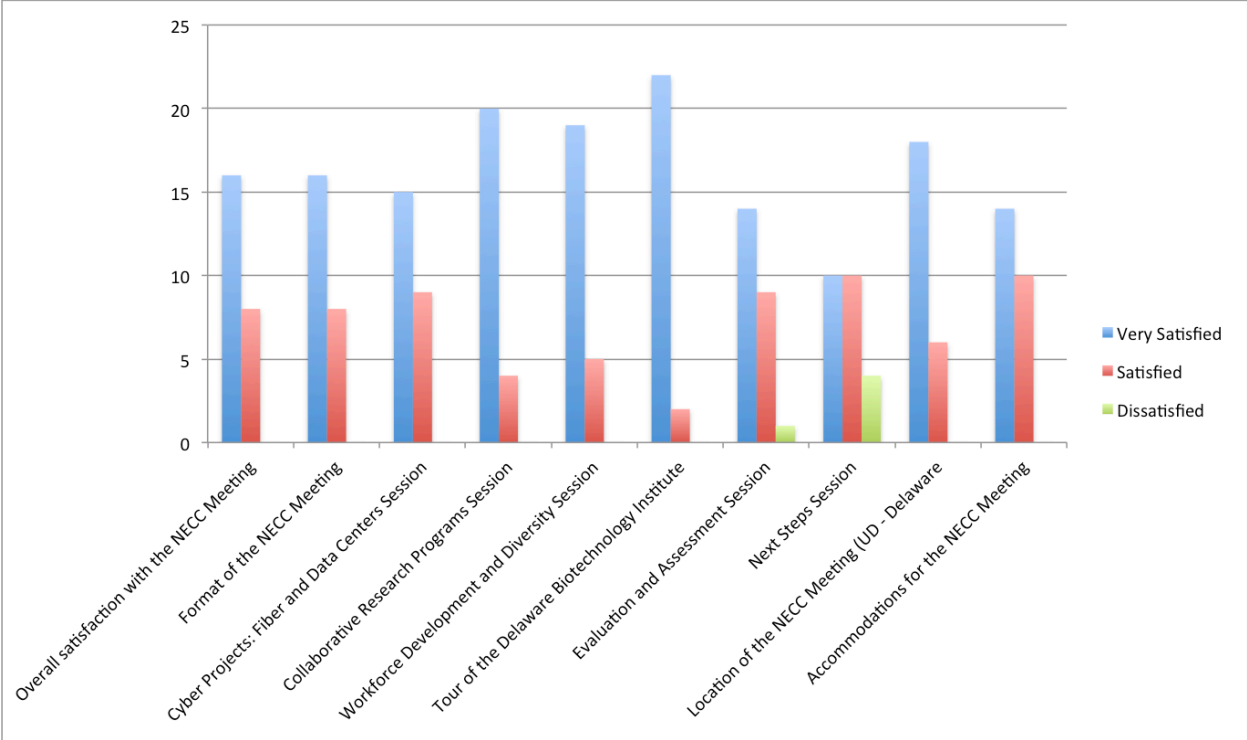


Fig 8: Graphical results of 2012 3<sup>rd</sup> Annual NECC Meeting evaluation indicate a successful event.