

Full Project Detail with Vermont Specific Details Included

Collaborative Research: North East Cyberinfrastructure Consortium NSF EPS# 0918284
Final Reporting period – September 1, 2011 – August 31, 2012

Report Sections:

Project Summary -

The North East Cyberinfrastructure Consortium has finished its third year of Track-2 funding. In this report we summarize our overall progress and progress for Year 3.

In 2006, we began to organize as the five North Eastern EPSCoR states (ME, NH, VT, RI, DE) around cyberinfrastructure. The box below describes the state of cyberinfrastructure in 2008 by which time we had developed the North East Cyberinfrastructure Consortium to position ourselves for grant opportunities that would help us to address our cyber deficits.

Summary of NECC - Regional Cyberinfrastructure Landscape and Barriers

- Compute resources seem adequate but network connectivity and cyberinfrastructure personnel are lacking.
- **All states have important science drivers that depend upon connectivity, but regional collaborations in areas of shared strength are lacking. The areas of highest potential for regional efforts are Marine and Watershed Environmental Research, and Bioinformatics.** (We describe a plan for catalyzing regional collaborations for these fields in Section III.)
- Visualization is well developed in Delaware, New Hampshire and Maine. The growing volume of data from simulations, computations and observatories is projected to outstrip the capacity of individual institutions in the region.
- Connectivity **is poor between the four New England consortium members and within Delaware.** ME, NH, RI and VT rely upon commodity, sub-gigabit services that are susceptible to single-point failures. Smaller institutions within DE (Delaware State University and Wesley College) need to establish redundant connection to regional networks. All consortium members need to strengthen the capacity and capability of cyber-knowledgeable faculty and staff. The combination of poor connectivity and personnel provides barriers to accessing national HPC resources and limits the potential of virtual organizations and collaborators.
- All consortium members are trying to address cyberinfrastructure needs by heavily leveraging resources through other grant sources such as IDeA funds and state funds.

The Track-2 collaborative proposal submitted in January 2009 was designed to address these barriers in order enable our researchers to access and share data and compute resources that will make them more competitive and allow their research to have a broader impact on society.

Summary of our goals:

- Provide cyberinfrastructure for research and education in the North East EPSCoR region by creating fiber networks within and between jurisdictions;
- Establish regional distributed data centers to support cyber-enabled research
- Develop human infrastructure to create a virtual organization for distributed bioinformatics and data analysis;
- Launch pilot cyber-enabled research projects to develop the regional expertise to analyze genomics data, especially deep sequencing data, in a distributed manner;
- Encourage pilot projects to foster collaborations around water research across the region for future collaborations;
- Establish collaborations on cyber education in order to foster cyber-knowledgeable workforce development, diversity and outreach;

- Leverage resources to accomplish our goals

General progress over three years:

Intellectual Merit:

- The fiber installation and upgrade projects are finalized in four of the five states (VT, ME, RI and DE); the NH installation is close to completion
- Data centers are now functional in Maine and Delaware
- We carried out collaborative projects in metagenomics of toxic algal blooms in the region and the little skate genome to build the virtual organization that allows us to analyze data remotely in a distributed and highly collaborative manner – learning and training opportunities
- Pilot project awards were made to inspire novel collaborations on water and cyber-enabled research in the NECC region
- NECC fosters novel collaborations that would not have happened without NECC and its cyberinfrastructure

Broader Impact:

- We collaborated on the Watershed Project to leverage the water education outreach programs in the Region for workforce development and diversity
- We trained many students and faculty in genome annotation using the Little Skate genome that was part of our research
- We created cybertools for educational use in training in bioinformatics and genome annotation in particular
- We have examples of small businesses retained because of the new fiber and promise of fiber

A sample of novel aspects (what we can do now that we could not do before):

Cyber-enabled research collaborations that require NECC resources.

1. The Open Access shared database resource for eukaryotic metagenetics research (NH, VT)
2. Virtual Organization for bioinformatics and data analysis (NEBC, 5 states)
3. EOS Webster – GIS and positional information overlaid on field sample data for water research (NH and VT)
4. Five state metagenome project on algal blooms that impact the state economies and that require the coordinated efforts of bioinformaticians working remotely – work flow, data storage, data management
5. Watershed Watch and Watershed projects collaborate in RI and NECC at large
6. A new NHCORE grant is developing a HPC network among the five states for work on large quantitative biological data sets; this network would not be possible without the NECC fiber network
7. In Vermont, a web based company chose to remain in Vermont because of the bandwidth now available from the vendor that installed the fiber network to Albany for NECC; other businesses now use the commodity fiber that was installed parallel to our NECC fiber

Intellectual Merit for Three Years of Progress-

The Little Skate Genome

The genome of *Leucoraja erinacea*, the Little Skate, is being sequenced in a collaborative effort among NECC members. 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.

The lack of detailed genomic information has held back research on the Little Skate. The only reported chondrichthyan genome sequence is a very low coverage draft of the non-elasmobranch elephant shark. Therefore, the NECC's completion of the Little Skate genome provides 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 draft assembly from NECC and underlying DNA sequences were instrumental in discoveries published in *Science* and *Proc. Natl. Acad. Sci.* (King et al., 2011a,b). Randall Dahn and collaborators at Mount Desert Island Biological Laboratory reported genomic deletion of the entire HoxC cluster in the Little Skate, (*Leucoraja erinacea*). Additional data suggested loss of the HoxC cluster in elasmobranch fishes and evidence of the deletion of a Hox cluster in vertebrates. The authors used the draft assembly of the Little Skate to confirm genomic deletion of the HoxC cluster rather than transcriptional silencing. Although the assembly is ongoing, the immediate utility of the data bolsters the ongoing efforts of the Little Skate Genome project.

Annotation of draft genomes is an integral part of genome sequencing projects. It is often a bottleneck to wide spread use of the genome due to the labor intensive nature of the work. In anticipation of this, the North East Bioinformatics Collaborative (NEBC, part of the NECC) implemented three Little Skate Genome Annotation Jamborees. These week-long, hands on intensive workshops were designed to provide the skills necessary to annotate a genome. The culmination of this effort was a five state contest in which former workshop participants were asked to independently annotate the mitochondrial genome of the Little Skate. This highly successful project resulted in near unanimous annotations, confirming the training received by all participants. This work resulted in a publication linking the workshops, workforce development and the concerted regional collaborative research effort behind them (Wang et al., 2012).

King BL, Gillis JA, Carlisle HR, Dahn RD. A natural deletion of the HoxC cluster in elasmobranch fishes. *Science*. 2011 Dec 16;334(6062):1517. PMID: 22174244

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

Wang Q, Arighi CN, King BL, Polson SW, Vincent J, Chen C, Huang H, Kingham BF, Page ST, Rendino MF, Thomas WK, Udway DW, Wu CH; North East Bioinformatics Collaborative Curation Team. Community annotation and bioinformatics workforce development in concert--Little Skate Genome Annotation Workshops and Jamborees. *Database (Oxford)*. 2012 Mar 20;2012:bar064. Print 2012. PMID: 22434832

Metagenomes of Cyanobacterial Blooms

Lakes, estuaries and bays in VT, NH, ME and RI are plagued by algae blooms composed of cyanobacteria species and, frequently, the toxins they produce. These effects of 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 micro-organisms 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 provide a baseline survey of algal blooms over time from five water bodies in the North East. 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.

Lakes subject to algal blooms were sampled from five geographic locations across the NECC over a fourteen week period in the summer of 2010. Samples were shipped to Vermont where they were analyzed for Microcystin levels by ELISA. 16S amplicons were sequenced from three time points in all locations using Roche 454 pyrosequencing. Additionally, three time points from Lake Champlain at Highgate Springs, Vermont were subjected to full metagenome sequencing on an Illumina HiSeq instrument resulting in 2.9B paired-end reads. A manuscript describing this metagenomic study is close to submission. Taxonomic profiling analysis of 16S amplicons and functional profiling of the Highgate Springs samples is still ongoing.

Broader Impacts for Three Years of Progress- Little Skate Annotation Workshops and Jamborees

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.

- 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>), and 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.



Fig 2: The May 2011 NECC Annotation Workshop at UD

The Genome Annotation workshop is intended to inspire a new generation of bioinformaticians. The May 2010 workshop attracted 35 attendees to UD – 16 males and 19 females. The October 2010 workshop held at MDIBL attracted 25 with 14 males and 11 females. The May 2011 workshop held at UD attracted 41 attendees with 29 males and 12 females.

- The mitochondrial annotation workshop

There were 29 participants total (17 male; 12 female) in the curation team:

DE: Daniel Nasko, Chandran Sabanayagam, Liang Sun Yue Wang (University of Delaware)

ME: Jacob Berninger, Stevey Mahar, Eric Tan, John J. Wilson (University of Maine at Machias) Vanessa Coats (University of Maine); Clare Bates Congdon, Jeffrey Ahearn Thompson, David J. Gagne (University of Southern Maine)

RI: Jimmy Adediran, Thomas Bregnard, Alison C Cleary, Scott Grandpre, Bethany Jenkins, Lauren Killea, Bradford Lefoley, Katherine Mccusker, Matthew Mokszycki, Megan O'Brien, J. Christopher Oceau, Steven Shelales, Edward Spinard, Jacob Stupalski, Linh Tran, Joselynn Wallace (University of Rhode Island)

VT: Brian Cunniff (University of Vermont)

- An additional series of virtually distributed Annotation Jamborees involved 29 trainees in annotating 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," co-authored by all Jamboree participants, as well as a presentation at the April 2012 *International Biocuration Conference* in Washington, DC.

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. Udvary, 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.
PMCID:PMC3308154

Cybertools

An additional aspect of the Little Skate genome sequencing project was the construction of cybertools to enable progress on this and future research collaborations between the NECC states. In any such collaboration between distant institutions communication is the key to success. The members of the Little Skate project utilized regular weekly meetings by multi-point videoconference (Polycom) between the various participants. This provided an additional connectivity among the researchers that is often lacking with less personal forms of communication (e.g. email, telephone), allowing them to function as a team. It proved particularly useful in planning of the Little Skate genome workshops.

The project also necessitated the establishment of tools to allow for data sharing and analysis between NECC partners. In order to facilitate early activities, a number of tools leveraged the Amazon S3/EC2 cloud storage/computing infrastructure. These tools included an interface for file sharing, genome homology searches (BLAST), and cross-referencing of genomic and transcriptomic sequence identities. With the establishment of the shared data center (SDC) at the University of Delaware (with live backup at the University of Maine), these tools have begun to migrate to that permanent hardware. Currently, the data sharing tools have completely relocated to the dedicated storage server at the SDC. BLAST server capabilities have also transitioned to the SDC and are residing on a repurposed six-node cluster provided by the University of Delaware. This BLAST cluster supported the annotation activities of the third skate workshop and will continue to evolve with this and future projects.

Programmatic Terms and Conditions –

Individual state’s Programmatic Terms and Conditions (PTC)

1) Broadening Participation over 3 Years

Vermont Specific NECC Participant Table:

	Number	% women	% URM
Total	43	54	15
Faculty	8	50	0
PostDoc	0	0	0
Graduate Students	5	40	0
Undergrads	30	57	21
HS Teachers	15	47	13
HS Students	30	63	13

**2) Institutional Engagement
NECC Wide Over Three Years**

Delaware

University of Delaware
Delaware State University
Delaware Technical Community College
Wesley College

Rhode Island

University of Rhode Island
Brown University

Maine

University of Maine
University of Southern Maine
University of Augusta
University of Maine Farmington
University of Maine Fort Kent
University of Maine Machias
University of Maine Presque Isle
Bates College
Bowdoin College
Colby College
College of the Atlantic
University of New England
Unity College

New Hampshire

University of New Hampshire
Dartmouth College
Plymouth State University

Vermont

University of Vermont
Johnson State College

Green Mountain College
Johnson State College
Lyndon State College
Norwich University
Saint Michael's College
Universidad Metropolitana, PR
University of Puerto Rico, PR

Vermont Specific Over Three Years

The VT EPSCoR Watershed Project hosted HS teams from ME, RI and DE at the Watershed Project Training Week, June 27, 2011 at **St. Michael's College** with Science Leader, Dr. Declan McCabe.

- The participating High Schools for the 2010-2011 session were:
 - Bellows Free Academy, St. Albans VT
 - Colchester High School, Chittenden VT
 - DeWitt Clinton High School, Bronx NY
 - Jose E. Aponte De La Torre School, Carolina, PR
 - Long Trail School, Bennington VT
 - Lyndon Institute, Lamoille VT
 - Milton High School, Franklin VT
 - Mt. Anthony Union High School, Bennington VT
 - North Country Union High School, Orleans VT
 - Oxbow High School, Orange VT
 - Rice Memorial High School, Chittenden VT
 - Twin Valley High School, Windham VT
 - University Gardens High School, San Juan PR
 - Whitcomb Jr/Sr High School, Windsor VT

- The participating baccalaureate institutions for the 2010-2011 session were:
 - Green Mountain College, VT
 - Johnson State College, VT
 - Lyndon State College, VT
 - Norwich University, VT
 - Saint Michael's College, VT
 - Universidad Metropolitana, PR
 - University of Puerto Rico, PR
 - University of Vermont, VT

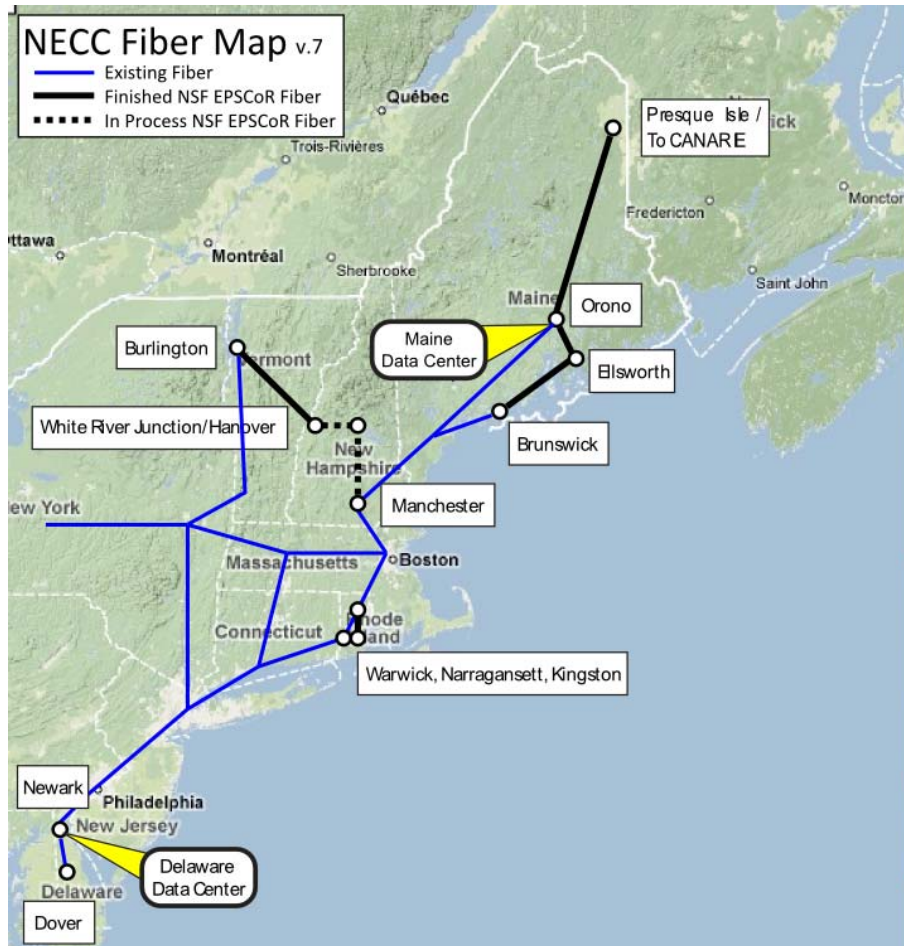
3) Progress on Program Elements Year 3 (NECC Wide): Status of Network Connectivity–

The upgraded fiber network is in place in Delaware. In Maine, the network has reached Canada and is connected to CANARIE. The Vermont fiber is complete to Albany and the New Hampshire border. In New Hampshire, the fiber build out is underway and is complete from UNH to Maine/Boston. The Rhode Island fiber network is complete. Details of these networks will be given in the individual state progress reports.

Of special Year 3 note: Maine Fiber

Company held a celebratory event to mark their completion of the Three Ring Binder build out.

Below is a link to local television coverage <http://www.wlbz2.com/news/article/216796/3/Three-Ring-Binder-up-and-running-in-Maine>



Leveraged opportunities and activities for NECC

The NECC has been designed to be highly leveraged. The first face-to-face meeting among INBRE PIs and EPSCoR PDs was held in Vermont in 2008, leading to a collaborative EPSCoR Track-2 proposal in January 2009 and simultaneous, complementary supplement requests to NIH NCRR in Spring 2009. In addition, RII C2 awards were made to Delaware and Rhode Island in September 2010 and to Vermont and Maine in September 2011. To date, NSF EPSCoR has funded \$10.75M in cyberinfrastructure improvements to NECC states in Track-2 and C2 awards.

There are \$8.6M in funds from the NIH NCRR and \$8.4M in funds from Track-2 to develop the fiber network, carry out cyber-enabled research and education, and other activities for broader impact and sustainability. Beyond these two funding sources, the stimulus funds to the NECC states have added value to the fiber projects in particular. These funds are described in the report from each state.

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	51.7
USDA	RUS		1.3			122.4
Other					12.3	

Vermont specific connectivity and leveraging - UVM finished installation of 12 10G Ethernet channels between Burlington, VT, and Albany, NY, in January, 2012, and is awaiting availability of 10G Internet2 service in Albany to make that connection. We are participating in discussions with UNH, UME, NEREN and Internet2 for the provision of that service. In the meantime, one of the 10G channels is in use to complete a 1G connection from Albany to New York City that provides redundant broadband connectivity for the UVM campus.

Fiber connectivity between UVM and Dartmouth, the transfer point for connections to UNH, UME, and Boston, was established in September, 2012. Final testing of the 12 10G Ethernet channels from Burlington, VT, to Hanover, NH, will be completed in mid-October, 2012, and one of the channels will be used as a 10G Ethernet connection from UVM to Dartmouth. The connection to UNH and University of Maine will await the completion of the USNH fiber link between Manchester and Hanover, likely in the summer of 2013. As a result of the CI investments in Vermont, connectivity for the flagship research university in the state, the University of Vermont (UVM), has improved dramatically. Originally connected via a combination of commodity and I2 services at 345 Mbps, UVM now has the capacity for multiple 10 Gbps connections to both Albany, NY, and Boston, MA, and is positioned for connectivity to U.S. UCAN (<http://www.usucan.org>).

Completion of the Albany leg was announced on 28 October 2010 with a press release that included UVM President Daniel Fogel and Senator Patrick Leahy (www.uvm.edu/~epscor/index.php?EventID=27). These efforts from the NSF and NIH complement NTIA and USDA broadband initiatives: to date, Vermont has been awarded \$51.7M in NTIA funds (<http://www2.ntia.doc.gov/vermont>) and \$122.4M in RUS funds (http://www.rurdev.usda.gov/supportdocuments/RBBreport_V5ForWeb.pdf).

Vermont joined the North East Research and Education Network (NEREN) in 2010 and has been a full, participating member since then. UVM and Vermont State Colleges (VSC) are represented on the NEREN Board, and UVM and VSC network architects participate actively in NEREN technology planning.

Vermont is also leveraging a C2 award (EPS 1107945) that connects the Vermont State College network to the new North East fiber network through UVM via a redundant 10G connection. In addition, the C2 award supports a cyber education specialist to help schools, museums, libraries and other educational institutions in the state connect to a newly established statewide High-speed research and education network, integrate existing Internet2 collaborative projects into their educational programs, and create their own broadband-enabled collaborations within the state. Finally, the C2 award provides funds for membership in the Internet2 Sponsored Educational Group Participant (SEGP) program for the statewide research and education network and the University of Vermont will continue to support this membership at the end of the C2 award.

Region-wide Cyberinfrastructure-enabled science and engineering projects – Progress Over 3 Years

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 North East. 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 1.

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. Processing included filtration of bulk water and extraction of algae, algae counting by three methods and preparation for DNA sequencing at the University of Delaware.

The NECC Shared Data Centers (SDCs) are established and have been in production for over a year. The lead developer, Marc Farnum Rendino, has worked with engineers at the University of Delaware 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.

Metagenome Project Collaborators

Water Sample Collection	Sebasticook Lake, ME	Kristin Ditzler (UME) Jasmine Saros (UME)
	Yawgoo & Trustom Ponds, RI	Linda Green (URI)
	Lake Winnepesaukee, NH	Jeff Scholoss (UNH) Jeff Haney (UNH)
	Highgate Springs, VT	Pat Pearson (UVM)
Laboratory Processing		Tim Hunter (UVM) Scott Tighe (UVM) Pat Pearson (UVM)
DNA Sequencing		Bruce Kingham (UD)
Bioinformatics & Data Management		James Vincent (UVM) Marc Farnham Rendino (UVM & UD) Chelsea Mitchell (UVM) Colin Delaney (SMC) Shawn Polson (UD)

UME = University of Maine; URI = University of Rhode Island; UNH = University of New Hampshire; UVM = University of Vermont; UD = University of Delaware; SMC = Saint Michael's College, Vermont

The Little Skate Genome Project:

The Little Skate and other elasmobranch fishes, such as the dogfish shark, have been informative model organisms for a variety of research fields. Of particular note are studies of kidney function and ion transport dating back to the 1920's that have advanced our understanding of renal physiology. A deeper understanding of genes and proteins involved in a shared biological process can be generated by comparing those from a widely used model organism (e.g., mouse or zebrafish) to the Little Skate.

At MDIBL, faculty are studying regeneration and have been comparing regenerative processes in a number of model organisms including zebrafish, *Polypterus senegalus* and axolotls (*A. mexicanum*). Several genes have been demonstrated previously to be important in regeneration including genes that pattern tissues such as the homeobox transcription factor gene family. This gene family is very unusual as it is confined into four linked clusters (*HOXA*, *HOXB*, *HOXC* and *HOXD*) and the order of gene expression is the same as the order along the chromosomal regions. Using transcriptome sequence data from the Little Skate that pre-date the genome project, Mr. King found that none of the *HOXC* cluster genes were expressed during embryonic development and hypothesized that the cluster may have been deleted in elasmobranch fishes. Using the skate genome sequence, we confirmed that the cluster is indeed missing from the genome. Interestingly, the *HOXC* cluster has also been reported missing from the genome of another elasmobranch fish, the catshark (*Scyliorhinus canicula*) (Oulion et al., 2010 and Oulion et al., 2011).

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* 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, and a Research Highlight, "Case of the missing cluster," in the journal *Nature*, also in December 2011.

References:

Oulion S, Debiais-Thibaud M, d'Aubenton-Carafa Y, Thermes C, Da Silva C, Bernard-Samain S, Gavory F, Wincker P, Mazan S, Casane D. Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. *Mol Biol Evol.* 2010 Dec;27(12):2829-38.

Oulion S, Borday-Birraux V, Debiais-Thibaud M, Mazan S, Laurenti P, Casane D. Evolution of repeated structures along the body axis of jawed vertebrates, insights from the *Scyliorhinus canicula* Hox code. *Evol Dev.* 2011 May;13(3):247-59.

Publications resulting:

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

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

Dr. Rebeka Merson (Rhode Island College) studies the Aryl Hydrocarbon Receptor (AHR) as it is a key transcription factor that is involved in cellular proliferation and differentiation and also specifically binds 2,3,7,8-tetrachlorodibenzo-p-dioxin, commonly known as dioxin. Using the skate genome sequence she has identified four AHR genes that have been studied previously in various shark species that have been useful in phylogenetic analysis of these proteins. She is also assembling the intron and exon sequences for all the AHR genes so that she can identify conserved regulatory elements among skates and other species. The regulation of AHR genes is poorly understood and given the differences in gene copy number among different vertebrate

classes, we seek to determine whether the function of these gene products is complementary, duplicative or novel by examining these conserved elements.

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 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.

Broader Impacts Vermont Specific Workshop Attendees Over Three Years:

Workshop I (May 2010)

James Vincent	University of Vermont
Meriel Brooks	Green Mountain College
Elizabeth Dolci	Johnson State College
Ryan Dundon	University of Vermont
Olivia McGee	University of Vermont
Ryan Joy	Johnson State College
Miranda Lapierre	Lyndon State College

Workshop II (October 1, 2010)

James Vincent	University of Vermont
Karen Hinkle	Norwich University
Brian Cunniff	University of Vermont

Workshop III (May 23-27, 2011)

James Vincent	University of Vermont
Ryan Joy	Johnson State College
Benjamin Kirchner	Johnson State College

Mitochondrial Annotation Workshop:

James Vincent	University of Vermont
Brian Cunniff	University of Vermont

Cyber-enabled Research Awards Region-Wide over Three Years:

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 was 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 administered the proposal evaluation process.

In 2010, the first RACER award was made to Dr. Thomas Kelly at the University of New Hampshire and Dr. James Vincent of the University of Vermont for work on “An open-access, shared database resource for eukaryotic metagenetics research.” This proposal establishes a database structure based upon the Community Cyberinfrastructure for Advanced Microbial

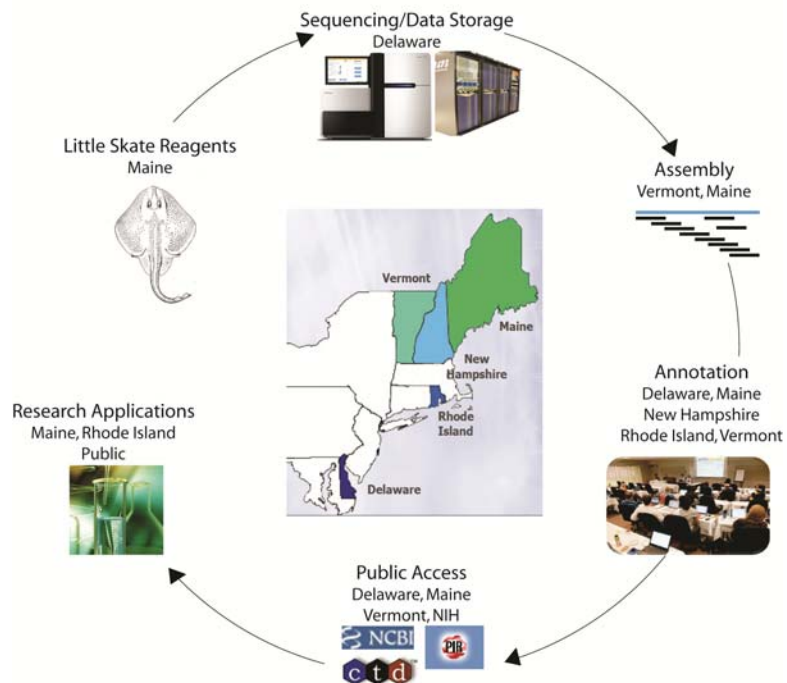
Ecology Research and Analysis (CAMERA) framework, but customized for the needs of scientists studying eukaryotic metagenetics. The project uses the NECC network and data centers to serve the needs of the eukaryote research community to create a data repository and bioinformatics tools resource that addresses the unique challenges of metagenetics analysis for eukaryotic organisms.

In 2011, three RACER proposals were submitted; two awards were made to a collaborative effort between investigators. Dr. Julia Daly, University of Maine, and Dr. Lori LaPlant, St. Anselms College, NH, are developing a cyber-enabled Northeastern Monitored Lake Temperature (MeLT) Network. Dr. Benjamin King's, Mount Desert Island Biological Laboratory, and Dr. Kevin Peterson of Dartmouth College are focused on the 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 (<http://www.necyberconsortium.org/?q=content/bioinformatics-collective-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, and to gradually allow greater public access to the data as it moves through the analysis pipeline.

Using this policy, the bulk of the skate genome sequence data was made publicly available at the NCBI on December 16, 2011 as analysis of the initial genome assembly was described in Mr. King's Science publication (King et al, Science (2011)). Genome sequence reads were deposited to the NCBI Short Read Archive under accession number SRA026856 and the initial genome assembly deposited to GenBank under accession number AESE000000000. The final assembly of the skate mitochondrial genome sequence was deposited to GenBank under accession number JQ034406 on Nov. 28, 2011. The annotation of this sequence was done collaboratively across the NECC states and published in Database (Wang et al, Database (2012)). The remainder of the skate genome sequence data are currently being analyzed to create a new genome assembly. The underlying genome sequence data and new genome assembly will be made publicly available following ongoing research.



Following the data sharing policy for the Little Skate, all data associated with the algal bloom metagenomics project is currently available through the NECC Shared Data Center. The raw sequencing data, as well as all analyses performed to date, are available to any researcher within the NECC. All sequencing data will be publicly released under NCBI Short Read Archive BioProject PRJNA176929. This is expected by the end of 2012 in anticipation of a manuscript in progress.

King BL, Gillis JA, Carlisle HR, Dahn RD. A natural deletion of the HoxC cluster in elasmobranch fishes. *Science*. 2011 Dec 16;334(6062):1517. PubMed PMID:22174244; PubMed Central PMCID: PMC3264428.

Wang Q, Arighi CN, King BL, Polson SW, Vincent J, Chen C, Huang H, Kingham BF, Page ST, Farnum Rendino M, Thomas WK, Udway DW, Wu CH; the North East Bioinformatics Collaborative Curation Team. Community annotation and bioinformatics workforce development in concert--Little Skate Genome Annotation Workshops and Jamborees. *Database (Oxford)*. 2012 Mar 20;2012(0):bar064. PubMed PMID: 22434832.

Diversity Plan and Workforce Development for NECC Over Three Years-

This year marked the pilot year of the Watershed Project, a collaborative effort among the five EPSCoR jurisdictions, to leverage our outreach and workforce development programs. The Watershed Project was modeled after the Vermont EPSCoR Streams Project, which engages high school teams and undergraduates in watershed research. In its inaugural year, 1 high school team from RI and 1 high school team from DE joined 12 VT, 1 NY and 2 Puerto Rico high school teams for a week of training June 28 – July 2, 2010 at Saint Michael's College in Colchester, Vermont. These teams then returned to their home states existing watershed education programs where they integrated their training experience and monitored local streams. Seven undergraduate interns from Delaware and Rhode Island joined 13 Vermont and 10 Puerto Rican undergraduates for a week of training June 1 – June 5, 2010 in Vermont. All undergraduate students then conducted an internship (summer of academic year) within their jurisdictions. The Watershed Project Spring Symposium, April 26, 2011, marked the culmination of the program where participants presented the results of their research through 11 oral presentations and 27 poster presentations. In Year 3 of NECC, the participants had access to the Watershed Data on the website for use in modeling projects in their home states.

The Genome Annotation workshop was intended to inspire a new generation of bioinformaticians. The May 2010 workshop attracted 35 attendees to UD – 16 males and 19 females. The October 2011 work shop in MDIBL attracted 25 attendees (14 males, 11 females). 41 attended the workshop in UD in May 2011, with 29 males and 12 females.

There were 30 participants in the Mitochondrial Annotation Jamboree. Participants ranged from undergraduates to faculty and came from all five NECC states. There were 4 from DE; 1 from NH; 8 from ME; 16 from RI and 1 from VT (18 male, 12 female). (Two instructors from VT and DE were trained at the prior Skate Genome Annotation Workshops.) The Mitochondrial Annotation materials have been incorporated into an undergraduate course, 'Practical Tools for Molecular Sequence Analysis', offered at the University of Rhode Island.

Vermont specifics for Three Years: Of the participants in the Watershed Project, 12 high school teams from Vermont, 1 from NY, and 2 from Puerto Rico were supported by Vermont, and 3 Vermont and 10 Puerto Rican undergraduates were supported by Vermont.

Vermonters Attending the Annotation Workshops:

Workshop 1: 1 faculty; 6 attendees; (3 male, 4 female)

Workshop 2: 1 faculty; 2 attendees; (2 male, 1 female)

Workshop 3: 1 faculty; 2 attendees; (3 males)

Mitochondrial workshop: 1 faculty; 1 students (2 males)

Dissemination and Communication -

The communication and dissemination plan for the NECC included the following efforts:

Public dissemination:

- Local print and web coverage extended beyond the NECC member states (VT, ME, NH, DE, RI) in Year 2 to coverage in the NCCR 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
 - 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
- Vermont Public Television ran video of the announcement so it is available for on-demand viewership.
- All media coverage is available on the VT NECC web site at
<http://www.uvm.edu/~epscor/necc/>.

NECC website: The website has been a collaborative effort. Rhode Island EPSCoR office hosted the web site and maintained a wiki for users to edit the content. In 2011 the web site was redesigned for efficiency.

Google Analytics was added to the site to monitor the number of page hits and where those visitors originate. From 11 September 2012 through 4 October 4 2012, the NECC website had a total of 78 visits with a total view of 244 pages. The page views consisted of 32 Homepage views, 33 Final Reports page views, 21 Information Page views with the remainder of the views distributed between the remaining pages. 66 (84%) of the website visits were from the United States with the majority of those coming from the East Coast. There were many visits from institutions outside the NECC and outside EPSCoR.

Communication among NECC Partners:

Each year, we held a meeting to review progress. Our third annual meeting was held on March 16, 2012 at the University of Delaware. 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. Members of NIH National Center for Research Resources (NCRR) observed part of the meeting by videoconferencing. Committees of the NECC met and reported to the group. Overall, we considered that we have made significant progress as an unprecedented consortium among five EPSCoR states. The conference agenda, list of participants, and minutes are attached in Appendix 1 of this report.

The NECC Year 2 Annual Meeting was hosted by Maine EPSCoR on March 15, 2011 at the University of Maine in Orono. Six states were represented (the five NECC states and a guest from Arkansas), with 41 participants including 7 faculty, 1 postdoc, 2 graduate students, 21 technical/professional staff, and 2 others from the private sector in attendance. In addition, NSF EPSCoR Program Director Jennifer Schopf was present in person, and NIH IDEa Program Director Fred Taylor, NIH Deputy Director Michael Sayre, plus 5 others from NIH participated via videoconferencing.



- News
- Information
- Committees
- Meetings
- Projects
- Solicitations
- Members



Meetings

Tags: [annual meeting](#) [maine](#) +

> Edited Mar 22, 2011 9:13 AM by Kate Wilson...

Meetings

Second Annual Regional Meeting

The meeting was held at the Wells Conference Center at the University of Maine on Tuesday, March 15, 2011.



What's Hot

No items tagged with hot.

Recent Changes

Meetings
2011-03-22T13:09+00:00

Regular NECC videoconferences of Executive Steering Committee and all other committees scheduled each month.

To address comments from our Program Officer in 2011 about the level of collaboration among NECC jurisdictions and participants:

The NECC participants have communicated extensively, mostly through videoconferencing, over the course of three years. The videoconferencing is used for regular executive committee meetings, technical committee meetings, weekly NEBC meetings, review panels for the RACER awards, conferences about progress of the research and outreach around the Little Skate Genome, water sampling and data analysis of the algal bloom metagenomes, and so on. We currently are meeting weekly about future grant opportunities to expand the collaborative research on water, sensors and sustainability – interests that cross all five of our states. Participants on these videoconferences include members from all five states.

The four RACER awards were made to collaborators from at least two NECC jurisdictions. In one case, there were researchers from 3 institutions and 2 states involved in the research.

The metagenomic, microbiome and Little Skate Genome projects are enormously collaborative. For the algal bloom metagenome, sampling was done by multiple researchers in four of the five states. Bioinformaticians from all five states divided up the jobs for these genomic analyses. Data Centers in Maine and Delaware are critical to this work. The researchers continue to access the data sets remotely for their work. Delaware also provides critical sequencing services that have generated the massive amounts of data that are at the heart of these projects.

In order to launch the shared data centers (SDCs), there was a great deal of work and communication of the Technical Committee members from all five states. The SDCs in Maine and Delaware worked intensively with Vermont's Dr. James Vincent and Mr. Marc Farnum-Rendino, who became a temporary employee of UD where he helped launch the data center in Delaware.

The broader impact programs also rely upon the energetic participation of members from all five NECC states. The annotation workshops have engaged students and faculty from all five states, and required the coordination and high level of commitment from the NEBC members from all five states.

While Vermont hosted the participants in the Watershed Project for training, all five states then worked with their undergraduates or high school teams to continue the water research in the context of their home state.

Last, the development of the fiber network in the NE required the coordinated submission of proposals to NSF and to NIH. There was a great deal of altruism in that the budgets did not divide the funds available by five; instead we based budgets on the most urgent needs for fiber, which meant that some states received less than one fifth of the budget.

Evaluation and Assessment Plan Over 3 Years-

Because funding for the NECC project is an admixture of funds from NIH NCRR and NSF EPSCoR funds, assessment and evaluation for the 5 NECC jurisdictions is done by a combination of INBRE and EPSCoR organizations. 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 next External Advisory Board evaluation occurs in April 2011 and the INBRE AAAS review in Fall 2011. On the EPSCoR side, the next RII AAAS review will occur in Fall 2011.
- Maine: evaluation via the INBRE External Advisory Committee will occur in August 2011 and by the AAAS in fall 2011. The EPSCoR Track-2 evaluation will occur as part of the Track-1 process.
- New Hampshire: assessment of the fiber transmission for the southern route occurred in May 2011.
- Rhode Island: INBRE supplement assessment occurred in 2010 and 2012.
- Vermont EPSCoR worked with Dr. Joy Livingston, Flint Springs Associates, to perform formative assessment for participants in the Annotation Jamborees. (See appendix for the summary of the jamboree evaluations.)
- Vermont: the AAAS panel in June 2011 provided formative assessment of the entire NECC program. This is part of the Vermont INBRE. The AAAS report is included as an appendix. The general AAAS recommendations: Regional collaboration is working well; Collaboration relies upon interconnectivity of state; RACERs should continue; States should continue to monitor progress. This AAAS evaluation of the NECC across the five states addresses the concern of our previous Program Officer about evaluation.

Sustainability Plan NECC Wide Over 3 Years-

At the annual meetings, we discussed the future of NECC and how 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 will continue this conversation and planning.

Part of our approach to sustainability is to foster cyber-enabled research through collaborative research projects funded first through pilot funds that we call RACERs as described above. New collaborations around water, sensors, and sustainability and resilience are currently being designed. Some of the ideas and researchers involved come from the RACER process. Additional researchers and proposed research directions come from all of our five states. We continue to have videoconferences with these researchers every 2 weeks even after the end of the Track-2 grant.

Management and Coordination Plan NECC Wide Over 3 Years-

The NECC executive committee continues to meet through videoconferencing on a regular basis. The executive committee has held 8 videoconferences in the last year. These conferences helped us update each other on progress, address issues, and welcome new members of the committee. We took recommendations from the Water Research Committee on RACER awards and the selection of RACER awardees and approved the year 2 RFP. On January 6, our videoconference included Dr. Jennifer Schopf who discussed the annual report of progress for the NSF Track-2 grant.

We had the opportunity to present our progress on the fiber network and cyber-enabled research at three venues in Year 2:

1. 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
2. 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
3. December 7, 2010: Judith Van Houten and Karl Steiner presented an over view of the NECC project to the EPSCoR/IDeA Foundation and Coalition meeting in Washington, DC
4. February 9, 2011: Kelvin Chu and Judith Van Houten presented an overview of NECC to representatives of Sandia National Lab
5. 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.

The other standing committees meet regularly, including the Technical Committee, Water Research Committee and NE Bioinformatics Collaborative. The Water Research Committee reviewed and recommended a RACER award in year 2 and the RFP for awards in Year 3 have been disseminated.

State Specific Unobligated Funds:

VT EPSCoR EPS-0918284

We have no unobligated funds remaining. This is the final report for this award.

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
Clemins	Patrick	Cyber Specialist, VT EPSCoR	University of Vermont
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 Services	University of Rhode Island

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**



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

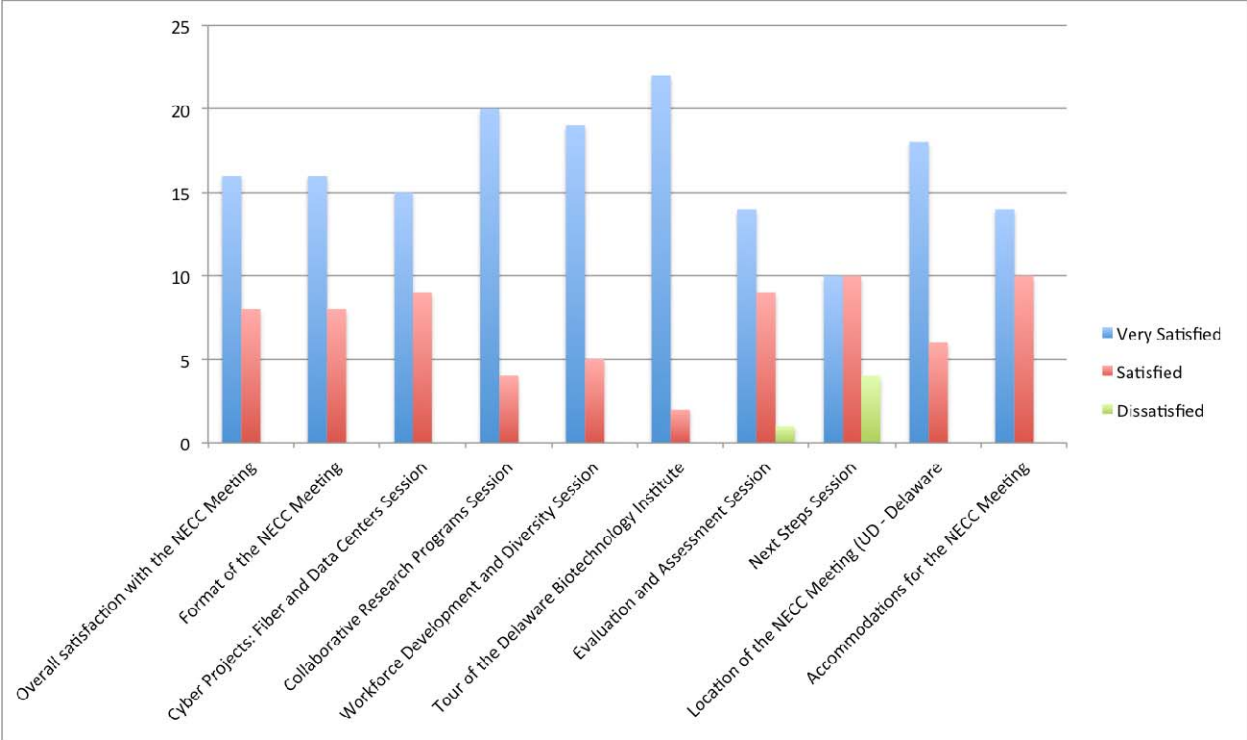


Fig 8: Graphical results of 2012 3rd Annual NECC Meeting evaluation indicate a successful event.

APPENDIX 2:

Outcome Report – Little Skate Workshop

Outcome Report of the Little Skate Workshop provided by DE EPSCoR
June 2012

The sequencing and annotation of the little skate (*Leucoraja erinacea*) genome has provided a venue for promoting research collaboration, bioinformatics infrastructure development, as well as bioinformatics training opportunities in our region. Several annotation workshops and jamborees hosted by NEBC have provided training to students and investigators in genome sequencing, annotation, and analysis. The SkateBase (<http://skatebase.org>) website has been developed to support this training activity and serve as a central portal for collaborative annotation and project dissemination.

To date 56 trainees have participated in three workshops, receiving both theoretical and hands-on training. An additional series of virtual distributed jamborees involved 29 trainees, who learned critical bioinformatics skills while annotating mitochondrial genome features. The project has produced many tools to support community genome annotation, data sharing, and dissemination. By coupling education and scientific objectives, this project has not only imparted important bioinformatics knowledge, but has already contributed to multiple high-impact publications.

We are applying the physical/virtual infrastructure and lessons learned from these activities to enhance and streamline the genome annotation workflow and expand our training activities. Feedback from participants in all three workshops was very positive overall. The feedback from each workshop has improved subsequent workshops, providing for a better learning experience for participants, such as shorter lectures on background material coupled with more extensive hands-on activity. Indeed coupling such hands-on training with annotation has proved very successful, fostering better understanding of the tools taught in lectures. The hands-on exercises with real-world problems provoked deeper thinking and strengthened understanding of abstract bioinformatics concepts.

During the first workshop hosted in Delaware, the participants were asked to fill out a survey form on-line at the end of each day. This format allowed the workshop organizers to adjust the lecture material and hands-on activities for the following day accordingly. For the third workshop, the survey was distributed on the final day with 13 participant completing survey forms. Almost all the participants agreed that the workshop provided them with new knowledge about the subject matter and taught valuable skills that they would be able to apply in their work. Overall, the workshop was considered valuable by all the participants.

Through this process of hosting and evaluating workshops and jamborees we have gained critical insights into efficiently hosting productive workshops and jamborees. Many of the lessons learned involve methods for encouraging participation in the workshops, providing adequate support (training materials and technical support) for hands-on activities, and promoting scientific utility and contribution beyond the workshops. Some of the recommendations and observations are described below.

Participation. i) if students are involved, it is important to carefully select the dates of the workshop to avoid conflicts with regular semester schedules, if possible avoid having the workshop during regular semesters; ii) for broader community participation it is important to offer incentives. In our case, some of these were full funding to attend the workshops including travel and lodging, free advanced training in bioinformatics, visits to state-of-art sequencing facilities, opportunity to disseminate participants work via poster presentations and interaction with multidisciplinary researchers, and a certificate at the end of the workshops.

Support. It is critical to the success of the annotation workshops and jamborees for the diverse participants that training materials include tutorials and clear annotation guidelines, as well as intuitive web-based annotation interface with analysis and visualization tools.

Scientific utility and contribution beyond the workshops. To stimulate community usage and contributions, the workshop training materials (e.g. SkateBase tools) are being integrated into the educational curricula across institutions to engage students through independent research or special topics courses. The direct integration of genome annotation into regular biological sciences courses provides effective means for student training and participation.

**APPENDIX 3:
AAAS REPORT**

North East Cyberinfrastructure Consortium

For this current visit, the AAAS panel was asked to review and provide guidance on the five-state North East cyberinfrastructure collaborative (NECC) in which VGN participates. Established in 2006 through Track-2 grants from NSF EPSCoR to programs in Delaware, Maine, New Hampshire, Rhode Island, and Vermont, the NECC aims to 1) identify and promote the shared use of research facilities across the region; and 2) assess and address cyber-infrastructure needs. The University of Vermont has been designated the lead institution for the NECC.

Funding for the NECC is heavily leveraged, with federal contributions from NSF and NIH, and is used to support the purchase and lighting of physical fiber, personnel to maintain the fiber network and enable collaborative research, and two regionally distributed data centers (in Delaware and Maine). In addition, the NECC has created a funding stream to support cyber-enabled pilot research projects involving collaboration between researchers throughout New England.

To better understand the NECC, its relationship to VGN, and its impact on the region, the AAAS panel met with stakeholders from the University of Vermont as well as members of the NECC Executive Committee from each of the participating 5 states.

NECC is currently building out the high-speed fiber optic backbone which will interconnect its member institutions, as well as provide high-speed connectivity to regional and national research networks. The new fiber will create dual, diverse paths through Vermont and New Hampshire, as well as additional paths in Maine and Rhode Island. Importantly, in addition to directly connecting the member institutions, the new fiber will provide high-speed connectivity to the broader Northeast Research and Education Network (NEREN) which includes the Northern Crossroads, OSHEAN, Nysernet, the University of Maine and others.

The network is in various stages of completion within the individual states. The fiber pathways through Vermont, which will complete the loop with New Hampshire, were in the process of being laid during the AAAS visit. New fiber in Maine will complete a Maine loop via the MDI Biological Lab, as well as provide a path to the northeastern border with Canada near Presque Isle. Maine's connectivity is highly leveraged from their BTOP funding and other state networking infrastructure. Rhode Island is also highly leveraging their existing OSHEANS network infrastructure, as well as a new pathway to Kingston, and their build-out is 70% complete.

The project is ending its second year of NSF Track-2 and NIH funding. Once the installation of the remaining fiber and loop is completed, the network will be capable of providing multiple pairs of 10 gigabit/second circuits, a huge increase in available bandwidth for all of the members. The partners are providing good leverage from a variety of funding sources, including the NSF and NIH, but also others, as well as

leveraging existent WAN infrastructure. Much of the current funding is being used to fund long-term leases of fiber. Additionally, most of the institutions will be utilizing the new NECC network paths to replace their other commercial circuits and redirecting their cost-savings into the NECC. This strategy of acquiring long-term leases and redirecting cost savings bodes well for the long term sustainability of the network. All of the institutions have committed to the network operation for at least ten years.

Although the NECC is still in its early stages, the project is already having significant impact locally and regionally. Numerous examples of the collaboration and cooperation between the parties were exhibited during the visit, including several “proof-of-principle” projects. These include the Little Skate Genome and the Blue-green Algae Metagenomics projects. During the meetings with the participants, it became clear that the foundations for this project had been set through numerous prior interactions and collaboration that have been taking place throughout the region, many of these being driven by INBRE and EPSCoR programs. Cooperation within each of the institutions was equally in evidence with good vertical coordination between the researchers, the EPSCoR and INBRE directors, the Bioinformatics facility, the CIO’s office, and senior university leadership.

Regional impacts

Once the NECC network is fully operational, it should dramatically increase the member institutions ability to engage in cyber-enabled research and education by providing significantly increased network capacity and the increased availability afforded by the redundant optical network paths. In addition to creating enhanced network connectivity, the NECC is creating benefits from the increased levels of collaboration and coordination between the participants. Even in these early stages, a number of collaborative projects have been initiated and appear to be working well.

a. Two examples of collaborative projects created by the NECC are the Little Skate Genomics project and the Blue-green Algae Metagenomics project. These projects resulted in collaborative research across several states, educational opportunities for a large number of students at multiple institutions and a new genomics data base published for the broader research community. These projects would not have been possible without the NECC. To foster collaborative research, NECC has created the Regional Awards for Cyber Enabled Research (RACER) program to fund small (~\$10K) pilot projects. Funded projects must involve investigators from multiple institutions. In its first year RACER funded the start-up of the metagenomics project. In year two RACER funded a UNH/UVM project to create a metagenetics data base for eukaryotes. This year six pre-proposals have been submitted for the current solicitation, and two proposals have been funded in year 3. The first proposal is to study microRNAs expressed in response to hypoxia in, *Fundulus heteroclitus*, and the second is to develop a cyber-enabled Northeastern Monitored Lake Temperature Network. Both are aligned with the NECC mission and support environmental or bioinformatics collaborations among investigators from at least two jurisdictions within the

NECC.

The NECC has afforded the opportunity to create a number of shared resource centers across the region, most notably a pair of data centers, one in Delaware and another in Maine. These data centers are providing researchers with access to large data storage arrays. NECC has developed cybertools to house the primary data storage in Delaware and to keep a live, replicated copy in Maine. The shared data centers played a key role in the Skate Genome and Metagenomics projects. Within the framework of NECC the partners have developed a data access and sharing plan which will be beneficial to all of the members. Once the new network connections become fully operational, the demand on the shared data centers is likely to increase substantially.

Additionally, Delaware has received an MRI grant to acquire a next-gen sequencer, in substantial part due to their participation in the NECC. This will be a valuable resource for all of the partners. Delaware's sequencing capabilities are playing a key role in many of the NECC projects, such as the Skate Genome.

As a result of the NECC, the partners are beginning to institute curricular changes within their individual institutions and states. There are several initiatives to bring bioinformatics training into the educational pipeline. Rhode Island is working with the Rhode Island School of Design (RISD) on the visualization of genomics data and working around the state to develop curricula that integrate these technologies. New Hampshire and Delaware are also developing programs to introduce bioinformatics into the classroom. The NECC Watershed project is engaging undergraduates and high school students across and beyond the region. Many students from multiple institutions were engaged through the workshops associated with the Skate Genome project. Although currently unfunded, the partners have proposed the development of a cyber-based, undergraduate bioinformatics course. This course would use the shared datacenter resource to analysis large genomic data sets and use video conferencing over the NECC network for distributed, remote teaching. Projects such as this which leverage so many aspects of the NECC-provided cyberinfrastructure capabilities are laudable.

All of the NECC team members that met with AAAS panel were very positive about the collaborations, as well as the impact that the project was having on their institutions. The NECC members at several levels meet regularly using video conferencing. The organization appears to be running well and being quite productive.

Vermont impacts

The new NECC fiber will have a dramatic impact on networking at UVM. Once complete the new fiber will afford two orders of magnitude increase in available bandwidth. This dramatically increases the opportunities for UVM to engage in cyber-based research and education. Beyond the local region, the new network will provide enhanced access to other network resources including Internet2, the NEREN, the

proposed U.S. UCAN pop in Albany, and NSF and NIH resources such as XSEDE, the next generation Teragrid.

Using NIH ARRA supplements, VGN has hired a bio-IT professional to develop the shared data center. A new full time position, as well as two full-time summer internships, has been created within the Bioinformatics Core to support large scale sequencing projects for collaborative research and to build capacity in the metagenomics area.

To fully benefit from the NECC capabilities, the networking infrastructure would need to extend throughout the state. Since Vermont does not currently have a state broadband network, this remains a challenge. Bringing advanced cyberinfrastructure to other educational institutions throughout the state, including the BPIs, would be an important, but very costly, undertaking.

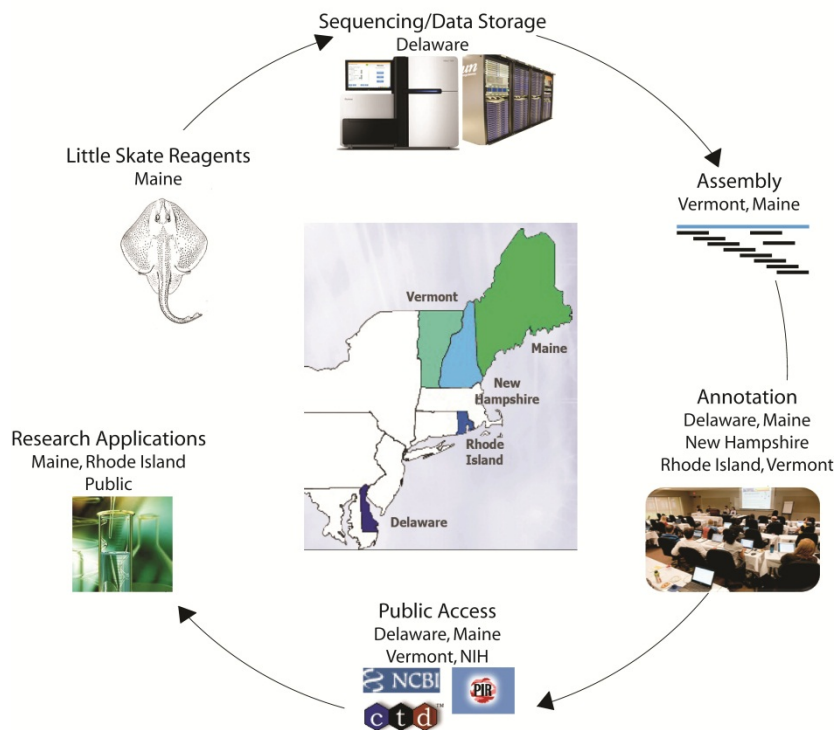
Recommendations:

- The RACER program to fund pilot research projects is very creative and offers a good proof of principle for the network. This program will be important to continue to spur collaboration, utility, and evolution of the network.
- Regional collaboration seems to be working well. However, the risk of interconnectivity is reliance on each state to do their part. It will be important to define how that will be monitored and enforced.
- VGN may consider asking Joy Livingston, the project evaluator, to attempt to assess the impacts of NECC.

APPENDIX 4: NECC Highlight – NSF EPS# 0918284

Data Sharing

Due to the high-level of interest in the skate genome, metagenomics sequence data and the nature of the North East Consortium (NECC) collaboration (ME, NH, DE, VT, RI), Mr. Ben King and Dr. Carolyn Mattingly of Mount Desert Island Biological Laboratory (MDIBL) developed a data access policy (<http://www.necyberconsortium.org/?q=content/bioinformatics-collective-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, and to gradually allow greater public access to the data as it moves through the analysis pipeline.



Using this policy, the bulk of the skate genome sequence data was made publicly available at the National Center for Biotechnology Information (NCBI) on December 16, 2011 as analysis of the initial genome assembly was described in Mr. King's Science publication (King et al, Science (2011)). Genome sequence reads were deposited to the NCBI Short Read Archive under accession number SRA026856 and the initial genome assembly deposited to GenBank under accession number AESE000000000. The final assembly of the skate mitochondrial genome sequence was deposited to GenBank under accession number JQ034406 on Nov. 28, 2011. The annotation of this sequence was done collaboratively across the NECC states and published in Database (Wang et al, Database (2012)). The remainder of the skate genome sequence data are currently being analyzed to create a new genome assembly. The underlying genome sequence data and new genome assembly will be made publicly available following ongoing research.

Following the data sharing policy for the Little Skate, all data associated with the algal bloom metagenomics project is currently available through the NECC Shared Data Center (SDC). The raw sequencing data, as well as all analyses performed to date, are available to any researcher within the NECC. All sequencing data will be publicly released under NCBI Short Read Archive BioProject PRJNA176929. This is expected by the end of 2012 in anticipation of a manuscript in progress.

*Wang Q, Arighi CN, King BL, Polson SW, Vincent J, Chen C, Huang H, Kingham BF, Page ST, Farnum Rendino M, Thomas WK, Udway DW, Wu CH; the North East Bioinformatics Collaborative Curation Team. Community annotation and bioinformatics workforce development in concert--Little Skate Genome Annotation Workshops and Jamborees. *Database (Oxford)*. 2012 Mar 20;2012(0):bar064. PubMed PMID: 22434832.

*Image created by Qinghua Wang wangq@dbi.udel.edu and published in this article